AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR024266

A Regularization Approach to Improve the Sequential Calibration of a Semidistributed Hydrological Model

A. de Lavenne[1,2], **V. Andréassian**[1], **G. Thirel**[1], **M.-H. Ramos**[1], **and C. Perrin**[1]

[1]HYCAR Research Unit, Irstea, Antony, France, [2]Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden

Key Points:
- Sequential calibration without regularization yields low performance in small ungauged catchments
- Sequential calibration constrained by regularization strategy helps improving model regionalization
- Regularization benefits from a priori parameter values of neighboring catchments

Supporting Information:
- Supporting Information S1
- Figure S1
- Figure S2
- Figure S3
- Figure S4

Correspondence to:
A. de Lavenne,
alban.delavenne@smhi.se

Citation:
de Lavenne, A., Andréassian, V., Thirel, G., Ramos, M.-H., & Perrin, C. (2019). A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resources Research*, *55*, 8821–8839. https://doi.org/10.1029/2018WR024266

Received 27 OCT 2018
Accepted 17 SEP 2019
Accepted article online 11 OCT 2019
Published online 11 NOV 2019

Abstract In semidistributed hydrological modeling, sequential calibration usually refers to the calibration of a model by considering not only the flows observed at the outlet of a catchment but also the different gauging points inside the catchment from upstream to downstream. While sequential calibration aims to optimize the performance at these interior gauged points, we show that it generally fails to improve performance at ungauged points. In this paper, we propose a regularization approach for the sequential calibration of semidistributed hydrological models. It consists in adding a priori information on optimal parameter sets for each modeling unit of the semidistributed model. Calibration iterations are then performed by jointly maximizing simulation performance and minimizing drifts from the a priori parameter sets. The combination of these two sources of information is handled by a parameter k to which the method is quite sensitive. The method is applied to 1,305 catchments in France over 30 years. The leave-one-out validation shows that, at locations considered as ungauged, model simulations are significantly improved (over all the catchments, the median KGE criterion is increased from 0.75 to 0.83 and the first quartile from 0.35 to 0.66), while model performance at gauged points is not significantly impacted by the use of the regularization approach. Small catchments benefit most from this calibration strategy. These performances are, however, very similar to the performances obtained with a lumped model based on similar conceptualization.

1. Introduction

1.1. The Challenge of Calibrating a Semidistributed Model

A spatialized hydrological model aims to provide "predictions everywhere" in gauged as well as ungauged catchments (Sivapalan et al., 2003). When different goals (economical, social, and environmental) have to be achieved, hydrological models are particularly useful if they are built to provide a reliable description not only at interior gauged stations but also at any location over the entire catchment. This could concern, for instance, the allocation of water in space and time (e.g., water reservoir management) or the design and implementation of ecosystem services (e.g., nature-based solutions for flood or water quality control).

To achieve these goals, the space is discretized in different modeling units such as regular grids (Samaniego et al., 2010), subcatchments (Lindström et al., 1997), hillslopes (Zehe et al., 2001), or hydrological response units (HRUs; Leavesley et al., 1983). These discretizations represent hypotheses of the underlying system behavior (Fenicia et al., 2016). For models that are not physically based (where parameters and states correspond to measurable quantities), these modeling units need to be calibrated. Compared to lumped models, this discretization in modeling units drastically increases the number of parameters that need to be estimated and it can lead to overparameterization and poor identifiability. Therefore, the calibration of a semidistributed hydrological model requires a well-defined strategy that makes use of all the available information in the catchment (discharge observations, catchment characteristics, prior knowledge on parameters, etc.).

1.2. Calibration Strategies for Semidistributed Models

A large number of studies investigated the different strategies to handle the spatial heterogeneity of the catchment and to better constrain the model calibration procedure. Many strategies are based on a common concept called regularization. It corresponds to a family of mathematical techniques that aim at introducing an additional constraint to solve an ill-posed problem or to prevent overfitting (Engl et al., 1996). In the context of hydrological modeling, these techniques aim at improving parameter realism, spatial consistency, and the stability of the optimization problem (Doherty & Skahill, 2006; Tonkin & Doherty, 2005). Most of

the time, these additional constraints are based on a regionalization strategy (note that the term "regionalization" is sometimes preferred over "regularization" to describe the strategy). Thanks to regularization coefficients (sometimes called "superparameters"), it is possible to relate model parameters to observable characteristics, and this additional knowledge will drive the calibration. In this way, only these superparameters need to be calibrated (reducing the number of unknowns), and parameter values are spatially distributed in order to be consistent with the prior spatial information.

In the following subsections, we summarize the different strategies to calibrate a semidistributed model into four different strategies. Some works could, however, be spread into these four groups, and most of them share the concept of regularization.

(1) Reducing the Number of Unknowns

By calibrating regularization parameters instead of model parameters, several authors face the overparametrization issue by directly reducing the number of unknowns. For instance, Pokhrel et al. (2008) developed a regularization relationship (a nonlinear equation based on three parameters: a multiplier, a power term, and an additive term) to estimate parameters mainly from soil property data (soil depth and curve number), greatly reducing the number of unknowns to be estimated from 858 to 33. The regionalization developed by Samaniego et al. (2010) allows to estimate the 28 parameters of each hydrological units of the mHM model by calibrating only 62 global parameters. Antonetti et al. (2017) also emphasize the need to make conceptual models more process-based in order to reduce the need for calibration.

(2) Regionalizing Model Parameters

Regularization is often based on a regionalization approach where the goal is to transfer information from gauged to ungauged locations (He et al., 2011). It is a common concept for calibration. When applied to model parameters, it aims to provide a parameter set for any model unit of the semidistributed model.

Abdulla and Lettenmaier (1997) calibrated the VIC model for each subcatchment independently and then related the calibrated model parameters with the catchment characteristics using regression methods. To avoid the risk of equifinality with this traditional approach, Hundecha and Bárdossy (2004) assumed a functional form of the relationship between the parameters of the HBV model and the catchment characteristics (such as land uses, soil type, slope, or size). Instead of calibrating the HBV parameters, they calibrated this relationship for many catchments simultaneously. Götzinger and Bárdossy (2007) tested the addition of two conditions to constrain functional relationships between catchment characteristics and model parameters: a monotony condition and a Lipschitz condition (where similar subcatchments were calibrated simultaneously). The combination of both constraints appears to be the most promising as it improves the consistency of the regression relationship without affecting the performance of the regionalization. Samaniego et al. (2010, 2017) also calibrated statistical relations between catchment predictors (e.g., soil texture, land cover or slope) and the parameters of the mHM model. They called the method "multiscale parameter regionalization" (MPR): By considering the spatial resolution of those descriptors, it aims to provide a performance independent from the model's spatial resolution.

The assessment of catchment similarity can also help model calibration. For instance, Hundecha et al. (2016) calibrated some parameters of the HYPE model at the European scale based on catchment classification according to their characteristics. A regionalization was then performed within each group. Piniewski et al. (2017) transferred model parameters from gauged to ungauged locations using hydrological similarity. Fenicia et al. (2016) defined their HRUs based on hydrological similarity.

(3) Constraining the Optimization

In the model building process, Gupta and Nearing (2014) emphasized the need to test more conceptual representations that could be based on constraining model parameters, state variables, and internal fluxes. This constraint should not be too restrictive in order to keep a sufficient degree of freedom while preserving parameter realism and spatial consistency (Mendoza et al., 2015). By using additional data in the procedure, regularization is one way of adding a constraint, which can come from different sources.

To constrain parameters and processes, the additional information can be based on expert knowledge (Antonetti & Zappa, 2018; Gharari et al., 2014) or on physical data. Soil moisture, for instance, either observed in situ or assessed by remotely sensed data, is often used as a constraint for calibration (Koren et al., 2008; Li et al., 2018). Recently, Nijzink et al. (2018) used combinations of remotely sensed

products to constrain parameters in the absence of streamflow data. The soil moisture products of AMSR-E, ASCAT, and the total water storage anomalies from GRACE were particularly relevant in determining feasible parameter sets. Khakbaz et al. (2012 with the SAC-SMA semidistributed model), and similar to Francés et al. (2007; with the TETIS fully distributed model), used a priori parameter sets previously defined by Koren et al. (2003) from NRCS-STATSGO soil data to define the spatial variability of the parameter values within a catchment. A common parameter for all modeling units is then automatically calibrated to correct this a priori using a linear relation that preserves the spatial coherence of the a priori. Similarly, Pokhrel and Gupta (2010) used a priori parameter values (again previously defined by soil data) in their regularization, which consists of nonlinear equations to facilitate the adjustment of the mean, variance, and shape of the parameter distributions. They showed that regularization improves performance at the basin outlet compared to a priori parameters. Bulygina et al. (2009) used the HOST database in the United Kingdom, which provides hydrological indices for different soil classifications, to constrain the parameter space and relate parameter values to land management.

Based on the regularization developed by Pokhrel et al. (2008), Pokhrel et al. (2012) used signatures to select more consistent parameter sets in a multiobjective calibration. Kelleher et al. (2017) addressed the issue of equifinality by proposing a hierarchical approach to reduce the number of behavioral sets based on different constraints such as signatures and several errors metrics.

(4) Multistage Estimation Approaches

Another common strategy that could be combined with regularization is to calibrate parameters step by step rather than all at once. In order to calibrate the HYPE model at global scale, Arheimer et al. (2019) followed a stepwise procedure by calibrating groups of parameters which regulate specific processes in representative gauged catchments. They observed similar performance between the catchments used for calibration and those kept for validation.

Stepwise calibration could also be done sequentially from upstream to downstream, by calibrating the successive intermediary modeling units that contribute to the total discharge. Several authors demonstrated the benefit of internal flow measurements for calibration (Andersen et al., 2001; Boscarello et al., 2013; Moussa et al., 2007; Wang et al., 2012). Some studies then looked for the best calibration strategy that would make use of those internal gauging stations (Feyen et al., 2008; Lerat et al., 2012; Wi et al., 2015): They compared lumped calibration to multisite calibration that is either based on a sequential calibration or a simultaneous calibration of all interior gauges.

Feyen et al. (2008) demonstrated that moving from a lumped (one parameter set used on every modeling units) to a sequential calibration approach improves the accuracy of the flow predictions, especially in the upstream subcatchments. Lerat et al. (2012) demonstrated that the multisite strategies clearly outperform the lumped strategy at interior gauges. However, when the interior gauges are not used in the optimization, the lumped calibration appears more robust than the multisite calibrations. Wi et al. (2015) also demonstrated that multisite approaches outperform a calibration based only on the basin outlet, but they showed that the simultaneous use of interior gauges improves the calibration compared to a sequential approach.

In the studies mentioned above, the number of interior points remains low (always less than 10 gauges), making the optimization issue less complex than if it was applied to a much larger number of gauges (such as national scale modeling or continental scale modeling). For this reason, Wi et al. (2015) also concluded on the need to bring additional spatial information in the calibration procedure to develop more robust estimates of spatially distributed parameter values.

A sequential calibration could be combined with other optimization procedures. Hughes et al. (2016) combined a sequential calibration of model parameters with a system-wide optimization of a subset of parameters related to reach losses and gains. More generally, in the work of Fenicia et al. (2016), the stepwise calibration does not only address model calibration but it also aims to build different modeling hypotheses following three steps. Firstly, the choice of spatial discretization into HRUs is based on regions of hydrologic similarity. Secondly, the choice of the model structure of each HRU is made using the SUPERFLEX framework (Fenicia et al., 2011). Finally, the space-time transferability of model parameters is evaluated.

1.3. Regularization for Prediction in Ungauged Catchments

While most of these strategies aim to maximize model performance at gauged stations (at the outlet of a catchment and sometimes at interior points), performance everywhere else inside the catchment (i.e., at

ungauged interior locations) has not been systematically addressed. Past studies have often reported interior points performance to be worse than performance at the outlet (Bandaragoda et al., 2004; Khakbaz et al., 2012; Reed et al., 2004). This has already been pointed out by Blöschl et al. (2013), who reported only a few attempts at prediction in ungauged catchments with semidistributed models by the end of the International Association of Hydrological Sciences decade on prediction in ungauged basins.

Regularization approaches have been evaluated mainly for improving calibration of the model at gauged stations. However, the technique has an important potential for prediction in ungauged catchments: the relationships established between model parameters and catchment characteristics can be generalized at any location within the catchment. Even though regularization aims to better describe the spatial variability of model parameters, its benefit is not always straightforward. Khakbaz et al. (2012) demonstrated that a lumped calibration applied to a semidistributed model structure performed better (at the outlet and at interior points) than a distributed calibration strategy based on regularization. Pokhrel and Gupta (2010) showed that their regularization based on a priori parameters helps improve the performance at the outlet but decreased the performance at interior ungauged points compared to the a priori parameters. Moreover, calibration using uniform parameters gave comparable performance at the outlet. This indicates that the different regularizations studied (multiplicative, additive, and nonlinear) were not particularly efficient at properly capturing spatial information and enabling satisfying regionalization. Conversely, Samaniego et al. (2010, 2017) demonstrated that their MPR technique also leads to a loss of performance (15% of the Nash-Sutcliffe Efficiency) at interior ungauged locations with respect to the performance obtained at the outlet. However, the MPR technique appears to be more robust than standard regionalization. It also resulted in an easier transferability of model parameters to ungauged locations, compared to calibration based on HRUs (Kumar et al., 2013).

Most of the studies in the literature use regularization strategies based on spatial physical descriptors of the catchment to constrain the identification of model parameters. However, in data-rich regions, the best regionalization strategies, are often obtained using spatial proximity. For instance, in Austria, Parajka et al. (2005) demonstrated better performance in regionalizing a semidistributed model using a kriging-based approach than using a physiographic similarity-based approach. In France, Oudin et al. (2008) showed that spatial proximity provides the best regionalization solution as a result of the high density of gauged stations.

1.4. Aims of the Paper

In this paper, we explore a regularization approach based on spatial proximity to robustly parameterize a semidistributed hydrological model at a countrywide scale. We chose to focus on the sequential calibration for its potential to provide good performance at interior points (Feyen et al., 2008; Lerat et al., 2012; Wi et al., 2015). However, this paper aims to address two issues with this strategy. First, the performance of sequential calibration at ungauged stations has received only limited attention compared to the other calibration strategies. One of the aims of this paper is to fill this gap by applying the methods on a large sample of nested catchments. Second, this strategy alone has one major weakness, as we have shown in a previous study (de Lavenne et al., 2016). The discharge contribution of modeling units to the main river tends to decrease from upstream to downstream relatively to the total discharge. The greater part of the water volume comes naturally from upstream through the river, and small downstream tributaries or hillslopes may represent only a low percentage of the main river discharge. Consequently, they have an insufficient impact on the total river discharge to identify their parameter values efficiently. The hydrological behavior of those downstream contributions is somehow masked by the behavior of the main river. Addressing this sensitivity issue is one of the principal objectives of this paper. We propose using the regularization paradigm to contribute an additional constraint that can facilitate parameter calibration in a semidistributed model.

This paper addresses the following questions: (1) Are the parameter sets estimated in a sequential calibration from upstream to downstream reliable at ungauged locations? (2) How much can regularization improve prediction in ungauged catchments within this kind of sequential calibration? (3) How does a lumped model behave in comparison to a semidistributed model based on the same conceptualization?

The paper is organized in six sections. The model and the data are presented in section 2. The proposed calibration strategy is described in section 3, together with the methodology to evaluate its performance. In section 4, the performance of the regularized calibration strategy is compared to the traditional sequential calibration and the impact on model parameters is evaluated. Finally, the discussion and conclusion sections highlight the usefulness of the proposed approach and its limits.

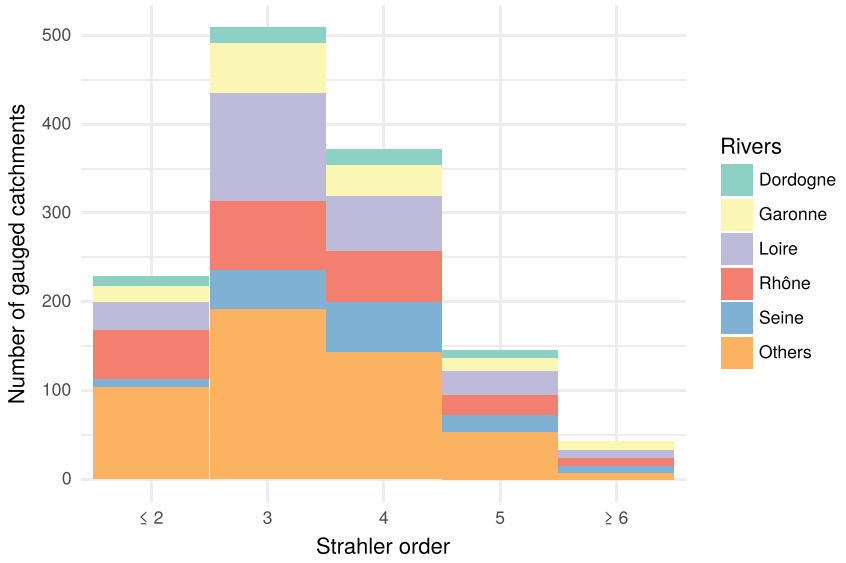

Figure 1. Distribution of the Strahler order values on the 1,305 catchments, according to their location within the main hydrographic basins.

2. Data and Models

2.1. Data

Model evaluation is done over 30 years of daily data using a split sample test (Klemeš, 1986), where two 15-year periods (1980–1994 and 1995–2010) were alternatively used for calibration and validation. For each subperiod, the five preceding years (i.e., 1975–1979 and 1990–1994, respectively) were used to initialize the internal states of the models.

Discharge series (Q) were extracted at a daily time step from the French Hydro database (hydro.eaufrance.fr) for the 1980–2010 period (Leleu et al., 2014). A total of 1,305 stations was used. Only catchments that are not regulated according to this database were selected. Corresponding catchment areas vary between 14 km^2

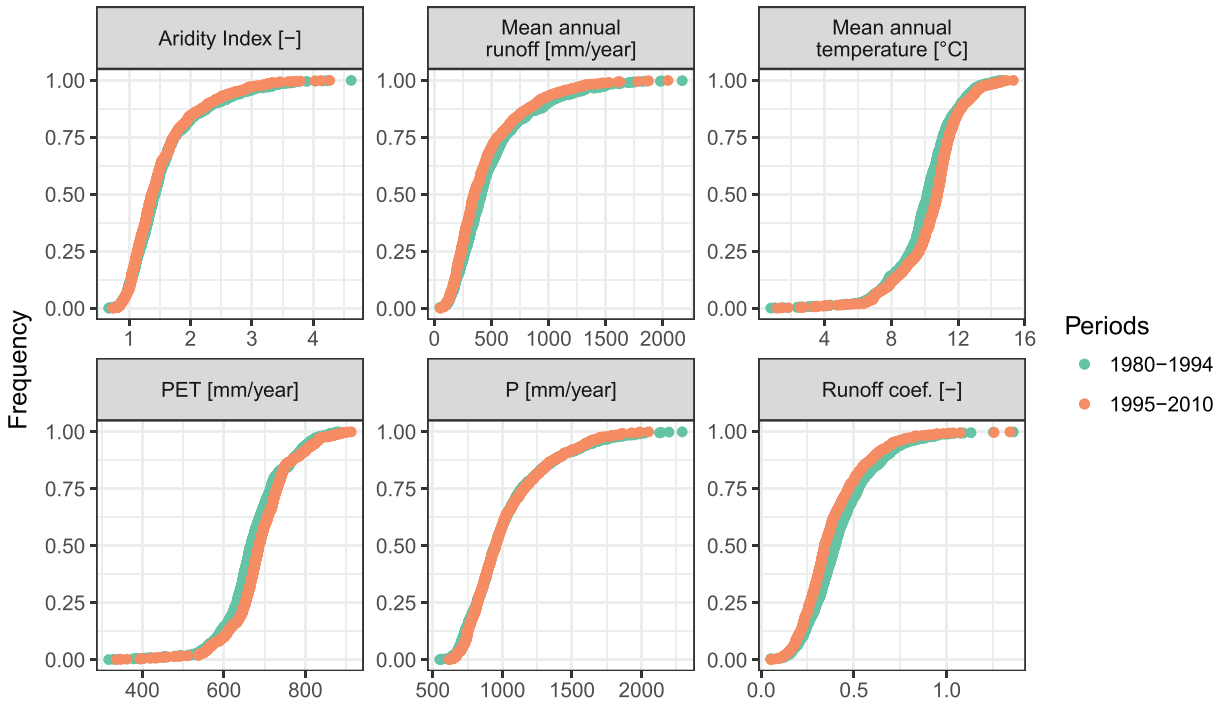

Figure 2. Distributions of several hydroclimatic descriptors on the 1,305 catchments for the two test periods.

Table 1
Parameters of the Lumped (GR5J) and Semidistributed (GRSD) Rainfall-Runoff Models With the CemaNeige Snow Module

Model parameter	GR5J	GRSD	Description	Units
$X1$	Free	Free	Production store maximum capacity	mm
$X2$	Free	Free	Intercatchment groundwater flow coefficient	mm/day
$X3$	Free	Free	Routing store maximum capacity	mm
$X4$	Free	Free	Time base of the unit hydrograph	day
$X5$	Free	Free	Threshold for intercatchment groundwater flow	—
C	—	Free	Average streamflow velocity	m/s
C_{TG}	Fixed	Fixed	Weighting coefficient of the snow thermal state	—
K_f	Fixed	Fixed	Degree day factor	$mm \cdot °C^{-1} \cdot day^{-1}$

to about 110,200 km^2 and the Strahler order (Strahler, 1964), from 1 to 8 (Figure 1). Only stations with less than 10% missing data were selected for the calibration-validation tests. This threshold was applied to each subperiod of the split sample test independently, that is, if it was exceeded for one subperiod, the catchment was then not considered in the tests.

Precipitation (P), potential evapotranspiration (PET), and temperature (T) data were extracted from the SAFRAN meteorological reanalysis produced by Météo-France (Vidal et al., 2010). It provides the spatialized climatic inputs needed by the models on an 8-km × 8-km square grid. The values of climatic variables at the scale of the modeling unit were obtained by overlaying the grid with the boundaries of each catchment. Figure 2 shows the distribution of a few hydroclimatic descriptors for the 1,305 catchments. It illustrates the wide hydroclimatic context explored in this study. Catchments showing the lowest temperature are mountainous catchments. Some difficulties in measuring climatic inputs and discharge output are observed for some catchments, particularly in karstic regions, when runoff coefficients are higher than one. The two periods are slightly different: the second period is warmer, with consequently higher PET values, which may explain the lower runoff coefficients.

2.2. GR5J and GRSD Daily Rainfall-Runoff Models

The GRSD semidistributed rainfall-runoff model was initially developed by Lobligeois et al. (2014). It is based on the GR5J lumped model (Figure 3a) proposed by Le Moine (2008), which has five free parameters to calibrate (Table 1). The main components of the model are two stores: a production store (maximum capacity $X1$ [mm]) and a routing store (maximum capacity $X3$ [mm]), which is filled by the output of a unit hydrograph (time base $X4$ [day]). Two other parameters, the groundwater exchange coefficient ($X2$ [mm/day]) and the threshold for groundwater exchange ($X5$ [—]), are used to quantify the intercatchment groundwater flows (IGFs). To account for snow accumulation and melt, the model is combined with the degree-day CemaNeige snow module (Valéry et al., 2014), which contains two additional parameters (C_{TG} [—] and K_f [$mm \cdot °C^{-1} \cdot day^{-1}$]). In this study, these parameters were not calibrated. They were fixed at their default values, respectively 0.2 and 4.5 $mm \cdot °C^{-1} \cdot day^{-1}$, as proposed by Valéry et al. (2014).

The semidistributed model GRSD discretizes the catchment into subcatchments that delineate the modeling units (Figures 3b and 3c). To capture the spatial variability of the meteorological inputs, a maximum size of the modeling unit is set at 250 km^2. The modeling units are delineated at all gauged points and then at as many ungauged points as needed to ensure that there is no modeling unit larger than the predetermined size. The lumped GR5J model is then applied on each modeling unit.

Each modeling unit receives its own meteorological inputs (rainfall P and PET) and uses a distinct parameter set (see section 3.1). Finally, the outflow of each modeling unit is routed to its downstream catchment using a linear lag propagation model (Bentura & Michel, 1997). Previous studies have shown that this propagation model gives a satisfactory level of efficiency compared to more sophisticated channel routing methods (Lobligeois et al., 2014). This routing functionality implies an additional free parameter (compared to the lumped model GR5J), which needs to be calibrated on each hydrological unit: the average flow velocity C (m/s).

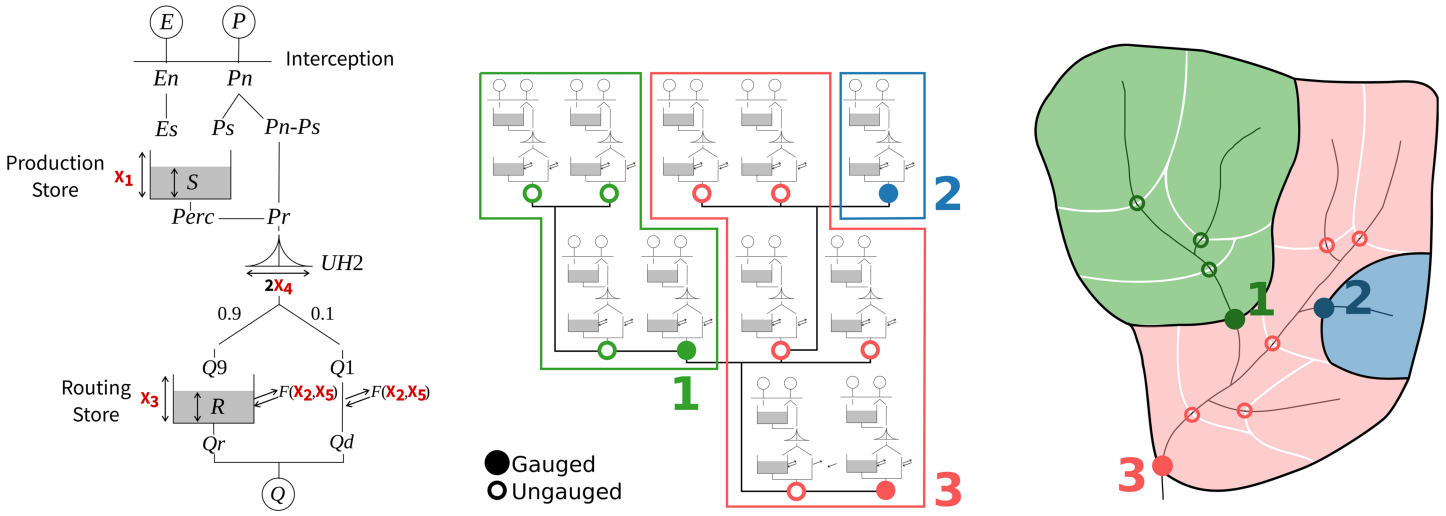

Figure 3. Schematic representation of (left) the GR5J lumped model structure and (middle) the GRSD semidistributed model for (right) the example catchment with three gauged stations (drainage points 1, 2, and 3). Headwater catchments (e.g., drainage Areas 1 and 2) and intermediary contributing areas (e.g., Drainage Area 3) are calibrated one after another, from upstream to downstream, by estimating the parameter sets of each of their modeling units.

Headwater catchments are the drainage area delineated by the first upstream gauged stations. They can be composed of several ungauged modeling units. The next downstream gauged stations delineate the intermediary contributing area (ICA) which corresponds to the additional drainage area that is not nested in any upstream gauged area (Figure 3). An ICA can also consist of several ungauged modeling units. These acronyms are summarized in Table 2.

The model is calibrated automatically by optimizing an objective function in two steps (see, e.g., Coron et al., 2017): (1) A screening is done by using three values for each parameters and by testing all their possible combinations; (2) a local optimization follows, starting from the parameter set that obtained the best results during the screening step. This optimization is done in a parameter space where each value is transformed to vary within the same range, which we later call the rescaled parameter space.

3. A Regularized Sequential Calibration

3.1. General Calibration Procedure

The proposed strategy is inspired by previous works on sequential calibration (Feyen et al., 2008; Lerat et al., 2012; Wi et al., 2015). The aim is not only to maximize performance at the outlet of a catchment but also to maximize performance at interior points through a multisite calibration: calibration is done at all gauged stations and is performed sequentially from upstream catchments to downstream catchments. This is done following four successive steps (Figure 4):

Step 1 Perform independent calibration of every HC with the same parameter set θ applied on each modeling unit (Catchments 1 and 2 in Figure 4).

Step 2 Estimate an a priori for each modeling unit in the next downstream ICA. Each modeling unit receives the parameter set of its closest upstream or neighboring catchments previously optimized during Step 1 (see section 3.2 for details).

Table 2
Definition of the Principal Acronyms

Acronyms	Definition
IGF	Intercatchment groundwater flow
ICA	Intermediary contributing area
HC	Headwater catchment

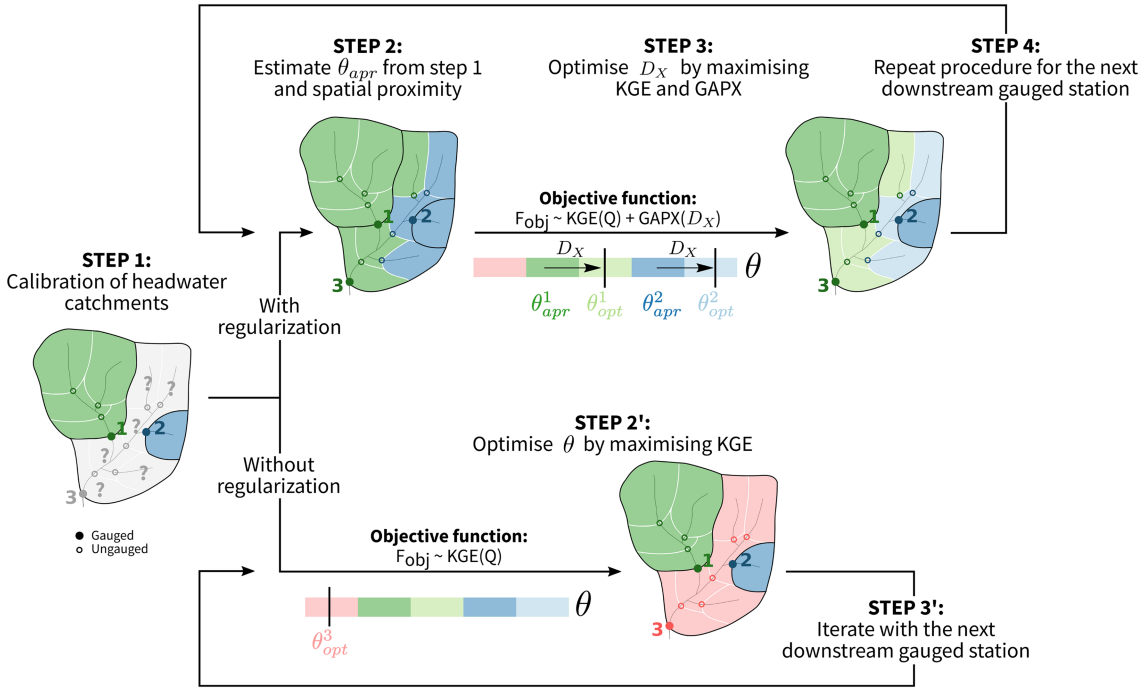

Figure 4. Calibration strategy to estimate the parameter set θ of each modeling unit within the intermediary contributing Area 3. The proposed regularization is compared to the original strategy without any regularization. The a priori parameter sets illustrated here are defined by spatial proximity (see section 3.2 for other a priori strategies).

Step 3 Calibrate a drift D_X from those a priori values using the objective function that maximize KGE at ICA's outlet and minimize this drift (according to *GAPX* criterion; see section 3.3).

Step 4 Repeat Steps 2 and 3 for the next downstream ICA and up to the last outlet. Optimized parameter sets of Step 3 can be used as an a priori for Step 2.

In contrast, a sequential calibration without regularization (Figure 4, Step 2') consists in applying a spatially homogeneous parameter set on each modeling unit in a way that it only aims at maximizing the simulation performance at the gauged points. Streamflow measurement is the only input of the calibration. Our regularization to this sequential calibration can be summarized by two major changes: (1) Calibration is driven by two sources of information (streamflow measurements and parameter's a priori values) instead of one (streamflow measurements), and (2) the spatial variability of parameter values within an ICA is assessed instead of applying one parameter set homogeneously over every modeling unit.

3.2. Defining an a Priori Parameter Set

The a priori parameter is our best guess on parameter values without using any streamflow measurement. This meets the topic of prediction in ungauged catchment. As discussed in section 1, spatial proximity between a gauged and an ungauged catchment has proved to be an efficient way of addressing this problem, at least in France and with this kind of conceptual and parsimonious model (Oudin et al., 2008).

For this reason, we estimate the a priori parameter from the previously optimized parameter set of the closest catchment. The rescaled Ghosh distance (de Lavenne et al., 2016) is used to compute the distance between catchments. Its ability to consider the nested structure of the catchments helps the identification of catchments that potentially have a similar hydrological behavior.

The definition of the a priori is flexible: Two other a priori strategies have also been tested. For the sake of simplicity, their description and their results will only be presented in the supporting information.

3.3. Objective Function for Calibration

The objective function (F_{obj}, equation (1)) is a weighted average (with a coefficient k) between (1) Kling-Gupta Efficiency (Gupta et al., 2009), noted *KGE* and computed on the root-squared discharges, and (2) the gap (noted GAPX, equation (2)) between the parameter set θ_{opt} obtained during the optimization process and its a priori value θ_{apr}. This gap is quantified as the Euclidean distance in the parameter space

Table 3
Drift Functions Used to Estimate a Parameter Set From Its a Priori Parameter Set by Applying D_X

Parameter	D_X	Units of D_X	Space
X_1	$D_X(1) = X_1 - X_{1,\text{apr}}$	mm	Rescaled
X_2	$D_X(2) = X_2 - X_{2,\text{apr}}$	mm/day	Rescaled
X_3	$D_X(3) = X_3 - X_{3,\text{apr}}$	mm	Rescaled
X_4	$D_X(4) = X_4^i / X_{4,\text{apr}}^i$ with $X_{4,\text{apr}}^i = X_{4,\text{apr}}(S/S_i)^{0.30}$	—	Real
X_5	$D_X(5) = X_5 - X_{5,\text{apr}}$	—	Rescaled
C	$D_X(6) = C_{\text{down}}/(C_{\text{up}} \cdot (d_{\text{up}} - d_{\text{down}}))$	m^{-1}	Real

Note. X and X_{apr} are, respectively, the parameter of the hydrological unit and its a priori value; X^i is the parameter at the lumped scale; S_i is the total drainage area; S is the area of the modeling unit; d_{up} and d_{down} are the hydraulic distances to the last outlet for upstream and downstream catchments, respectively (the two catchments are the one used as a priori and the one currently calibrated); C_{up} and C_{down} are their streamflow velocity.

(defined by n parameters and the range θ_{range} in which parameters are allowed to vary). It is weighted by the KGE performance provided by θ_{apr} on its donor catchment (with a lower limit set at 0). This weighting aims to reduce the risk of using an inefficient parameter set to constrain the optimization. We tested eight values of the coefficient k (0, 0.05, 0.10, 0.15, 0.25, 0.5, 0.75, and 0.9) to quantify the strength of the GAPX constraint.

$$F_{\text{obj}}(\theta_{\text{opt}}) = (1-k) \cdot KGE(\theta_{\text{opt}}) + k \cdot GAPX(\theta_{\text{opt}}) \tag{1}$$

$$GAPX(\theta_{\text{opt}}) = 1 - \sqrt{\sum_{i=1}^{n} \left(\frac{\theta_{\text{apr}}^i - \theta_{\text{opt}}^i}{\theta_{\text{range}}} \right)^2} \cdot \max(0, KGE(\theta_{\text{apr}})) \tag{2}$$

3.4. Estimate Parameter Values From Their a Priori Values

Each calibration iteration provides a new D_X vector. This vector is applied to the a priori parameter set in order to estimate the parameter values that will be used for the simulation. This drift D_X from the a priori is done differently depending on the parameter considered (Table 3). For four parameters ($X1, X2, X3$, and $X5$), it is done in a rescaled space (i.e., where all parameters vary within the same range) by applying a simple addition. For parameters $X4$ and C, since their values are expected to increase when moving from upstream to downstream as the drainage area gets bigger (de Lavenne et al., 2016; Lobligeois, 2014; Poncelet, 2016), the drift is constrained in order to respect this assumption. This is done using the catchment total drainage area (for $X4$) and the hydraulic distance d from the source (for C). Supporting information provides one example of the use of D_X vector within an ICA.

3.5. Use of Upstream Discharge Observations

In the context of simulation in ungauged catchments, the semidistributed model has an advantage over the lumped model: all upstream modeling units that are gauged do not necessarily have to be regionalized (because this may generate large errors), and upstream observations, when available, can be directly routed in order to estimate discharges downstream (Feyen et al., 2008; Lerat et al., 2013). Only the modeling units of the ICA would then have to be calibrated. In this study, we evaluated two simulation options:

1. Upstream *simulated* discharge is routed downstream (noted as "Rout. sim.")
2. Upstream *observed* discharge is routed downstream (noted as "Rout. obs."). In case of gaps in discharge observations, they are filled with their corresponding simulations.

It would also be possible to route discharge observations only during the calibration procedure (Step 3, section 3.1) and then route upstream simulations for the final simulation. This strategy leads to very similar simulation performance compared to always routing simulated discharges, so it will not be presented.

3.6. Leave-One-Out Evaluation

We evaluated our calibration strategy through a leave-one-out procedure. In this approach, every gauged catchment is successively considered ungauged. It means that we assume that no flow data are available for

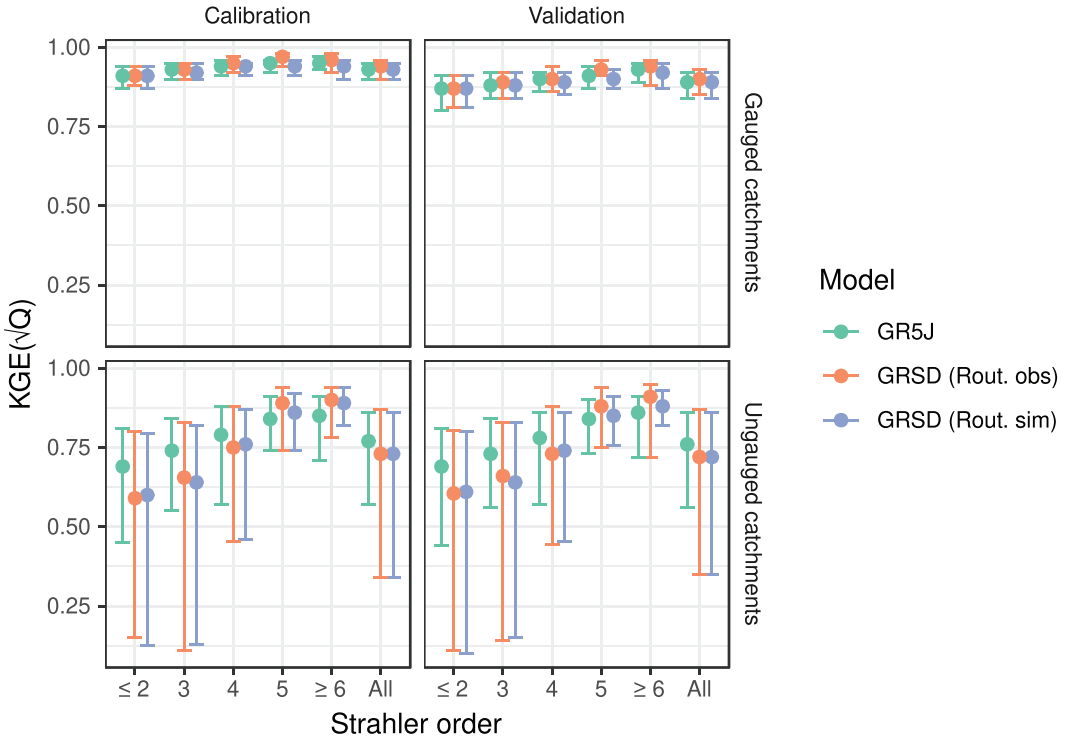

Figure 5. Distribution of KGE performance at gauged and pseudo-ungauged catchments computed on root-squared discharge using the calibration strategy without regularization. Dots are the median values, and bars are delimited by the first and third quartiles. KGE values are presented according to the Strahler order of the catchments.

model calibration. But for model evaluation, the flow simulations obtained at this target point are compared to the flow observations. To avoid confusion, we will use the terminology "pseudo-ungauged" catchment to describe this situation where measurements are used for model evaluation but not for model calibration.

Considering a given gauged catchment as ungauged modifies the configuration of the sequential calibration: the leave-one-out evaluation procedure requires a new downstream calibration to evaluate each gauged station. For instance, in Figure 4, if catchment 2 is considered ungauged, it will now belong to ICA 3. A new calibration of this different ICA 3 has then to be done. It will assign a parameter to catchment 2 as if it was ungauged.

As a reference benchmark, the lumped model GR5J is also regionalized to provide a simulation in pseudo-ungauged catchments. A spatial proximity-based approach is implemented to facilitate the comparison with GRSD: the parameter set of the closest gauged catchment according to the rescaled Ghosh distance (see section 3.2) is transferred and applied at those pseudo-ungauged locations.

4. Results

4.1. Performance of the Calibration Strategy Without Regularization

Sequential calibration of GRSD is first compared to GR5J without any regularization strategy. As a point of comparison, and before implementing regularization, three simulations are compared: (1) the GR5J lumped model, calibrated independently on each of the 1,305 catchments; (2) the GRSD model with routing of upstream observations; and (3) the GRSD model with routing of upstream simulations.

The calibration strategy without regularization provides very good and robust performance at gauged stations: the split sample test highlights a median KGE criterion around 0.95 in calibration and around 0.90 in validation for all model configurations (Figure 5). The semidistributed GRSD model does not outperform the lumped GR5J model. The benefits of using upstream observations (instead of upstream simulations) appear essentially on larger catchments (Strahler order above 5, Figure 5).

The performance at pseudo-ungauged catchments is much lower with the semidistributed model compared to the lumped model. Despite the use of upstream observations, GRSD does not outperform GR5J. This

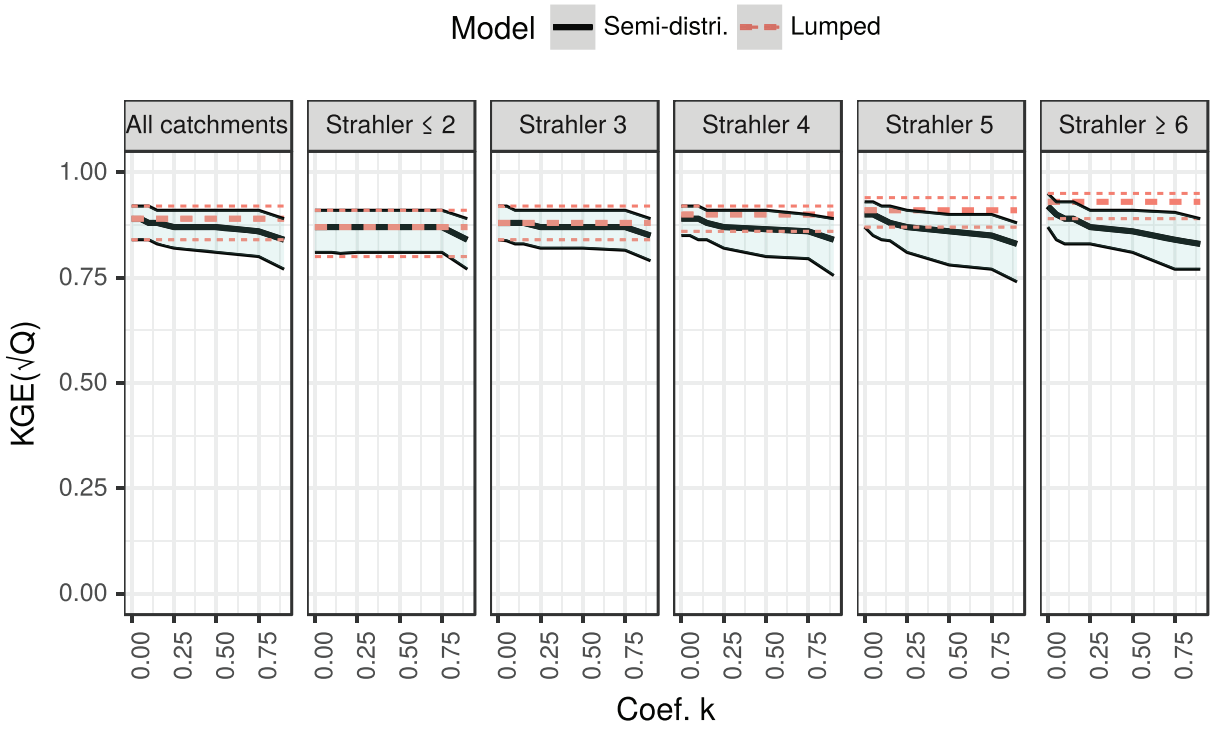

Figure 6. Distribution of performance during the validation periods for gauged catchments using different k regularization coefficients (equation (1)). Thick lines are the median values, and thin lines are the first and third quartiles of KGE values (solid lines are for GRSD and dashed lines for GR5J).

difference is particularly significant on small catchments. For larger catchments with Strahler order above 5, the semidistributed model performs slightly better. Lower performances on smaller catchments were expected and are usually observed (Reed et al., 2004).

These results demonstrate that the sequential calibration strategy without regularization is relatively efficient at gauged locations, but it may result in poor performance for ungauged points. This result actually motivated the calibration strategy proposed here.

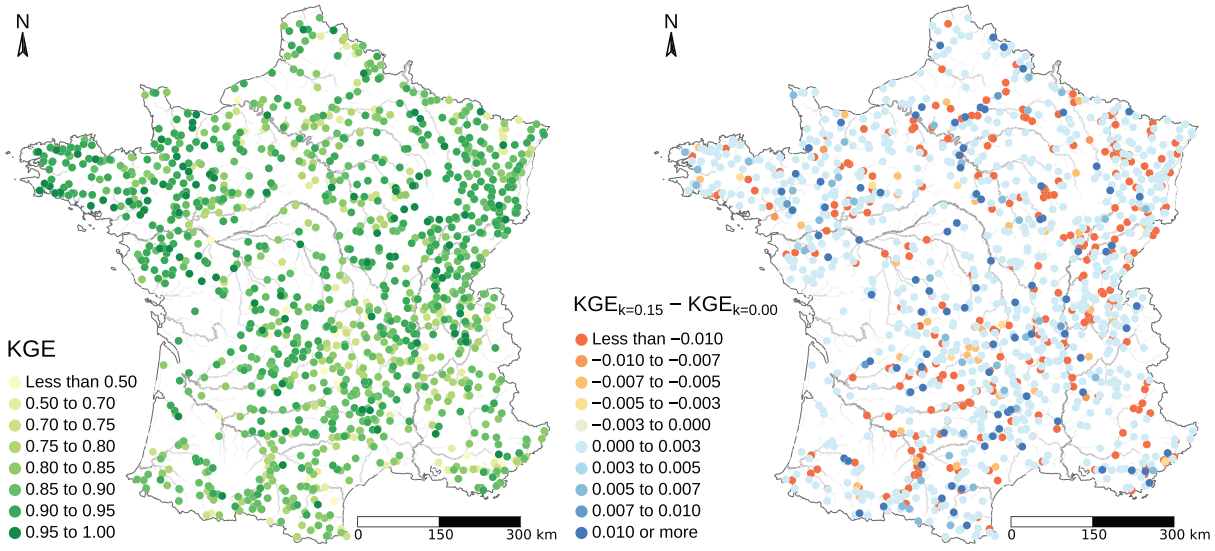

Figure 7. Performance in validation at gauged catchments using regularization ($k = 0.15$, left) and comparison with performance without regularization ($k = 0$, right). A Student's t test highlights no significant difference of performance during the validation periods.

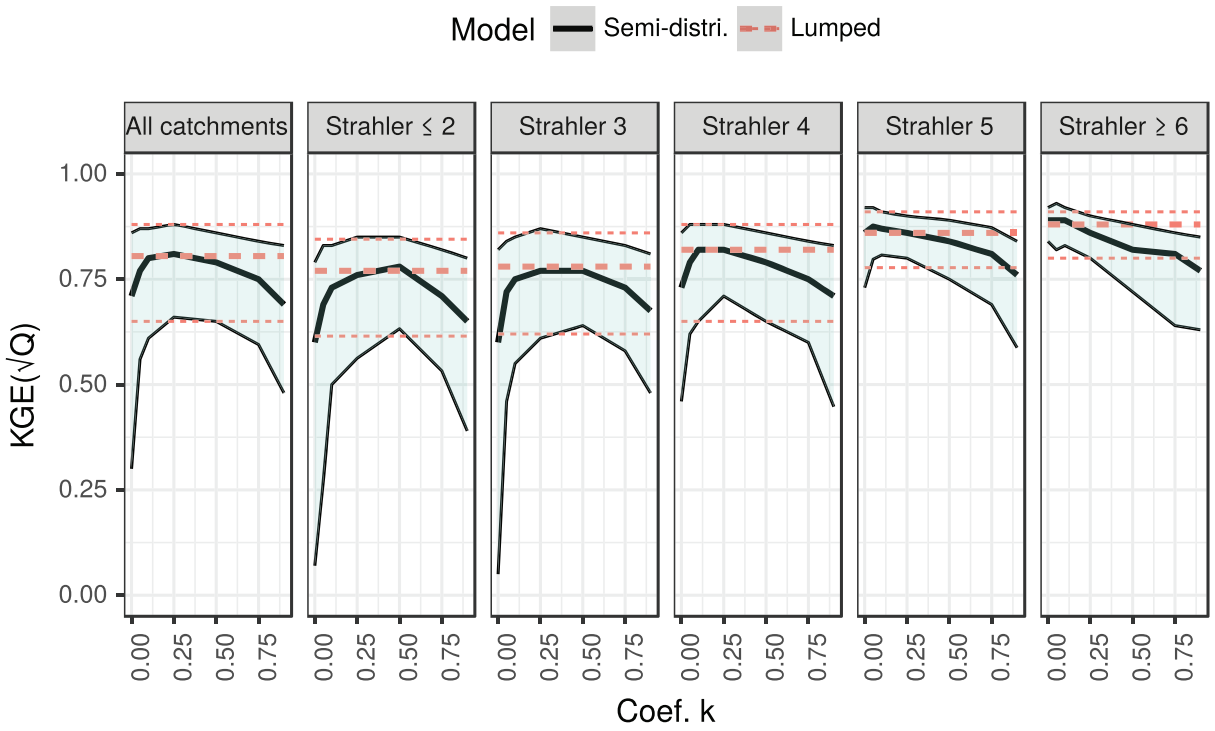

Figure 8. Distribution of performance over the entire period for pseudo-ungauged catchments using different k coefficients of the regularization (equation (1)). Thick lines are the median values, and thin lines are the first and third quartiles of KGE values (solid lines are for GRSD and dashed lines for GR5J).

4.2. Impact of Regularization on the Performance at Gauged Stations

Regularization adds a constraint to calibration of GRSD, and the magnitude of this constraint is governed by the coefficient k (equation (1)). It aims at improving the spatial consistency at the cost of a decrease in the weight given to the KGE value during its maximization. This is illustrated in Figure 6, where KGE values are presented for different values of the coefficient k when using the semidistributed model (GRSD). It is compared to the lumped model (GR5J), where no regularization is used. The loss of performance is limited on small catchments (Strahler order below 2) but is higher for larger catchments, where the median value

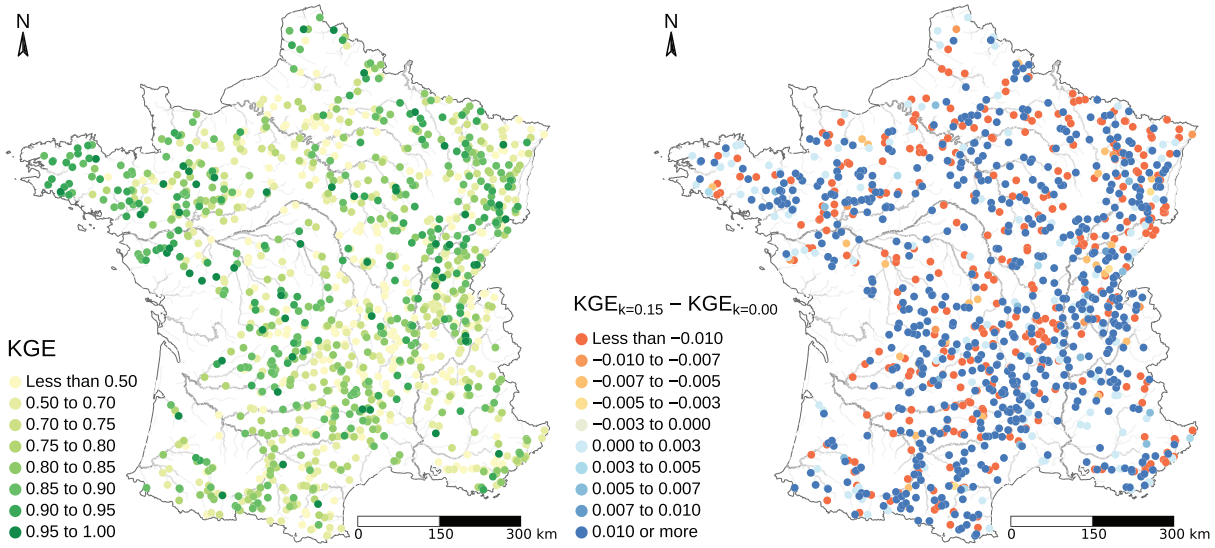

Figure 9. Performance of discharge simulation by the semidistributed model at points considered ungauged (leave-one-out evaluation) with regularization (k = 0.15, left) and comparison with performance without regularization (k = 0, right). At ungauged points, the regularization has a positive effect for 55% of the catchments, a negative effect for 27%, and no effect for 12%.

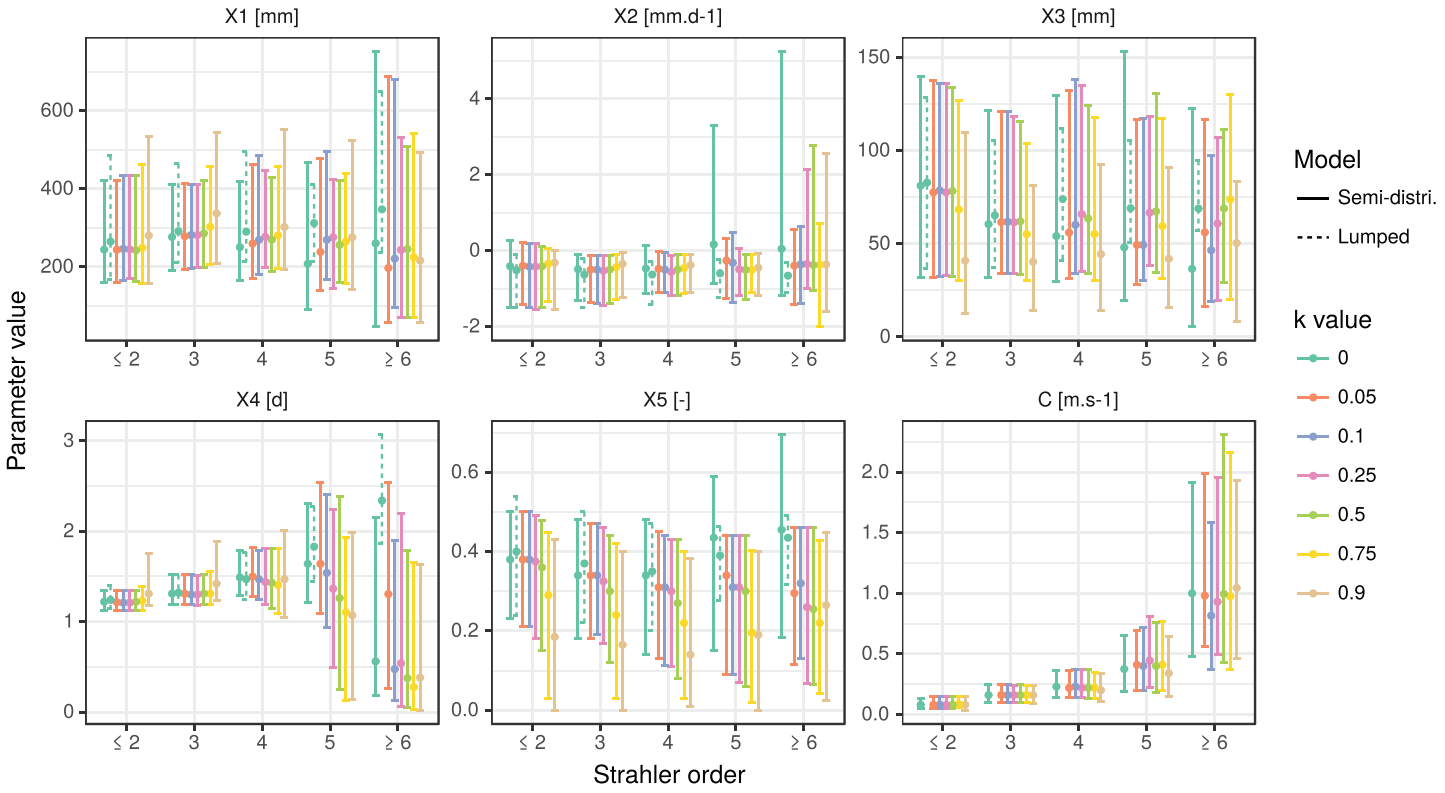

Figure 10. Distributions of the parameter values according to the coefficient k and the different Strahler orders of the catchments when using upstream discharge simulation.

decreases with k and interquartile ranges increase (Figure 6). The results show that the coefficient k should be as low as possible to keep the best possible performance at gauged stations.

Model performance is similar across the country without any particular region showing significantly lower performance (Figure 7). Even if performance tends to be higher without regularization, a Student's t test highlights no significant difference (p value = 0.6861) on the KGE at gauged stations when regularization is used. This is done by comparing the KGE when $k = 0$ to the KGE when $k = 0.15$ over the validation periods. Some increase and decrease of the KGE could, however, be observed at a small number of locations on the maps of Figure 7.

4.3. Impact of Regularization on the Performance at Ungauged Stations

While regularization has, on average, a negative impact at gauged stations (Figure 6), performance during the leave-one-out evaluation is generally improved by regularization (Figure 8). Median values of KGE, and more particularly the first quartile values, are greatly improved: The median KGE value increases from 0.75 to 0.82 (9% improvement) and the first quartile from 0.35 to 0.66 (88% improvement).

The effect of the constraint, quantified by an increased k coefficient, follows a hyperbole with an optimum. This maximum performance is reached with $k = 0.25$, on average, over all catchments, with a small variation with catchment size: the optimal k coefficient is larger for small catchments than for large catchments. This can be related to previous results on gauged catchments (section 4.2), where a high k coefficient also had a more negative effect on large catchments.

With the optimal k coefficient, performance at pseudo-ungauged catchments is as good as the regionalized lumped model. For Strahler order below 3, the lumped model slightly outperforms the semidistributed model, whereas it is slightly better for catchments with Strahler order above 4. For both models, catchment size is an important driver of performance, with large catchments showing, on average, higher performance than small catchments (Figure 8).

The spatial variability of KGE is much higher when stations are considered ungauged, compared to when they are considered gauged (Figure 9 in comparison to Figure 7): The standard deviation increases from

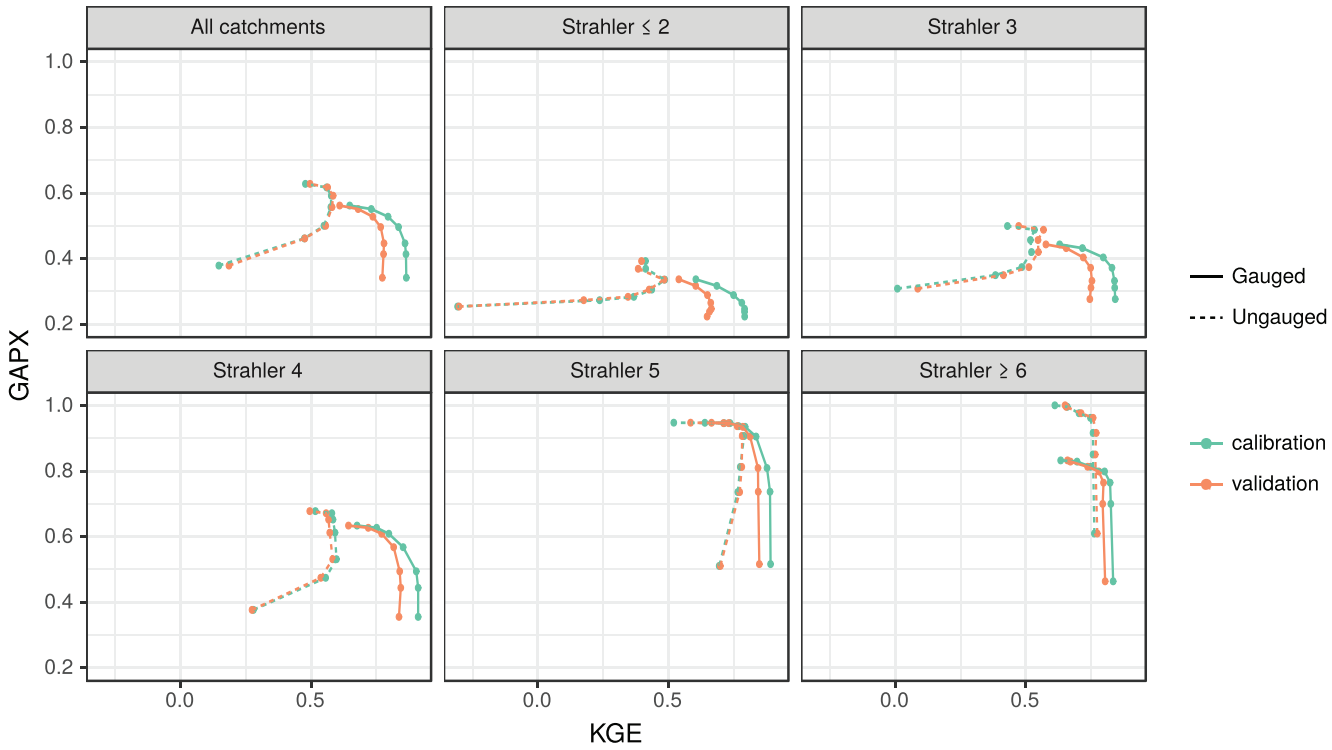

Figure 11. Values of the two components of the objective function, *GAPX* and *KGE* (see equation (1)), when routing upstream simulated discharges in the semidistributed model GRSD. The results are averaged over all catchments studied according to the Strahler order of the catchments and for different coefficients k (the 8 points correspond to coefficient k of 0, 0.10, 0.15, 0.25, 0.5, 0.75, and 0.9).

0.35 (during validation period) to 1.73 (during leave-one-out). No particular spatial organization could be identified.

A Student's t test highlights a significant improvement of the KGE at pseudo-ungauged points, when a regularization is used (p value = 0.00059). This is done by comparing the KGE when $k = 0$ to the KGE when $k = 0.15$. Even if the regularization generally has a positive impact on performance at pseudo-ungauged stations (55% of the catchments), the map in Figure 9 shows that regularization could also have negative effects on some catchments (27% of the catchments).

4.4. Impact of Regularization on Parameter Values

Figure 10 illustrates the impact of regularization on parameter values and their variability. It compares the lumped model and the semidistributed model according to the Strahler order of the catchments and to the k coefficient of the regularization (equation (1)). Regularization can be expected to have an effect on parameter values, especially for downstream modeling units where the parameter identifiability is lower (de Lavenne et al., 2016). The regularization is thus expected to drive the calibration, as explained in section 1. We verified this hypothesis for the $X2$ (intercatchment groundwater flow coefficient) and $X3$ (capacity of routing store) parameters: For Strahler orders above 6, the variability (quantified by the interquartile) is much lower as soon as one constraint is used.

Compared to the lumped model, the variability of model parameters across catchments is higher with the semidistributed model. Although parameters generally remain within the same range of values for both models, the variability of $X4$ (unit hydrograph time base) appears to be much larger. The coefficient k also tends to reduce $X4$ values. This is also verified on the velocity parameter C, where a clear increase from upstream to downstream is verified and where k tends to reduce velocity.

4.5. Objective Function Analysis

The objective function seeks a compromise between two criteria, GAPX and KGE (equation (1)). Figure 11 shows how these two criteria vary according to the coefficient k. The KGE performance on gauged catchments generally decreases slowly with increasing values of coefficient k, while performance on

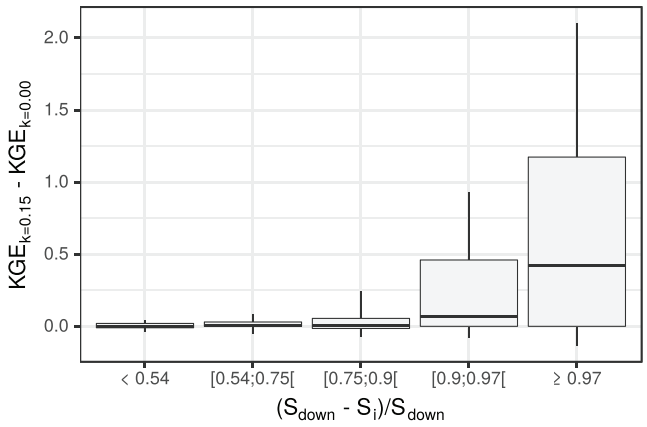

Figure 12. Performance improvement of discharge simulation brought by regularization (KGE difference when $k = 0.15$ and $k = 0$) at points considered ungauged (leave-one-out evaluation) and according to the difference of drainage area with the next downstream gauged catchment. S_i is the total drainage area where the performance is evaluated; S_{down} is the total drainage area of the further downstream gauged catchment. Classes on the x axes are defined by quantiles.

pseudo-ungauged catchments (by leave-one-out evaluation) increases rapidly. This is particularly true for small catchments; also, performance can slightly increase during validation. On average, over all catchments, the KGE performance at pseudo-ungauged catchments keeps improving for k values above 0.1 but at the cost of a small decrease in performance at gauged stations. The performance at gauged catchments decreases drastically for k coefficients above 0.5.

The performance at pseudo-ungauged stations is robust, with very similar performance during calibration and validation. The difference in performance between gauged and ungauged catchments also tends to decrease when catchment size increases. This means that the performance at ungauged large catchments should be very close to what a calibrated model would have provided.

The benefit of the regularization method decreases with catchment size and is always negative (no increase in performance) for Strahler orders above 6. It is likely that the discharge of the ICA for large catchments is very low compared to upstream contributions, so the effect of the calibration is difficult to observe.

For a given value of k, the GAPX criterion is often higher at pseudo-ungauged catchments than at gauged catchments. This means that during the leave-one-out evaluation, the optimized parameter set stays closer to its a priori. This might come from the fact that, during the leave-one-out evaluation, the ICA is larger, so the calibration applies the move given by D_X in parameter space (see section 3.4) to more modeling units (with potentially a different a priori). In this configuration, D_X has to remain low (and so GAPX high) in order to keep a consistent parameter set on every hydrological unit. The same reason could explain why the GAPX performance increases with catchment size.

5. Discussion

The comparison between the lumped and the semidistributed models used in this study does not highlight a clear steady benefit of the semidistributed model in terms of performance at gauged locations. These results corroborate the findings of several other studies showing that the benefits of semidistributed models are not straightforward (Hughes et al., 2014; Lobligeois et al., 2014; Pokhrel & Gupta, 2011). One possible reason concerns the spatial and temporal resolution of the model. The daily time step and the 250 km^2 modeling unit size might not be small enough to capture the spatiotemporal variability of the physical phenomena. Investigating smaller resolutions might be necessary (Ficchí et al., 2016; Obled et al., 2009).

Lumped and semidistributed models do not follow exactly the same objectives. A semidistributed modeling framework aims to build a spatially consistent model, where every gauged station has its performance optimized at the same time. The lumped model optimizes each catchment independently and does not address the spatiotemporal consistency of discharge simulations and observations (e.g., the impact of measurement errors on downstream simulations and local exchange with groundwater).

Small catchments benefit the most from the regularized sequential calibration strategy proposed here. Figure 12 provides insights to understand in which situations these small catchments benefit from the regularization. It appears that this improvement is particularly important when the drainage area of the next downstream gauged catchment is much larger than the drainage area of the ungauged catchment that needs to be calibrated. In a sequential calibration without regularization, every modeling units of the ICA are calibrated only from the downstream performance. Without regularization, this optimized parameter set does not fit well every modeling units of the ICA.

Small catchments are more directly concerned by two issues presented in section 1, which are (1) the regionalization strategy and (2) the downstream identifiability issue (de Lavenne et al., 2016). Concerning the first issue, the regionalization is improved by using the rescaled Ghosh distance to find an a priori parameter set around the neighboring gauged catchments, so small tributaries do not systematically retrieve the

optimized parameter set from large downstream gauged catchments (as it is the case without regularization). The second issue occurs most particularly when ICA located on lower reaches have to get extreme parameter values in order to have a little effect on the total discharge: Their contribution is, relatively to the upstream discharge volume, very small. Indeed, we observed a larger variability of parameter values at lower reaches (Figure 10). This issue is addressed in this paper by having an objective function that will always be sensitive to parameter changes: if calibration iterations do not improve simulation performance, the optimal parameter set will be the a priori parameter set. This prevents calibration from reaching inconsistent parameter sets that would be applied to every modeling unit of the ICA (including small ungauged modeling units that would be significantly impacted).

Starting from an a priori and progressively adjusting the parameter values by adding new information may be seen as a typical Bayesian process (Thiemann et al., 2001). However, this work is not based on a probability analysis or on any assumption on the statistical distribution of this information. Within this perspective, it would be interesting to investigate a more formal Bayesian approach, where parameter distributions would be transferred instead of a single parameter set.

One weakness of the regularized sequential calibration strategy is that the movement in the parameter space by using D_X does not explicitly account for parameter interactions. Indeed, D_X is applied similarly from every a priori parameter set during the calibration of the ICA. However, the sensitivity of a given parameter might depend on the value of the others, so D_X might affect each a priori differently. This issue is partly addressed by applying D_X, in the rescaled parameter space where parameters vary in similar ranges (see section 3.4). Moreover, a simple Euclidean distance was used here to compute the distance and move the a priori parameter set in the parameter space. Alternative distances (e.g., Manhattan distance or any goodness of fit criterion) could also be investigated.

Although we only presented the first option for the a priori parameter set (based on spatial proximity), we tested the two others (see supporting information). Between the three a priori options tested, the best is the parameter set of the closest catchment previously calibrated. Spatial proximity is an efficient proxy for identifying catchments with similar hydrological behavior (Oudin et al., 2008). Without defining and quantifying this similarity explicitly, it is based on implicit reasons that are based on a common territory and a common climate shared among catchments that are close to each other.

If the result of a previous calibration iteration is used as an a priori for the next one, it leads to a propagation of information in the sequential calibration. The sensitivity of how upstream parameter sets influence downstream parameter sets has not been addressed in this paper but, as well as the impact of upstream simulation (Hughes et al., 2014; Lerat et al., 2013), could be an interesting topic for future work. The two other a priori options are not affected by this issue because each catchment's a priori is independent of the others. Even if only one option for the a priori parameter set was used at a time in this study, one perspective for future work could be to use several a priori options together.

6. Conclusion

This paper revisited the generic sequential calibration of a semidistributed model in order to improve model performance at ungauged catchments. Additional information (a priori information) is used to constrain the calibration: The calibration procedure drifts away from this prior knowledge only if it is justified by observations. In practice, the procedure uses the following:

- an objective function that combines the maximization of a performance criterion and the minimization of the distance from the a priori parameter set;
- drift functions that describe how each parameter set moves away from its own a priori in the parameter space; and
- a priori information that can result from different regionalization strategies (regression from catchment characteristics or transfer of parameters based on spatial proximity).

The strategy was implemented over 1,305 catchments in France using the semidistributed GRSD model. The choice on the a priori parameters remains flexible (two other strategies are presented in the Supplementary Material). This paper focuses on a priori given by spatial proximity analysis where a priori parameter sets correspond to previous calibration of the closest catchment.

The paper first addressed the issue of the performance of a sequential calibration for ungauged catchments. The leave-one-out evaluation showed that, without any regularization, the performance at small pseudo-ungauged catchment is usually low. Despite relatively good performance at gauged stations, if the sequential calibration is not constrained, it leads to an unreliable calibration for many hydrological units.

Second, the paper focused on the benefits of the regularization to address this issue. We demonstrated that the regularized calibration strategy significantly improves the performance at pseudo-ungauged catchments (over all the catchments, the median KGE is increased from 0.75 to 0.83 and the first quartile from 0.35 to 0.66). More particularly, small pseudo-ungauged catchments were those that benefit most from this strategy. This improvement can be reached based on a compromise on the performance at gauged stations (especially for catchments with a Strahler order above 6, with a decrease from 0.91 to 0.87). However, this decrease is not significant over all the catchments.

Finally, the paper aimed at comparing the semidistributed model to a lumped model based on a similar conceptualization. Regularization enables to reach very similar performances between the lumped and the semidistributed model. This is observed at gauged and pseudo-ungauged locations. It highlights that the benefit of the semidistributed model is not straightforward, at least with this chosen spatiotemporal resolution for the semidistributed model (daily time step and average size of modeling units set to 250 km^2).

In summary, the calibration strategy presented in this paper is able to generally keep the advantages of a sequential calibration (performance optimization of interior gauged stations and reduction of the number of parameters to optimize at one time) while, additionally, providing more consistent simulations for the ungauged subcatchments. Therefore, the regularized sequential calibration provides a more robust spatial assessment of water resources throughout the catchment and, particularly, when simulating hydrological variables over large catchments, at a regional, countrywide, or continental scales.

Acknowledgments
The authors wish to thank the Associate Editor Juliane Mai, as well as Justin Hughes and two anonymous reviewers for their constructive comments which helped improve the overall quality of the article. The first author was partially funded by the French Waterboard "Agence de l'Eau Rhin-Meuse" within the MOSARH21 project (Project 15C92002). The support of the TOSCA/CNES program (SPAWET project) is also acknowledged. Météo-France and SCHAPI are acknowledged for making climatic and hydrological data available for this study. Discharge time series are available from the French Hydro database (hydro.eaufrance.fr).

References

Abdulla, F. A., & Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of Hydrology*, *197*(1), 230–257. https://doi.org/10.1016/S0022-1694(96)03262-3

Andersen, J., Refsgaard, J. C., & Jensen, K. H. (2001). Distributed hydrological modelling of the senegal river basin—Model construction and validation. *Journal of Hydrology*, *247*(3), 200–214. https://doi.org/10.1016/S0022-1694(01)00384-5

Antonetti, M., Scherrer, S., Kienzler, P. M., Margreth, M., & Zappa, M. (2017). Process-based hydrological modelling: The potential of a bottom-up approach for runoff predictions in ungauged catchments. *Hydrological Processes*, *31*(16), 2902–2920. https://doi.org/10.1002/hyp.11232

Antonetti, M., & Zappa, M. (2018). How can expert knowledge increase the realism of conceptual hydrological models? A case study based on the concept of dominant runoff process in the Swiss Pre-Alps. *Hydrology and Earth System Sciences*, *22*(8), 4425–4447.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2019). Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. *Hydrology and Earth System Sciences Discussions*, *2019*, 1–34. https://doi.org/10.5194/hess-2019-111

Bandaragoda, C., Tarboton, D. G., & Woods, R. (2004). Application of topnet in the distributed model intercomparison project. *Journal of Hydrology*, *298*(1), 178–201. https://doi.org/10.1016/j.jhydrol.2004.03.038

Bentura, P. L., & Michel, C. (1997). Flood routing in a wide channel with a quadratic lag-and-route method. *Hydrological Sciences Journal*, *42*(2), 169–189. https://doi.org/10.1080/02626669709492018

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. H. G. (2013). *Runoff prediction in ungauged basins. Synthesis across processes, places and scales*. Cambridge, UK: Cambridge University Press.

Boscarello, L., Ravazzani, G., & Mancini, M. (2013). Catchment multisite discharge measurements for hydrological model calibration. *Procedia Environmental Sciences*, *19*, 158–167. https://doi.org/10.1016/j.proenv.2013.06.018

Bulygina, N., McIntyre, N., & Wheater, H. (2009). Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis. *Hydrology and Earth System Sciences*, *13*(6), 893–904. https://doi.org/10.5194/hess-13-893-2009

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, *94*, 166–171. https://doi.org/10.1016/j.envsoft.2017.05.002

de Lavenne, A., Skøien, J. O., Cudennec, C., Curie, F., & Moatar, F. (2016). Transferring measured discharge time series: Large-scale comparison of top-kriging to geomorphology-based inverse modeling. *Water Resources Research*, *52*, 5555–5576. https://doi.org/10.1002/2016WR018716

de Lavenne, A., Thirel, G., Andréassian, V., Perrin, C., & Ramos, M.-H. (2016). Spatial variability of the parameters of a semi-distributed hydrological model. *Proceedings of the International Association of Hydrological Sciences*, *373*, 87–94. https://doi.org/10.5194/piahs-373-87-2016

Doherty, J., & Skahill, B. E. (2006). An advanced regularization methodology for use in watershed model calibration. *Journal of Hydrology*, *327*(3), 564–577. https://doi.org/10.1016/j.jhydrol.2005.11.058

Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems* (Vol. 375). Dordrecht: Kluwer Academic Publishers.

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, *47*, W11510. https://doi.org/10.1029/2010WR010174

Fenicia, F., Kavetski, D., Savenije, H. H. G., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, *52*, 954–989. https://doi.org/10.1002/2015WR017398

Feyen, L., Kalas, M., & Vrugt, J. A. (2008). Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization. *Hydrological Sciences Journal*, *53*(2), 293–308. https://doi.org/10.1623/hysj.53.2.293

Ficchí, A., Perrin, C., & Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *Journal of Hydrology*, *538*, 454–470. https://doi.org/10.1016/j.jhydrol.2016.04.016

Francés, F., Vélez, J. I., & Vélez, J. J. (2007). Split-parameter structure for the automatic calibration of distributed hydrological models. *Journal of Hydrology*, *332*(1), 226–240. https://doi.org/10.1016/j.jhydrol.2006.06.032

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & Savenije, H. H. G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, *18*(12), 4839–4859. https://doi.org/10.5194/hess-18-4839-2014

Götzinger, J., & Bárdossy, A. (2007). Comparison of four regionalisation methods for a distributed hydrological model. *Journal of Hydrology*, *333*(2), 374–384. https://doi.org/10.1016/j.jhydrol.2006.09.008

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., & Nearing, G. S. (2014). Debates—The future of hydrological sciences: A (common) path forward? using models and data to learn: A systems theoretic perspective on the future of hydrological science. *Water Resources Research*, *50*, 5351–5359. https://doi.org/10.1002/2013WR015096

He, Y., Bárdossy, A., & Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, *15*(11), 3539–3553. https://doi.org/10.5194/hess-15-3539-2011

Hughes, J. D., Dutta, D., Kim, S. S. H., Vaze, J., & Podger, G. M. (2014). The effect of stream network simplification and gauge error on river model performance. *Australasian Journal of Water Resources*, *18*(2), 151–161. https://doi.org/10.1080/13241583.2014.11465447

Hughes, J. D., Kim, S. S. H., Dutta, D., & Vaze, J. (2016). Optimization of a multiple gauge, regulated river-system model. a system approach. *Hydrological Processes*, *30*(12), 1955–1967. https://doi.org/10.1002/hyp.10752

Hundecha, Y., Arheimer, B., Donnelly, C., & Pechlivanidis, I. (2016). A regional parameter estimation scheme for a pan-European multi-basin model. *Journal of Hydrology: Regional Studies*, *6*, 90–111. https://doi.org/10.1016/j.ejrh.2016.04.002

Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1), 281–295. https://doi.org/10.1016/j.jhydrol.2004.01.002

Kelleher, C., McGlynn, B., & Wagener, T. (2017). Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding. *Hydrology and Earth System Sciences*, *21*(7), 3325–3352. https://doi.org/10.5194/hess-21-3325-2017

Khakbaz, B., Imam, B., Hsu, K., & Sorooshian, S. (2012). From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. *Journal of Hydrology*, *418-419*, 61–77. https://doi.org/10.1016/j.jhydrol.2009.02.021

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13–24.

Koren, V., Moreda, F., & Smith, M. (2008). Use of soil moisture observations to improve parameter consistency in watershed calibration. *Physics and Chemistry of the Earth, Parts A/B/C*, *33*(17), 1068–1080. https://doi.org/10.1016/j.pce.2008.01.003

Koren, V., Smith, M., & Duan, Q. (2003). Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. In Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau & R. Turcotte (Eds.), *Calibration of Watershed Models*. https://doi.org/10.1002/9781118665671.ch18

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*, 360–379. https://doi.org/10.1029/2012WR012195

Le Moine, N. (2008). The topographic catchment seen from the groundwater body: A way to improve the performances and realism of rainfall-runoff models? (Doctoral dissertation), UPMC, Paris, France. https://irsteadoc.irstea.fr/cemoa/PUB00025628 (348 p.)

Leavesley, G. H., Lichty, R. W., Troutman, B. M., & Saindon, L. G. (1983). Precipitation-runoff modeling system: User's manual. *Water-resources investigations report*, *83*, 4238.

Leleu, I., Tonnelier, I., Puechberty, R., Gouin, P., Viquendi, I., Cobos, L., et al. (2014). Re-founding the national information system designed to manage and give access to hydrometric data. *La Houille Blanche*, *1*, 25–32. (in French) https://doi.org/10.1051/lhb/2014004

Lerat, J., Andréassian, V., Perrin, C., Vaze, J., Perraud, J. M., Ribstein, P., & Loumagne, C. (2012). Do internal flow measurements improve the calibration of rainfall-runoff models?. *Water Resources Research*, *48*, W02511. https://doi.org/10.1029/2010WR010179

Lerat, J., Dutta, D., Kim, S., Hughes, J., & Vaze, J. (2013). Reducing propagation of uncertainty in river system modelling by optimal use of streamflow data. In *Proceedings of the 20th international congress on modelling and simulation*, Adelaide, Australia, pp. 1–6.

Li, Y., Grimaldi, S., Pauwels, V. R. N., & Walker, J. P. (2018). Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *Journal of Hydrology*, *557*, 897–909. https://doi.org/10.1016/j.jhydrol.2018.01.013

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, *201*(1), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

Lobligeois, F. (2014). Mieux connaître la distribution spatiale des pluies améliore-t-il la modélisation des crues? Diagnostic sur 181 bassins versants français (https://pastel.archives-ouvertes.fr/tel-01134990), AgroParisTech, Paris, France. 310 p.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrology and Earth System Sciences*, *18*(2), 575–594. https://doi.org/10.5194/hess-18-575-2014

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, *51*, 716–728. https://doi.org/10.1002/2014WR015820

Moussa, R., Chahinian, N., & Bocquillon, C. (2007). Distributed hydrological modelling of a mediterranean mountainous catchment—Model construction and multi-site validation. *Journal of Hydrology*, *337*(1-2), 35–51. https://doi.org/10.1016/j.jhydrol.2007.01.028

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., et al. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, *54*, 8332–8362. https://doi.org/10.1029/2017WR021895

Obled, C., Zin, I., & et Benoît Hingray (2009). Optimal space and time scales for parsimonious rainfall-runoff models. *La Houille Blanche*, *5*, 81–87.

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

Parajka, J., Merz, R., & Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, *9*(3), 157–171. https://doi.org/10.5194/hess-9-157-2005

Piniewski, M., Szcześniak, M., Kardel, I., Berezowski, T., Okruszko, T., Srinivasan, R., et al. (2017). Hydrological modelling of the Vistula and Odra river basins using SWAT. *Hydrological Sciences Journal*, *62*(8), 1266–1289. https://doi.org/10.1080/02626667.2017.1321842

Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*, *46*, W01505. https://doi.org/10.1029/2009WR008066

Pokhrel, P., & Gupta, H. V. (2011). On the ability to infer spatial catchment variability using streamflow hydrographs. *Water Resources Research*, *47*, W08534. https://doi.org/10.1029/2010WR009873

Pokhrel, P., Gupta, H. V., & Wagener, T. (2008). A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research*, *44*, W12419. https://doi.org/10.1029/2007WR006615

Pokhrel, P., Yilmaz, K. K., & Gupta, H. V. (2012). Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *Journal of Hydrology*, *418-419*, 49–60. https://doi.org/10.1016/j.jhydrol.2008.12.004

Poncelet (2016). Du bassin au paramétre : jusqu'oú peut-on régionaliser un modéle hydrologique conceptuel ? (Doctoral dissertation), UPMC, Paris, France. Retrieved from https://tel.archives-ouvertes.fr/tel-01529196 (366 p.)

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., & DMIP Participants (2004). Overall distributed model intercomparison project results. *Journal of Hydrology*, *298*(1-4), 27–60. (The Distributed Model Intercomparison Project (DMIP)) https://doi.org/10.1016/j.jhydrol.2004.03.031

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. https://doi.org/10.1029/2008WR007327

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., et al. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, *21*(9), 4323–4346. https://doi.org/10.5194/hess-21-4323-2017

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., et al. (2003). IAHS decade on predictions in ungauged basins (pub), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, *48*(6), 857–880. https://doi.org/10.1623/hysj.48.6.857.51421

Strahler, A. N. (1964). Quantitative geomorphology of drainage basins and channel networks. In V. T. Chow (Ed.), *Handbook of applied hydrology* (pp. 4–39/4–76). New York: McGraw-Hill.

Thiemann, M., Trosset, M., Gupta, H., & Sorooshian, S. (2001). Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research*, *37*(10), 2521–2535. https://doi.org/10.1029/2000WR900405

Tonkin, M. J., & Doherty, J. (2005). A hybrid regularized inversion methodology for highly parameterized environmental models. *Water Resources Research*, *41*, W10412. https://doi.org/10.1029/2005WR003995

Valéry, A., Andréassian, V., & Perrin, C. (2014). 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2—Sensitivity analysis of the cemaneige snow accounting routine on 380 catchments. *Journal of Hydrology*, *517*, 1176–1187.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., & Soubeyroux, J.-M. (2010). A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology*, *30*(11), 1627–1644. https://doi.org/10.1002/joc.2003

Wang, S., Zhang, Z., Sun, G., Strauss, P., Guo, J., Tang, Y., & Yao, A. (2012). Multi-site calibration, validation, and sensitivity analysis of the MIKE SHE model for a large watershed in Northern China. *Hydrology and Earth System Sciences*, *16*(12), 4621–4632. https://doi.org/10.5194/hess-16-4621-2012

Wi, S., Yang, Y. C. E., Steinschneider, S., Khalil, A., & Brown, C. M. (2015). Calibration approaches for distributed hydrologic models in poorly gaged basins: Implication for streamflow projections under climate change. *Hydrology and Earth System Sciences*, *19*(2), 857–876. https://doi.org/10.5194/hess-19-857-2015

Zehe, E., Maurer, T., Ihringer, J., & Plate, E. (2001). Modeling water flow and mass transport in a loess catchment. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, *26*(7), 487–507. https://doi.org/10.1016/S1464-1909(01)00041-7