Multi-gauge Hydrological Variational Data Assimilation: Regionalization Learning with Spatial Gradients using Multilayer Perceptron and Bayesian-Guided Multivariate Regression

Ngo Nghi Truyen Huynh, Pierre-André Garambois, François Colleoni, Benjamin Renard, Hélène Roux

Multi-gauge Hydrological Variational Data Assimilation: Regionalization Learning with Spatial Gradients using Multilayer Perceptron and Bayesian-Guided Multivariate Regression

Ngo Nghi Truyen Huynh[1], Pierre-André Garambois[1,*], François Colleoni[1], Benjamin Renard[1], Hélène Roux[2]

[1]*INRAE, Aix-Marseille Université, RECOVER, 3275 Route Cézanne, 13182 Aix-en-Provence, France*
[2]*Institut de Mécanique des Fluides de Toulouse (IMFT), Université de Toulouse, CNRS, 31400 Toulouse, France*

*corresponding author: `pierre-andre.Garambois@inrae.fr`

Abstract—**Tackling the difficult problem of estimating spatially distributed hydrological parameters, especially for floods on ungauged watercourses, this contribution presents a novel seamless regionalization technique for learning complex regional transfer functions designed for high-resolution hydrological models. The transfer functions rely on: (i) a multilayer perceptron enabling a seamless flow of gradient computation to employ machine learning optimization algorithms, or (ii) a multivariate regression mapping optimized by variational data assimilation algorithms and guided by Bayesian estimation, addressing the equifinality issue of feasible solutions. The approach involves incorporating the inferable regionalization mappings into a differentiable hydrological model and optimizing a cost function computed on multi-gauge data with accurate adjoint-based spatially distributed gradients.**

Keywords—**Variational Data Assimilation, Distributed Hydrological Modeling, Artificial Neural Networks, Bayesian Estimation, Hydrological Regionalization**

I. INTRODUCTION

Regardless of the improvements made in hydrological forward models and available data, hydrological calibration remains a challenging ill-posed inverse problem faced with the equifinality (Beven, 2001) of feasible solutions. Most calibration approaches aim to estimate spatially uniform model parameters for a single gauged catchment, resulting in piecewise constant discontinuous parameters fields for adjacent catchments. Moreover, these calibrated parameter are not transferable to ungauged locations, which represents the majority of the global land surface (Fekete & Vörösmarty, 2007; Hannah et al., 2011). Therefore, prediction in ungauged basins remains a key challenge in hydrology (Hrachowitz et al., 2013).

Regionalization approaches are employed to estimate hydrological model parameters in ungauged locations by transferring hydrological information from gauged locations. In early studies, the predominant method for regionalization involved individually calibrating catchments and then using multiple regression or interpolation techniques to transfer the calibrated parameter sets from gauged to ungauged locations (Abdulla & Lettenmaier, 1997; Seibert, 1999; Parajka et al., 2005; Razavi & Coulibaly, 2013; Parajka et al., 2013). This process can be referred to as post-regionalization (Samaniego et al., 2010). However, post-regionalization approaches are limited to lumped parameters by catchment, thus ignoring within-catchment variabilities (Samaniego et al., 2010; Razavi & Coulibaly, 2013). Furthermore, they are generally faced with the issue of equifinal parameter sets and hence equifinal estimated transfer laws, while spatial proximity is more adapted to densely gauged river networks and regions (Oudin et al., 2008; Reichl et al., 2009). A simultaneous regionalization approach, which involves optimizing a mapping between physical descriptors and model parameters (cf. Parajka et al. (2005); Götzinger & Bárdossy (2007)), is able to overcome most of the aforementioned problems and can be referred as "pre-regionalization". Typically, a Multiscale Parameter Regionalization (MPR) method, combining descriptors upscaling and pre-regionalization function in form of multilinear regressions, implemented within a spatially distributed multiscale hydrological model (mHm), has been proposed by Samaniego et al. (2010), and later applied to other gridded hydrological models in several applicative studies (e.g., Mizukami et al. (2017); Beck et al. (2020)). In all the above studies, state of the art optimization algorithms are used, especially Shuffle Complex Evolution algorithm (SCE) (Duan et al., 1992) in Mizukami et al. (2017) or Distributed Evolutionary Algorithms (DEAP) (Fortin et al., 2012) in Beck et al. (2020). Nevertheless, those optimization algorithms are limited to low-dimensional controls, which imposes the use of a limited number of descriptors in lumped multivariate pre-regionalization mappings, and thus restricts the capability to fully exploit the large amount of information available from multiple data sources with flexible formulations and adequate spatial rigidity.

In Huynh et al. (2023), efficient pre-regionalization algorithms have been proposed for spatially distributed hy-

drological modeling based on descriptors-to-parameters mappings with neural networks or multivariate regressions in a variational data assimilation framework. Despite the strong spatial constrain and regularizing effect introduced via pre-regionalization mappings, some sensitivity to prior remains in context of equifinality (model structural equifinality plus spatial equifinality) and its inference is explored here using the Bayesian weighting approach proposed in Chelil et al. (2022); Gejadze et al. (2022).

In this work, we present a novel seamless regionalization method for learning the pre-regionalization mapping between physical data and conceptual parameters of spatially distributed hydrological models using information from multi-gauge river flow observations and high-resolution physical descriptors. We explore two approaches to infer the pre-regionalization mapping:

- Bayesian-Guided Multivariate Regional Regression (BGM2R): a multivariate polynomial regression approach, which combines high-dimensional optimization algorithms guided by a Bayesian estimation on the first guess;
- Artificial Neural Network Regionalization (ANNR) enabling a "seamless flow of gradient computation" and employing machine learning optimizers.

The proposed algorithms are implemented in the SMASH platform (see online documentation and tutorials at https://smash.recover.inrae.fr) available on public GitHub (https://github.com/DassHydro-dev/smash).

II. METHODOLOGY

The full forward model and the optimization process are schematized in Figure 1.

A. Forward Model and Cost Function

Let us consider observed discharge time series $Q_g^*(t)$ at N_G observation cells of coordinates $x_g \in \Omega$, $g \in 1, .., N_G$ with $N_G \geq 1$. For each observation cell, the corresponding gauged upstream sub-catchment is denoted Ω_g so that $\Omega_{ung} = \Omega \setminus \left(\cup_{g=1}^{N_G} \Omega_g \right)$ is the remaining ungauged part of the whole spatial domain Ω. Then, the rainfall and potential evapotranspiration fields are respectively denoted as $\boldsymbol{P}(x,t)$ and $\boldsymbol{E}(x,t)$, $\forall x \in \Omega$. The classical forward model $\mathcal{M}_{rr}$ is a dynamic operator projecting the input fields $\boldsymbol{P}(x,t)$ and $\boldsymbol{E}(x,t)$, given an input drainage plan $\mathcal{D}_\Omega(x)$, onto the discharge field $Q(x,t)$ and states fields $\boldsymbol{h}(x,t)$ written as a multivariate function:

$$(\boldsymbol{h}, Q)(x,t) = \mathcal{M}_{rr}[\mathcal{D}_\Omega(x), \boldsymbol{P}(x,t'), \boldsymbol{E}(x,t'),$$
$$\boldsymbol{h}(x,0), \boldsymbol{\theta}(x), t], \forall(x,t') \in \Omega \times [0,t]$$
(1)

where $\boldsymbol{\theta}$ is the N_θ-dimensional vector of model parameters 2D fields that we aim to estimate regionally with the new algorithms proposed below, and $\boldsymbol{h}$ is the N_S-dimensional vector of internal model states. In this study, the distributed hydrological model $\mathcal{M}_{rr}$ is a parsimonious GR-like conceptual structure with the parameters vector $\boldsymbol{\theta}(x) =$

$(c_p(x), c_{ft}(x), k_{exc}(x), l_r(x))^T$, $\forall x \in \Omega$, which is the "gr-b" structure presented in Colleoni et al. (2023).

Now, the full forward model $\mathcal{M}$ is composed of the distributed hydrological model $\mathcal{M}_{rr}$ on top of which is applied a pre-regionalization operator $\mathcal{F}_R$ to estimate hydrological parameters $\boldsymbol{\theta}$ such that:

$$\mathcal{M} = \mathcal{M}_{rr}\left[\, . \, , \, \boldsymbol{\theta}(x) = \mathcal{F}_R(\boldsymbol{D}(x), \boldsymbol{\rho})\right], \forall x \in \Omega \quad (2)$$

This allows to constrain spatially and explain these spatial fields of conceptual model parameters $\boldsymbol{\theta}(x)$ from physical descriptors $\boldsymbol{D}(x)$. The pre-regionalization operator $\mathcal{F}_R$ being a descriptor-to-parameters mapping, with $\boldsymbol{D}$ the N_D-dimensional vector of physical descriptor maps covering Ω, and $\boldsymbol{\rho}$ the vector of tunable regionalization parameters that will be defined later.

A calibration cost function is defined in order to measure the misfit between simulated and observed discharge time series, respectively denoted $Q_g(t)$ and $Q_g^*(t)$, for $g \in 1..N_G$ gauged cells. A convex differentiable objective function is classically defined as follows: $J = J_{obs} + \gamma J_{reg}$, with J_{obs} the observation term that measures the difference between observed and simulated quantities and J_{reg} a regularization term weighted by $\gamma > 0$. The observation term is $J_{obs} = \sum_{g=1}^{N_G} w_g J_g^*$ with w_g a physical weighting function, J_g^* a local quadratic metric "at the station" (e.g., $1 - NSE$) involving the response of the direct model. Thus, J_{obs} depends on the control vector $\boldsymbol{\rho}$ through the direct model $\mathcal{M}$. The multi-site calibration corresponds to $N_G > 1$ while $N_G = 1$ is a classical calibration on a single station where $w_1 = 1$. For $N_G > 1$, several physical weighting expressions w_g are possible with the single constraint that $\sum_{g=1}^{N_G} w_g = 1$. In this work, we simply use use $w_g = \frac{1}{N_G}$ for multiple gauges calibration. The regional optimization problem writes as follows:

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} J(\boldsymbol{\rho}) \quad (3)$$

B. Regional Calibration with BGMR

In this case, the pre-regionalization mapping $\mathcal{F}_\mathcal{R} \equiv \mathcal{P}$ with tunable parameter $\boldsymbol{\rho}$ consists in a multivariate polynomial regression between input physical descriptors $\boldsymbol{D}(x)$ and hydrological model parameters $\boldsymbol{\theta}(x, \boldsymbol{D}, \boldsymbol{\rho}) \coloneqq \mathcal{P}(\boldsymbol{D}(x), \boldsymbol{\rho})$ such that:

$$\theta_k(x, \boldsymbol{D}, \rho_k) \coloneqq s_k\left(\alpha_{k,0} + \sum_{d=1}^{N_D} \alpha_{k,d} D_d^{\beta_{k,d}}(x)\right),$$
$$\forall k \in [1..N_\theta], \forall x \in \Omega$$
(4)

with $s_k(.)$ a Sigmoid-based transformation imposing bound constraints in the direct hydrological model. The lower and upper bounds are assumed to be spatially uniform for each parameter field θ_k of the hydrological model. The optimization of the control vector $\boldsymbol{\rho} \equiv \left[(\rho_k)_{k=1}^{N_\theta}\right]^T \equiv \left[\left(\alpha_{k,0}, (\alpha_{k,d})_{d=1}^{N_D}\right)_{k=1}^{N_\theta}\right]^T$, that is solving problem 3, is performed using the L-BFGS-B algorithm (Zhu et al., 1997),
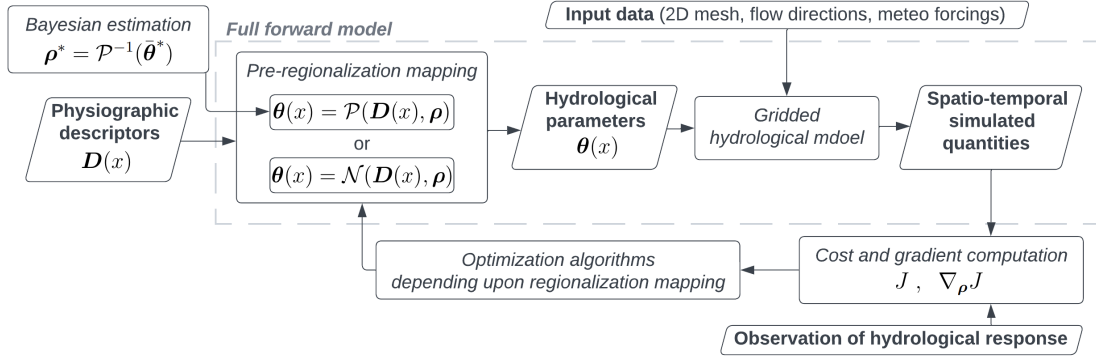
Fig. 1. Flowchart of inverse algorithms for the full forward model consisting of a GR-based gridded hydrological model (spatio-temporal regular grid at $1\ \mathrm{km}^2$ and 1 h) and a pre-regionalization mapping.

adapted to high-dimensional controls, without bound constraints on the $\alpha_{k,.}$, whereas the exponents $\beta_{k,d}$ is simply fixed to 1 (multi-linear pre-regionalization). This algorithm requires the gradient of the cost function with respect to the sought parameters $\nabla_{\boldsymbol{\rho}}J$. This gradient is computed by solving the adjoint model, which is obtained by automatic differentiation using the Tapenade engine (Hascoet & Pascual, 2013). The entire process is implemented in the SMASH Fortran source code, where the full forward model $\mathcal{M} \equiv \mathcal{M}_{rr}\left(.,\mathcal{P}\left(.\right)\right)$ is a composition of both the hydrological model and the polynomial descriptors-to-parameters mapping. The convergence criterion involves reaching a maximum iteration limit or meeting conditions related to cost function change or gradient magnitude.

It is worth noting that determining a background value $\boldsymbol{\rho}^*$ is important for the convergence of this algorithm. It is used as a starting point for the optimization, and is defined from a spatially uniform prior $\bar{\boldsymbol{\theta}}^*$ as $\boldsymbol{\rho}^* \equiv \left[\alpha_{k,0} = s_k^{-1}\left(\bar{\theta}_k^{\,*}\right), \left(\alpha_{k,d} = 0, \beta_{k,d} = 1\right)\right]^T, \forall (k, d) \in [1..N_\theta] \times [1..N_D]$, where $s_k^{-1}(z) = \ln\left(\frac{z - l_k}{u_k - z}\right)$ is the inverse Sigmoid. The spatially uniform low-dimensional (LD) prior $\bar{\boldsymbol{\theta}}^*$ is determined considering the cost function without pre-regionalization, i.e., $\mathcal{M} \equiv \mathcal{M}_{rr}$ and $\boldsymbol{\rho} := \bar{\boldsymbol{\theta}}$, and classically using a global optimization algorithm (SBS in Michel (1989)).

A Bayesian-like estimator (cf. Gejadze et al. (2022); Chelil et al. (2022)) is used to look at the mean of the posterior distribution $f\left(\boldsymbol{\theta}|\boldsymbol{Q}^*\right)$ that is more stable in context of equifinality than searching its mode (inverse problem 3, maximum a posteriori probability (MAP) search is the essence of variational data assimilation). A prior probability distribution $f_{\bar{\theta}}$ is used to generate a sample of spatially uniform parameter sets $\boldsymbol{\theta}_i, \forall i \in 1..N$ within the hypercube defined by parameters bounds $[l_k, u_k], \forall k \in 1..N_\theta$. The likelihood function is defined as: $\mathcal{L}_i^\alpha = e^{-2^\alpha\left(J_i/J_{\min}-1\right)^2}$, where α is a parameter controlling the decay rate of this function that compares the value of $J_i = J\left(\boldsymbol{\theta}_i\right)$ to J_{min} the minimum value of J_i over the sample of N parameter sets. The posterior ensemble mean

and variance are computed as follows:

$$\bar{\boldsymbol{\theta}}^{*,\alpha} = \tfrac{1}{K}\sum_{i=1}^{N}\left(\mathcal{L}_\alpha^i \cdot \boldsymbol{\theta}_i \odot f_{\bar{\theta}}\left(\boldsymbol{\theta}_i\right)\right)$$

$$\mathrm{Var}\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right) = \tfrac{1}{K}\sum_{i=1}^{N}\left(\mathcal{L}_\alpha^i \cdot \left(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}^{*,\alpha}\right) \odot \left(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}^{*,\alpha}\right) \odot f_{\bar{\theta}}\left(\boldsymbol{\theta}_i\right)\right) \tag{5}$$

where $K = \sum_{i=1}^{N}\mathcal{L}_\alpha^i \cdot f_{\bar{\theta}}\left(\boldsymbol{\theta}_i\right)$ and "$\odot$" denotes the Hadamard product - simple scalar product between vectors here but usable with higher dimensional controls. The parameter α is determined using the L-curve approach, considering a parametric curve $\left\{J\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right), D^\alpha\right\}, \alpha = -1, ..., 10$, where $D^\alpha = \left(\mathrm{Var}\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right)\right)^{-1} \odot \left(\bar{\boldsymbol{\theta}}^{*,\alpha} - \bar{\boldsymbol{\theta}}^0\right) \odot \left(\bar{\boldsymbol{\theta}}^{*,\alpha} - \bar{\boldsymbol{\theta}}^0\right)$ is the probabilistic (Mahalanobis) distance between the estimate $\bar{\boldsymbol{\theta}}^{*,\alpha}$ and the average prior $\bar{\boldsymbol{\theta}}^0 = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\theta}_i$. The value of α is sought in a L-curve "corner" such that it minimizes both $J\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right)$ and D^α.

C. Regional Calibration with ANNR

In this case, an ANN-based regional mapping $\mathcal{F}_{\mathcal{R}} \equiv \mathcal{N}$, consisting of a multilayer perceptron, aims to learn the descriptors-to-parameters mapping such that:

$$\boldsymbol{\theta}(x, \boldsymbol{D}, \boldsymbol{\rho}) := \mathcal{N}\left(\boldsymbol{D}(x), \boldsymbol{W}, \boldsymbol{b}\right), \forall x \in \Omega \tag{6}$$

where $\boldsymbol{W}$ and $\boldsymbol{b}$ are respectively weights and biases of the neural network, whose output layer consists in a scaling transformation based on the Sigmoid function in order to impose bound constraints on each hydrological parameters. The regional control vector $\boldsymbol{\rho} \equiv [\boldsymbol{W}, \boldsymbol{b}]^T$ is optimized by Algorithm 1, that uses spatial gradients computed by the adjoint model to minimize the cost function $J(\boldsymbol{\rho}) = J\left(Q^*, \mathcal{M}_{rr}(. , \boldsymbol{\theta} = \mathcal{N}(\boldsymbol{D}, \boldsymbol{\rho}))\right)$ in the present case.

The cost function depends on the forward model $\mathcal{M} \equiv \mathcal{M}_{rr}\left(., \mathcal{N}\left(.\right)\right)$, which is composed of two components in its numerical implementation: (i) an ANN implemented in Python, which produces the output $\boldsymbol{\theta}$ served as input for (ii) the hydrological model $\mathcal{M}_{rr}$ implemented in Fortran. To optimize J, we need its gradients with respect to $\boldsymbol{\rho}$.

Algorithm 1 The proposed back-propagation at each training iteration with "opt_func" denotes the update function of the optimizer used (e.g., Adam), and "hyper_param" denotes its hyper parameters (e.g., learning rate)

Input: descriptors $\boldsymbol{D}(x)$, initial/pre-updated weights and biases $\boldsymbol{\rho} = (\rho_1, ..., \rho_{N_L})$

$\boldsymbol{\theta} \leftarrow \left[(\mathcal{N}(\boldsymbol{D}(x), \boldsymbol{\rho}))_{x \in \Omega} \right]^T$ ▷ Forward propagation

$\nabla A \leftarrow \nabla_{\boldsymbol{\theta}} J = \left(\frac{\partial J}{\partial \theta_1}, ..., \frac{\partial J}{\partial \theta_{N_\theta}} \right)$ ▷ Initialize gradient accumulation

for $j = N_L..1$ **do**

$\quad \frac{\partial J}{\partial \rho_j} \leftarrow \left(\frac{\partial \boldsymbol{\theta}}{\partial \rho_j} \right)^T \nabla A$ ▷ Compute gradients

$\quad \nabla A \leftarrow \nabla A . W_j^T$ ▷ Update gradient accumulation

$\quad \rho_j \leftarrow \text{opt\_func} \left(\frac{\partial J}{\partial \rho_j}, \rho_j, \text{hyper\_param} \right)$ ▷ Update weights and biases

end for

return $\boldsymbol{\rho} = (\rho_1, ..., \rho_{N_L})$ ▷ Updated weights and biases

---

The main technical difficulty here is achieving a "seamless flow of gradients" through back-propagation. To overcome this, we divide the gradients into two parts. First, $\nabla_{\boldsymbol{\theta}} J$ can be computed via the automatic differentiation applied to the Fortran code corresponding to $\mathcal{M}_{rr}$. Then, $\nabla_{\boldsymbol{\rho}} \boldsymbol{\theta}$ is simply obtained by analytical calculus applicable given the explicit architecture of the ANN, consisting of a multilayer perceptron. The convergence criteria is simply determined by reaching the maximum number of training iterations.

III. RESULTS

A. Numerical experiment

The proposed alorithms are tested on a highly challenging regionalization case from Huynh et al. (2023): a high-resolution regional modeling of a flash flood prone area located in the South-East of France, with heterogeneous physical properties including karstic areas. Multiple gauges downstream of nested and independent catchments are simultaneously considered, enabling multi-gauge optimization. A total of 11 gauged catchments are employed as "donor" catchments for calibration, while 9 other catchments are treated as pseudo-ungauged for spatial validation to assess regionalization capabilities of the proposed algorithms. In this study, a set of 7 physical descriptors (see Table I) available over the whole French territory is used to learn the regional transfer functions.

TABLE I
PHYSICAL DESCRIPTORS USED FOR PRE-REGIONALIZATION METHODS.

Descriptor	Type	Unit	Source
Slope	Topography	m	EU-DEM, Copernicus (2016)
Drainage density	Morphology	-	Organde et al. (2013)
Karst index	Influence	%	Caruso et al. (2013)
Woodland percentage	Land use	%	CLC European Union (2012)
Urbanization rate	Land use	%	CLC European Union (2012)
Soil water storage	Hydrogeology	mm	Poncelet (2016)
Soil moisture storage	Hydrogeology	%	Odry (2017)

In the following, we compare and analyze: (i) local uniform $\boldsymbol{\rho} \equiv \bar{\boldsymbol{\theta}}$ and full spatially distributed $\boldsymbol{\rho} \equiv \boldsymbol{\theta}(x)$ calibrations for each gauges, that are respectively under- and over-parameterized hydrological optimization problems, but are served as reference performances ("Uniform (local)" and "Distributed (local)"); multigauge regional calibrations with (ii) lumped model parameters $\boldsymbol{\rho} \equiv \bar{\boldsymbol{\theta}}$ which somehow represents "level 0" regionalization ("UR"); (iii) a multivariate linear mapping (i.e., $\boldsymbol{\rho} \equiv \left[\alpha_{k,0}, (\alpha_{k,d}, 1) \right]^T$) using a first guess obtained by global optimization algorithm ("M2R"), or guided by a Bayesian estimation ("BGM2R"); and (iv) a multilayer perceptron (i.e., $\boldsymbol{\rho} \equiv \left[\boldsymbol{W}, \boldsymbol{b} \right]^T$) ("ANNR").

Two study periods, namely P1 (August 2011 – August 2015) and P2 (August 2015 – August 2019), are considered for split sample testing. A two-fold cross-temporal calibration approach is employed, where the models are calibrated on one period and validated on the other period. In each case, we consider three types of validation: spatial validation (performance in pseudo-ungauged catchments during the calibration period), temporal validation (performance in gauged catchments during the validation period), and spatio-temporal validation (performance in pseudo-ungauged catchments during the validation period).

B. Regionalization performances and analysis

The performance of all calibration and regionalization methods is presented in Figure 2. Unsurprisingly, spatially uniform calibration (UR) leads to limited performance in calibration and poor performance in regionalization, especially when compared to the reference local spatially distributed calibration that is overparameterized. The pre-regionalization methods, which incorporate information from multi-gauge discharge as well as physical descriptor maps, all yield relatively satisfying performances in calibration, temporal validation, and spatio-temporal validation at pseudo-ungauged sites (median NSE scores higher than 0.4 when calibrated on P1). The regionalization approach based on ANN (ANNR) achieves the best results for both gauged and pseudo-ungauged catchments.

Regarding the determination of prior parameter sets for the multi-linear pre-regionalization mapping, the Bayesian estimation approach (LDB-FG) demonstrates fairly good performance, comparable to that obtained with the global heuristic algorithm (SBS-FG), in calibration and spatial validation, with only minor differences in temporal validation. We believe that this is reflective of the importance of exploring a Bayesian approach for the definition of the cost function, which would enable intrinsic weighting of model misfits to different gauged hydrological behaviors. Moreover, when considering the performances of M2R and BGM2R on P1 (upper sub-figure of Figure 2), the Bayesian approach exhibits markedly higher performance in calibration and validations, while relatively similar performances are observed on P2 (lower sub-figure of Figure 2). This difference may be attributed to variations in hydrological information on P1 and potentially higher data errors, which have a lesser impact on the Bayesian approach.
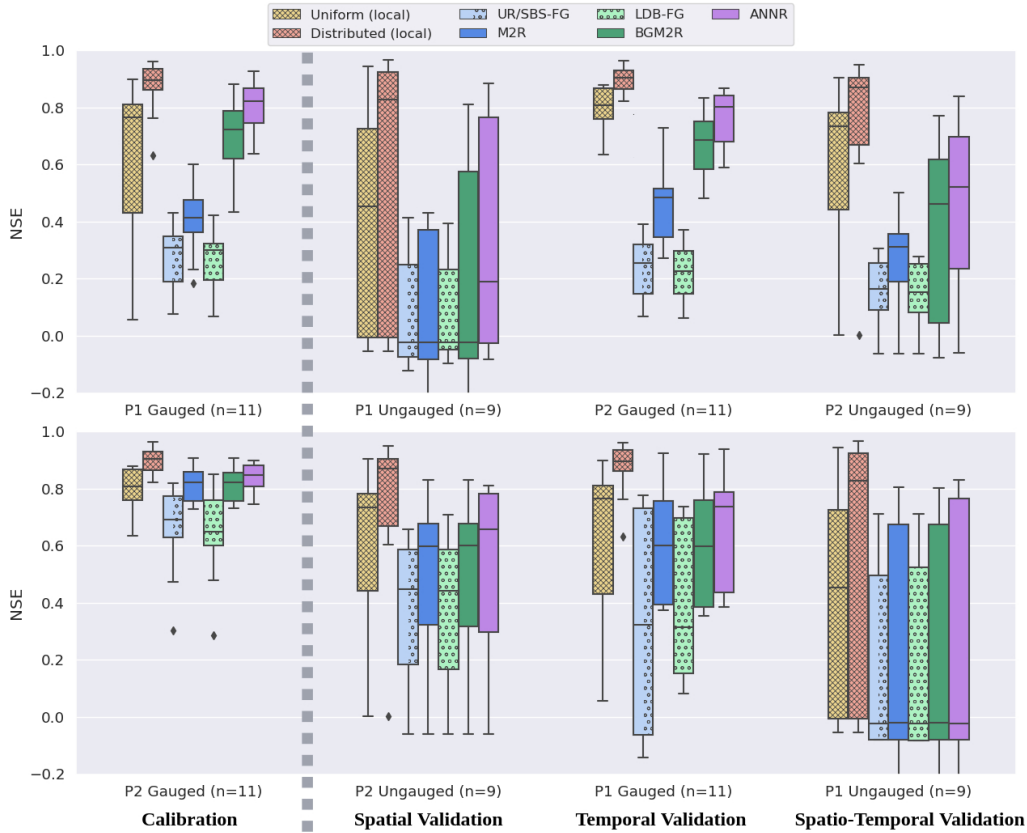
Fig. 2. Performance of all calibration and regionalization methods when calibrated on P1 (upper sub-figure) and on P2 (lower sub-figure). BGM2R uses the first guess obtained by Low-Dimensional Bayesian estimation (LDB-FG), while the first guess of M2R is obtained by the SBS algorithm (SBS-FG), which also represents the solution of the regionalization method with lumped parameters (UR). n denotes the number of study catchments.

Table II represents several statistical quantities of the distributed parameter maps obtained through different regionalization approaches. All methods result in distinct parameter maps and varying levels of temporal stability (see Figure 3). The ANNR leads to the most robust inference over P1 and P2, with remarkably stable average parameter values as well as spatial standard deviation over time. The priors inferred with SBS-FG or LDB-FG exhibit slight differences and also lead to a different optimum during pre-regionalization for P1. Interestingly, the opposite trend is observed for P2, where data uncertainty and model adequacy might be better, resulting in similar functioning points after regionalization despite substantially different priors determined with SBS-FG or LDB-FG.

Last but not least, during calibration on P1, M2R and BGM2R lead to a negative exchange coefficient ($k_{exc} < 0$), despite starting from priors with positive exchange values (3.04 for SBS-FG (P1) and 2.9 for LDB-FG (P1)); AANR also lead to negative exchange. This intriguing result, of reaching systematically significant negative exchange, is particularly noteworthy because the exchange coefficient directly impacts mass conservation. These findings relate to those on flash floods water balance sensitivity and regionalization based on geological descriptors presented by Garambois et al. (2015) for

catchments in the same and nearby areas. Their event process-oriented and conservative model required an increase in modeled soil volume, while pedological and geological descriptors provided valuable constraining information, especially in the context of flash floods.

IV. CONCLUSION

A Bayesian calibration algorithm has been tested in this study, on top of our Hybrid Variational Data Assimilation Parameter Regionalization (HVDA-PR) approach enabling seamless regionalization in hydrology. The methods were tested in a challenging flash flood-prone area in the South-East of France, characterized by diverse physical properties and hydrological responses. Overall, the methods demonstrated satisfactory performance in several aspects: (i) accurate modeling of discharge at both gauged and pseudo-ungauged sites, and (ii) effective identification of conceptual parameters and extraction of information from physical descriptors. Notably, the ANN-based regionalization method outperformed other approaches in terms of discharge accuracy and parameter stability. Bayesian prior estimation exhibited good performance and relative robustness, even in challenging cases like calibration on the period P1, where data uncertainty and model inadequacy were assumed to be higher. While the Bayesian method is computationally more demanding than traditional

| Method (Cal period)	c_p	c_{ft}	k_{exc}	l_r
UR/*SBS-FG* (P1)	2000	470.88	3.04	55.12
UR/*SBS-FG* (P2)	322.68	239.65	-3.18	45.53
M2R (P1)	2000 (2000, 0)	113.57 (397.78, 436.27)	-0.15 (-3.8, 8.35)	52.79 (83.63, 79.27)
M2R (P2)	398.64 (658.28, 623.15)	338.41 (433.16, 351.08)	-0.53 (-1.56, 5.29)	94.16 (100.66, 75.71)
LDB-FG (P1)	*1917.74*	*527.77*	*2.9*	*72.32*
LDB-FG (P2)	*138.53*	*517.6*	*-13.04*	*53.71*
BGM2R (P1)	76.02 (700.93, 869.81)	599.93 (540.72, 404.87)	-13.33 (-10.0, 9.94)	63.47 (92.67, 87.33)
BGM2R (P2)	396.71 (651.86, 607.05)	348.38 (439.31, 353.62)	-0.57 (-1.47, 4.98)	95.65 (100.98, 78.13)
ANNR (P1)	369.87 (700.95, 737.21)	393.52 (492.71, 334.31)	-3.63 (-6.92, 7.58)	56.63 (71.04, 62.52)
ANNR (P2)	404.13 (593.4, 477.85)	510.9 (444.66, 220.29)	-2.23 (-2.12, 2.34)	62.72 (67.28, 32.5)
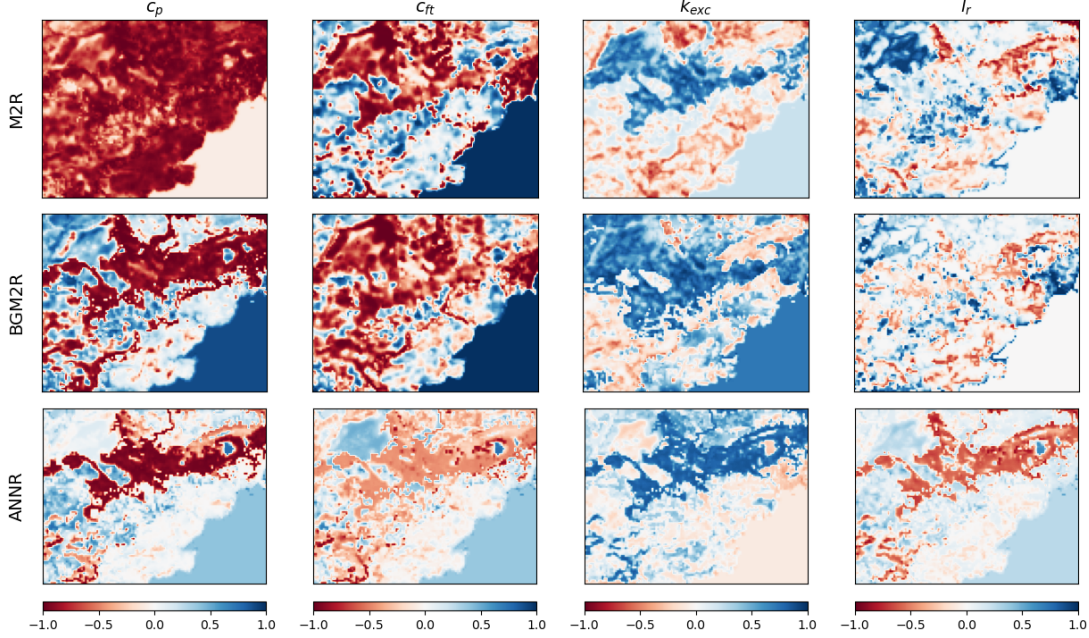
Fig. 3. Normalized temporal stability over the periods P1 and P2 of the optimal distributed parameters $\frac{\hat{\theta}_k^{(P2)} - \hat{\theta}_k^{(P1)}}{u_k - l_k}, k \in [1..N_\theta]$, for the three regionalization approaches using physical descriptors.

low-dimensional calibration algorithms, it can be efficiently parallelized. Moreover, the Bayesian approach can be extended to higher-dimensional contexts, such as determining semi-distributed priors and exploring spatial equifinality using our variational data assimilation algorithms. Interestingly, in contrast to the ANN, the regression methods provided insights into more complex modeling situations and potential data-model discrepancies. This highlights the importance of maintaining both "classical" approaches and AI-based solutions in research and applications, particularly in the continuous development of physically and mathematically interpretable methodologies.

REFERENCES

Abdulla, F. A., & Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of hydrology*, *197*(1-4), 230–257.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I., & Wood, E. F. (2020). Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *Journal of Geophysical Research: Atmospheres*, *125*(17), e2019JD031485.

Beven, K. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, *5*(1), 1–12. doi: 10.5194/hess-5-1-2001

Caruso, A., Guillot, A., & Arnaud, P. (2013). Notice sur les indices de confiance de la méthode shyreg-débit–définitions et calculs. In *Aix en provence: Irstea, convention dgpr/snrh.*

Chelil, S., Oubanas, H., Henine, H., Gejadze, I., Malaterre, P. O., & Tournebize, J. (2022). Variational data assimilation to improve subsurface drainage model parameters. *Journal of Hydrology*, 128006.

Colleoni, F., Huynh, N. N. T., Garambois, P.-A., Jay-Allemand, M., & Villenave, L. (2023). *Smash documentation.* Retrieved from https://smash.recover.inrae.fr (Version: 0.4.2, Release date: 2023-05-23)

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015-1031. doi: https://doi.org/10.1029/91WR02985

Fekete, B. M., & Vörösmarty, C. J. (2007). The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS publ*, *309*, 129–136.

Fortin, F.-A., Rainville, F.-M. D., Gardner, M.-A., Parizeau, M., & Gagné, C. (2012). Deap: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, *13*(70), 2171–2175.

Garambois, P.-A., Roux, H., Larnier, K., Labat, D., & Dartus, D. (2015). Parameter regionalization for a process-oriented distributed model dedicated to flash floods. *Journal of Hydrology*, *525*, 383–399.

Gejadze, I., Malaterre, P.-O., Oubanas, H., & Shutyaev, V. (2022). A new robust discharge estimation method applied in the context of swot satellite data processing. *Journal of Hydrology*, *610*, 127909.

Götzinger, J., & Bárdossy, A. (2007). Comparison of four regionalisation methods for a distributed hydrological model. *Journal of Hydrology*, *333*(2-4), 374–384.

Hannah, D. M., Demuth, S., van Lanen, H. A., Looser, U., Prudhomme, C., Rees, G., ... others (2011). Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200.

Hascoet, L., & Pascual, V. (2013). The tapenade automatic differentiation tool: principles, model, and specification. *ACM Transactions on Mathematical Software (TOMS)*, *39*(3), 1–43.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., ... Cudennec, C. (2013). A decade of predictions in ungauged basins (pub)—a review. *Hydrological Sciences Journal*, *58*(6), 1198-1255. doi: 10.1080/02626667.2013.803183

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., Renard, B., Roux, H., Demargne, J., ... Javelle, P. (2023). Learning pre-regionalization of a differentiable high-resolution hydrological model with spatial gradients. *HAL preprint hal-04145059*. Retrieved from https://hal.inrae.fr/hal-04145059

Michel, C. (1989). Hydrologie appliquée aux petits bassins ruraux. *Hydrology handbook (in French), Cemagref, Antony, France*.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., ... Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research*, *53*(9), 8020–8040.

Odry, J. (2017). *Prédétermination des débits de crues extrêmes en sites non jaugés : régionalisation de la méthode par simulation shyreg* (Doctoral dissertation). Retrieved from http://www.theses.fr/2017AIXM0424 (Thèse de doctorat dirigée par Arnaud, Patrick Géosciences de l'environnement. Hydrologie Aix-Marseille 2017)

Organde, D., Arnaud, P., Fine, J.-A., Fouchier, C., Folton, N., & Lavabre, J. (2013). Régionalisation d'une méthode de prédétermination de crue sur l'ensemble du territoire français: la méthode shyreg. *Revue des Sciences de l'Eau*, *26*(1), 65–78.

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*(3).

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.

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

Poncelet, C. (2016). *Du bassin au paramètre : jusqu'où peut-on régionaliser un modèle hydrologique conceptuel ?* (Doctoral dissertation). Retrieved from http://www.theses.fr/2016PA066550 (Thèse de doctorat dirigée par Andréassian, Vazken et Oudin, Ludovic Hydrologie Paris 6 2016)

Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of hydrologic engineering*, *18*(8), 958–975.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., & Chiew, F. H. S. (2009). Optimization of a similarity measure for estimating ungauged streamflow. *Water Resources Research*, *45*(10). doi: https://doi.org/10.1029/2008WR007248

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5).

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and forest meteorology*, *98*, 279–293.

Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, *23*(4), 550-560. Retrieved from http://dblp.uni-trier.de/db/journals/toms/toms23.html#ZhuBLN97