

Learning Regionalization within a Differentiable High-Resolution Hydrological Model using Accurate Spatial Cost Gradients

Ngo Nghi Truyen Huynh¹, Pierre-André Garambois¹, François Colleoni¹, Benjamin Renard¹,
Hélène Roux², Julie Demargne³, and Pierre Javelle¹

¹INRAE, Aix-Marseille Université, RECOVER, 3275 Route Cézanne, 13182 Aix-en-Provence, France

²Institut de Mécanique des Fluides de Toulouse (IMFT), Université de Toulouse, CNRS, 31400 Toulouse, France

³HYDRIS Hydrologie, Parc Scientifique Agropolis II, 2196 Boulevard de la Lironde, 34980 Montferrier-sur-Lez, France

Correspondence: Ngo Nghi Truyen Huynh (ngo-nghi-truyen.huynh@inrae.fr)

Abstract. Estimating spatially distributed hydrological parameters in ungauged catchments poses a challenging regionalization problem and requires imposing spatial constraints given the sparsity of discharge data. A possible approach is to search for a transfer function that quantitatively relates physical descriptors to conceptual model parameters. This paper introduces a Hybrid Data Assimilation and Parameter Regionalization (HDA-PR) approach incorporating learnable regionalization mappings, based on either multivariate regressions or neural networks, into a differentiable hydrological model. It enables the exploitation of heterogeneous datasets across extensive spatio-temporal computational domains within a high-dimensional regionalization context, using accurate adjoint-based gradients. The inverse problem is tackled with a multi-gauge calibration cost function accounting for information from multiple observation sites. HDA-PR was tested on high-resolution, hourly and kilometric regional modeling of two flash-flood-prone areas located in the South of France. In both study areas, the median Nash-Sutcliffe efficiency (NSE) scores ranged from 0.52 to 0.78 at pseudo-ungauged sites over calibration and validation periods. These results highlight a strong regionalization performance of HDA-PR, improving NSE by up to 0.57 compared to the baseline model calibrated with lumped parameters, and achieving a performance comparable to the reference solution obtained with local uniform calibration (median NSE from 0.59 to 0.79). Multiple evaluation metrics based on flood-oriented hydrological signatures are also employed to assess the accuracy and robustness of the approach. The regionalization method is amenable to state-parameter correction from multi-source data over a range of time scales needed for operational data assimilation, and it is adaptable to other differentiable geophysical models.

1 Introduction

Irrespective of their type and complexity, hydrological models are more or less empirical and uncertain representations of multiscale coupled hydrological processes whose observability is limited. Hydrological model parameters are in general effective quantities that cannot be directly measured. Instead, they are typically inferred through a calibration procedure aimed primarily at obtaining satisfactory streamflow simulations (e.g., Beven (2001); Kirchner (2006); Gupta et al. (2006); Vrugt et al. (2008)). In most cases, this optimization problem is a difficult ill-posed inverse problem faced with the equifinality (Beven, 2001) of feasible solutions, which can be further interpreted as model structural equifinality and spatial equifinality in the context of

spatially sparse observations compared to model controls (see for example discussions in Garambois et al. (2020)). Most calibration approaches enable the estimation of spatially uniform model parameters for a single gauged catchment, but this leads to piecewise constant and discontinuous parameters fields for adjacent catchments. Moreover, parameter sets determined through calibration are not transferable to ungauged locations although the latter represent the majority of the global land surface (Fekete and Vörösmarty, 2007; Hannah et al., 2011). Therefore, prediction in ungauged basins remains a “grand challenge” (Sivapalan, 2003) in hydrology (Hrachowitz et al., 2013), which hinders the development of effective high-resolution models adapted to the simulation of hydrological extremes in a context of high data uncertainty (e.g., for Mediterranean flash floods in Garambois et al. (2015); Jay-Allemand et al. (2022b)).

The estimation of hydrological model parameters in ungauged regions is performed with so-called regionalization approaches that exploit and transfer hydrological information from gauged to ungauged catchments using various descriptors of catchments physical properties (see reviews in Blöschl et al. (2013); Samaniego et al. (2010); Hrachowitz et al. (2013); Beck et al. (2020)). The most widely used approach in early studies for regionalization involved independent catchment-by-catchment calibrations, followed by multiple regression or interpolation techniques to transfer the calibrated parameter sets from gauged to ungauged locations (Abdulla and Lettenmaier, 1997; Seibert, 1999; Parajka et al., 2005; Razavi and Coulibaly, 2013; Parajka et al., 2013) and can be called post-regionalization (Samaniego et al., 2010). This approach presumes that the variability of calibrated model parameters through the catchments have relations based, for instance, on spatial proximity (Widén-Nilsson et al., 2007; Oudin et al., 2008), and physical or climatic similarity (Oudin et al., 2010; Beck et al., 2016). Statistical learning methods have also been applied in post-regionalization to explore the relationships between physical descriptors and calibrated parameter sets at gauged locations (e.g., Saadi et al. (2019); Wang et al. (2023)). However, post-regionalization approaches are limited to lumped parameters, thus ignoring within-catchment variabilities (see reviews in Samaniego et al. (2010); Razavi and Coulibaly (2013)), except when calibrated parameters correspond to tuning factors of physical pedotransfer functions or hydraulic frictions correspondence tables (see Garambois et al. (2015) for details). The identification of transfer functions in post-regionalization is complicated by the uncertainty of estimated parameter sets, while spatial proximity is more adapted to densely gauged river networks and regions (Oudin et al., 2008; Reichl et al., 2009). Moreover, incorporating a statistical learning process, especially unsupervised learning approaches, in the post-regionalization step can exacerbate existing issues such as parameter biases induced by data measurement errors (Kavetski et al., 2006). A regionalized calibration simultaneously exploiting the information of multiple gauges, within spatial clusters defined a priori from descriptors, is performed in Huang et al. (2019) over Norway using climatic similarity. The parameters calibrated over multiple gauges of a climatic zone are applied to ungauged catchments of the same zone. This approach does not account for hydrological heterogeneity within the catchments or within the regional clusters determined by physical similarity, which can have a major impact on the forecasting of extreme floods in particular (Garambois et al., 2015; Jay-Allemand et al., 2022b).

The simultaneous regionalization approach involves optimizing a transfer function between physical descriptors and model parameters (cf. Hundscha and Bárdossy (2004); Götzinger and Bárdossy (2007); Bastola et al. (2008); Samaniego et al. (2010)). In this case and contrarily to post-regionalization, the descriptors-to-parameters mapping is the first optimizable operator of the forward hydrological model. It enables to overcome most of the aforementioned problems and has been applied in several stud-

ies. For instance, it has been used for regionalizing semi-distributed models such as: HBV in Hundscha and Bárdossy (2004) or in Göttinger and Bárdossy (2007) who introduced monotonicity and Lipschitz condition into the optimization problem to constrain the inferred spatial fields; TOPMODEL in Bastola et al. (2008) who used an Artificial Neural Network (ANN)-based mapping between catchment descriptors and model parameters (both their value and their uncertainty as quantified by a posterior covariance matrix). A multiscale parameter regionalization (MPR) method, combining descriptors maps, upscaling functions and regionalization transfer functions in the form of multivariate mappings from descriptors, and implemented within a spatially distributed multiscale hydrological model (mHm), has been proposed by Samaniego et al. (2010), and later applied to over 400 European catchments at 0.25° spatial resolution in Rakovec et al. (2016). This approach imposes a spatial regularization effect that is needed when working with spatially distributed hydrological models and spatially sparse discharge data (see regularization for semi-lumped model calibration from multiple nested gauges in De Lavenne et al. (2019)). The MPR method from Samaniego et al. (2010) has also been used with other gridded hydrological models in large sample applications. For example, Mizukami et al. (2017) calibrate the VIC model at a resolution of 0.125° over 531 headwater catchments (area $< 2,000 \text{ km}^2$) in the contiguous US area, using a lumped regionalization approach. Another example is Beck et al. (2020), who calibrate the HBV model at 0.05° resolution over 4,229 headwater catchments (area $< 5,000 \text{ km}^2$) worldwide. In their study, they categorize the catchments into three climatic groups and perform tenfold cross-validation using 90% of the gauged catchments. While these studies applied MPR deterministically, in Lane et al. (2021), the MPR method is applied within the generalized likelihood uncertainty estimation (GLUE) framework, with a high-resolution HRU model (DECIPHeR framework) at daily time resolution over a large sample of 437 catchments in the UK. However, the routing module in this study is calibrated separately with a simple random sampling approach. In Mizukami et al. (2017), the runoff routing model is a gamma distribution function with two parameters that were “directly calibrated for each basin”. Therefore, those regionalization approaches essentially focus on runoff production, at a daily time step for mostly headwater catchments whose characteristic response/concentration time scale might be shorter. The same remark can be made for Beck et al. (2020) who work at daily time step on headwater catchments, simply without modeling routing.

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), or the GLUE framework with a random sampling approach in Lane et al. (2021). Those algorithms are applicable with low-dimensional controls only, which limits the affordable number of descriptors and the spatialization of regional transfer parameters (that are lumped in all methods above), and more importantly, which limits the affordable complexity of the regionalization operator. Nevertheless, gradient-based algorithms are efficient approaches for solving high dimensional inverse problems, and their potential has been demonstrated in optimization of spatial parameters of hydraulic models (Monnier et al., 2016), or spatially distributed hydrological models (Castaings et al., 2009; Jay-Allemand et al., 2020). They crucially need accurate estimates of cost gradients, i.e. gradients of the cost (objective) function with respect to the sought parameters, which can be spatialized and of large dimension. Such gradients can be computed with an adjoint model for example obtained by source code differentiation in Castaings et al. (2009); Monnier et al. (2016); Jay-Allemand et al. (2020). Note that a key property of neural networks is their differentiability, which makes them compatible with variational data assimilation frame-

works (Monnier et al., 2016) and hence suggests them as promising candidate to derive flexible regionalization functions. In the context of hydrological modeling of ungauged basins, achieving such a flexible regionalization is desirable to adequately represent the multi-scale variabilities of the physical system. It would also maximize the extraction of information from large sets of physical descriptors and hydrological response observations, while accounting for data and modeling uncertainties.

A novel approach called HDA-PR (Hybrid Data Assimilation and Parameter Regionalization) is presented in this article. HDA-PR relies on seamless regional optimization algorithms for learning complex transfer functions between physical descriptors and conceptual parameters of spatially distributed hydrological models, applicable at high-resolution with spatial constraints of various rigidity to address the spatial equifinality issue. It is designed to exploit the informative content of massive heterogeneous datasets over large spatio-temporal computational domains, and is therefore adapted to solving high-dimensional inverse problems. Our approach leverages information from multi-site river flow observations and high-resolution data on a 1 km² and 1 h resolution grid, relying on the original combination of the following ingredients:

- Learnable regionalization functions via the introduction into the direct hydrological model of an explicit tunable mapping between heterogeneous physical descriptors and spatially distributed unknown conceptual parameters. This mapping allows estimating parameter values while imposing a constraint on their spatial variability, via the use of physical descriptors and a priori knowledge. Multivariate polynomial regressions and neural networks are employed to learn such a complex nonlinear descriptors-parameters mapping.
- A differentiable spatially distributed hydrological model into which the regionalization operators have been implemented. This enables the computation of accurate, spatially-distributed gradients of the calibration cost (objective) function, with respect to the sought regionalization parameters, which can be of high dimension. Obtaining accurate gradients for such high-dimensional parameters is crucially needed for optimization algorithms.

The original combination of the above ingredients amounts to introducing regionalization transfer functions into a variational data assimilation (VDA) algorithm (similar to the tunable differentiable mappings in hydraulic VDA algorithms (Monnier et al., 2016; Garambois et al., 2020)) dedicated to spatially distributed hydrological modeling and high-dimensional inverse problems. This has seldom been investigated especially for regional hydrological learning from multi-site data. The strength of HDA-PR lies in its capability to learn complex relation between physical descriptors and conceptual parameters of spatially distributed models in the context of structural and spatial parametric equifinality. Additionally, our approach aims at ensuring that the hybrid data assimilation algorithm, which integrates an explainable learning process, produces results that can be physically interpreted (see Larnier and Monnier (2020); Höge et al. (2022); Fablet et al. (2021); Althoff et al. (2021)). It is able to enhance calibration scores with deep learning from large heterogeneous datasets while maintaining the interpretability of physics-based hydrological models. The evaluation procedure adopted in this work considers challenging regionalization problems with multi-gauge settings in flash-flood-prone areas and multiple evaluation metrics including flood event hydrological signatures (Huynh et al., 2023). We address the following aspects of the HDA-PR approach: (i) performance at gauged and ungauged sites; (ii) factors determining the performance; and (iii) spatial patterns of the regionalized parameters in relation to information extraction from physical descriptors.

The remaining sections of this paper are organized as follows: Section 2 describes the HDA-PR algorithms and the SMASH spatially distributed hydrological assimilation platform into which they have been implemented. In Section 3, we present the case studies based on two contrasted areas and analyze the performance of HDA-PR using different regionalization mappings. Subsequently, in Section 4, we discuss compelling findings based on the results from the previous section. Finally, in Section 5, we conclude our work and outline potential future research directions.

2 Forward-Inverse Algorithms

This section presents the forward model and inverse algorithms of the proposed HDA-PR method. An algorithm flowchart is provided in Figure 1 to help in global understanding.

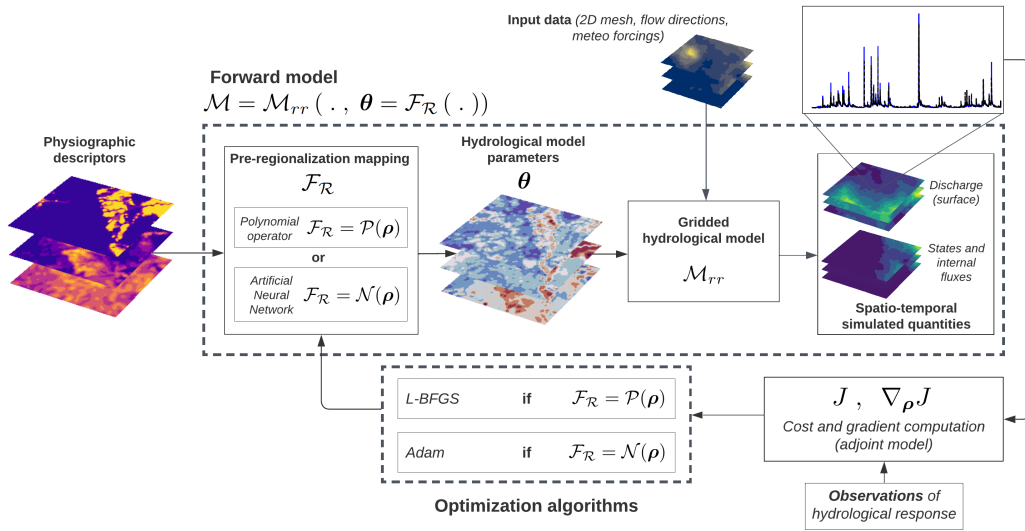


Figure 1. Flowchart of the forward-inverse algorithm used in HDA-PR. The forward hydrological model is a gridded model (spatio-temporal regular grid at 1 km² and 1 h) using GR operators.

First, the differentiable forward model consists in: (i) a parsimonious and robust GR-like conceptual hydrological model structure (Perrin et al., 2003) that is spatially distributed and differentiable (Jay-Allemand et al., 2020); and (ii) regionalization operators, consisting of either multivariate polynomial regressions or neural networks, for mapping descriptors onto hydrological model parameters. The calibration cost function adapted to multi-site (and potentially multi-source) observations is then defined. The inverse optimization algorithms, that use the spatially distributed gradients of the cost function with respect to model parameters, and are capable of dealing with high-dimensional inverse problems such as encountered with tunable parameters of regionalization descriptors-parameters mappings, are then detailed.

The core strength of HDA-PR is the use of differentiable descriptors-to-parameters transfer functions, especially in the form of neural networks, and the capability to automatically compute accurate cost gradients. The latter enables the use of

gradient-based variational optimization algorithms with high-dimensional regional parameter vectors. The method is applicable to any differentiable forward model as well as to multi-source heterogeneous datasets, and hence constitutes a powerful data assimilation framework.

2.1 Forward Model with Regionalization

First, let $\Omega \subset \mathbb{R}^2$ denote a 2D spatial domain that can contain multiple catchments, both gauged and ungauged, with a minimum of one gauged catchment, and $t > 0$ the physical time. In what follows, the vector of spatial coordinates over Ω is denoted x . The number of active cells within the spatial domain Ω is noted N_x . A 2D flow directions map \mathcal{D}_Ω is obtained from terrain elevation processing and will be used for runoff routing, with the only condition that a unique point in the regular mesh \mathcal{T}_Ω has the highest drainage area.

Consider observed discharge time series $Q_g^*(t)$ at N_G observation cells of coordinates $x_g \in \Omega$, $g = 1..N_G$ ($N_G \geq 1$). For each observation cell, the corresponding gauged upstream sub-catchment is noted Ω_g so that $\Omega_{ung} = \Omega \setminus \left(\bigcup_{g=1}^{N_G} \Omega_g \right)$ is the remaining ungauged part of the whole spatial domain Ω . Note that this definition is suitable for the general regionalization case dealing with spatially independent and/or nested gauged catchments.

Then, the forward model \mathcal{M} can be defined as a multivariate function obtained by partially composing a hydrological model \mathcal{M}_{rr} with a regionalization operator \mathcal{F}_R to compute hydrological parameters θ such that:

$$\mathcal{M} = \mathcal{M}_{rr}(\cdot, \theta = \mathcal{F}_R(\cdot)) \quad (1)$$

Let us now introduce and detail the hydrological model and the regionalization operator along with their input variables.

The rainfall and potential evapotranspiration fields are respectively noted $\mathbf{P}(x, t)$ and $\mathbf{E}(x, t)$, $\forall x \in \Omega$. The hydrological model \mathcal{M}_{rr} is a dynamic operator projecting the input fields $\mathbf{P}(x, t)$ and $\mathbf{E}(x, t)$, given an input drainage plan $\mathcal{D}_\Omega(x)$, onto the discharge field $\mathbf{Q}(x, t)$ and states fields $\mathbf{h}(x, t)$ such that, $\forall (x, t') \in \Omega \times [0, t]$:

$$\mathbf{U}(x, t) = (\mathbf{h}, \mathbf{Q})(x, t) = \mathcal{M}_{rr}[(\mathcal{D}_\Omega, \theta)(x); (\mathbf{P}, \mathbf{E})(x, t'), \mathbf{h}(x, 0), t] \quad (2)$$

where θ is the N_θ -dimensional vector of model parameters 2D fields that we aim to estimate regionally with the new algorithms proposed below, and \mathbf{h} is the N_S -dimensional vector of internal model states. In this study, the distributed hydrological model \mathcal{M}_{rr} is a parsimonious GR4-like conceptual structure (Perrin et al., 2003), which is the spatialized ‘‘S-GR4’’ structure presented in Colleoni et al. (2023). The hydrological parameters vector $\forall x \in \Omega$ is:

$$\theta(x) = (c_p(x), c_{ft}(x), k_{exc}(x), l_r(x))^T$$

where the four spatially varying parameter fields are the capacity of the production reservoir (c_p in [mm]), the capacity of the transfer reservoir (c_{ft} in [mm]), the parameter (k_{exc} in [mm/dt]) of the non-conservative water exchange flux, and the linear routing parameter (l_r in [min]).

In order to constrain and explain these spatial fields of conceptual model parameters $\theta(x)$ from descriptors $\mathbf{D}(x)$, we introduce a regionalization operator \mathcal{F}_R that is a descriptors-to-parameters mapping such that:

$$\theta(x) = \mathcal{F}_R(\mathbf{D}(x), \rho), \forall x \in \Omega \quad (3)$$

with \mathbf{D} the N_D -dimensional vector of physical descriptor maps covering Ω , and $\boldsymbol{\rho}$ the vector of tunable regionalization parameters that is defined below.

Two types of regionalization operators are used in HDA-PR (see Figure 1):

1. A set \mathcal{P} of multivariate polynomial regression operators for each parameter of the forward hydrological model (Equation 2):

$$\begin{aligned}\boldsymbol{\theta}(x, \mathbf{D}, \boldsymbol{\rho}) &:= \mathcal{P}(\mathbf{D}(x), \boldsymbol{\rho}) \equiv \left[(\theta_k(x, \mathbf{D}, \boldsymbol{\rho}_k))_{k=1}^{N_\theta} \right]^T, \forall x \in \Omega; \\ \theta_k(x, \mathbf{D}, \boldsymbol{\rho}_k) &:= s_k \left(\alpha_{k,0} + \sum_{d=1}^{N_D} \alpha_{k,d} D_d^{\beta_{k,d}}(x) \right), \forall k \in [1..N_\theta]\end{aligned}\quad (4)$$

with $s_k(z) = l_k + (u_k - l_k) / (1 + e^{-z})$, $\forall z \in \mathbb{R}$, a transformation based on a Sigmoid function with values in $]l_k, u_k[$, thus imposing bound constraints in the direct hydrological model such that $l_k < \theta_k(x) < u_k$, $\forall x \in \Omega$. The lower and upper bounds l_k and u_k , associated to each parameter field θ_k of the hydrological model (Equation 2) are assumed spatially uniform for simplicity here. The regional control vector to be estimated in this case is:

$$\boldsymbol{\rho} \equiv \left[(\boldsymbol{\rho}_k)_{k=1}^{N_\theta} \right]^T \equiv \left[\left(\alpha_{k,0}, (\alpha_{k,d}, \beta_{k,d})_{d=1}^{N_D} \right)_{k=1}^{N_\theta} \right]^T \quad (5)$$

2. An ANN denoted \mathcal{N} , consisting of a multilayer perceptron, aimed at learning the descriptors-to-parameters mapping such that:

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

where \mathbf{W} and \mathbf{b} are respectively weights and biases of the neural network composed of N_L dense layers. The architecture of the neural network and the forward propagation is detailed in Appendix B and Equation B2. Note that an output layer consisting of a scaling transformation based on the Sigmoid function (cf. Equation B1) enables to impose $l_k < \theta_k(x) < u_k$, $\forall x \in \Omega$, i.e., bound constraints on the k^{th} -hydrological parameters. The regional control vector in this case is:

$$\boldsymbol{\rho} \equiv [\mathbf{W}, \mathbf{b}]^T \equiv \left[(\mathbf{W}_j, \mathbf{b}_j)_{j=1}^{N_L} \right]^T \quad (7)$$

For each regionalization operator (Equation 4 or 6), the regional calibration problem consists in optimizing (in a sense defined below) the regionalization control $\boldsymbol{\rho}$ (Equation 5 or 7) that can be of relatively high dimension since it is proportional to the number of descriptors (N_D), the number of model parameters (N_θ), and the degree of spatialization of the regional controls. Optimization algorithms adapted to the high-dimensional problems of interest, taking advantage of accurate spatially distributed gradients computation with the adjoint of the forward model, are detailed thereafter. Importantly, note that by definition of the mathematical model and given the numerical implementation rules followed, the forward model is differentiable. This is a necessary condition for computing cost gradients with respect to spatially distributed hydrological parameters and obtain those of regional controls, as needed for solving the optimization problem. This is a key idea and property of our proposed algorithms.

The numerical resolution of the ODE (Ordinary Differential Equation)-based operator of the forward hydrological model (Equation 2) relies on an explicit expression of its solution, approximated on the regular mesh \mathcal{T}_Ω of constant step dx with a fixed time step dt . All physical descriptors are mapped onto model grid for simplicity here.

Note that adding an upscaling operator after the regionalization scheme (as done in Samaniego et al. (2010)) is feasible in HDA-PR under the condition that it is differentiable (at least numerically), and is a potentially interesting topic for further research, as is improving observation operators. In both cases one could use algebraic expressions or neural networks in the HDA-PR assimilation framework.

2.2 Calibration Cost Function

A calibration cost function is defined to measure the misfit between simulated and observed discharge time series, respectively noted $Q_g(t)$ and $Q_g^*(t)$, for $g \in 1..N_G$ gauged cells. In order to measure the discrepancy between observed and simulated quantities from multiple observation sites, we consider the cost function:

$$J = \sum_{g=1}^{N_G} w_g J_g^* \quad (8)$$

with w_g a weighting function explained afterwards, J_g^* a local quadratic metric “at the station”, here $1 - NSE$ or $1 - KGE_2$ (see Appendix A). This cost function is a differentiable and convex function, involving the response of the direct model. It depends on the control vector ρ through the direct model \mathcal{M} (Equation 1) composed of the regionalization operator \mathcal{F}_R (Equation 3) and the direct hydrological model \mathcal{M}_{rr} (Equation 2).

The multi-site calibration corresponds to $N_G > 1$ while $N_G = 1$ is the classical single-gauge calibration. For $N_G > 1$, the weighting w_g is defined such that $\sum_{g=1}^{N_G} w_g = 1$ and is simply set here as $w_g = \frac{1}{N_G}$, which represents the average cost over multiple gauges.

2.3 Regional Calibration Algorithms

The inverse problem is written as the following convex optimization problem of the regional control vector ρ :

$$\hat{\rho} = \arg \min_{\rho} J(\mathcal{U}(\rho)) \quad (9)$$

The regional calibration aims to (i) reduce the misfit between observed and simulated discharges at spatially sparse gauging stations, as evaluated by Equation 8, while (ii) determining hydrological parameter maps $\theta(x)$, relevant for modeling at ungauged sites, thus benefiting from the information extracted from physical descriptors and the spatial constraint induced by the regional transfer functions whose parameters ρ are optimized. The regionalization operator can be expressed as either (i) a multi-polynomial mapping $\mathcal{F}_R \equiv \mathcal{P}$ (Equation 4), or (ii) an artificial neural network $\mathcal{F}_R \equiv \mathcal{N}$ (Equation 6). In both cases, the regional control vector ρ to optimize is large, and gradient based optimization methods adapted to high-dimensional inverse problems are employed.

2.3.1 Optimization Algorithm for Polynomial Regionalization

In this case, the forward model includes the polynomial descriptors-to-parameters mapping (Equation 4), i.e., $\mathcal{F}_R \equiv \mathcal{P}$ and the regional control vector is:

$$\boldsymbol{\rho} := [\alpha_{k,0}, (\alpha_{k,d}, \beta_{k,d})]^T, \forall (k, d) \in [1..N_\theta] \times [1..N_D]$$

The optimization problem, represented in Equation 9, is solved using the L-BFGS-B algorithm (limited-memory Broyden–Fletcher–Goldfarb–Shanno bound-constrained) (Zhu et al., 1997). This algorithm is specially adapted to high-dimensional parameter spaces, and in this study, there are no bound constraints on the values of $\alpha_{k,\cdot}$, whereas the exponents $\beta_{k,d}$ are simply sought between 0.5 and 2. 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 and Pascual, 2013). The entire process is implemented in the SMASH Fortran source code, where the full forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(\cdot, \mathcal{P}(\cdot))$ is a composition of both the hydrological model and the polynomial descriptors-to-parameters mapping.

The background value $\boldsymbol{\rho}^*$, used as a starting point for the optimization, is set using a spatially uniform solution $\bar{\boldsymbol{\theta}}^*$, which is obtained by a simple global optimization algorithm (Michel, 1989) of the inverse problem (Equation 9) where $\mathcal{M} \equiv \mathcal{M}_{rr}$ and $\boldsymbol{\rho} := \bar{\boldsymbol{\theta}}$, as follows:

$$\boldsymbol{\rho}^* \equiv [\alpha_{k,0} = s_k^{-1}(\bar{\theta}_k^*), (\alpha_{k,d} = 0, \beta_{k,d} = 1)]^T, \forall (k, d) \in [1..N_\theta] \times [1..N_D]$$

where $s_k^{-1}(z) = \ln\left(\frac{z - l_k}{u_k - z}\right)$ is the inverse Sigmoid.

The termination criterion is determined based on the satisfaction of at least one of the following criteria:

- Maximum number of iterations;
- Cost function criterion: $\frac{J^{(i)} - J^{(i+1)}}{\max(|J^{(i)}|, |J^{(i+1)}|, 1)} \leq \epsilon \times 10^6$ (e.g., $\epsilon \approx 2.22 \times 10^{-16}$);
- Gradient criterion: $\|\nabla_{\boldsymbol{\rho}} J^{(i)}\|_\infty \leq 10^{-12}$

where $J^{(i)}$, $\|\nabla_{\boldsymbol{\rho}} J^{(i)}\|_\infty$ are respectively the cost value and its projected gradient at iteration i , and ϵ represents the machine precision.

2.3.2 Optimization Algorithm for Neural Network-based Regionalization

In this case, the forward model includes a descriptors-to-parameters mapping performed with a neural network, i.e., $\mathcal{F}_R \equiv \mathcal{N}$ and the regional control vector is $\boldsymbol{\rho} := [\mathbf{W}, \mathbf{b}]^T$. The optimization problem (Equation 9) can typically be solved using Adam optimization algorithm (Kingma and Ba, 2014), an efficient stochastic gradient descent algorithm able to adapt the learning rate based upon the first and the approximation of the second moments of the gradients for fast convergence, and only requiring first order gradients of the cost function. In the present case, the cost function writes as:

$$J(\mathbf{U}(\boldsymbol{\rho})) = J(Q^*, \mathcal{M}_{rr}(\cdot, \boldsymbol{\theta} = \mathcal{N}(\mathbf{D}, \boldsymbol{\rho}))) \quad (10)$$

This formulation of the cost function highlights its dependency on the forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(\cdot, \mathcal{N}(\cdot))$, which is composed of two components in its numerical implementation: (i) an ANN implemented in Python, which produces the output θ used as input by (ii) the hydrological model \mathcal{M}_{rr} implemented in Fortran. In order to optimize J , its gradients with respect to ρ are required. The main technical difficulty here is to achieve a “seamless flow of gradients” through back-propagation. To overcome this, we divide the gradients into two parts and apply the chain rule with analytical derivation and numerical code differentiation (cf. hybrid VDA course in Monnier (2021) and references therein). First, $\nabla_{\theta} J$ can be computed via the automatic differentiation applied to the Fortran code corresponding to \mathcal{M}_{rr} . Then, $\nabla_{\rho} \theta$ is simply obtained by analytical calculus applicable given the explicit architecture of the ANN, consisting of a multilayer perceptron. Finally, the two gradients can be combined as $\nabla_{\rho} J = \nabla_{\theta} J \cdot \nabla_{\rho} \theta$. The termination criterion is determined by a specified number of iterations in the optimization algorithm. A detailed explanation of the network architecture, backward propagation, and the optimization process can be found in Appendix B.

3 Data and Numerical Experiment

3.1 Study Area and Experimental Design

The performance of HDA-PR using various regional optimization algorithms is evaluated based on high-resolution regional modeling of two flash-flood-prone areas located in the South of France (Figure 2). They are characterized by contrasted physical properties and catchments behaviors. The modeling approach is applied to each regional domain that contains multiple gauges downstream of both nested and independent catchments, used together in the optimization process through the multi-site cost function.

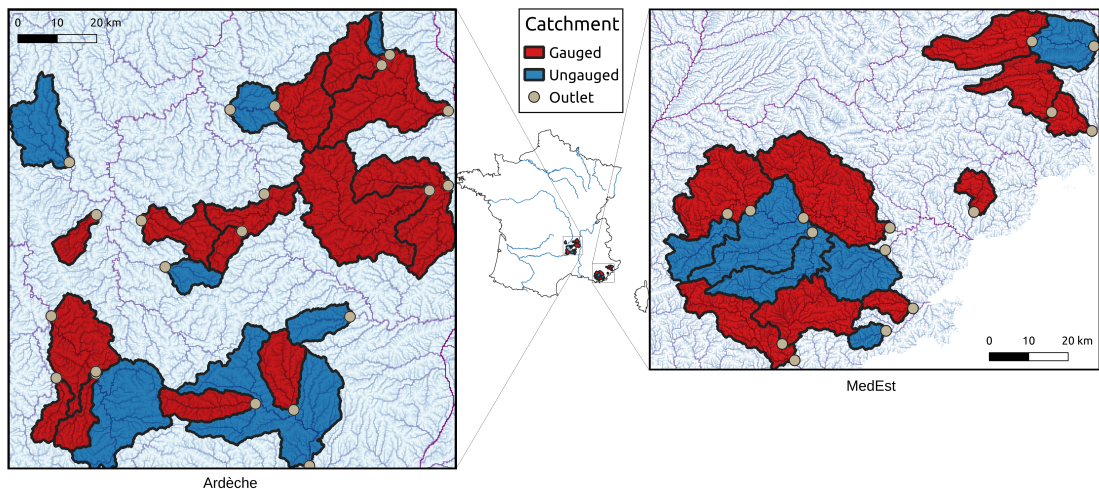


Figure 2. Two study areas on the map on France: Ardèche (left) and MedEst (right). For each area, the gauged catchments (red) are used for multi-site calibration while the pseudo-ungauged catchments (blue) are used for spatial validation.

In order to examine the spatio-temporal extrapolation capabilities of HDA-PR, an a priori partition of the available discharge stations is made into calibration sites and pseudo-ungauged catchments for validation. We selected as gauged stations for regionalization those with good local model performances (i.e., “donor” catchments with potentially lower modeling error). Discharge time series at gauged sites are also split into a calibration and a validation period. A set of 7 physical descriptors (Table 1) available over the whole French territory is used following Odry (2017) and Jay-Allemand et al. (2022b). Note that this setup is sufficient to assess the regionalization performance of the proposed algorithms while keeping the present article concise. The issue of selecting the most relevant information (multi-source observations of hydrological responses and physical descriptors) is intentionally left for future research since it requires several upgrades to the algorithms (dedicated observation operators, descriptors selection layers, etc.). It is worth noting that prior to the optimization process, all descriptors are standardized between 0 and 1 through min-max scaling.

Table 1. Descriptors used as input data for regionalization methods.

Notation	Type	Description	Unit	Source
d_1	Topography	Slope	°	Odry (2017)
d_2	Morphology	Drainage density	-	Organde et al. (2013)
d_3	Influence	Percentage of basin area in karst zone	%	Caruso et al. (2013)
d_4	Land use	Forest cover rate	%	CLC European Union (2012)
d_5	Land use	Urban cover rate	%	CLC European Union (2012)
d_6	Hydrogeology	Potential available water reserve	mm	Poncelet (2016)
d_7	Hydrogeology	High storage capacity basin rate	%	Finke et al. (1998)

The first zone is located in the Eastern Mediterranean region and is called “MedEst”. 9 gauged catchments are used for calibration while 6 others are considered as pseudo-ungauged for validation. The second zone is located in the Ardèche region and called “Ardèche”. 14 catchments are considered as gauged while 7 are considered as pseudo-ungauged. Both study areas constitute challenging cases, with contrasted hydrological properties including steep topography and very heterogeneous soils and bedrock (e.g., Garambois et al. (2015)). They are affected by intense rainfall that trigger non linear flash flood responses. Moreover, the MedEst area is the most difficult to model because of the significant proportion of karstic zones. The selection of catchments is based on the availability of long time series with high quality of observed flow and limited anthropogenic impacts. The SMASH model is run on a $dx = 1$ km spatial grid at $dt = 1$ h time step. It is forced by: (i) observed rainfall grids based on hourly ANTILOPE J+1 radar-gauge rainfall reanalysis from Météo-France (Champeaux et al., 2009); (ii) potential evapotranspiration (PET) estimated using the formula of Oudin et al. (2005); and (iii) temperature data from SAFRAN reanalysis produced by Météo-France on a 8×8 km² spatial grid (Quintana-Seguí et al., 2008) downscaled to a 1×1 km² spatial grid.

For each study area, we perform multi-site regional calibration methods using gauged catchments. The chosen calibration metric is the NSE, computed using data from multiple gauges over a ten-year period (2006-2016). The following calibration methods are compared:

- Local calibrations for each gauge, both with spatially uniform (i.e., $\rho \equiv \bar{\theta}$) and full spatially distributed controls (i.e., $\rho \equiv \theta(x)$), that are respectively under and over parameterized hydrological optimization problems. These represent reference performances, denoted “Uniform (local)” and “Distributed (local)”.

And multi-gauge regional calibrations with:

- lumped model parameters (i.e., $\rho \equiv \bar{\theta}$) which somehow represents “level 0” regionalization, denoted “Uniform (regionalization)”;
- a multivariate linear mapping (i.e., $\rho \equiv [\alpha_{k,0}, (\alpha_{k,d}, 1)]^T$), denoted “Multi-linear (regionalization)”;
- a multivariate polynomial mapping (i.e., $\rho \equiv [\alpha_{k,0}, (\alpha_{k,d}, \beta_{k,d})]^T$), denoted “Multi-polynomial (regionalization)”;
- a multilayer perceptron (i.e., $\rho \equiv [W, b]^T$), denoted “ANN (regionalization)”.

To ensure robust validation, model performances are assessed in terms of spatial validation, temporal validation, and spatio-temporal validation, with a validation period covering two years from 2016 to 2018. Various evaluation metrics are also used, including multiple hydrological signatures-based metrics from Huynh et al. (2023).

3.2 Regional Learning Performance

In this section, the regional learning performance is analyzed with primary focus on the MedEst region, which represents the most challenging study area. First, Figure 3 suggests that using lumped model parameters $\bar{\theta}$ leads to a poor performance in simulating discharges, whereas the other three regional learning methods result in similar signals with remarkably improved performance, in both the gauged and the pseudo-ungauged catchments.

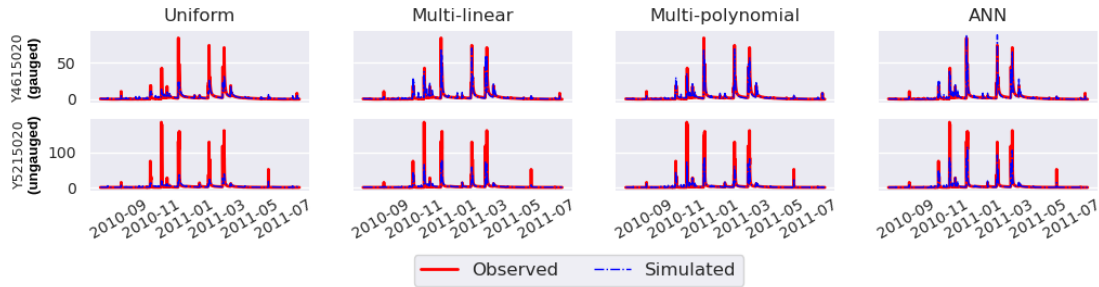


Figure 3. Study area: MedEst. One year of observed and simulated discharges (in m^3/s) obtained with four multi-gauge regional calibration methods at one gauged (Y4615020) and one pseudo-ungauged catchment (Y5215020).

Given the complexity and heterogeneity of the region, it is unsurprising that lumped model parameters regionalization is unable to accurately reproduce contrasted hydrological responses. Figure 4 demonstrates that the three regionalization methods incorporating information from physical descriptors lead to largely superior performance when compared to the Uniform baseline for the calibration catchments. For example, the NSE scores increase from 0.3 to more than 0.7 in catchment Y5615030,

or from 0.4 to more than 0.75 in catchment Y5615010. In addition, these three methods yield relatively high scores in the pseudo-ungauged catchments used for spatial validation (e.g., in Y5215020 where the NSE score has been improved from 0.36 up to 0.83 in the case of ANN). Note that some stations have many missing data: for instance, 32 % of the data are missing for catchment Y5106610, and 21 % for catchment Y5005210. This may affect the NSE scores and at least partly explain the poor validation performances for all methods.

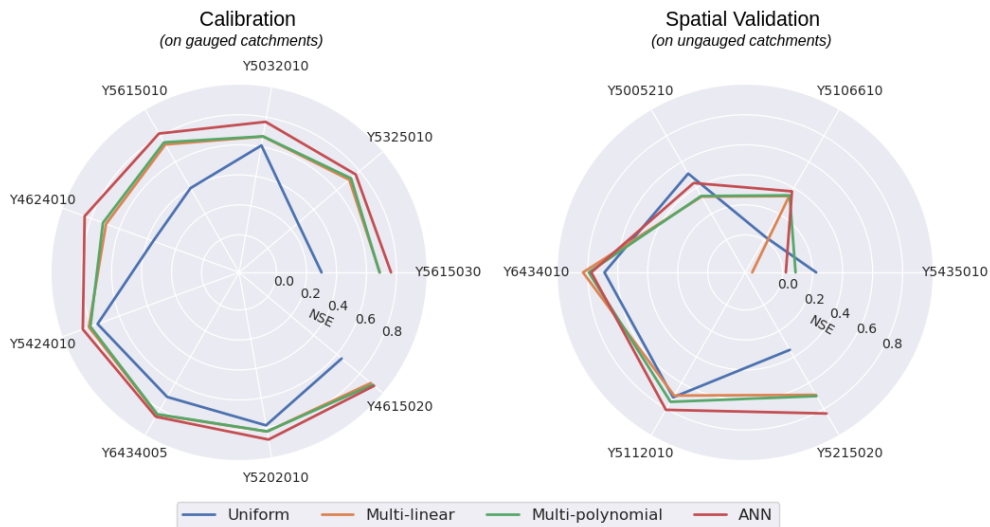


Figure 4. Study area: MedEst. Radial plots of the NSE (optimal value = 1) in gauged catchments (left) and pseudo-ungauged catchments (right) during the calibration period for four multi-gauge regional calibration methods.

We now focus on the global results in both study areas, Ardèche and MedEst, as presented in Figure 5. This figure summarizes the NSE scores for the two local calibration methods (spatially uniform and spatially distributed calibrations) and the four regionalization methods. A rapid interpretation reveals that in the Ardèche area, making the model parameters vary in space seems unnecessary, given the already good performance of the lumped parameters (Uniform) methods in the context of a less complex area. However, it is worth noting that the three regionalization methods also achieve satisfying results, with the median efficiencies (NSE) exceeding 0.65 in both gauged and pseudo-ungauged catchments during the validation period. This suggests that while not necessary, using a non-uniform regionalization approach does not deteriorate the performance in spatial or temporal validation. In contrast and as evidenced in the preceding figures, significant disparities in efficiency scores are observed among the regionalization methods in the MedEst area, where the multivariate regressions and ANN outperform the uniform regionalization baseline calibrated with lumped model parameters. The three methods demonstrate promising results, achieving median NSE scores above 0.72, 0.51, 0.76 and 0.61 (compared to 0.6, 0.41, 0.2 and 0.16 for the baseline) respectively, across calibration, spatial validation, temporal validation and spatio-temporal validation. Specifically, when considering temporal validation, the ANN reaches a median NSE score of approximately 0.9 with eight out of nine stations exceeding a score of 0.8, which is comparable to the performance of the local calibration reference with spatially distributed controls.

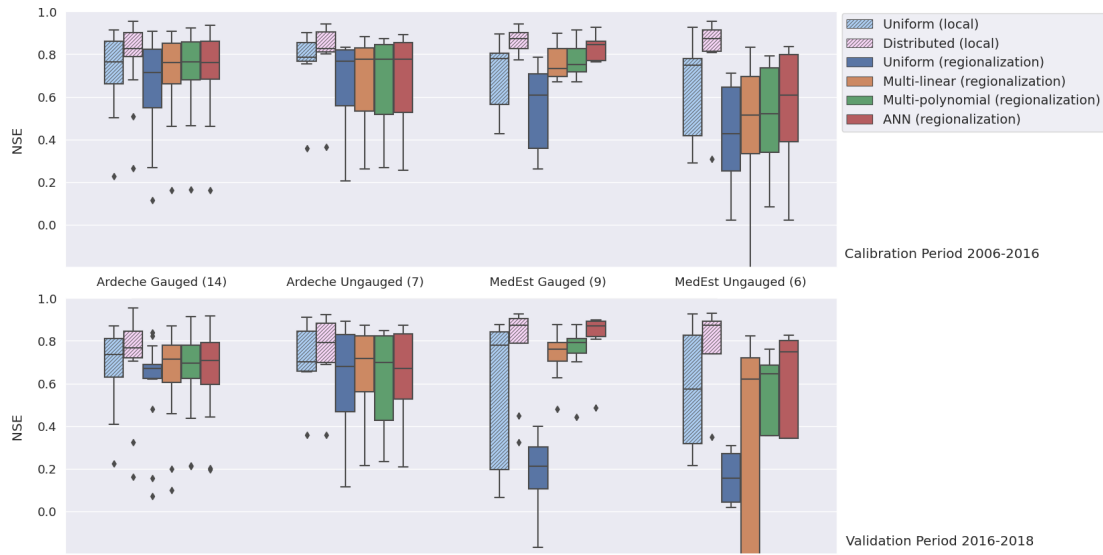


Figure 5. Comparison of NSE scores across gauged and pseudo-ungauged catchments during the calibration (top) and validation (bottom) periods for two local calibration methods and four regionalization methods in both study areas. The numbers in parentheses indicate the number of catchments included in each boxplot.

In order to obtain a more robust evaluation criterion adapted to flood modeling, we consider validation in terms of multiple evaluation metrics based on hydrological signatures for flood events, which are computed via an automated segmentation algorithm proposed by Huynh et al. (2023). Turning the focus back to MedEst, the three non-uniform regionalization methods, with a particular focus on the ANN, demonstrate their ability to outperform the uniform regionalization (lumped model parameters), in both temporal and spatio-temporal validation, as shown in Figure 6. The relative errors (median over flood events) of simulated signatures using these three regionalization methods are around 0.2 compared to more than 0.8 for the uniform baseline in temporal validation, and between 0.2 and 0.6 compared to more than 0.7 for the baseline in spatio-temporal validation. Especially, in the case of ANN, the error in simulating the peak flow is drastically reduced from 0.85 and 0.91 (for the baseline) to 0.28 and 0.43, respectively, in temporal and spatio-temporal validation. It is noteworthy that these flood signatures-based metrics were not included in the cost function during the calibration process, which further supports the robustness and power of the regionalization methods, particularly the one based on the ANN. It thus underscores the potential of such approaches for enhancing flash flood forecasting system.

Finally, Figure 7-b shows the parameter maps obtained with the three non-uniform regionalization methods, that can be physically interpreted. In general, the spatial patterns of the drainage density (d_2) in Figure 7-a are easily observable in the distributed parameter maps in the two regression cases, as evidenced by strong correlations between this descriptor and the four model parameters. Such correlations can be quantified using one-to-one parameter-descriptor correlation matrices (Figure 7-c). The relations between the model parameters and physical descriptors are not limited to linear and polynomial forms in the ANN case, leading to parameter maps that are quite distinct from the regression-based ones. Furthermore, it is noteworthy that

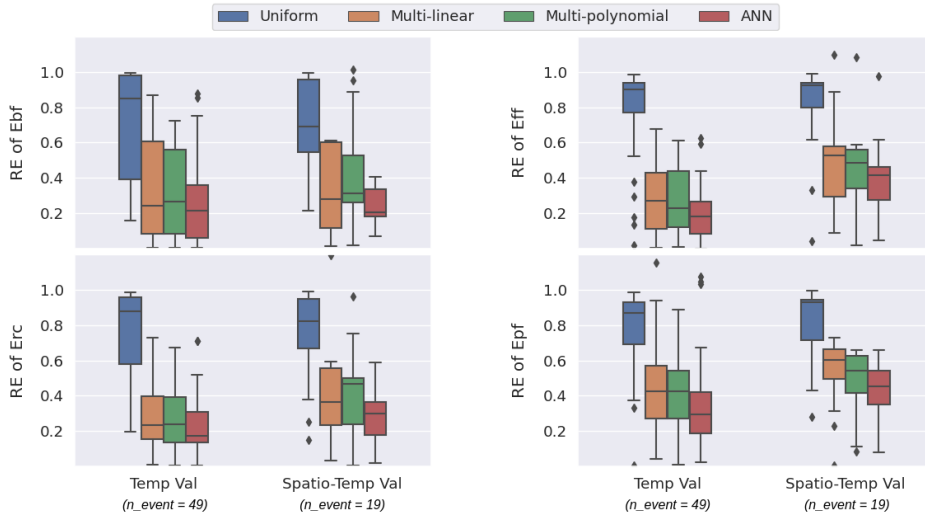


Figure 6. Study area: MedEst. Relative error (RE, optimal value = 0) of 4 flood event signatures: base flow (Ebf), flood flow (Eff), runoff coefficient (Erc) and peak flow (Epf), evaluated using $n_{event} = 49$ flood events at gauged catchments for temporal validation (Temp Val) and $n_{event} = 19$ flood events at pseudo-ungauged catchments for spatio-temporal validation (Spatio-Temp Val).

the ANN can identify stronger correlations (compared to both regression cases) for the routing parameter (l_r) and the transfer parameter (c_{ft}) with the karst index (d_3). For Ardèche, parameter maps show a reduced variability (see Figure C4-b), and the ANN identifies maps and relations between descriptors and parameters which are similar to those found in both regression cases (see Figure C4-c). This finding suggests that the ANN can effectively approximate the multi-linear regression when dealing with areas like Ardèche, where it does not necessitate more intricate mappings. This demonstrates the robustness and flexibility of the ANN, which proves to be efficient not only in complex study areas but also in simpler ones.

The corresponding figures in this section depicting the results in Ardèche can be found in Appendix C. The following section focuses on results more specific to the optimization process and provide insights on the use of multivariate regressions and ANN in both study areas.

4 Discussions

Learning the spatial variability of conceptual hydrological parameters may be difficult to achieve with simple regionalization methods (e.g., based on multi-linear regression). However, while complex regional mappings can reduce the misfit between observed and simulated hydrological responses, their ability to produce physically interpretable results may be questioned. This section presents compelling findings and insightful observations obtained from the calibration using HDA-PR, with a particular focus on the learning process in the case of ANN.

When calibrating a model with gradient-based optimization algorithms, it is important to discuss the descent of the cost function. This analysis enables understanding how optimization algorithms converge towards the global or local minimum

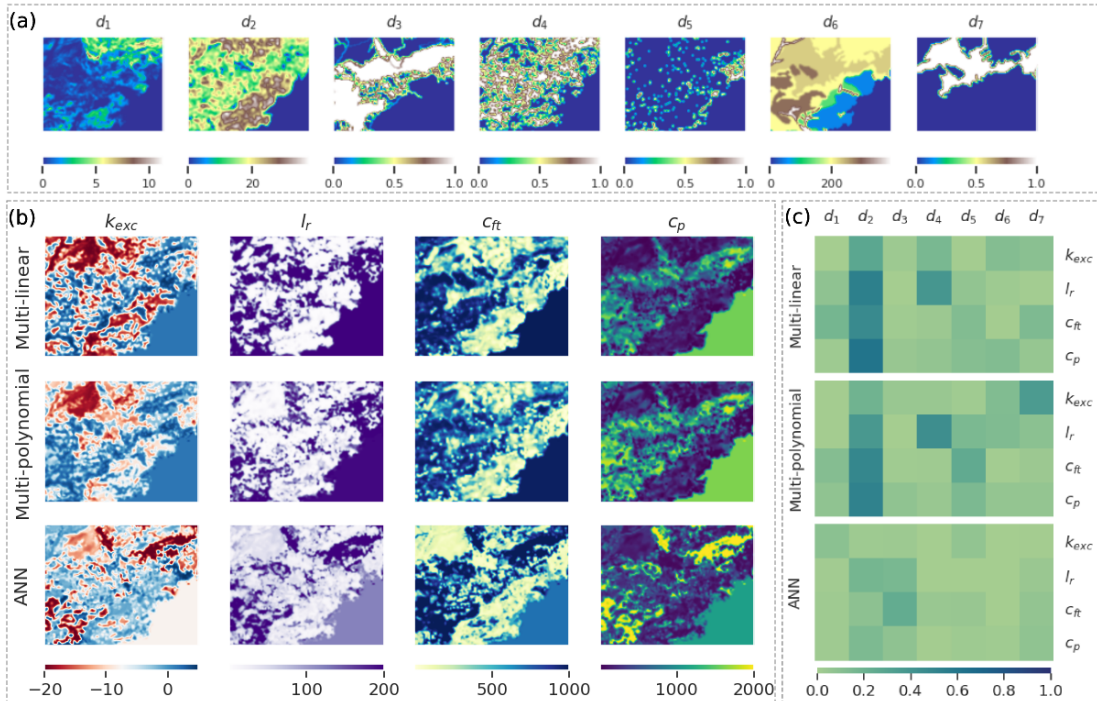


Figure 7. Study area: MedEst. Sub-figure a: Maps of input descriptors (d_1 - d_7), whose information is provided in Table 1. Sub-figure b: Calibrated hydrological parameters (k_{exc} , l_r , c_{ft} , c_p) maps for three non-uniform regionalization methods. Sub-figure c: Linear correlation between descriptor and parameter for the three regionalization methods.

of the cost function, and identifying potential trade-offs between model flexibility and overparameterization, in addition to validation results. Similar results are observed in both study areas, including the descent of the cost function J and its projected gradient $\nabla_{\rho} J$, which are represented in Figure 8 for the MedEst area.

It is apparent that the cost functions in the two regression methods start from a more optimal point (approximately 0.46) than the ANN since they use a uniform background solution obtained by a global optimization method, as mentioned in 2.3.1. Furthermore, they converge after around 200 iterations and remain monotonous throughout the optimization process. The polynomial regression approach achieves a slightly lower cost than the linear approach, due to the fact that it has nearly twice the number of parameters (Table 2), while the ANN, with a significantly larger number of parameters, can achieve a lower cost despite starting from a higher cost. Moreover, the ANN cost function in the left panel of Figure 8 is not monotonous throughout iterations: it shows localized increases for several instances within the first 100 iterations.

The variability of the projected cost gradient shown in the right panels of Figure 8 suggests that the surface of the cost function is more complex for ANN than for the two regressions. This is confirmed by the gradient values in the right panels of Figure 8: multi-linear regression lead to projected gradients that are marginally higher than those of multi-polynomial regression but much higher than those of ANN. This discrepancy can be attributed to the complexity of the regionalization mapping in each

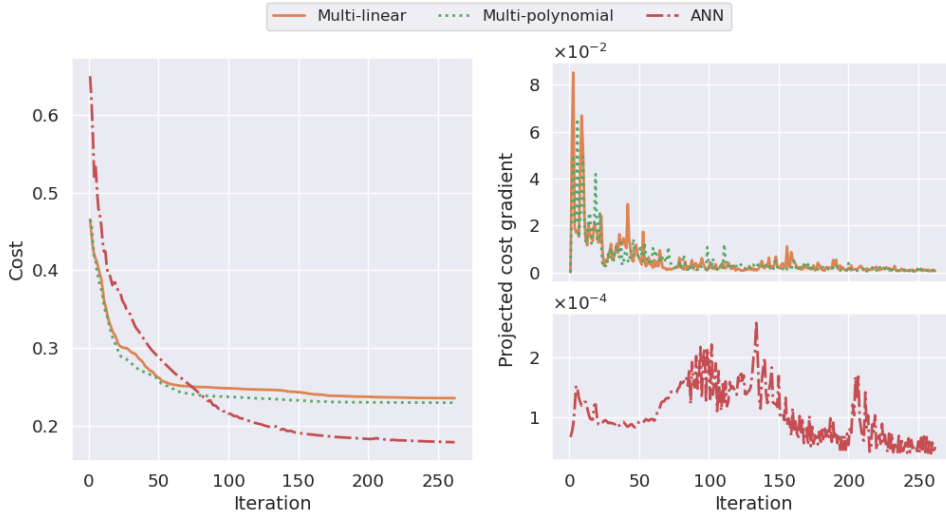


Figure 8. Study area: MedEst. The descent of the cost $J = 1 - NSE$ (left) and projected gradient $\|\nabla_{\rho} J\|_{\infty}$ (right) for the three non-uniform regionalization methods.

Table 2. Number of parameters in the control vector for the four regionalization methods studied, where $N_{\theta} = 4$ and $N_D = 7$. A detailed calculation of the total number of parameters for the ANN is provided in Table B1. The last two columns show the ratios between the number of parameters in the four methods studied and the fully distributed calibration method (i.e., $\rho \equiv \theta(x)$), which was overparameterized and unable to perform regionalization.

Mapping	Coefficient	Exponent	Total parameters	% (MedEst)	% (Ardèche)
Uniform	$N_{\theta} \cdot 1 = 4$	0	4	0.03	0.04
Multi-linear	$N_{\theta}(N_D + 1) = 32$	0	32	0.21	0.29
Multi-polynomial	$N_{\theta}(N_D + 1) = 32$	$N_{\theta} \cdot N_D = 28$	60	0.39	0.54
ANN			6276	40.3	56.48

Fully distributed: $N_{\theta} \cdot N_x$. Remind that N_x is the number of active cells within the spatial domain Ω .

case, with the ANN being the most complex. Moreover, jumps in the projected gradient (bottom right panel) suggest that the solutions of the ANN can explore different paths and avoid getting stuck in local minima: there are four instances (around iterations 5, 90, 130, and 210) where significant changes occur in the control vector space (biases and weights) to locate the optimum. This property could be essential in tackling equifinality and reach robust global optimum even with different starting points.

To alleviate the vanishing gradient problem inherent in the ANN, we employed several techniques commonly used in the machine learning community. First, we applied Xavier initialization (Glorot and Bengio, 2010) to the weights, maintaining a reasonable magnitude of the gradients. Second, we utilized the ReLU activation function or its variants in the hidden layers, enabling the gradient to flow more freely through the network. Third, we varied the number of hidden layers between 2 and 4,

striking a balance between network flexibility and exacerbation of the vanishing gradient problem. Ultimately, we employed a relatively high initial learning rate (e.g., 0.005) to prevent the gradients from shrinking excessively during training.

One can also inquire about the way the different regionalization mappings transfer information from physical descriptors to hydrological parameters. For instance, how can we ensure that HDA-PR effectively uses the physical patterns provided by descriptor maps to constrain the estimation of meaningful spatial parameter maps? To address this question, it is worth noting that the “safest” approach is to use multi-linear regression, which corresponds to a simple weighted average of the descriptor patterns. In the case of multi-polynomial regression, the risk of losing physical properties may arise when the polynomial degree is unbounded. To mitigate this risk, this study imposed bounds on the polynomial degree, $0.5 \leq \beta_{k,d} \leq 2$, as mentioned earlier in 2.3.1. The ANNs, however, pose the most complicated case, where the control vector (that is, the weights and biases) consists of numerous parameters that are difficult to physically constrain. Our hands-on experience indicates that a multilayer perceptron with two or three hidden layers is sufficient for learning the parameters of a parsimonious conceptual distributed hydrological model without under- or over-extracting the physical information of the input descriptors. Note that the number of neurons in each layer must be reasonable, which should not exceed $\sqrt{N_D \cdot N_x}$ based on our experiments.

Regarding regionalization over larger areas, such as for large basins or at country scales, for dealing with significant physical heterogeneity, an increased flexibility in the regional mapping might be needed. This can be achieved through the use of spatialized regional controls, for example as done in regional calibration for catchment clusters determined with a similarity measure (Huang et al., 2019). In the proposed HDA-PR framework, the definition of transfer functions allows considering flexible mappings, as well as spatialized regional controls through masked descriptor maps, for each hydrological parameter independently or jointly. This enables the exploration of transfer functions on a clustering of the spatial domain, for example into sub-regions or hydrological response units (HRU). This would certainly be necessary to circumvent the rigidity of the linear or multi-linear mappings, but maybe not for a flexible one such as the ANN. In this work, the capabilities of HDA-PR have been successfully demonstrated in a high-dimensional and challenging high-resolution flash-flood modeling context. Determining effective physical descriptor sets from large databases as well as finding optimal spatial flexibility represent interesting research avenues for constructing optimal regional data assimilation approaches.

5 Conclusions

A Hybrid Data Assimilation and Parameter Regionalization (HDA-PR) approach has been introduced in this study. We investigated the potential of incorporating learnable regionalization mappings, including multi-variate polynomial regressions and neural networks, into a differentiable high-resolution hydrological model. To the best of our knowledge, we present the first implementation of ANNs within this context, enabling a seamless regionalization in hydrology. Effective optimization algorithms capable of performing high-dimensional optimizations from multi-source data have been obtained with:

- effective regional transfer functions of adaptable complexity, enabling the use of information from heterogeneous data sources, with flexible formulations and various degrees of spatial rigidity;

- a differentiable forward hydrological model, embedding the regional mappings, that enables accurate computation of spatially distributed gradients of the multi-gauge cost function - which is crucially needed in the context of sparse observations (i.e., cost evaluation locations), and relatively small gradient values and spatial variability;
- optimization algorithms, adapted to high dimensional problems, with seamless flow of cost gradients, especially when combined with physical descriptors and spatial gradients, which efficiently enhance the transferability of geophysical properties from gauged to ungauged locations.

HDA-PR has been tested on two very challenging regional optimization cases from multi-gauge discharge and descriptors maps, with a high-resolution conceptual hydrological modeling at 1 km^2 . The results obtained on both zones, and especially on the most challenging MedEst case, highlight the effectiveness of HDA-PR that utilizes physical descriptors, surpassing the performance of a uniform regionalization method with lumped model parameters. Notably, the ANN exhibited superior performance, even comparable to the reference benchmarks, thereby establishing its remarkable capability in challenging modeling scenarios as well as its capacity to collapse to a simpler mapping in less challenging ones. The median NSE scores of all HDA-PR methods (including multivariate regressions and ANN) are greater than 0.75 in calibration in both study areas, whereas the scores in spatio-temporal validation exceed 0.6 in MedEst (versus 0.85 for the the local distributed calibration which represents the upper limit) and exceed 0.65 in Ardèche (versus 0.8). Various flood event signatures computed from Huynh et al. (2023) are also used as validation metrics to demonstrate the robustness of HDA-PR, where the three regionalization methods using descriptors outperform the uniform regionalization method. For instance, when considering spatio-temporal validation across 19 flood events in MedEst, these three methods yield median relative errors of the peak flow below 0.6, whereas the uniform method yields errors exceeding 0.9.

This research and the proposed algorithms open several perspectives. Immediate work focuses on: (i) the testing and improvement of HDA-PR for application at national scales and on other continents; (ii) study of effective descriptor selection along with multi-gauge cost functions explicitly accounting for data uncertainties, and optimal spatial clustering of regional controls, for example into HRU; (iii) study of a global Bayesian estimator to improve the first-guess determination, especially with the multi-polynomial mapping. Adding a learnable descriptors selection and ingestion layer on top of the regionalization transfer functions would enable the exploration of even larger databases including categorical data. HDA-PR can be extended to state and composite parameters-states optimization which could be very interesting for multi-scale DA and real time model correction from multi-source and multi-site data. Finally, the method is transposable to regionalization of differentiable integrated hydrological-hydraulic networks models (e.g., Pujol et al. (2022)) and could be used to explore regionalization potential from cocktails of in-situ and satellite data, including the forthcoming SWOT (Surface Water and Ocean Topography satellite mission) observations of water surfaces variabilities of worldwide larger rivers. In general, its applicability extends beyond hydrological models and can be adapted to other geophysical models.

Appendix A: Metrics

Denote $Q(t)$ and $Q^*(t)$ being the simulated and observed discharge time series. The hydrological cost functions studied are:

- observation cost function based on the Nash-Sutcliffe Efficiency (NSE):

$$1 - NSE = \frac{\sum_{t=t^*}^T (\mathbf{Q}^*(t) - \mathbf{Q}(t))^2}{\sum_{t=t^*}^T (\mathbf{Q}^*(t) - \overline{\mathbf{Q}^*})^2}$$

- observation cost function based on the Kling-Gupta Efficiency (KGE):

$$1 - KGE_2 = a_1 (r(\mathbf{Q}^*(t), \mathbf{Q}(t)) - 1)^2 + a_2 (\beta(\mathbf{Q}^*(t), \mathbf{Q}(t)) - 1)^2 + a_3 (\alpha(\mathbf{Q}^*(t), \mathbf{Q}(t)) - 1)^2$$

with r , β and α being respectively measures of the correlation, bias and variability of observation with respect to simulated discharge time series; $\sum_{i=1}^3 a_i = 1$. This function is quadratic and differentiable.

Appendix B: Incorporating ANN into the differentiable hydrological model

This appendix details the neural network design and the derivation of hydrological cost gradients for the ANN-based regionalization algorithm.

A simple ANN denoted \mathcal{N} , consisting of N_L fully connected (dense) layers, intends to learn the descriptors-to-parameters field mapping in the 2D spatial domain, from $\mathbf{D}(x) \in \mathbb{R}^{N_D}$ to $\boldsymbol{\theta}(x) \in \mathbb{R}^{N_\theta}, \forall x \in \Omega$ (Figure B1).

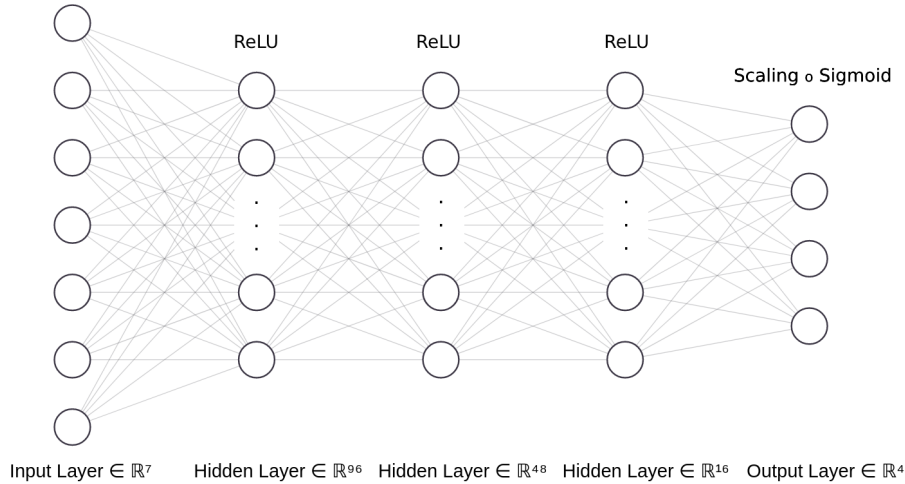


Figure B1. The architecture of the ANN consists of three hidden layers followed by the ReLU activation function and an output layer that uses the Sigmoid activation function in combination with a scaling function. In this particular case, we have $N_D = 7$, $N_L = 4$ and $N_\theta = 4$. The calculation of the total number of trainable parameters for this architecture is detailed in Table B1.

Let us consider an ensemble of layers where each layer is associated with its weight \mathbf{W}_j and bias \mathbf{b}_j . Then, an input I of each layer is mapped to the input of the next layer by a linear function $\phi_j(I) = \mathbf{W}_j I + \mathbf{b}_j$, and followed by the ReLU activation function denoted δ , except for the last layer, which is followed by the Sigmoid activation function denoted σ , ensuring that

Table B1. Number of parameters of the ANN where $N_D = 7$, $N_L = 4$ and $N_\theta = 4$.

	Hidden layer 1	Hidden layer 2	Hidden layer 3	Output layer
Input shape	$(N_D,)$	$(96,)$	$(48,)$	$(16,)$
Number of neurons	96	48	16	N_θ
Number of parameters	$N_D \cdot 96 + 96 = 768$	$96 \cdot 48 + 48 = 4656$	$48 \cdot 16 + 16 = 784$	$16 \cdot N_\theta + N_\theta = 68$
Total parameters: 6276.				

its outputs are between 0 and 1. Now an output $\mathbf{O}_x = \sigma \circ \phi_{N_L}(\cdot, x) \in [0, 1]^{N_\theta}$ of the last layer is mapped to the range of the hydrological model parameters by a differentiable scaling function s :

$$\boldsymbol{\theta}(x) = s(\mathbf{O}_x) = l + (u - l) \odot \mathbf{O}_x \quad (\text{B1})$$

where $l = (l_1, \dots, l_{N_\theta})$ and $u = (u_1, \dots, u_{N_\theta})$ with the lower and upper bounds $l_k \in \mathbb{R}$ and $u_k \in \mathbb{R}$, assumed spatially uniform, defining the bound constraints of $\theta_k(x), \forall (k, x) \in [1..N_\theta] \times \Omega$, in the direct hydrological model. The notation “ \odot ” denotes the Hadamard product.

Noting $\Psi_j \equiv \begin{cases} \delta \circ \phi_j, j = 1..N_L - 1 \\ \sigma \circ \phi_j, j = N_L \end{cases}$, the forward propagation of the neural network \mathcal{N} is defined as Equation B2.

$$\boldsymbol{\theta}(x) = \mathcal{N}(\mathbf{D}(x), \cdot) = s \circ \Psi_{N_L} \circ \Psi_{N_L-1} \circ \dots \circ \Psi_1(\mathbf{D}(x)), \forall x \in \Omega. \quad (\text{B2})$$

Here, the notation “ \circ ” denotes the function composition operator.

Recall that our objective is the calibration problem of Equation 9 with respect to the regional control vector $\boldsymbol{\rho} := [\mathbf{W}, \mathbf{b}]$, using the cost function of Equation 10. In such manner, different variants of stochastic gradient descent algorithm are used and thus require the gradients of the cost function with respect to the weights and biases $\frac{\partial J}{\partial \boldsymbol{\rho}_j}$ for each layer, where $\boldsymbol{\rho}_j := [\mathbf{W}_j, \mathbf{b}_j]$. Since the forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(\cdot, \mathcal{N}(\cdot))$ with $\boldsymbol{\theta}$ being both the output of \mathcal{N} and the input of \mathcal{M}_{rr} , we can write $\frac{\partial J}{\partial \boldsymbol{\rho}_j} = \frac{\partial J}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\rho}_j}$. Then these two gradients are obtained as follows:

- The gradients $\frac{\partial J}{\partial \boldsymbol{\theta}}$ of the cost function with respect to the hydrological model parameters, computed by solving the numerical adjoint model of \mathcal{M}_{rr} ;
- The gradients $\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\rho}_j}$ of the network output with respect to the weight and bias, computed using the chain rule of composite functions of \mathcal{N} .

Eventually, the backward propagation for updating the weights and biases, using for instance Adam optimizer, is described in Algorithm B1.

Appendix C: Calibration results in case of Ardèche

This section provide several calibration/validation results obtained in the Ardèche area.

Algorithm B1 Adapted back-propagation using Adam optimizer

▷ Randomly initialized weights and biases $\boldsymbol{\rho}^{(0)} = (\boldsymbol{\rho}_1^{(0)}, \dots, \boldsymbol{\rho}_{N_L}^{(0)})$

▷ Number of training iterations N_{ite}

for $i = 1..N_{ite}$ **do**

▷ Forward propagation over the spatial domain $\boldsymbol{\theta} \leftarrow \left[\left(\mathcal{N} \left(\mathbf{D}(x), \boldsymbol{\rho}^{(i-1)} \right) \right)_{x \in \Omega} \right]^T$

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

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

▷ Gradient computation $\frac{\partial J}{\partial \boldsymbol{\rho}_j} \leftarrow \left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\rho}_j} \right)^T \nabla A$

▷ Updated gradient accumulation $\nabla A \leftarrow \nabla A \cdot \left[W_j^{(i-1)} \right]^T$

▷ Updated weights and biases $\boldsymbol{\rho}_j^{(i)} \leftarrow \boldsymbol{\rho}_j^{(i-1)} - \eta \frac{m^{(i)}}{(1-\beta_1) \left(\sqrt{\frac{v^{(i)}}{1-\beta_2}} + \epsilon \right)}$ where:

$$m^{(i)} \leftarrow \beta_1 m^{(i-1)} + (1-\beta_1) \frac{\partial J}{\partial \boldsymbol{\rho}_j} \left(\boldsymbol{\rho}_j^{(i-1)} \right)$$

$$v^{(i)} \leftarrow \beta_2 v^{(i-1)} + (1-\beta_2) \left(\frac{\partial J}{\partial \boldsymbol{\rho}_j} \left(\boldsymbol{\rho}_j^{(i-1)} \right) \right)^2$$

$\beta_1 = 0.9$ and $\beta_2 = 0.999$ are the decay rates for first and second moments of gradients

$\epsilon = 10^{-8}$ is a small scalar

η is the learning rate that is a tuning parameter determining the step size of the optimization problem

end for

end for

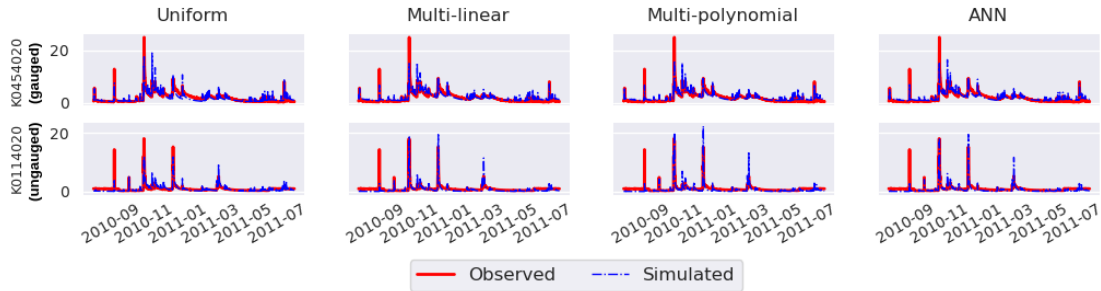


Figure C1. Study area: Ardèche. One year of observed and simulated discharges (in m^3/s) obtained with four multi-gauge regional calibration methods at one gauged (K0454020) and one pseudo-ungauged catchment (K0114020).

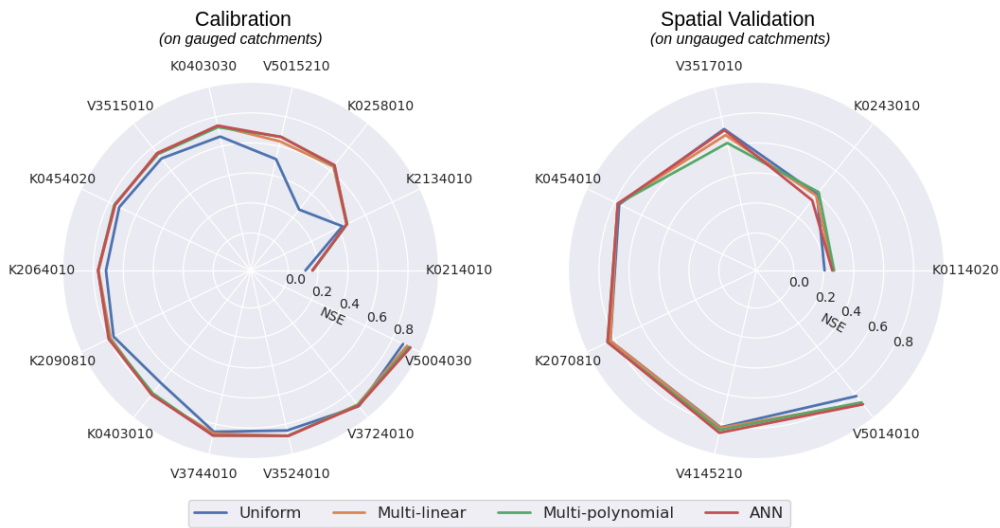


Figure C2. Study area: Ardèche. Radial plots of the NSE (optimal value = 1) in gauged catchments (left) and pseudo-ungauged catchments (right) during the calibration period for four multi-gauge regional calibration methods.

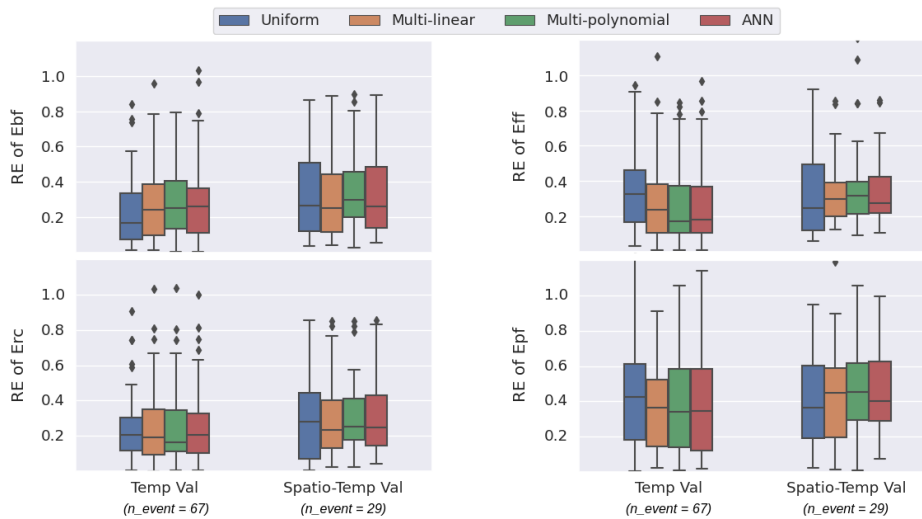


Figure C3. Study area: Ardèche. Relative error (RE, optimal value = 0) of 4 flood event signatures: base flow (Ebf), flood flow (Eff), runoff coefficient (Erc), peak flow (Epf), evaluated using $n_{event} = 67$ flood events at gauged catchments for temporal validation (Temp Val) and $n_{event} = 29$ flood events at pseudo-ungauged catchments for spatio-temporal validation (Spatio-Temp Val).

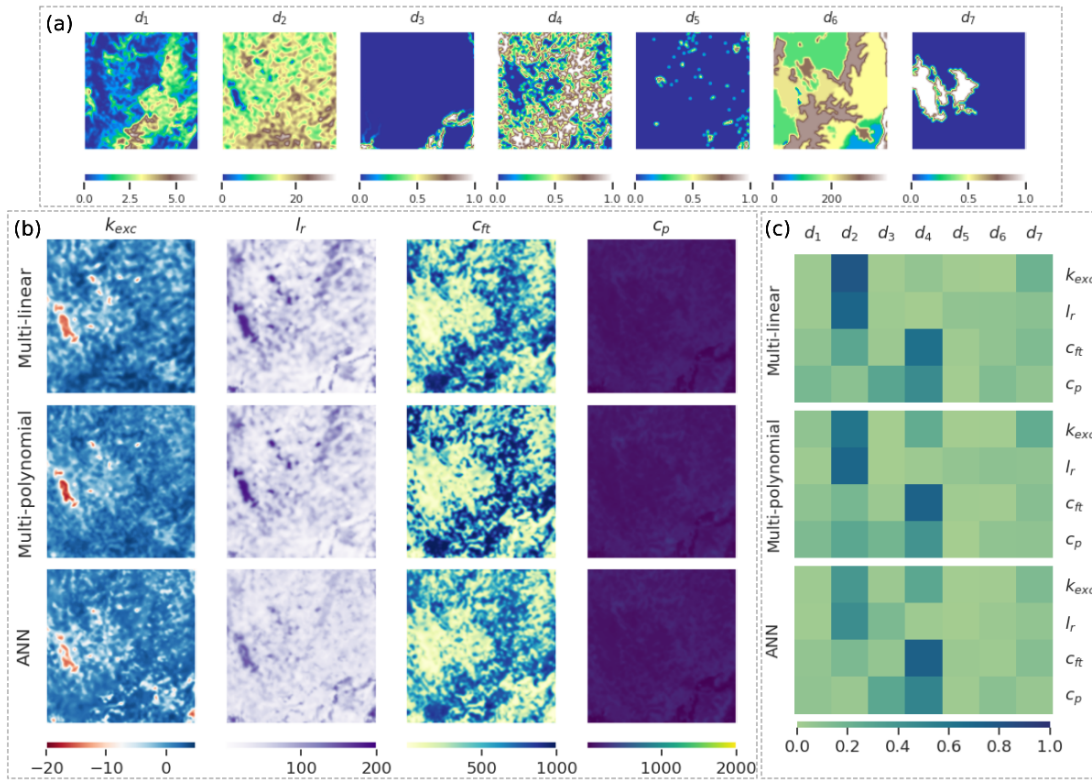


Figure C4. Study area: Ardèche. Sub-figure a: Maps of input descriptors (d_1 - d_7), whose information is provided in Table 1. Sub-figure b: Calibrated hydrological parameters (k_{exc} , l_r , c_{ft} , c_p) maps for three non-uniform regionalization methods. Sub-figure c: Linear correlation between descriptor and parameter for the three regionalization methods.

Code availability. The proposed algorithms were implemented into the SMASH platform, which is interfaced in Python (Jay-Allemand et al., 2022a). The SMASH source code is licensed under GPL-3.0 and can be accessed at: <https://github.com/DassHydro-dev/smash>. Additionally, the code for conducting the numerical experiments and analysis in this study is licensed under MIT and provided at: <https://github.com/nghi-truyen/Regionalization-Learning>.

Acknowledgements. The authors greatly acknowledge SCHAPI-DGPR and Météo-France for providing data used in this study; Killian Pujol-Nicolas and Maxime Jay-Allemand for their contributions in the preliminary stages of this work; Etienne Leblois from INRAE Riverly (Lyon) for fine terrain elevation processing at multiple scale over French territory. This work was supported by funding from SCHAPI-DGPR, ANR grant ANR-21-CE04-0021-01 (MUFFINS project, “MULTIscale Flood Forecasting with INnovating Solutions”), and NEPTUNE European project DG-ECO.

References

- Abdulla, F. A. and Lettenmaier, D. P.: Development of regional parameter estimation equations for a macroscale hydrologic model, *Journal of hydrology*, 197, 230–257, 1997.
- Althoff, D., Rodrigues, L. N., and da Silva, D. D.: Addressing hydrological modeling in watersheds under land cover change with deep learning, *Advances in Water Resources*, 154, 103965, 2021.
- Bastola, S., Ishidaira, H., and Takeuchi, K.: Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe, *Journal of Hydrology*, 357, 188–206, <https://doi.org/10.1016/j.jhydrol.2008.05.007>, 2008.
- Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, *Water Resources Research*, 52, 3599–3622, 2016.
- Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I., and Wood, E. F.: Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031485, 2020.
- Beven, K.: How far can we go in distributed hydrological modelling?, *Hydrology and Earth System Sciences*, 5, 1–12, <https://doi.org/10.5194/hess-5-1-2001>, 2001.
- Blöschl, G., Sivapalan, M., Wagener, T., Savenije, H., and Viglione, A.: *Runoff prediction in ungauged basins: synthesis across processes, places and scales*, Cambridge University Press, 2013.
- Caruso, A., Guillot, A., and Arnaud, P.: Notice sur les indices de confiance de la méthode Shyreg-Débit-Définitions et Calculs, in: *Aix en Provence: IRSTEA, Convention DGPR/SNRH*, 2013.
- Castaigns, W., Dartus, D., Le Dimet, F.-X., and Saulnier, G.-M.: Sensitivity analysis and parameter estimation for distributed hydrological modeling: potential of variational methods, *Hydrology and Earth System Sciences*, 13, 503–517, <https://doi.org/10.5194/hess-13-503-2009>, 2009.
- Champeaux, J.-L., Dupuy, P., Laurantin, O., Soulan, I., Tabary, P., and Soubeyroux, J.-M.: Les mesures de précipitations et l'estimation des lames d'eau à Météo-France: état de l'art et perspectives, *La Houille Blanche*, pp. 28–34, 2009.
- Colleoni, F., Huynh, N. N. T., Garambois, P.-A., Jay-Allemand, M., and Villenave, L.: SMASH documentation, <https://smash.recover.inrae.fr>, version: 0.4.2, Release date: 2023-05-23, 2023.
- De Lavenne, A., Andréassian, V., Thirel, G., Ramos, M.-H., and Perrin, C.: A regularization approach to improve the sequential calibration of a semidistributed hydrological model, *Water Resources Research*, 55, 8821–8839, 2019.
- Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, 28, 1015–1031, <https://doi.org/10.1029/91WR02985>, 1992.
- Fablet, R., Chapron, B., Drumetz, L., Mémin, E., Pannekoucke, O., and Rousseau, F.: Learning variational data assimilation models and solvers, *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002572, 2021.
- Fekete, B. M. and Vörösmarty, C. J.: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements, *IAHS publ*, 309, 129–136, 2007.
- Finke, P., Hartwich, R., Dudal, R., Ibanez, J., Jamagne, M., King, D., Montanarella, L., and Yassoglou, N.: *Geo-referenced soil database for Europe. Manual of procedures*, version 1.0, European Communities, 1998.
- Fortin, F.-A., Rainville, F.-M. D., Gardner, M.-A., Parizeau, M., and Gagné, C.: DEAP: Evolutionary Algorithms Made Easy, *Journal of Machine Learning Research*, 13, 2171–2175, <http://jmlr.org/papers/v13/fortin12a.html>, 2012.

- Garambois, P.-A., Roux, H., Larnier, K., Labat, D., and Dartus, D.: Parameter regionalization for a process-oriented distributed model dedicated to flash floods, *Journal of Hydrology*, 525, 383–399, 2015.
- Garambois, P.-A., Larnier, K., Monnier, J., Finaud-Guyot, P., Verley, J., Montazem, A.-S., and Calmant, S.: Variational estimation of effective channel and ungauged anabranching river discharge from multi-satellite water heights of different spatial sparsity, *Journal of Hydrology*, 581, 124409, <https://doi.org/10.1016/j.jhydrol.2019.124409>, 2020.
- Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, in: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, *JMLR Workshop and Conference Proceedings*, 2010.
- Göttinger, J. and Bárdossy, A.: Comparison of four regionalisation methods for a distributed hydrological model, *Journal of Hydrology*, 333, 374–384, 2007.
- Gupta, H. V., Beven, K. J., and Wagener, T.: Model calibration and uncertainty estimation, *Encyclopedia of hydrological sciences*, 2006.
- Hannah, D. M., Demuth, S., van Lanen, H. A., Looser, U., Prudhomme, C., Rees, G., Stahl, K., Tallaksen, L. M., et al.: Large-scale river flow archives: importance, current status and future needs., *Hydrological Processes*, 25, 1191–1200, 2011.
- Hascoet, L. and Pascual, V.: The Tapenade automatic differentiation tool: principles, model, and specification, *ACM Transactions on Mathematical Software (TOMS)*, 39, 1–43, 2013.
- Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., and Fenicia, F.: Improving hydrologic models for predictions and process understanding using Neural ODEs, *Hydrology and Earth System Sciences Discussions*, pp. 1–29, 2022.
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrological Sciences Journal*, 58, 1198–1255, <https://doi.org/10.1080/02626667.2013.803183>, 2013.
- Huang, S., Eisner, S., Magnusson, J. O., Lussana, C., Yang, X., and Beldring, S.: Improvements of the spatially distributed hydrological modelling using the HBV model at 1 km resolution for Norway, *Journal of Hydrology*, 577, 123585, <https://doi.org/10.1016/j.jhydrol.2019.03.051>, 2019.
- Hundecha, Y. and Bárdossy, A.: 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, 281–295, 2004.
- Huynh, N. N. T., Garambois, P.-A., Colleoni, F., and Javelle, P.: Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to Mediterranean floods, *Journal of Hydrology*, 625, 129992, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.129992>, 2023.
- Jay-Allemand, M., Javelle, P., Gejadze, I., Arnaud, P., Malaterre, P.-O., Fine, J.-A., and Organde, D.: On the potential of variational calibration for a fully distributed hydrological model: application on a Mediterranean catchment, *Hydrology and Earth System Sciences*, 24, 5519–5538, 2020.
- Jay-Allemand, M., Colleoni, F., Garambois, P.-A., Javelle, P., and Julie, D.: SMASH - Spatially distributed Modelling and ASSimilation for Hydrology: Python wrapping towards enhanced research-to-operations transfer, *IAHS*, <https://hal.archives-ouvertes.fr/hal-03683657>, 2022a.
- Jay-Allemand, M., Demargne, J., Garambois, P.-A., Javelle, P., Gejadze, I., Colleoni, F., Organde, D., Arnaud, P., and Fouchier, C.: Spatially distributed calibration of a hydrological model with variational optimization constrained by physiographic maps for flash flood forecasting in France, *Tech. rep.*, *Copernicus Meetings*, <https://doi.org/10.5194/iahs2022-166>, 2022b.

- Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water resources research*, 42, 2006.
- Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, <https://arxiv.org/abs/1412.6980>, 2014.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resources Research*, 42, <https://doi.org/10.1029/2005WR004362>, 2006.
- Lane, R. A., Freer, J. E., Coxon, G., and Wagener, T.: Incorporating uncertainty into multiscale parameter regionalization to evaluate the performance of nationally consistent parameter fields for a hydrological model, *Water Resources Research*, 57, e2020WR028393, 2021.
- Larnier, K. and Monnier, J.: Hybrid Neural Network–Variational Data Assimilation algorithm to infer river discharges from SWOT-like data, *Nonlinear Processes in Geophysics Discussions*, pp. 1–30, 2020.
- Michel, C.: Hydrologie appliquée aux petits bassins ruraux, *Hydrology handbook* (in French), Cemagref, Antony, France, 1989.
- Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models, *Water Resources Research*, 53, 8020–8040, 2017.
- Monnier, J.: Variational Data Assimilation and Model Learning, <https://hal.science/hal-03040047>, lecture, 2021.
- Monnier, J., Couderc, F., Dartus, D., Larnier, K., Madec, R., and Vila, J.-P.: Inverse algorithms for 2D shallow water equations in presence of wet dry fronts: Application to flood plain dynamics, *Advances in Water Resources*, 97, 11–24, <https://doi.org/10.1016/j.advwatres.2016.07.005>, 2016.
- Odry, J.: Prédétermination des débits de crues extrêmes en sites non jaugés : régionalisation de la méthode par simulation SHYREG, Ph.D. thesis, <http://www.theses.fr/2017AIXM0424>, thèse de doctorat dirigée par Arnaud, Patrick Géosciences de l’environnement. Hydrologie Aix-Marseille 2017, 2017.
- Organde, D., Arnaud, P., Fine, J.-A., Fouchier, C., Folton, N., and Lavabre, J.: Régionalisation d’une méthode de prédétermination de crue sur l’ensemble du territoire français: la méthode SHYREG, *Revue des Sciences de l’Eau*, 26, 65–78, 2013.
- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2 Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling, *Journal of hydrology*, 303, 290–306, 2005.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resources Research*, 44, 2008.
- Oudin, L., Kay, A., Andréassian, V., and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, *Water Resources Research*, 46, 2010.
- Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrology and Earth System Sciences*, 9, 157–171, 2005.
- Parajka, J., Viglione, A., Rogger, M., Salinas, J., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins–Part 1: Runoff-hydrograph studies, *Hydrology and Earth System Sciences*, 17, 1783–1795, 2013.
- Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *Journal of hydrology*, 279, 275–289, 2003.
- Poncelet, C.: Du bassin au paramètre : jusqu’où peut-on régionaliser un modèle hydrologique conceptuel ?, Ph.D. thesis, <http://www.theses.fr/2016PA066550>, thèse de doctorat dirigée par Andréassian, Vazken et Oudin, Ludovic Hydrologie Paris 6 2016, 2016.

- Pujol, L., Garambois, P.-A., and Monnier, J.: Multi-dimensional hydrological–hydraulic model with variational data assimilation for river networks and floodplains, *Geoscientific Model Development*, 15, 6085–6113, <https://doi.org/10.5194/gmd-15-6085-2022>, 2022.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France, *Journal of Applied Meteorology and Climatology*, 47, 92, <https://doi.org/10.1175/2007JAMC1636.1>, 2008.
- Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, *Journal of Hydrometeorology*, 17, 287 – 307, <https://doi.org/10.1175/JHM-D-15-0054.1>, 2016.
- Razavi, T. and Coulibaly, P.: Streamflow prediction in ungauged basins: review of regionalization methods, *Journal of hydrologic engineering*, 18, 958–975, 2013.
- Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S.: Optimization of a similarity measure for estimating ungauged streamflow, *Water Resources Research*, 45, <https://doi.org/10.1029/2008WR007248>, 2009.
- Saadi, M., Oudin, L., and Ribstein, P.: Random forest ability in regionalizing hourly hydrological model parameters, *Water*, 11, 1540, 2019.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, 46, 2010.
- Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, *Agricultural and forest meteorology*, 98, 279–293, 1999.
- Sivapalan, M.: Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrological Processes*, 17, 3163–3170, <https://doi.org/10.1002/hyp.5155>, 2003.
- Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, 44, 2008.
- Wang, W., Zhao, Y., Tu, Y., Dong, R., Ma, Q., and Liu, C.: Research on parameter regionalization of distributed hydrological model based on machine learning, *Water*, 15, 518, 2023.
- Widén-Nilsson, E., Halldin, S., and Xu, C.-y.: Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation, *Journal of Hydrology*, 340, 105–118, 2007.
- Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization., *ACM Trans. Math. Softw.*, 23, 550–560, <http://dblp.uni-trier.de/db/journals/toms/toms23.html#ZhuBLN97>, 1997.