HAL
open science

Learning Regionalization Using Accurate Spatial Cost Gradients Within a Differentiable High-Resolution Hydrological Model: Application to the French Mediterranean Region

Ngo Nghi Truyen Huynh, Pierre-André Garambois, François Colleoni, Benjamin Renard, Hélène Roux, Julie Demargne, Maxime Jay-Allemand, Pierre Javelle

To cite this version:

Ngo Nghi Truyen Huynh, Pierre-André Garambois, François Colleoni, Benjamin Renard, Hélène Roux, et al.. Learning Regionalization Using Accurate Spatial Cost Gradients Within a Differentiable High-Resolution Hydrological Model: Application to the French Mediterranean Region. Water Resources Research, 2024, 60 (11), 10.1029/2024WR037544 . hal-04145059v3

HAL Id: hal-04145059

https://hal.inrae.fr/hal-04145059v3

Submitted on 1 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted manuscript. **Version of record:** https://doi.org/10.1029/2024WR037544

Learning Regionalization using Accurate Spatial Cost Gradients within a Differentiable High-Resolution Hydrological Model: Application to the French Mediterranean Region

Ngo Nghi Truyen Huynh[1], Pierre-André Garambois[1], François Colleoni[1], Benjamin Renard[1], Hélène Roux[2], Julie Demargne[3], Maxime Jay-Allemand[3], Pierre Javelle[1]

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

Key Points:

- Novel approach for regional calibration of a distributed hydrologic model using learnable and non-linear descriptors-to-parameters mappings
- Original combination of numerical adjoint model and neural network Jacobian: accurate gradients enable high-dimensional optimization
- Extensive case study in flash-flood-prone Mediterranean region shows effective regionalization of high-resolution model with neural network

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 multi-linear regression or artificial neural networks (ANNs), into a differentiable hydrological model. This approach demonstrates how two differentiable codes can be linked and their gradients chained, enabling the exploitation of heterogeneous data sets 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 126 flash-flood-prone catchments in the French Mediterranean region. The results highlight a strong regionalization performance of HDA-PR especially in the most challenging upstream-to-downstream extrapolation scenario with ANN, achieving median Nash-Sutcliffe efficiency (NSE) scores from 0.6 to 0.71 for spatial, temporal, spatio-temporal validations, and improving NSE by up to 30% on average compared to the baseline model calibrated with lumped parameters. Multiple evaluation metrics based on flood-oriented hydrological signatures also indicate that the use of an ANN leads to better performances than a multi-linear regression in a validation context. ANN enables to learn a non-linear descriptors-to-parameters mapping which provides better model controllability than a linear mapping for complex calibration cases.

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 low-dimensional spatially uniform model parameters for a single gauged catchment, which may lead to piecewise constant and discontinuous parameters fields for several adjacent gauged catchments treated separately in optimization for instance. Moreover, parameter sets determined through calibration are not transferable to un-

Correspondence to: N.N.T. Huynh and P.-A. Garambois,
ngo-nghi-truyen.huynh@inrae.fr, pierre-andre.Garambois@inrae.fr

gauged locations although the latter represent the majority of the global land surface (Fekete & 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. (2024)).

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 regionalization studies involves independent catchment-by-catchment calibrations, followed by 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) and can be called post-regionalization (Samaniego et al., 2010). This approach presumes that the variability of calibrated model parameters through the catchments is related, for instance, to 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, including machine learning models and artificial neural networks (ANNs), have also been applied in post-regionalization to explore the relationships between physical descriptors and calibrated parameter sets at gauged locations (e.g., Bastola et al. (2008); 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 pedotranfer 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 mostly applicable 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., 2006a). 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 forecasting, in particular for extreme floods (Garambois et al., 2015; Jay-Allemand et al., 2024).

The simultaneous regionalization approach involves optimizing a transfer function between physical descriptors and model parameters (cf. Hundecha 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 trainable operator of the forward hydrological model. This approach enables overcoming most of the aforementioned problems and has been applied in several studies. For instance, it has been used for regionalizing semi-distributed models such as HBV in Hundecha and Bárdossy (2004) or in Götzinger and Bárdossy (2007) who introduced monotonicity and Lipschitz condition into the optimization problem to constrain the inferred spatial fields. A multiscale parameter regionalization (MPR) method, combining descriptors maps, spatial upscalings functions and regionalization transfer functions in the form of multivariate mappings from descriptors, has been proposed by Samaniego et al. (2010). This method is implemented within a spatially distributed multiscale hydrological model (mHM) and later applied to over 400 European catchments at 0.25° spatial resolution in Rakovec et al. (2016). The MPR approach, providing consistent (seamless) parameter and flux fields across scales (Samaniego et al., 2017), imposes a spatial regularization effect through a strong constraint in the forward model (upscaling laws and regionalization transfer functions). Such a regularization is needed when working with spatially distributed hydrological models and spatially sparse discharge data leading to overparameterized optimization problems. Likewise, De Lavenne et al. (2019) discussed regularization and calibration from nested gauges for semi-lumped models and proposed a sequential optimization strategy from upstream sub-basins to downstream basins with upstream parameters relaxation. In the case of a fully distributed model calibrated with a variational data assimilation (VDA) algorithm, overparameterization issues are typically addressed using classical regularization with a background term (Jay-Allemand et al., 2020) or with a physiographic regularization term (Jay-Allemand et al., 2024) in the cost function. This approach induces weak constraints on the optimization problem and is not sufficient for effective spatial extrapolation to ungauged basins.

The MPR method of 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 $< 2,000$ 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 $< 5,000$ 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 proposed by Coxon et al. (2019)) 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 are "directly calibrated for each basin". The same remark can be made for Beck et al. (2020), with HBV discharge modeling at a daily time step on headwater catchments without any routing modeling. Therefore, those regionalization studies focus on lumped rainfall-runoff modeling at a daily time step for mostly headwater catchments whose characteristic response time scale might be shorter, or on more complex spatially distributed land surface modeling (LSM, including soil moisture and evaporation modeling in addition to river discharge) applied at regional or country scale still at a daily time step (mHM, e.g., Boeing et al. (2022), or VIC).

In all the above studies, state of the art optimization or sampling algorithms are used, especially the Shuffle Complex Evolution algorithm (SCE) (Duan et al., 1992) in Mizukami et al. (2017), or the 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. Higher complexity regionalization mappings achieved with neural networks have been shown promising for daily lumped modeling over large samples with pure long short-term memory (LSTM) hydrological model structure (e.g., Kratzert et al. (2018); Hashemi et al. (2022)), or hybrid structures including physical laws (e.g., Jiang et al. (2020); Feng et al. (2022); Tsai et al. (2021)). Nevertheless, the need for spatially distributed modeling and for higher spatio-temporal resolution (e.g., hourly and kilometric scales) is crucial to capture spatial variability in rainfall, land use, soil properties, and hydrological processes which lumped models fail to represent (e.g., Clark et al. (2015); Melsen et al. (2016)). The integration of an ANN-based regionalization operator into a spatially distributed differentiable hydrological model with routing remains a difficulty. Even though this comes with the challenge of high-dimensional parameter spaces, which significantly increases the complexity and computational cost of the optimization process, gradient-based algorithms are efficient approaches for solving such high-dimensional inverse problems, and their potential has been demonstrated in optimization of spatial parameters of differentiable hydraulic models (Monnier et al., 2016), or spatially distributed differentiable hydrological models (Castaings et al., 2009; Jay-Allemand et al., 2020; Colleoni et al., 2022). 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 Fortran source code differentiation in Castaings et al. (2009); Monnier et al. (2016); Jay-Allemand et al. (2020); Colleoni et al. (2022). Note that a key property of neural networks is their differentiability, which makes them compatible with VDA frameworks that rely on differentiable geophysical models (cf. Monnier et al. (2016); Feng et al. (2022); Song et al. (2024)). Moreover, the intrinsic capability of neural networks to extract multi-level information from large data set (LeCun et al., 2015) and to perform non-linear multivariate regressions (e.g., Gemperline et al. (1991)) makes them promising candidates to learn effective non-linear descriptors-to-parameters regionalization functions for spatially distributed differentiable hydrological modeling.

A novel approach called HDA-PR (Hybrid Data Assimilation and Parameter Regionalization) is presented in this article. The word "hybrid" refers to the incorporation of machine learning methods into a deterministic hydrological model for parameter regionalization. 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 data sets 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 km^2 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 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 non-linear descriptors-to-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 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 relations 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 (Larnier & 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 data sets while maintaining their physical interpretability.

The research questions investigated in this article are the following:

- Does the embedding of neural network-based regionalization transfer functions within a spatially distributed differentiable hydrological model enable using VDA to extract relevant information from physical descriptors and discharge time series at multiple gauges? And does it provide effective spatial constraints to avoid spatial overparameterization in the inverse problem, leading to effective and interpretable regional maps of conceptual hydrological parameters for modeling gauged and ungauged catchments?
- How do neural networks compare, in terms of modeling performances in calibration and spatio-temporal extrapolation, to a simple multi-gauge calibration approach with spatially uniform model parameters or to a linear multivariate regression on physical descriptors, that is, transfer functions of an equivalent complexity to the one used by, for example, Beck et al. (2020)?

To assess the proposed HDA-PR approach and to study the research questions above, the evaluation procedure adopted in this work considers challenging regionalization problems over a relatively large set of 126 flash-flood prone catchments in the French Mediterranean region, for an hourly and kilometric distributed hydrological model. It is a high-resolution hydrological modeling problem compared to existing studies with daily rainfall-runoff or LSM models for regional to continental coverage. Moreover, it is a difficult case because of very sudden and non-linear hydrological responses with significant spatial variabilities (e.g., Garambois et al. (2015)). Performances are assessed by means of multiple evaluation metrics including flood event hydrological signatures (Huynh, Garambois, Colleoni, & Javelle, 2023). We address the following aspects of the HDA-PR approach: (i) performance at gauged and ungauged sites during calibration and validation time periods; (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 calibration setups 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.

First, the differentiable forward model consists in: (i) a parsimonious and robust spatially distributed GR-like conceptual hydrological model structure, composed of GR hydrological operators ("Génie Rural" lumped models cf. Perrin et al. (2003)) applied at pixel scale for runoff generation, and a simple pixels-to-pixel routing scheme, the whole hydrological model being differentiable (Jay-Allemand et al., 2020; Colleoni et al., 2022); 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, are capable of dealing with high-dimensional inverse problems such as encountered with tunable parameters of regionalization descriptors-to-parameters mappings.

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 data sets, and hence constitutes a powerful data assimilation framework.

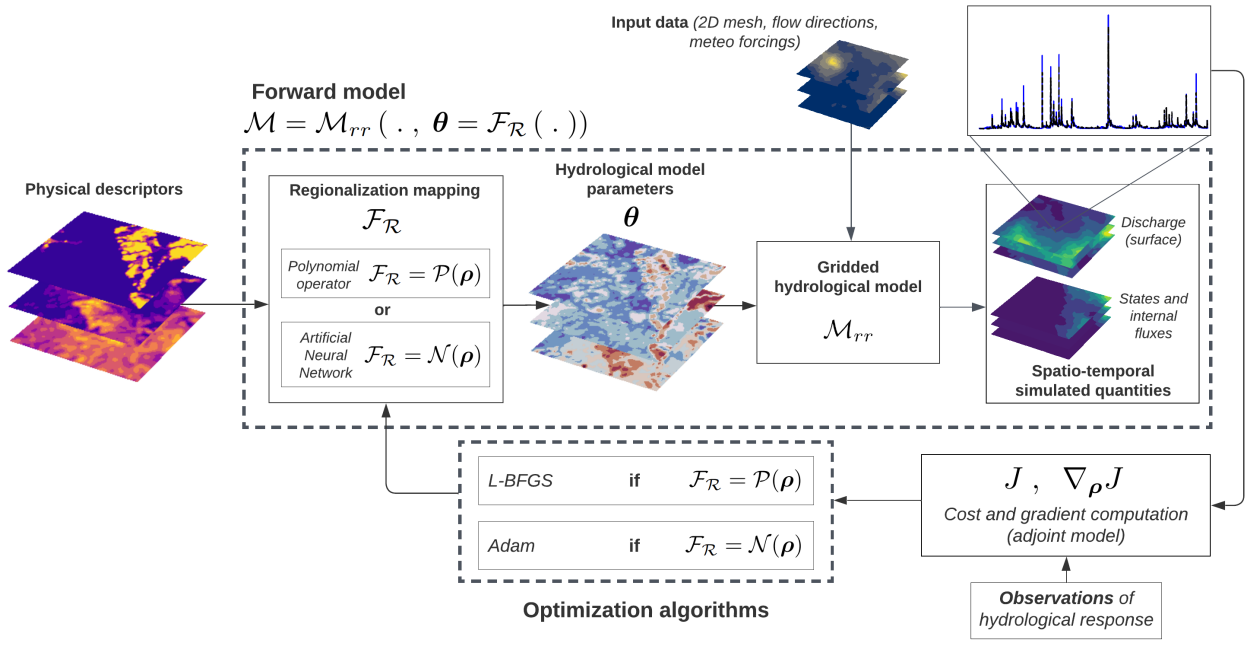

Figure 1. Flowchart of the forward-inverse algorithm used in HDA-PR. The forward hydrological model is a gridded model (spatio-temporal regular grid at 1 km^2 and 1 h) using GR operators (Perrin et al., 2003).

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 $\mathcal{D}_\Omega$ 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 $\mathcal{T}_\Omega$ has the highest drainage area.

Consider observed discharge time series $\boldsymbol{Q}_g^*(t)$ 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 $\Omega_{ung} = \Omega \setminus \left(\cup_{g=1}^{N_G} \Omega_g\right)$ 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 $\boldsymbol{\theta}$ such that:

$$\mathcal{M} = \mathcal{M}_{rr}\,(\,.\,,\ \boldsymbol{\theta} = \mathcal{F}_{\mathcal{R}}\,(\,.\,)) \tag{1}$$

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 $\boldsymbol{P}\,(x,t)$ and $\boldsymbol{E}\,(x,t)$, $\forall x \in \Omega$. The hydrological 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 an output field $\boldsymbol{U}(x,t)$, which comprises the discharge field $\boldsymbol{Q}\,(x,t)$ and state fields $\boldsymbol{h}\,(x,t)$. This is such that, for all $(x,t') \in \Omega \times [0,t]$:

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

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 GR4-like conceptual structure (Perrin et al., 2003), which is the spatialized "S-GR4" structure presented in the documentation of SMASH (Colleoni et al., 2024). The hydrological parameters vector $\forall x \in \Omega$ is:

$$\boldsymbol{\theta}\,(x) = (c_p(x), c_t(x), k_{exc}(x), l_{l_r}(x))^T$$

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_t in [mm]), the parameter (k_{exc} in [mm/dt]) of the non-conservative water exchange flux, and the linear routing parameter (l_{l_r} in [min]).

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

$$\boldsymbol{\theta}\left(x\right) = \mathcal{F}_R(\boldsymbol{D}\left(x\right), \boldsymbol{\rho}), \forall x \in \Omega \tag{3}$$

with $\boldsymbol{D}$ 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):

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

   with $s_k(z) = l_k + (u_k - l_k)/\left(1 + e^{-z}\right), \forall z \in \mathbb{R}$, 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:

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

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

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

   where $\boldsymbol{W}$ and $\boldsymbol{b}$ are respectively weights and biases of the neural network 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 the imposition of bound constraints on the kth-hydrological parameters, i.e., $l_k < \theta_k(x) < u_k$, $\forall x \in \Omega$. The regional control vector in this case is:

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

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 (differentiable hydrological model, cf. Jay-Allemand et al. (2020), combined with embedded neural networks or polynomial regionalization functions which are differentiable) and given the numerical implementation rules followed, the forward numerical 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 ordinary differential equation (ODE)-based operator of the forward hydrological model (Equation 2) relies on an explicit expression of its solution, approximated on the regular mesh $\mathcal{T}_\Omega$ 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:

$$J = \sum_{g=1}^{N_G} w_g J_g^* \tag{8}$$

with w_g a weighting function explained afterwards, J_g^* a local quadratic metric "at the station", here $1 - NSE$ or $1 - KGE_2$ (the latter being only used in validation in the case study of section 3, see Appendix A). This NSE-based calibration cost function is a quadratic differentiable and convex function, involving the response of the direct model. It depends on the control vector $\boldsymbol{\rho}$ 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 $\sum_{g=1}^{N_G} w_g = 1$.

2.3 The VDA Algorithm

The VDA aims at estimating the unknown input parameter $\boldsymbol{\rho}$ of the descriptors-to-parameters transfer functions $\mathcal{F}_R$ embedded into the hydrological model $\mathcal{M}_{rr}$ and predicting hydrological parameters $\boldsymbol{\theta}(x), \forall x \in \Omega$ from physical descriptors maps $\boldsymbol{D}(x)$, by minimizing the discrepancy between the modeled and observed discharges at multiple gauges. The cost function $J(\boldsymbol{U}(\boldsymbol{\rho}))$ to optimize depends on the regional control $\boldsymbol{\rho}$ through the response $\boldsymbol{U}(\boldsymbol{\rho}) = \mathcal{M}_{rr}(., \boldsymbol{\theta} = \mathcal{F}_R(\boldsymbol{\rho}))$ of the forward model that combines the hydrological model $\mathcal{M}_{rr}$ and the regionalization operator $\mathcal{F}_R$. The VDA inverse problem is written as the following convex optimization problem of the control vector $\boldsymbol{\rho}$:

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} J(\boldsymbol{U}(\boldsymbol{\rho})) \tag{9}$$

The calibration of this regional control $\boldsymbol{\rho}$ aims to (i) reduce the misfit between observed and simulated discharges at spatially sparse gauging stations, as evaluated by Equation 8, while (ii) determining the hydrological parameter maps $\boldsymbol{\theta}(x)$ for discharge modeling at any ungauged sites, thereby benefiting from the information extracted from physical descriptors $\boldsymbol{D}(x)$, and spatial constraints induced by the regional transfer functions, whose parameters $\boldsymbol{\rho}$ are being 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). For both formulations of the regionalization operator, the regional control vector $\boldsymbol{\rho}$ to optimize is large, and gradient-based optimization methods adapted to high-dimensional inverse problems are employed and detailed thereafter.

Note that the inverse problem 9 is a VDA optimization problem in the sense that it seeks to optimally combines observations (here discharge series $Q^*_{g=1..N_G}$ at N_G multiple gauges) with a model (here of Equation 1). While the NSE-based cost function defined earlier is not built upon explicit probabilistic assumptions, it can be shown (Kavetski et al., 2006a) that the minimization in Equation 9 is equivalent to maximizing the posterior density $p(\boldsymbol{\rho}|\boldsymbol{Q}^*)$ of the parameter $\boldsymbol{\rho}$ given observations $\boldsymbol{Q}^*$ under the assumptions of independent and identically distributed Gaussian errors at each gauge, spatially independent errors and no prior information. This probabilistic interpretation highlights a set of assumptions that could certainly be improved upon, and opens the way for more advanced probabilistic models that would explicitly recognize the various sources of uncertainty affecting the model and the surrounding data (forcings and responses). This is further discussed in section 4.

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 the high-dimensional parameter space, and in this study, there are no bound constraints on the values of $\alpha_{k,.}$, 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 a single run of the adjoint model, which is obtained by automatic differentiation (AD) using the Tapenade engine (Hascoet & Pascual, 2013). The entire process is implemented in the SMASH Fortran source code, where the full forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(., \mathcal{P}(.))$ 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:

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

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

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

- Maximum number of iterations;
- Cost function criterion: $\frac{J^{(i)}-J^{(i+1)}}{\max\left(|J^{(i)}|,|J^{(i+1)}|,1\right)} \leq \epsilon \times 10^6$ (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 $\boldsymbol{\rho} := [\boldsymbol{W}, \boldsymbol{b}]^T$. The optimization problem (Equation 9) can typically be solved using the Adam optimization algorithm (Kingma & 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 the first order gradients of the cost function. In the present case, the cost function writes as:

$$J\left(\boldsymbol{U}\left(\boldsymbol{\rho}\right)\right) = J\big(\boldsymbol{Q}^*, \mathcal{M}_{rr}(.\,,\ \boldsymbol{\theta} = \mathcal{N}(\boldsymbol{D}, \boldsymbol{\rho}))\big) \tag{10}$$

This formulation of the cost function highlights its dependency 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 $\boldsymbol{\theta}$ used as input by (ii) the hydrological model $\mathcal{M}_{rr}$ implemented in Fortran. In order to optimize J, its gradients with respect to $\boldsymbol{\rho}$ 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, $\nabla_{\boldsymbol{\theta}} J$ can be computed via the automatic differentiation (AD) 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. Finally, the two gradients can be combined as $\nabla_{\boldsymbol{\rho}} J = \nabla_{\boldsymbol{\theta}} J . \nabla_{\boldsymbol{\rho}} \boldsymbol{\theta}$, where the AD-based Fortran code and the neural network are connected via a Python-Fortran interface (here with f90wrap wrapper (Kermode, 2020)). The background value $\boldsymbol{\rho}^*$ in this case is randomly initialized using a specific method, which will be discussed later. The termination criterion is determined by a specified number of training epochs 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 in terms of discharge modeling but also in terms of its capability to extract information from data to estimate conceptual model parameters is analyzed over the French Mediterranean region. The performance of three regionalization methods is compared: a simple multi-gauge calibration approach for spatially uniform model parameters (Uniform), and two variants of HDA-PR. This first variant (Multi-linear) uses linear multivariate regression on physical descriptors, i.e., transfer functions of the same complexity as those of Beck et al. (2020). The second variant (ANN) uses a multilayer perceptron, which corresponds to a more complex mapping.

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 km^2 spatial grid (Quintana-Seguí et al., 2008) downscaled to a 1×1 km^2 spatial grid. The evaluation is performed on a flash-flood-prone area known as the Mediterranean arc (and called ArcMed hereafter), situated in the South of France, as indicated in Figure 2. Covering an approximate land area of 100,000 km^2, ArcMed comprises 126 flash-flood-prone catchments, including both nested and independent ones, representing about 26,000 km^2 of combined drainage area. These catchments have been selected based on the availability of long time series with high-quality observed flow data and minimal anthropogenic impacts. ArcMed is known for its diverse hydrological properties and contrasted catchment behaviors, including steep topography and very heterogeneous soils and bedrock (e.g., Garambois et al. (2015)). The area is affected by intense rainfall events that trigger non-linear flash flood responses, presenting a very challenging modeling scenario, in particular due to the significant presence of karstic zones. This evaluation data set is quite extensive considering the high-resolution, which is made necessary by the fast and small-scale non-linear processes characterizing the flash floods occurring in the area. The resulting spatio-temporal computational domain is quite large ($> 26,000$ pixels $\times\, 35,000$ time steps). By comparison, regionalization studies in the literature often make use of larger regions and more numerous catchments, but also of much coarser models typically running at a daily time step and coarser spatial resolution.

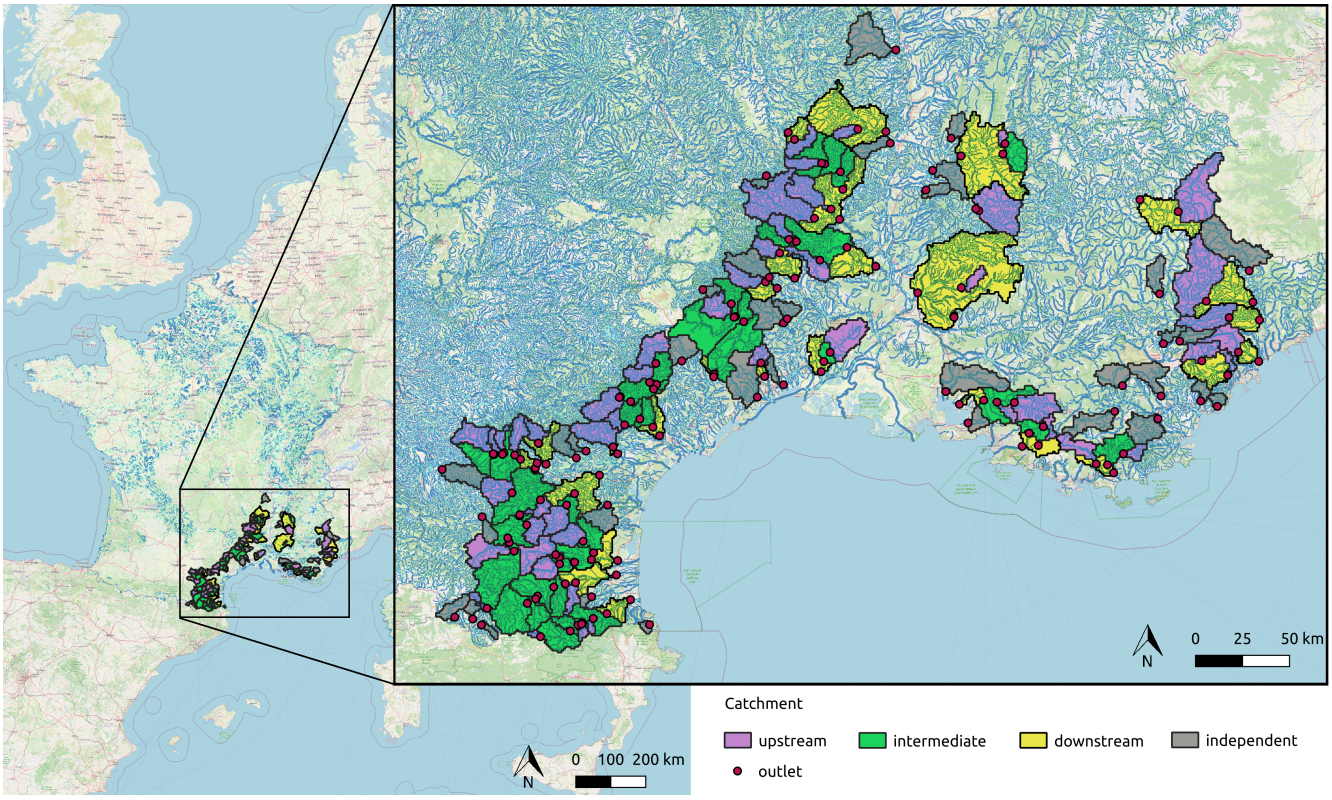

Figure 2. Map of France highlighting the ArcMed study area, covering 150,000 km^2 (100,000 km^2 excluding sea), comprising 126 catchments categorized as 38 catchments located upstream, 33 intermediate catchments, 24 catchments positioned downstream, and 31 independent catchments, representing a total drainage area of 26,000 km^2.

A set of 7 physical descriptors (Table 1) available over the whole French territory is used following Odry (2017) and Jay-Allemand et al. (2024). 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 for multi-source observations of hydrological responses and the most adequate descriptors layers is intentionally left for future research since it requires additional complementary modules to the proposed framework. 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	%	CORINE Land Cover (2012)
d_5	Land use	Urban cover rate (including artificial and non-vegetated areas)	%	CORINE Land Cover (2012)
d_6	Hydrogeology	Potential available water reserve	mm	Poncelet (2016)
d_7	Hydrogeology	High storage capacity basin rate	%	Finke et al. (1998)

To evaluate the performance of HDA-PR, we employ two spatial cross-validation setups which involve (i) calibrating the model on 38 upstream gauges (upstream calibration) and (ii) calibrating the model on 24 downstream gauges (downstream calibration). The upstream calibration poses a greater challenge due to the smaller catchments' areas, resulting in a lower integrative effect on non-linear hydrological processes and potentially more hydrological variabilities within and between basins. Consequently, this setup represents a more demanding interpolation scenario compared to the downstream calibration, which involves larger basins, as demonstrated later with

calibration results. These two calibration setups, along with a temporal validation scheme based on a two-period split, will facilitate the study of information extraction from discharge time series and from physical descriptors. For each calibration setup, we apply the different multi-site regional calibration methods using the set of gauges from the calibration catchments, while the remaining gauges from the validation catchments are used for spatial and spatio-temporal validation purposes. The weighting in Equation 8 can be set as $w_g = \frac{1}{N_G}$, representing the average cost over multiple gauges, since the gauges used for calibration in each setup, have the same number of observations (up to sporadic missing values), and share the same nature (either downstream or upstream), thus have similar information accumulation along the flow paths. The chosen calibration metric is the NSE, computed using data from multiple gauges over a four-year period P1 (August 2016 to July 2020) with a one-year warm-up period (August 2016 to July 2017). The following calibration methods are compared:

- Local calibrations for each gauge, both with spatially uniform (i.e., $\boldsymbol{\rho} \equiv \bar{\boldsymbol{\theta}}$, lumped parameters for each gauge) and full spatially distributed parameters (i.e., $\boldsymbol{\rho} \equiv \boldsymbol{\theta}(x)$), which are respectively under- and over-parameterizations of the studied spatially distributed hydrological model. These approaches represent reference or benchmark performances, denoted as "Uniform (loc)" and "Distributed (loc)".
- Multi-gauge regional calibration approaches with:
  - lumped model parameters over the whole domain (i.e., $\boldsymbol{\rho} \equiv \bar{\boldsymbol{\theta}}$), representing "level 0" regionalization, denoted as "Uniform (reg)";
  - a multivariate linear mapping (i.e., $\boldsymbol{\rho} \equiv [\alpha_{k,0}, (\alpha_{k,d}, 1)]^T$), referred to as "Multi-linear (reg)", which represents the classical regionalization mapping with the same complexity as the transfer functions of Beck et al. (2020) and without upscaling laws;
  - a multilayer perceptron (i.e., $\boldsymbol{\rho} \equiv [\boldsymbol{W}, \boldsymbol{b}]^T$), denoted as "ANN (reg)", representing the core novelty of the HDA-PR framework.

Note that the multivariate linear mapping above is a particular case of the polynomial mapping presented in section 2.1, obtained by forcing exponents $\beta_{k,d}$ to one. General polynomial mappings have been studied (see Huynh, Garambois, Colleoni, Renard, Roux, Demargne, and Javelle (2023)) but are not further considered in this paper for the sake of conciseness. To ensure robust validation, model performances are assessed in terms of spatial, temporal, and spatio-temporal validations, with a temporal validation period P2 covering two years from August 2020 to July 2022. Various evaluation metrics, including multiple hydrological signature-based metrics presented in Huynh, Garambois, Colleoni, and Javelle (2023), are also employed.

3.2 Regional Learning Performance and Computational Efficiency

In this section, the regional learning performance is analyzed over the whole ArcMed region. Before a thorough evaluation of performance is performed, Figure 3 provides several examples of observed and simulated discharges at randomly selected gauges, and compares them to observations. For these examples, using lumped model parameters $\bar{\boldsymbol{\theta}}$ leads to a poor performance in simulating discharges. By contrast, the other two regional learning methods result in improved performance, in both calibration and validation catchments. Given the complexity and heterogeneity of the region, it is unsurprising that lumped model parameters regionalization is unable to accurately reproduce such contrasted hydrological responses.

Figure 4 focuses on global NSE scores (see Figure 5 for KGE scores), i.e., the calibration metric, over the whole ArcMed region, for the two local calibration methods (spatially uniform and spatially distributed calibrations) and the three regionalization methods (Uniform, Multi-linear, and ANN). Overall, it suggests that the regionalization methods incorporating information from physical descriptors and imposing model parameters spatialization (Multi-linear and ANN) lead to superior performance when compared to the spatially Uniform baseline. This trend is particularly noticeable in the more challenging calibration-validation scenario using only upstream gauges for calibration. Furthermore, the enhanced efficiency of ANN-based regionalization is evident from global performance comparisons against the Multi-Linear approach, considering both upstream and downstream basin sets used in regional optimization.

In calibration ("Cal"), median NSE scores for the upstream calibration on 38 catchments are 0.7 and 0.75 for Multi-linear and ANN regionalization methods, representing a marked improvement compared to 0.57 for the Uniform approach. By contrast, for the downstream calibration, the median NSE scores are much closer: 0.8 and 0.81 for Multi-linear and ANN regionalization methods versus 0.76 for the Uniform approach. This indicates that the improvement obtained with the former two regionalization methods is smaller for the downstream calibration than for the upstream one. Moreover, the NSE performance for the Uniform approach is already quite close to that of the reference local calibrations in the downstream case, leaving little room for improvement with Multi-linear and ANN regionalization methods. Overall, these results suggest that the multi-gauge calibration problem poses a more challenging interpolation issue on upstream gauges than on downstream gauges. Thus, the regional calibration problem incorporating descriptors is more demanding on smaller-sized catchments with lower integrative effects and

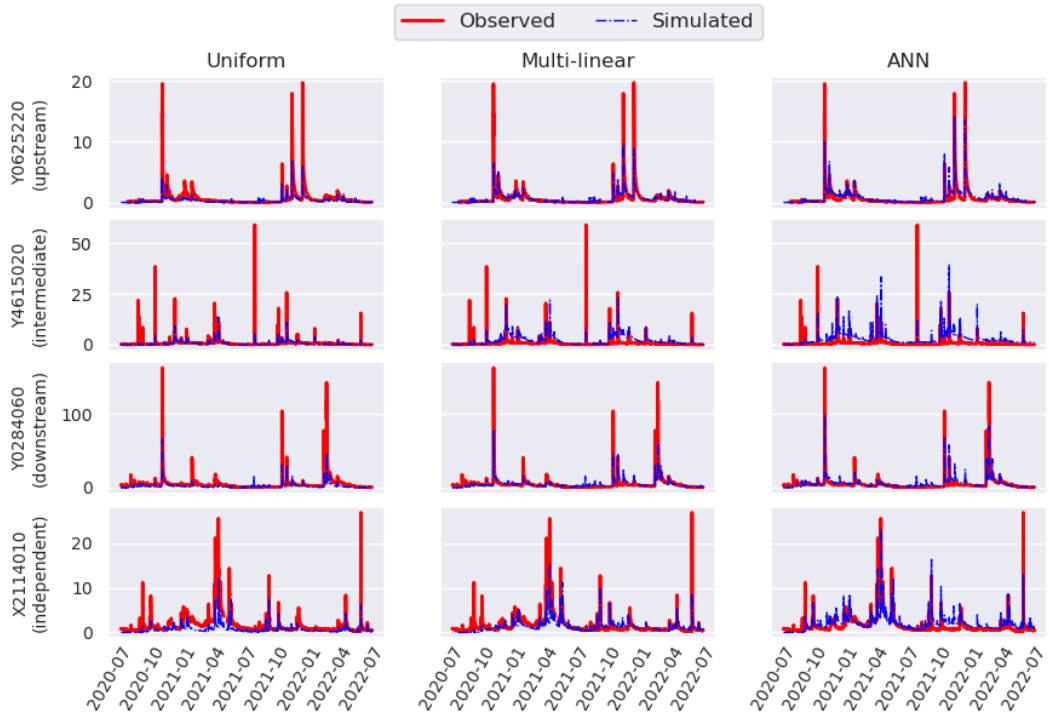

Figure 3. Observed and simulated discharges (in m^3/s) at several locations during the validation period P2, using three multi-gauge regional calibration methods (columns) within the first calibration setup, where upstream catchments are used for calibration on P1. The first row hence corresponds to temporal validation assessments, while the subsequent three rows correspond to spatio-temporal validation assessments.

potentially higher non-linear responses across basins. In terms of temporal validation ("T_Val") for both upstream and downstream cases, the same ranking as in calibration is observed with the three regionalization methods. Specifically, the ANN achieves a median NSE score of 0.69 (resp. 0.71) compared to 0.6 and 0.44 (resp. 0.69 and 0.68) for Multi-linear and Uniform in the upstream calibration setup (resp. downstream calibration setup). In spatial validation ("S_Val"), which assesses the potential of regionalization methods for ungauged basins, the ANN demonstrates the best performance in the upstream calibration setup, while similar performances (similar medians and interquartile ranges) of the three regionalization methods are observed in the downstream setup. Moving on to the most complex evaluation for the extrapolation capability of HDA-PR, spatio-temporal validation ("S-T_Val"), results consistently show the efficiency of the ANN, improving the NSE scores by 0.11 (resp. 0.18) in upstream setup and by 0.02 (resp. 0.04) in downstream setup compared to Multi-linear (resp. Uniform). In summary, multi-linear regression leads to slightly better median performances in spatial validation, while ANN performs best in temporal and spatio-temporal validations in the downstream calibration setup. However, in the upstream calibration case, the ANN demonstrates superior performance in calibration and for all validation scenarios. This result highlights the capability of ANN to outperform (at least in terms of scoring metrics up to this point) classical regionalization approaches using lumped model parameters and multi-linear regression in the most challenging extrapolation scenario (i.e., upstream calibration).

We now delve into a more detailed analysis of NSE performances in validation catchments categorized by their nature (upstream, downstream, intermediate, or independent) for both calibration setups: upstream and downstream, as illustrated in Figure 6 (see Figure 7 for KGE scores). In the downstream calibration setup, no clear ranking appears. Regarding the upstream calibration setup, the median NSE is systematically higher with ANN regionalization, with less spread (interquartile range) in all basin classes except for the 33 intermediate catchments, which show a slightly smaller first quartile compared to Multi-linear. Moreover, significantly higher performances are achieved with ANN in spatio-temporal validation. This is particularly notable on a relatively limited number of smaller-sized catchments in the calibration, indicating an effective extraction of information from both discharge data and descriptors. This effectiveness becomes evident when the model is tested on the validation catchments, which are more numerous, may or may not be nested and have varied areas. Overall, the performances obtained with both upstream and downstream descriptor-based calibrations show the relevance of regionalization with ANN. Indeed, ANN performs best in the upstream setup. ANN performance is not always the highest in the downstream

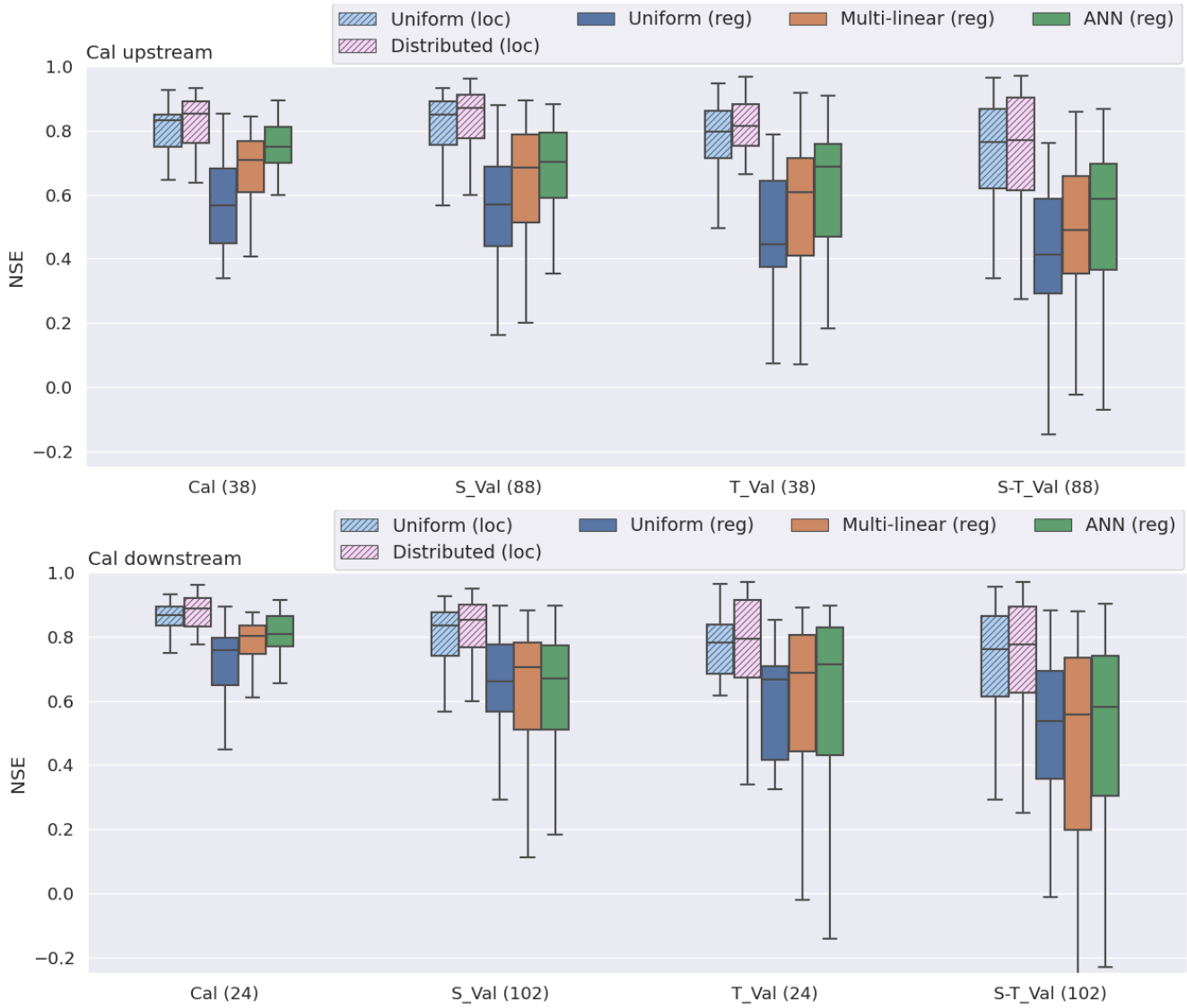

Figure 4. Boxplots of NSE scores (optimal value = 1) across calibration and validation catchments for both calibration setups which are upstream (top) and downstream (bottom), compared to reference solutions obtained by local calibration methods (Uniform (loc) and Distributed (loc)). From left to right: results are displayed for calibration catchments on period P1 ("Cal"), validation catchments on P1 for spatial validation ("S_Val"), calibration catchments on P2 for temporal validation ("T_Val"), and validation catchments on P2 for spatio-temporal validation ("S-T_Val"). The numbers in parentheses indicate the count of catchments included in each boxplot.

setup, but it remains very similar to other regionalization approaches. The complex and highly-parameterized nature of the ANN hence does not seem to result in a deterioration of performances in validation, even when this complexity may not be necessary.

To quantify how much the ANN-based regionalization approach improves the model performance compared to multi-linear regression and spatially uniform parameters regionalization, we compute the improvement rate in terms of NSE of the ANN relative to the other methods. We focus on the upstream calibration setup and evaluate improvement rates of ANN versus the two simpler approaches (Uniform and Multi-linear) in spatial and spatio-temporal validation, as illustrated in the maps of Figure 8. Comparing to Uniform, the ANN yields positive improvement rates at more than 80% of validation catchments on P1, with a median improvement rate of around 0.25 (see the boxplot in the top left panel of Figure 8). Transitioning to P2, the variation of improvement rates among catchments becomes more significant. Catchments that exhibited positive improvement rates during P1 generally maintain positive or enhanced rates, whereas some of those with negative rates either worsen or remain unchanged. Although the median improvement rate of NSE reaches 0.3, the interquartile range and whiskers visibly

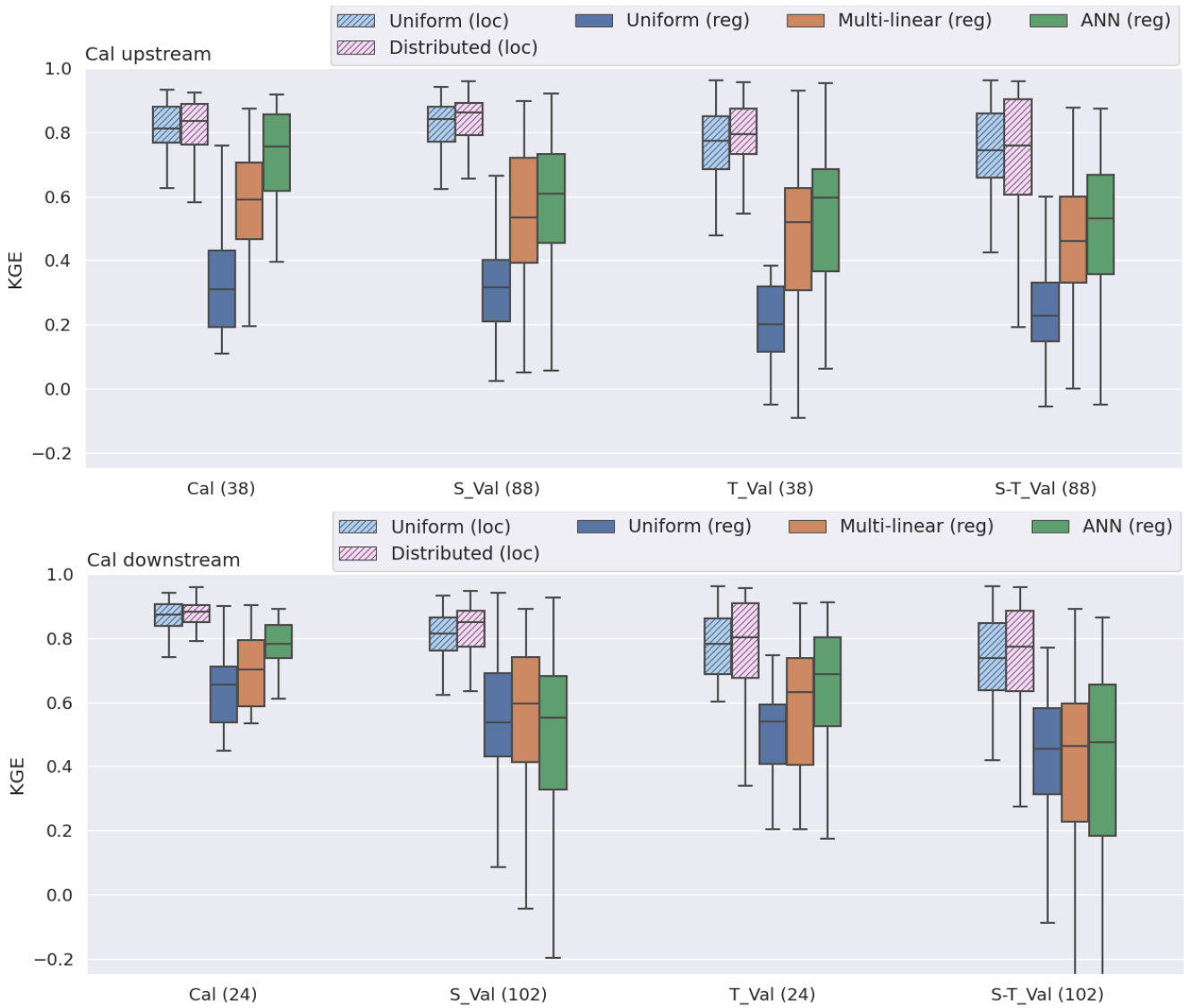

Figure 5. Boxplots of KGE scores (optimal value = 1) across calibrated (with NSE cost function) and validation catchments for both calibration setups (upstream (top) and downstream (bottom)), compared to reference solutions obtained by local calibration methods (Uniform (loc) and Distributed (loc)). From left to right: results are displayed for calibration catchments on period P1 ("Cal"), validation catchments on P1 for spatial validation ("S_Val"), calibration catchments on P2 for temporal validation ("T_Val"), and validation catchments on P2 for spatio-temporal validation ("S-T_Val"). The numbers in parentheses indicate the count of catchments included in each boxplot.

expand, as indicated by the boxplot in the bottom left panel of Figure 8. Regarding the improvement rates of ANN versus Multi-linear, we observe similar trends in the variation of improvement rates among catchments across the two periods. While the number of catchments with high improvement rates is generally reduced, more than 65% of catchments still exhibit positive rates on both periods. Nevertheless, it is apparent that some catchments (situated in the southeast region of the map and known for their complex hydrological behaviours including karstic effects and more difficulties in hydrological modeling) display negative improvement rates of the ANN versus both Uniform and Multi-linear approaches. This discrepancy may stem from difficulties in extrapolating effective model parameters from the considered physical descriptors at these locations with the ANN. This could be improved by using other descriptors and/or stronger constraints, for example, on the sensitive exchange parameter k_{exc} as discussed later.

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, Garambois, Colleoni, and Javelle (2023). This evaluation

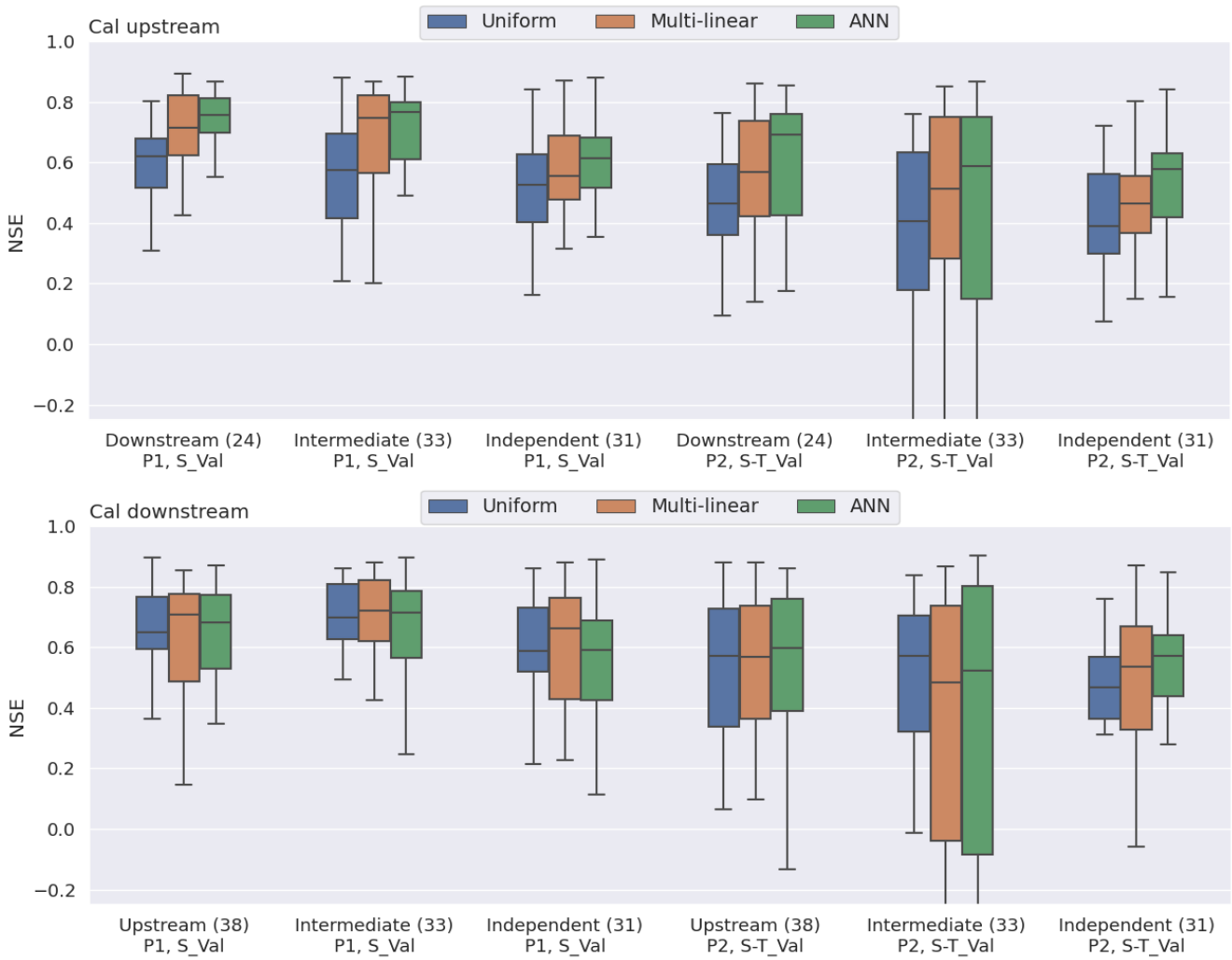

Figure 6. Comparison of NSE scores (optimal value = 1) across validation catchments categorized by their nature (upstream, downstream, intermediate, or independent) for both calibration setups (upstream (top) and downstream (bottom)). From left to right: results are displayed for validation catchments on period P1 for spatial validation ("S_Val") and validation catchments on P2 for spatio-temporal validation ("S-T_Val"). The numbers in parentheses indicate the count of catchments included in each boxplot.

is based on the relative error of three flood event signatures: flood runoff coefficient (Erc), flood flow (Eff), and peak flow (Epf), computed on a total of 1,737 flood events selected from the entire study area over a six-year period from August 2016 to July 2022. In the upstream calibration setup, the two non-uniform regionalization methods, and in particular the ANN, demonstrate their ability to outperform the uniform regionalization (using lumped model parameters) in both calibration and validation, as illustrated in Figure 9. For example, the relative errors (median over flood events) of simulated peak flows using ANN are around 0.4, compared to over 0.6 (for the uniform baseline) and 0.5 (for Multi-linear) in spatio-temporal validation (upstream calibration setup). In the downstream calibration setup, although the ANN does not markedly outperform the Multi-linear and Uniform regionalizations, it still achieves the best scores in terms of flood event signatures in both calibration and validation. For example, in spatio-temporal validation of the peak flow, ANN achieves a median relative error of 0.47, while Uniform and Multi-linear have median relative errors of 0.58 and 0.52. 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 ANN. It thus underscores the potential of HDA-PR for enhancing flash flood forecasting systems.

Finally, we discuss the computational efficiency of HDA-PR in terms of memory allocation and computation time. The SMASH numerical code, which incorporates the proposed HDA-PR algorithms, enables parallel computations using the openMP library (Dagum & Menon, 1998) and simple decomposition of the spatial domain, thus

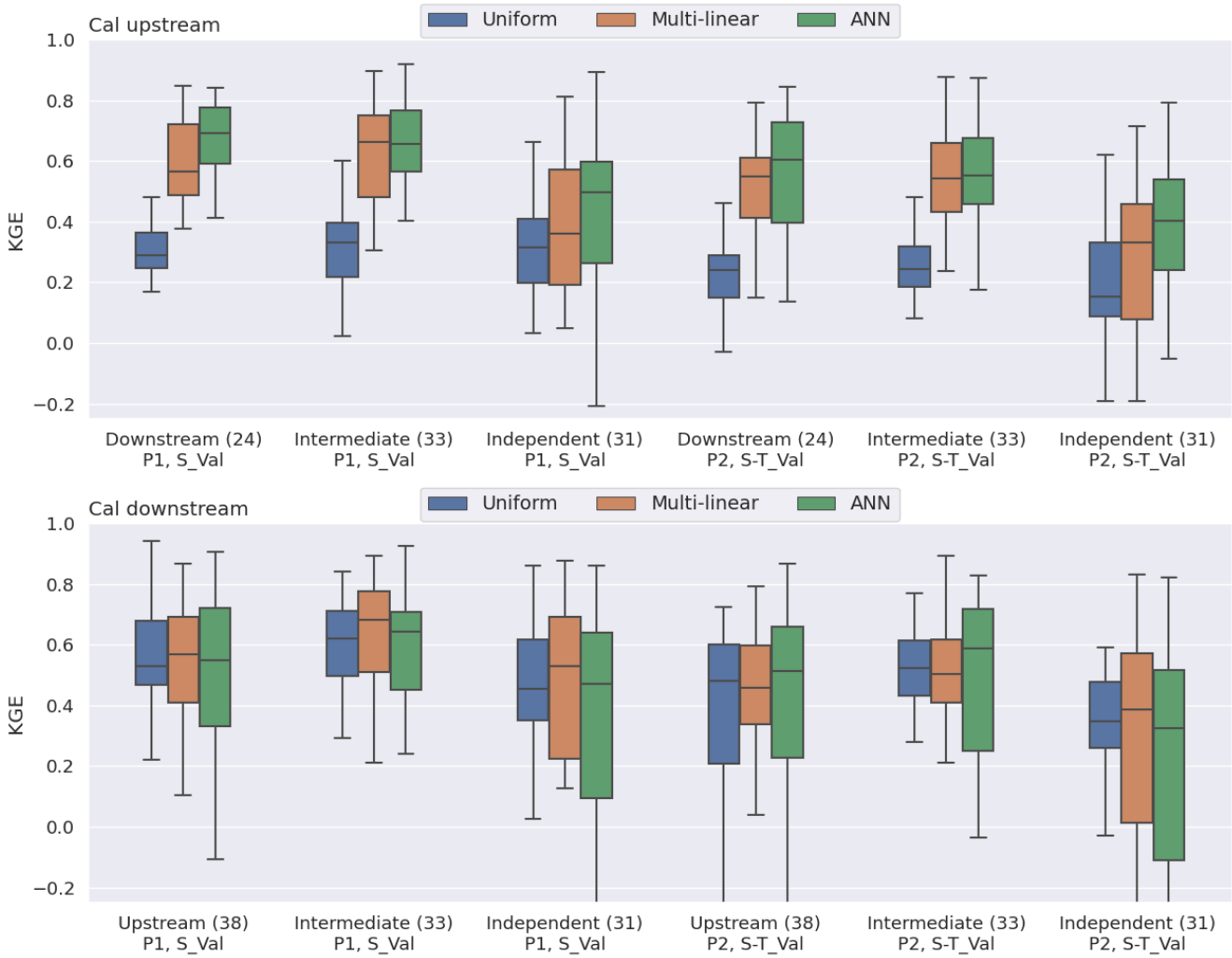

Figure 7. Comparison of KGE scores (optimal value = 1) across validation catchments categorized by their nature (upstream, downstream, intermediate, or independent) for both calibration setups (upstream (top) and downstream (bottom)). From left to right: results are displayed for validation catchments on period P1 for spatial validation ("S_Val") and validation catchments on P2 for spatio-temporal validation ("S-T_Val"). The numbers in parentheses indicate the count of catchments included in each boxplot.

reducing CPU computational time. Additionally, we have implemented sparse rainfall storage to mitigate memory consumption, which is crucial for running the adjoint model (whose memory usage has also been improved using a checkpointing technique) at each evaluation of the gradient, especially across expansive spatio-temporal domains. Note that the gridded atmospheric data constitutes a significant portion of memory usage across all methods. The CPU and memory computational costs of the different calibration runs over a relatively large computational domain are given in Table 2, demonstrating that the application of the method on larger domains such as at a country scale is feasible.

Table 2. Computation time for optimizing different regionalization mappings, where $N_\theta = 4$ and $N_D = 7$. Each calibration approach is performed on an AMD Opteron(TM) Processor 6276 at 2.5 GHz, running in parallel across 6 CPUs. The values are presented for the downstream calibration setup, with similar performance for the upstream case.

Mapping	Number of parameters	Memory usage (GB)	Optimization algorithm	Iterations	Time (h)
Uniform	4	10.97	SBS	11	16.94
Multi-linear	32	10.98	L-BFGS-B	250	157.24
ANN	6276	11.29	Adam	350	184.45

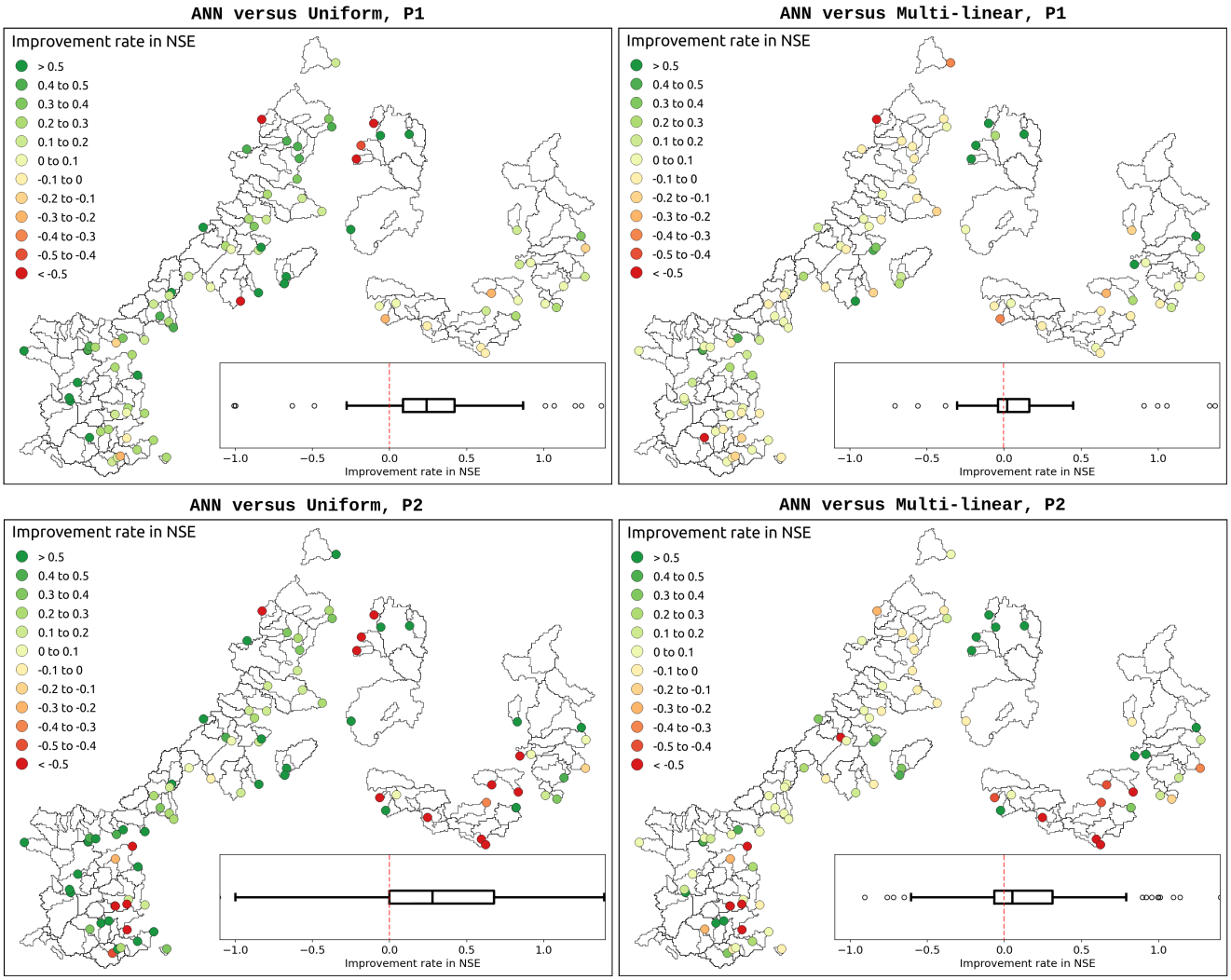

Figure 8. Comparison of NSE improvement rates in spatial and spatio-temporal validation for HDA-PR with ANN mapping versus uniform (left) and multi-linear (right) mappings, utilizing models calibrated with upstream catchments. Results are shown across validation catchments positioned downstream, intermediate, and independent for spatial validation on period P1 (top) and spatio-temporal validation on period P2 (bottom). The improvement rate at a catchment i is calculated as $r^i = \frac{NSE^i_{\text{ANN}} - NSE^i_{\square}}{|NSE^i_{\square}|}$, where $\square$ represents either multi-linear or uniform mapping.

3.3 Interpretation of the Learning Process and Inferred Parameters

A key feature of the proposed HDA-PR algorithm pertains to the differentiability of the forward spatially distributed model, including the ANN or Multi-Linear regionalization mappings (cf. section 2.1), which allows obtaining the gradient $\nabla_{\boldsymbol{\theta}} J$ of the cost function with respect to the hydrological model parameters. This gradient facilitates the computation of $\nabla_{\boldsymbol{\rho}} J$ necessary for optimizing the regional control vector $\boldsymbol{\rho}$ using algorithms tailored for high-dimensional inverse problems (cf. section 2.3), while maintaining a "physical" interpretation. Spatial maps of the cost gradient with respect to hydrological model parameters (which also represents the local sensitivity of the model parameters), at a background value $\boldsymbol{\theta}^* = \mathcal{F}_R(., \boldsymbol{\rho}^*)$ corresponding to the control value $\boldsymbol{\rho}^*$ at the first iteration of the optimization process (cf. section 2.3.1 and section 2.3.2), are shown in Figure 10: $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}^*)(x) = \left(\frac{\partial J(\boldsymbol{\theta}^*)}{\partial c_p}, \frac{\partial J(\boldsymbol{\theta}^*)}{\partial c_t}, \frac{\partial J(\boldsymbol{\theta}^*)}{\partial k_{exc}}, \frac{\partial J(\boldsymbol{\theta}^*)}{\partial l_{l_r}}\right)(x))$. The figure displays the maps for the downstream calibration setup, chosen for enhanced visualization compared to smaller basin sizes utilized in the upstream calibration setup. For both regionalization mappings (ANN or Multi-Linear), the initial cost gradient (sensitivity) at different points $\boldsymbol{\theta}^*$ in model parameter space for each mapping, indicates a high sensitivity to the non-conservative exchange parameter k_{exc}, which is coherent with global and local sensitivity analysis from Huynh, Garambois, Colleoni, and Javelle (2023) on a comparable model structure without regionalization mapping. Interestingly, the gradient maps (for the first optimization iteration) of the forward model containing descriptors-to-parameters mappings show the spatial

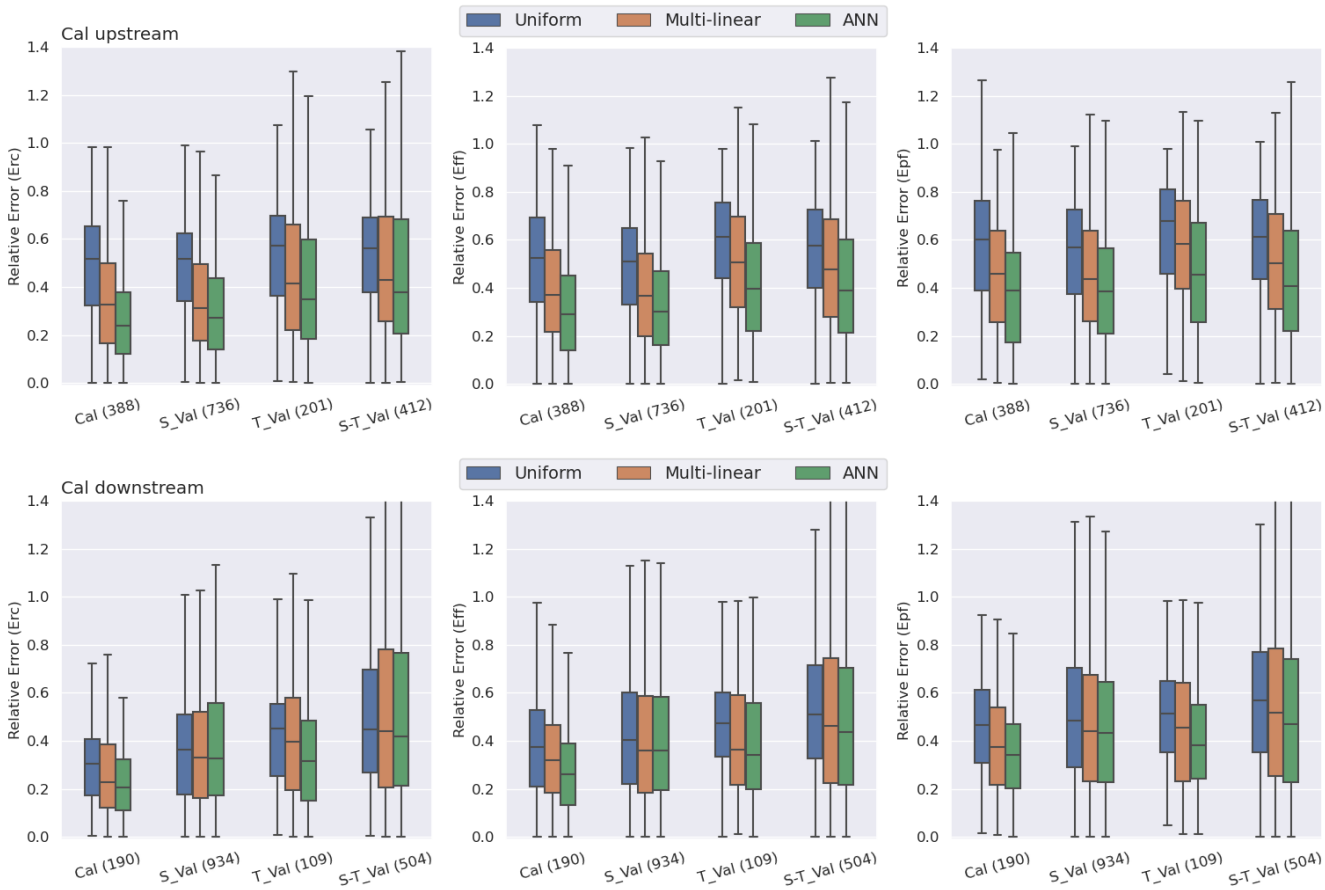

Figure 9. Relative error (optimal value = 0) of three flood event signatures—flood runoff coefficient (Erc), flood flow (Eff), and peak flow (Epf)—for both calibration setups: upstream (top) and downstream (bottom). The evaluation is based on a total of 1,737 flood events, with 1,124 selected during the calibration period P1 and 613 selected during the validation period P2, across 126 catchments. From left to right: results are displayed for calibration catchments on P1 ("Cal"), validation catchments on P1 for spatial validation ("S_Val"), calibration catchments on P2 for temporal validation ("T_Val"), and validation catchments on P2 for spatio-temporal validation ("S-T_Val"). The numbers in parentheses indicate the count of flood events included in each boxplot.

trends that will be obtained in the following optimization iteration, since a positive gradient should result in a parameter increase and conversely (if $\nabla_{\boldsymbol{\rho}}\boldsymbol{\theta}$ is of the same sign as $\nabla_{\boldsymbol{\theta}}J$ plotted here since $\nabla_{\boldsymbol{\rho}}J = \nabla_{\boldsymbol{\theta}}J.\nabla_{\boldsymbol{\rho}}\boldsymbol{\theta}$). Note that due to the high dimension of $\boldsymbol{\rho}$ and its non-physical meaning for ANN weights and biases, $\nabla_{\boldsymbol{\rho}}\boldsymbol{\theta}$ cannot be plotted in a simple manner and, more importantly, cannot be "physically" interpreted like $\nabla_{\boldsymbol{\theta}}J$. Moreover, these maps of $\nabla_{\boldsymbol{\theta}}J$ interestingly contain the footprint of physical descriptor patterns (along with the footprint of model structure and multi-site discharge observations and cost function formulation); they indicate sensitivity "hot spots", which could be useful in future works for tailoring descriptor processing and choice, as well as for studying other information sources, tailoring cost functions, parameter bounds, and regularizations of the forward-inverse problem.

When calibrating a model with gradient-based optimization algorithms, it is also important to discuss the descent of the cost function. This analysis enables understanding how optimization algorithms converge towards the global or local minimum of the cost function, and identifying potential trade-offs between model flexibility and overparameterization, thereby completing validation results. The descent of the cost function J is represented in Figure 11. It is apparent that the cost functions in the Multi-linear case start from a more optimal point than the ANN (approximately 0.37 in downstream calibration setup and 0.58 in upstream calibration setup) since they use a uniform background solution obtained by a global optimization method, as mentioned in section 2.3.1. Furthermore, they converge after around 200 iterations and remain monotonous throughout the optimization process. The ANN, with a significantly larger number of parameters (Table 2), can achieve a lower cost despite starting from a higher cost. Moreover, the ANN cost function in Figure 11 is not monotonous throughout training epochs: it shows localized increases for several instances within the first 100 epochs. This result can be attributed to the complexity

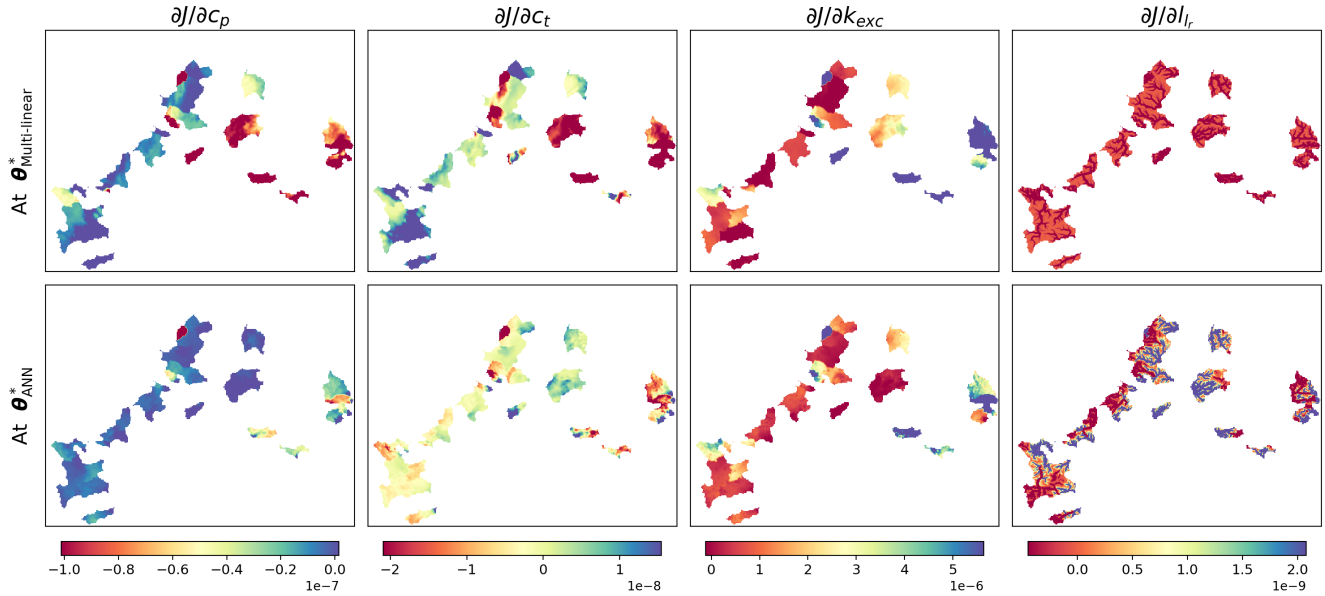

Figure 10. Spatially distributed cost gradients $\nabla_{\boldsymbol{\theta}} J\left(\boldsymbol{\theta}^*_{\square}\right)$ to conceptual hydrological parameters used in the optimization process (of $\boldsymbol{\rho}$ using $\nabla_{\boldsymbol{\rho}} J = \nabla_{\boldsymbol{\theta}} J . \nabla_{\boldsymbol{\rho}} \boldsymbol{\theta}$) showed for downstream calibration (for the sake of visibility) at initial points $\boldsymbol{\theta}^*_{\text{Multi-linear}} := \mathcal{P}(., \boldsymbol{\rho}^*_{\text{Multi-linear}})$ and $\boldsymbol{\theta}^*_{\text{ANN}} := \mathcal{N}(., \boldsymbol{\rho}^*_{\text{ANN}})$, where $\boldsymbol{\rho}^*_{\text{Multi-linear}}$ (resp. $\boldsymbol{\rho}^*_{\text{ANN}}$) is the background value of the multi-linear regression (resp. ANN) mapping. Note that the cost gradient values are obtained for the active cells within the gauged partition, i.e., cells contributing to the simulated discharge at gauges (downstream here) where the observation cost function J is evaluated.

of the regionalization mapping in each case, suggesting that the surface of the cost function is more complex for ANN than for multi-linear mapping. Moreover, the Adam algorithm used to optimize the ANN enables the exploration of different paths to locate the optimum, therefore avoiding getting stuck in local minima, where significant changes occur in the control vector space (biases and weights). This property could be essential in tackling equifinality and reach robust global optimum even with different starting points.

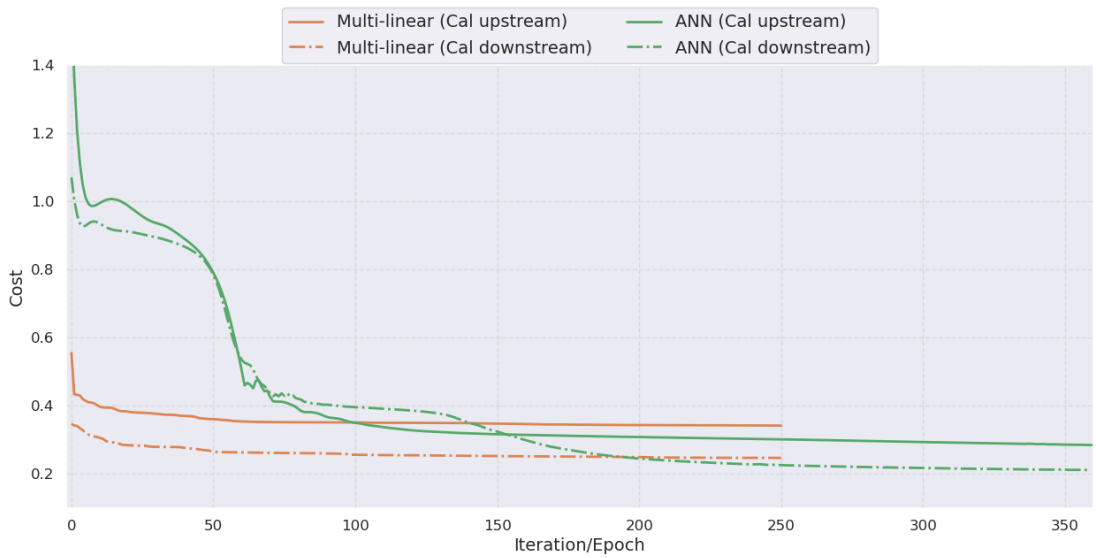

Figure 11. The descent of the cost function $J = 1 - NSE$ for both Multi-linear and ANN approaches across both calibration setups.

Finally, Figure 12-b shows the optimized parameter maps obtained with the two non-uniform regionalization methods across both calibration setups, that can be physically interpreted. In the downstream calibration setup, the spatial patterns of the drainage density descriptor (d_2) and the potential water reserve descriptor (d_6) in Figure 12-a are easily observable in the distributed parameter maps in the multi-linear regression case, as evidenced by strong correlations between these descriptors and the capacity of the production store (c_p), the capacity of the transfer store (c_t), and the non-conservative water exchange (k_{exc}). Such correlations can be quantified using one-to-one parameter-descriptor correlation matrices as depicted in Figure 12-c. Correlations are smaller in the ANN case than in the Multi-linear case: this suggests that the parameter maps derived with the former go beyond a simple linear combination of descriptors, and hence include some degree of non-linearity.

Overall, the obtained regional hydrological parameter maps, together with the performances in interpolation and spatio-temporal extrapolation cases (as shown in section 3.2), demonstrate the relevance of the intrinsic spatial constraint introduced by the descriptors-to-parameters mappings. These mappings result in a regularizing effect for the overparameterized calibration problem in distributed hydrological modeling from sparse discharge data.

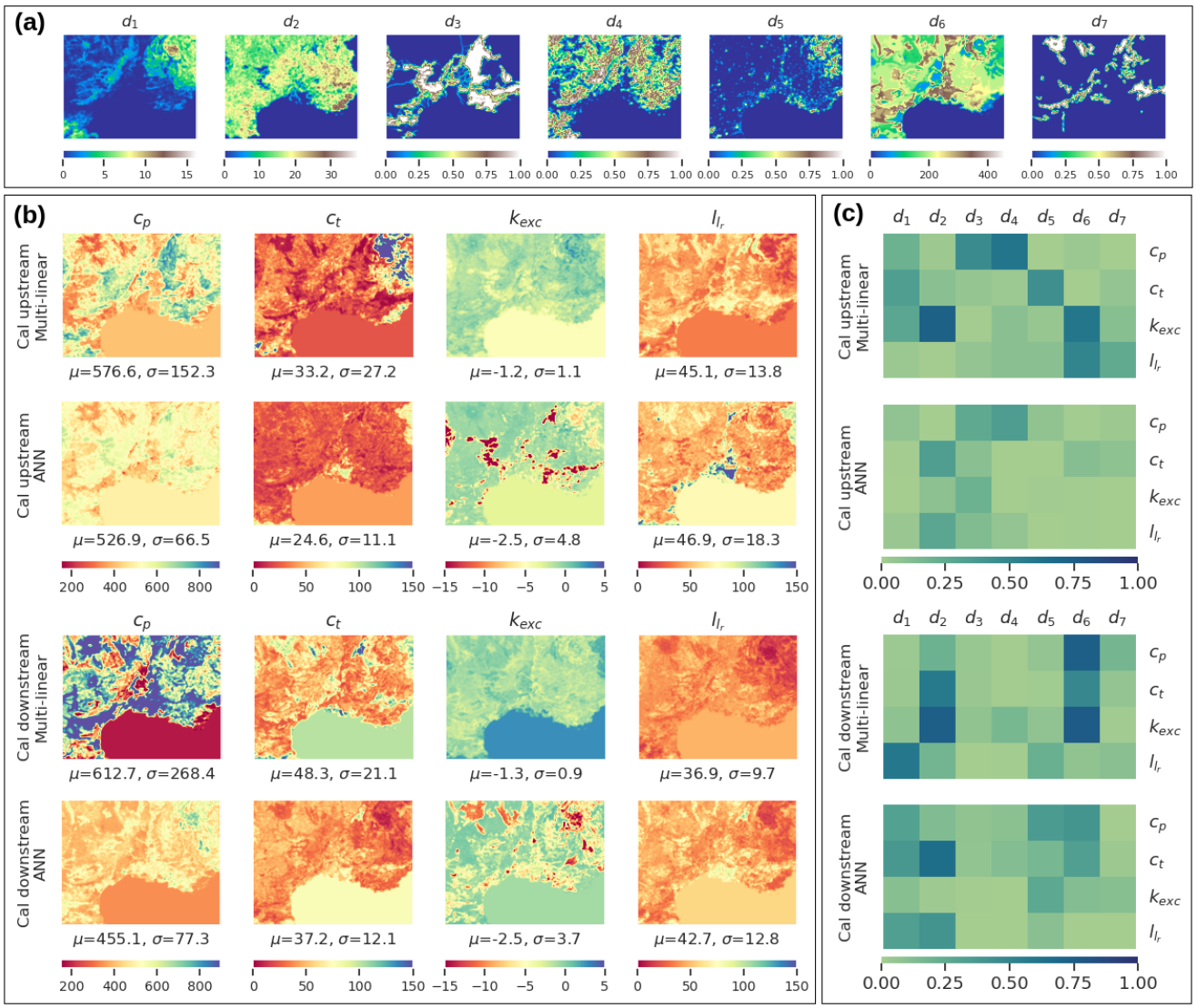

Figure 12. Analysis of input descriptors and output model parameters for two regionalization methods (Multi-linear and ANN) across both calibration setups (upstream and downstream): (a) Spatial distribution of physical descriptors (d_1-d_7), details provided in Table 1; (b) Spatial distribution of calibrated hydrological parameters (c_p, c_t, k_{exc}, l_{l_r}) with μ and σ indicating their spatial average and standard deviation over the active cells within the spatial domain Ω; (c) Linear correlation between descriptors and parameters over the active cells within the spatial domain Ω.

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. The various regionalization mappings define the way information is transferred from physical descriptors to hydrological parameters. It is important to ensure that HDA-PR effectively uses the physical patterns provided by descriptor maps to constrain the estimation of meaningful 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. The case of multi-polynomial regression was not included in the case study of this paper, but was tested by Huynh, Garambois, Colleoni, Renard, Roux, Demargne, and Javelle (2023) in a sub-region of the French Mediterranean area. In this case, the risk of losing physical properties may arise when the polynomial degree is unbounded. To mitigate this risk, we propose imposing bounds on the polynomial degree, $0.5 \leq \beta_{k,d} \leq 2$. ANNs, however, pose the most complicated scenario, 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, and should not exceed $\sqrt{N_D.N_x}$ based on our experiments. Next, to alleviate the vanishing gradient problem inherent to the ANN, we employed several techniques commonly used in the machine learning community. First, we applied Xavier initialization (Glorot & 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., from 0.003 to 0.005) to prevent the gradients from shrinking excessively during training.

While the ANN approach showed superior performance in the upstream calibration setup, its performance did not exhibit significant improvement in the downstream scenario, where the Uniform approach already yielded quite satisfactory results compared to local calibration references. It is worth noting that while "extreme" setups such as upstream-only and downstream-only calibrations are useful in the academic context of this paper, mixed setups that combine both, as well as include independent catchments, are likely to be employed in practice. For example, a calibration setup with a random selection of catchments was tested in Huynh, Garambois, Colleoni, Renard, Roux, Demargne, and Javelle (2023); Huynh, Garambois, Colleoni, Renard, and Roux (2023) for a subset of the 126 catchments presented in this paper. Furthermore, the performance of calibration setups that use nested gauges could be improved by adapting the cost function for better selecting and weighting information between calibration gauges and flow signal parts (e.g., with differentiable flood signatures-based algorithm for calibration (Huynh, Garambois, Colleoni, & Javelle, 2023)), as well as by accounting for data and structural uncertainties in this cost function (Kuczera et al., 2010; Renard et al., 2010).

Regarding regionalization over larger areas, such as for large basins or at country scales, an increased flexibility in the regional mapping might be needed to deal with significant physical heterogeneity. For instance, one may wish to use different transfer functions for different clusterings of the spatial domain (e.g., into sub-regions or hydrological response units (HRU)). 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, mappings that vary with regions or HRU can be obtained by including indicators of such regions/HRUs in the list of descriptors. This additional flexibility would certainly be necessary to circumvent the rigidity of the multi-linear mapping, but maybe not for a flexible mapping such as the ANN one. Another interesting research avenue is to develop an automatic identification of effective physical descriptors from large databases as well as identification of optimal spatial flexibility for constructing effective regional data assimilation approaches.

Applying HDA-PR to a non-differentiable hydrological model, which is the most common case encountered, may not be possible as the method requires accurate gradients of the simulated hydrological response with respect to the model parameters. This becomes particularly demanding for spatially distributed models, where the parameter space is high-dimensional and the gradient values are typically small as found here. In such cases, obtaining accurate gradients is essential, and these can be obtained by solving the numerical adjoint model, as implemented in our SMASH platform. However, to our knowledge, it would be possible to approach model output gradients to its parameters using a finite differences approach (e.g., Gupta and Razavi (2018)) in the case of weakly non-linear model and low-dimensional parameters, such as for a parsimonious and lumped model. Yet, even in this case, the gradient accuracy may still be insufficient for successful application of HDA-PR. A promising way to extend HDA-PR to other traditional, mathematically differentiable model structures is to implement them within platforms that support forward numerical model differentiation, such as SMASH or PyTorch (Ansel et al., 2024).

This approach still requires a differentiable hydraulic routing model (e.g., Pujol et al. (2022)), whose gradients can be combined with the spatialized hydrological model gradients using the chain rule, or requires the development of a differentiable, GPU-compatible, spatially distributed routing model. It is also worth noting that many hydrological models are based on ordinary differential equations (ODEs) or partial differential equations (PDEs) (e.g., Kavetski et al. (2006b)), and differentiability can often be achieved by making slight modifications to non-differentiable operators. For instance, replacing a unit hydrograph with a differentiable Nash-cascade, as done in the "GR4 state-space" model (Santos et al., 2018), is one such approach.

Finally, an important research avenue is the development of a probabilistic framework to quantify the uncertainties surrounding the application of HDA-PR. Several challenges need to be addressed for this purpose, some of them being related to the spatialized nature of the hydrologic model, while others are more specific to the large dimensionality induced by the regionalization mappings, in particular the ANN one. First, accounting for the uncertainty affecting the forcing and response data used in calibration requires a modification of the cost function based on data error models. Such models need to be parsimonious (e.g., to represent spatialized rainfall errors using a controlled number of parameters, Mustafa et al. (2018)), yet recognize the specificities of some data acquisition procedures (for instance, the partly systematic nature of rating curve errors affecting streamflow, Horner et al. (2018)). Second, since the structural errors made by the hydrologic model are varying in space, a probabilistic model describing structural uncertainty would need to be regionalized, as is the hydrologic model itself. Third, estimation tools need to be able to cope with the high dimensionality of the estimation problem. For instance, the use of Markov chain Monte Carlo (MCMC) methods will certainly require methods that can take advantage of the differentiable implementation of HDA-PR (Hoffman & Gelman, 2014) and of parallel computing capabilities (Laloy & Vrugt, 2012; Syed et al., 2022). Alternatively, the search of regionalization mappings that retain the flexibility provided by ANNs while being more parsimonious is another promising perspective.

5 Conclusion

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 multivariate 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 learning of a non-linear multivariate mapping in the case of the ANN, and providing effective spatial constraint of varied flexibility;
- 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;
- 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 thoroughly evaluated with two calibration setups in the French Mediterranean region with a high-resolution spatio-temporal hydrological modeling approach using multi-gauge discharge and descriptors maps. The results obtained on both calibration setups, and especially on the most challenging calibration scenario using only upstream gauges, highlight the effectiveness of HDA-PR that utilizes physical descriptors, surpassing the performance of a uniform regionalization method with lumped model parameters. Moreover, the ANN exhibited superior performance compared with multi-linear regression, that is, an approach with the same complexity as transfer functions used in previous studies on descriptors-based regionalization with a seamless calibration scheme. The NSE scores of HDA-PR with ANN in temporal validation surpass 0.7 for both calibration setups, which is a fairly good performance compared to the reference benchmarks of around 0.78 (obtained in local calibration), thereby establishing its remarkable capability in challenging modeling scenarios (i.e., upstream calibration) as well as its capacity to collapse to a simpler mapping in less challenging ones (i.e., downstream calibration). Various flood event signatures are also used as validation metrics to demonstrate the robustness of HDA-PR, where the two regionalization methods using descriptors outperform the uniform regionalization method. Particularly, the ANN-based mapping yields the best performance for all validation scenarios based on flood event signatures. Interestingly, the ANN enables extracting information from physical descriptors and learning non-linear multivariate descriptors-to-parameters mappings while providing better model controllability than a linear mapping for complex calibration cases. The exploration of more complex data set and neural networks architecture, on top of other differentiable flow models, represent a very promising axis for further research.

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, such as its operational application to estimate the regional hydrological model (at a 15-min temporal resolution) for the national flash flood early warning system called Vigicrues Flash in France (Javelle et al., 2019; Piotte et al., 2020); (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) information selection with a signature-based cost function (Huynh, Garambois, Colleoni, & Javelle, 2023) and spatial weighting strategies for multi-site discharge over river networks; (iv) implementation of a framework to quantify predictive uncertainty at both gauged and ungauged sites.

HDA-PR can be extended to state and composite parameters-states optimization which could be very interesting for multi-scale data assimilation and real time model correction from multi-source and multi-site data. Adding a learnable feature extraction layer from images (CNN) or time series (LSTM) on top of the regionalization transfer functions would enable the exploration of larger and richer databases including time varying data. Finally, HDA-PR 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 unprecedented SWOT (Surface Water and Ocean Topography satellite mission) observations of water surfaces variabilities of worldwide larger rivers. In general, the applicability of HDA-PR extends beyond hydrological models and can be adapted to other geophysical models.

A Metrics

Let $\boldsymbol{Q}(t)$ and $\boldsymbol{Q}^*(t)$ be the simulated and observed discharge time series. The hydrological cost functions studied are:

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

$$1 - NSE = \frac{\sum_{t=t^*}^{T} \left(\boldsymbol{Q}^*(t) - \boldsymbol{Q}(t)\right)^2}{\sum_{t=t^*}^{T} \left(\boldsymbol{Q}^*(t) - \bar{\boldsymbol{Q}}^*\right)^2}$$

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

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

  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.

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 $\boldsymbol{D}(x) \in \mathbb{R}^{N_D}$ to $\boldsymbol{\theta}(x) \in \mathbb{R}^{N_\theta}, \forall x \in \Omega$ (Figure B1).

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	$N_D.96 + 96 = 768$	$96.48 + 48 = 4656$	$48.16 + 16 = 784$	$16.N_\theta + N_\theta = 68$

Total parameters: 6276.

Let us consider an ensemble of layers where each layer is associated with its weight $\boldsymbol{W}_j$ and bias $\boldsymbol{b}_j$. Then, an input I of each layer is mapped to the input of the next layer by a linear function $\phi_j(I) = \boldsymbol{W}_j I + \boldsymbol{b}_j$, 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 its outputs are between 0 and 1. Now an output $\boldsymbol{O}_x = \sigma \circ \phi_{N_L}(., x) \in [0, 1]^{N_\theta}$ of the last layer is mapped to the range of the hydrological model parameters by a differentiable scaling function s:

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

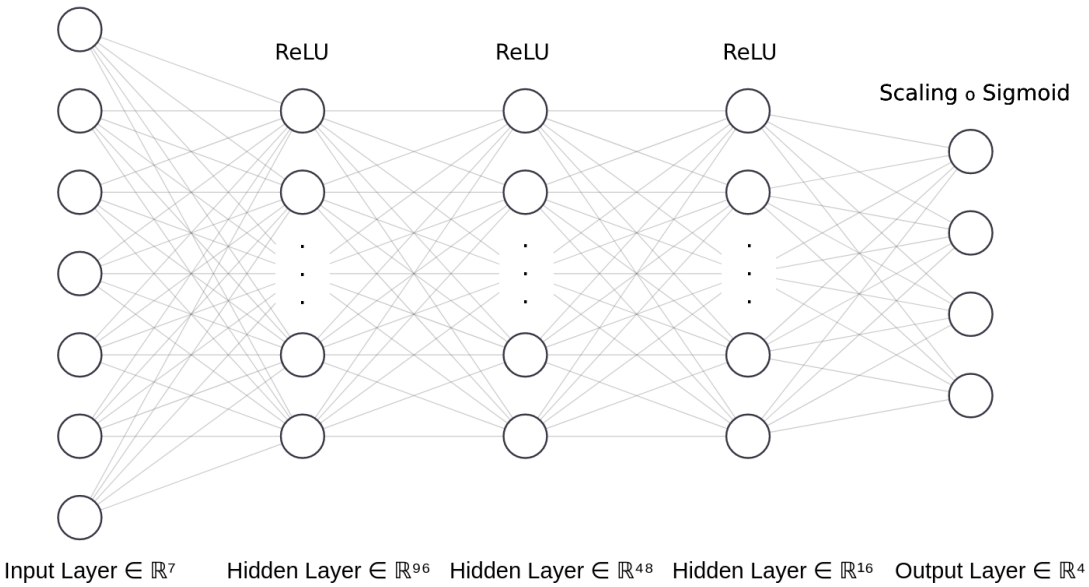

Figure B1. The architecture of the ANN consists of three hidden layers followed by the ReLU activation function and an output layer that uses the Sigmoid activation function in combination with a scaling function. In this particular case, we have $N_D = 7$, $N_L = 4$ and $N_\theta = 4$. The total number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

where $l = (l_1, ..., l_{N_\theta})$ and $u = (u_1, ..., u_{N_\theta})$ 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.

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

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 $\boldsymbol{\rho} := [\boldsymbol{W}, \boldsymbol{b}]$, 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 $\boldsymbol{\rho}_j := [\boldsymbol{W}_j, \boldsymbol{b}_j]$. Since the forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(., \mathcal{N}(.))$ 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 1.

Open Research

Data Availability Statement. The data set, Version 0.2, that supports this study comprises preprocessed data sourced from SCHAPI-DGPR and Météo-France, and is available at https://doi.org/10.5281/zenodo.10901472 (Huynh & Colleoni, 2024).

Software Availability Statement. The proposed algorithms in the study were implemented into the SMASH source code, Version 1.0.0-rc2.21, which is preserved at https://doi.org/10.5281/zenodo.8219280 (Colleoni et al., 2024), available via GNU-3 license and developed openly at https://github.com/DassHydro/smash. Additionally, the code for conducting the numerical experiments and analysis, Version 0.2.1, is preserved at https://doi.org/10.5281/zenodo.11520804 (Huynh, 2024), available via MIT license and developed openly at https://github.com/nghi-truyen/Regionalization-Learning.

Algorithm 1 Adapted back-propagation using Adam optimizer

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

▷ Number of training epochs N_{epo}

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

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

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

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 $\nabla A \leftarrow \nabla A. \left[W_j^{(i-1)}\right]^T$

▷ 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

Acknowledgments

The authors greatly acknowledge Killian Pujol-Nicolas for his contribution in the preliminary stage 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., & Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of hydrology*, *197*(1-4), 230–257.

Althoff, D., Rodrigues, L. N., & da Silva, D. D. (2021). Addressing hydrological modeling in watersheds under land cover change with deep learning. *Advances in Water Resources*, *154*, 103965.

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., ... Chintala, S. (2024). Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *Proceedings of the 29th acm international conference on architectural support for programming languages and operating systems, volume 2* (p. 929–947). Association for Computing Machinery. doi: 10.1145/3620665.3640366

Bastola, S., Ishidaira, H., & Takeuchi, K. (2008). Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving topmodel and basins across the globe. *Journal of Hydrology*, *357*(3), 188-206. doi: 10.1016/j.jhydrol.2008.05.007

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). doi: 10.1029/2019JD031485

Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, *52*(5), 3599–3622.

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

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

Boeing, F., Rakovec, O., Kumar, R., Samaniego, L., Schrön, M., Hildebrandt, A., ... Marx, A. (2022). High-resolution drought simulations and comparison to soil moisture observations in germany. *Hydrology and Earth System Sciences*, *26*(19), 5137–5161. doi: 10.5194/hess-26-5137-2022

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.*

Castaings, W., Dartus, D., Le Dimet, F.-X., & Saulnier, G.-M. (2009). Sensitivity analysis and parameter estimation for distributed hydrological modeling: potential of variational methods. *Hydrology and Earth System Sciences*,

13(4), 503–517. doi: 10.5194/hess-13-503-2009

Champeaux, J.-L., Dupuy, P., Laurantin, O., Soulan, I., Tabary, P., & Soubeyroux, J.-M. (2009). Les mesures de précipitations et l'estimation des lames d'eau à météo-france: état de l'art et perspectives. *La Houille Blanche*(5), 28–34.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... others (2015). A unified approach for process-based hydrologic modeling: 1. modeling concept. *Water Resources Research*, *51*(4), 2498–2514.

Colleoni, F., Garambois, P.-A., Javelle, P., Jay-Allemand, M., & Arnaud, P. (2022). Adjoint-based spatially distributed calibration of a grid gr-based parsimonious hydrological model over 312 french catchments with smash platform. *EGUsphere*, *2022*, 1–37. doi: 10.5194/egusphere-2022-506

Colleoni, F., Huynh, N. N. T., Benjamin, R., De Fournas, T., El Baz, A., & Garambois, P.-A. (2024, April). *Smash: Version 1.0.0-rc2.21.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.8219280 doi: 10.5281/zenodo.8219280

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J. M., Howden, N. J. K., ... Woods, R. (2019). Decipher v1: Dynamic fluxes and connectivity for predictions of hydrology. *Geoscientific Model Development*, *12*(6), 2285–2306. doi: 10.5194/gmd-12-2285-2019

Dagum, L., & Menon, R. (1998). Openmp: an industry standard api for shared-memory programming. *IEEE Computational Science and Engineering*, *5*(1), 46-55. doi: 10.1109/99.660313

De Lavenne, A., Andréassian, V., Thirel, G., Ramos, M.-H., & Perrin, C. (2019). A regularization approach to improve the sequential calibration of a semidistributed hydrological model. *Water Resources Research*, *55*(11), 8821–8839.

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: 10.1029/91WR02985

Fablet, R., Chapron, B., Drumetz, L., Mémin, E., Pannekoucke, O., & Rousseau, F. (2021). Learning variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems*, *13*(10).

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.

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resources Research*, *58*(10), e2022WR032404. doi: 10.1029/2022WR032404

Finke, P., Hartwich, R., Dudal, R., Ibanez, J., Jamagne, M., King, D., ... Yassoglou, N. (1998). *Geo-referenced soil database for europe. manual of procedures, version 1.0.* European Communities.

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. doi: 10.5555/2503308.2503311

Garambois, P.-A., Larnier, K., Monnier, J., Finaud-Guyot, P., Verley, J., Montazem, A.-S., & Calmant, S. (2020). Variational estimation of effective channel and ungauged anabranching river discharge from multi-satellite water heights of different spatial sparsity. *Journal of Hydrology*, *581*, 124409. doi: 10.1016/j.jhydrol.2019.124409

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.

Gemperline, P. J., Long, J. R., & Gregoriou, V. G. (1991). Nonlinear multivariate calibration using principal components regression and artificial neural networks. *Analytical Chemistry*, *63*(20), 2313–2323.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

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

Gupta, H. V., Beven, K. J., & Wagener, T. (2006). Model calibration and uncertainty estimation. *Encyclopedia of hydrological sciences*.

Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research*, *54*(11), 8692–8717.

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.

Hashemi, R., Brigode, P., Garambois, P.-A., & Javelle, P. (2022). How can we benefit from regime information to make more effective use of long short-term memory (lstm) runoff models? *Hydrology and Earth System Sciences*, *26*(22), 5793–5816. doi: 10.5194/hess-26-5793-2022

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., & Fenicia, F. (2022). Improving hydrologic models for predictions and process understanding using neural odes. *Hydrology and Earth System Sciences Discussions*, 1–29.

Horner, I., Renard, B., Le Coz, J., Branger, F., McMillan, H. K., & Pierrefeu, G. (2018). Impact of Stage Measurement Errors on Streamflow Uncertainty. *Water Resources Research*, *54*(3), 1952–1976. doi: 10.1002/2017WR022039

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

Huang, S., Eisner, S., Magnusson, J. O., Lussana, C., Yang, X., & Beldring, S. (2019). Improvements of the spatially distributed hydrological modelling using the hbv model at 1 km resolution for norway. *Journal of Hydrology*, *577*, 123585. doi: 10.1016/j.jhydrol.2019.03.051

Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of hydrology*, *292*(1-4), 281–295.

Huynh, N. N. T. (2024, June). *Regionalization-learning: Version 0.2.1.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.11520804 doi: 10.5281/zenodo.11520804

Huynh, N. N. T., & Colleoni, F. (2024, March). *Regionalization learning data: Version 0.2.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.10901472 doi: 10.5281/zenodo.10901472

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., & Javelle, P. (2023). Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to mediterranean floods. *Journal of Hydrology*, *625*, 129992. doi: 10.1016/j.jhydrol.2023.129992

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., Renard, B., & Roux, H. (2023). Multi-gauge hydrological variational data assimilation: Regionalization learning with spatial gradients using multilayer perceptron and bayesian-guided multivariate regression. *arXiv*. Retrieved from https://arxiv.org/abs/2307.02497 doi: 10.48550/arXiv.2307.02497

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., Renard, B., Roux, H., Demargne, J., & Javelle, P. (2023). Learning regionalization within a differentiable high-resolution hydrological model using accurate spatial cost gradients. *arXiv*. Retrieved from https://arxiv.org/abs/2308.02040v1 doi: 10.48550/arXiv.2308.02040

Javelle, P., Saint-Martin, C., Garandeau, L., & Janet, B. (2019). Flash flood warnings: Recent achievements in france with the national vigicrues flash system. *United Nations Office for Disaster Risk Reduction, Contributing Paper to the Global Assessment Report on Disaster Risk Reduction (GAR 2019)*, *60*.

Jay-Allemand, M., Demargne, J., Garambois, P.-A., Javelle, P., Gejadze, I., Colleoni, F., ... Fouchier, C. (2024). Spatially distributed calibration of a hydrological model with variational optimization constrained by physiographic maps for flash flood forecasting in france. *Proceedings of IAHS*, *385*, 281–290. doi: 10.5194/piahs-385-281-2024

Jay-Allemand, M., Javelle, P., Gejadze, I., Arnaud, P., Malaterre, P.-O., Fine, J.-A., & Organde, D. (2020). On the potential of variational calibration for a fully distributed hydrological model: application on a mediterranean catchment. *Hydrology and Earth System Sciences*, *24*(11), 5519–5538.

Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving ai system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. *Geophysical Research Letters*, *47*(13), e2020GL088229.

Kavetski, D., Kuczera, G., & Franks, S. W. (2006a). Bayesian analysis of input uncertainty in hydrological modeling: 1. theory. *Water resources research*, *42*(3).

Kavetski, D., Kuczera, G., & Franks, S. W. (2006b). Calibration of conceptual hydrological models revisited: 1. overcoming numerical artefacts. *Journal of Hydrology*, *320*(1), 173-186. (The model parameter estimation experiment) doi: 10.1016/j.jhydrol.2005.07.012

Kermode, J. R. (2020, may). f90wrap: an automated tool for constructing deep python interfaces to modern fortran codes. *Journal of Physics: Condensed Matter*, *32*(30), 305901. doi: 10.1088/1361-648X/ab82d2

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. Retrieved from https://arxiv.org/abs/1412.6980 doi: 10.48550/arXiv.1412.6980

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, *42*(3). doi: 10.1029/2005WR004362

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (lstm) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. doi: 10.5194/hess-22-6005-2018

Kuczera, G., Renard, B., Thyer, M., & Kavetski, D. (2010). There are no hydrological monsters, just models and observations with large uncertainties! *Hydrological Sciences Journal*, *55*(6), 980–991. doi: 10.1080/02626667.2010.504677

Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resources Research*, *48*(1). doi: 10.1029/2011wr010608

Lane, R. A., Freer, J. E., Coxon, G., & Wagener, T. (2021). Incorporating uncertainty into multiscale parameter regionalization to evaluate the performance of nationally consistent parameter fields for a hydrological model. *Water Resources Research*, *57*(10).

Larnier, K., & Monnier, J. (2020). Hybrid neural network–variational data assimilation algorithm to infer river discharges from swot-like data. *Nonlinear Processes in Geophysics Discussions*, 1–30.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. doi: 10.1038/nature14539

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., & Uijlenhoet, R. (2016). Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-alpine basin. *Hydrology and Earth System Sciences*, *20*(6), 2207–2226.

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.

Monnier, J. (2021, November). *Variational Data Assimilation and Model Learning* [Master]. France. Retrieved from https://hal.science/hal-03040047 (Lecture)

Monnier, J., Couderc, F., Dartus, D., Larnier, K., Madec, R., & Vila, J.-P. (2016). 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. doi: 10.1016/j.advwatres.2016.07.005

Mustafa, S. M. T., Nossent, J., Ghysels, G., & Huysmans, M. (2018, September). Estimation and Impact Assessment of Input and Parameter Uncertainty in Predicting Groundwater Flow With a Fully Distributed Model. *Water Resources Research*, *54*(9), 6585–6608. doi: 10.1029/2017WR021857

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).

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2 towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, *303*(1-4), 290–306.

Oudin, L., Kay, A., Andréassian, V., & Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, *46*(11).

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.

Perrin, C., Michel, C., & Andrèassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, *279*(1-4), 275–289.

Piotte, O., Montmerle, T., Fouchier, C., Belleudy, A., Garandeau, L., Janet, B., ... Organde, D. (2020). Les évolutions du service d'avertissement sur les pluies intenses et les crues soudaines en france. *La Houille Blanche*, *106*(6), 75–84. doi: 10.1051/lhb/2020055

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)

Pujol, L., Garambois, P.-A., & Monnier, J. (2022). Multi-dimensional hydrological–hydraulic model with variational data assimilation for river networks and floodplains. *Geoscientific Model Development*, *15*(15), 6085–6113. doi: 10.5194/gmd-15-6085-2022

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., ... Morel, S. (2008, January). Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. *Journal of Applied Meteorology and Climatology*, *47*(1), 92. doi: 10.1175/2007JAMC1636.1

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., ... Samaniego, L. (2016). Multiscale and multivariate evaluation of water fluxes and states over european river basins. *Journal of Hydrometeorology*, *17*(1), 287 - 307. doi: 10.1175/JHM-D-15-0054.1

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: 10.1029/2008WR007248

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5). doi: 10.1029/2009WR008328

Saadi, M., Oudin, L., & Ribstein, P. (2019). Random forest ability in regionalizing hourly hydrological model parameters. *Water*, *11*(8), 1540.

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

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

Santos, L., Thirel, G., & Perrin, C. (2018). Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the gr4 model using state-space gr4 (version 1.0). *Geoscientific Model Development*, *11*(4), 1591–1605.

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

Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology. *Hydrological Processes*, *17*(15), 3163-3170. doi: 10.1002/hyp.5155

Song, Y., Knoben, W. J., Clark, M. P., Feng, D., Lawson, K., Sawadekar, K., & Shen, C. (2024). When ancient numerical demons meet physics-informed machine learning: adjoint-based gradients for implicit differentiable modeling. *Hydrology and Earth System Sciences*, *28*(13), 3051–3077.

Syed, S., Bouchard-Côté, A., Deligiannidis, G., & Doucet, A. (2022, April). Non-Reversible Parallel Tempering: A Scalable Highly Parallel MCMC Scheme. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *84*(2), 321–350. doi: 10.1111/rssb.12464

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., ... Shen, C. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature communications*, *12*(1), 5988.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. *Water Resources Research*, *44*(12).

Wang, W., Zhao, Y., Tu, Y., Dong, R., Ma, Q., & Liu, C. (2023). Research on parameter regionalization of distributed hydrological model based on machine learning. *Water*, *15*(3), 518.

Widén-Nilsson, E., Halldin, S., & Xu, C.-y. (2007). Global water-balance modelling with wasmod-m: Parameter estimation and regionalisation. *Journal of Hydrology*, *340*(1-2), 105–118.

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. doi: 10.1145/279232.279236