Learning Pre-Regionalization of a Differentiable High-Resolution Hydrological Model with Spatial Gradients

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

Learning Pre-Regionalization of a Differentiable High-Resolution Hydrological Model with Spatial Gradients

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:

- Extending the multiscale parameter regionalization (MPR) technique to spatially learnable parameters regionalization
- New Hybrid Variational Data Assimilation Parameter Regionalization (HVDA-PR) approach
- Effective regionalization of high-resolution Mediterranean flash flood model

Abstract

Estimating spatially distributed hydrological model parameters on ungauged catchments poses a challenging regionalization problem, especially when searching for a transfer function that relates quantitatively physical descriptors to conceptual hydrological parameters, and imposing spatial constraints needed given sparse constraining discharge data. This paper introduces a Hybrid Variational Data Assimilation Parameter Regionalization (HVDA-PR) approach extending the multiscale parameter regionalization (MPR) technique. HVDA-PR leverages spatially distributed cost gradients to infer complex transfer functions designed for high-resolution hydrological models. The key components of HVDA-PR involve incorporating learnable regionalization mappings, which consist of either multivariate regressions or neural networks, into a differentiable hydrological model. This enables the exploitation of the informative content of heterogeneous datasets across extensive spatio-temporal computational domains, particularly in high-dimensional regionalization, with adapted optimization algorithms and accurate adjoint-based gradients. The inverse problem was tackled with a multi-gauge calibration cost function accounting for information from multiple observation sites. HVDA-PR was tested on high-resolution, hourly and kilometric regional modeling of two flash flood prone study areas located in the South of France. In both study areas, the median NSE scores of HVDA-PR ranged from 0.52 to 0.78 at pseudo-ungauged sites over calibration and validation periods, which exhibited strong regionalization performance, improving NSE by up to 0.57 compared to the baseline regionalization model calibrated with lumped parameters, and achieving a comparable level to the reference solution obtained from local uniform calibration (median NSE from 0.59 to 0.79). Multiple validation metrics based on hydrological flood-scale signatures are employed to assess the accuracy and robustness of the approach. The ability to produce physically explainable parameter maps from the physical descriptors demonstrates the effectiveness of HVDA-PR, in addition to the impressive scores obtained in calibration, and in validation in time and at ungauged sites. The

Corresponding author: Pierre-André Garambois, pierre-andre.garambois@inrae.fr

regionalization method is amenable to state-parameter correction from multi-source data, at multiple time scales such as for operational data assimilation, and it is transposable to other differentiable geophysical models.

1 Introduction

Whatever their status and complexity, hydrological models are more or less empirical and uncertain representations of multiscale coupled hydrological processes whose observability is limited. Hydrological model parameters, by definition, are 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 analyzed into model structural equifinality and spatial equifinality in context of spatially sparse observations compared to model controls (see for example discussions in Garambois et al. (2020)). Most calibration approaches enable to estimate spatially uniform model parameters for a single gauged catchment, hence piecewise constant discontinuous parameters fields for adjacent catchments. Moreover, parameter sets determined through calibration are not transferable to ungauged locations which yet represents 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 (cf. Hrachowitz et al. (2013)), especially in the search of effective high-resolution spatially distributed models in high uncertainty context, such as for hydrological extremes modeling (e.g., for Mediterranean flash floods in Garambois et al. (2015) and Jay-Allemand, Demargne, et al. (2022)).

The estimation of hydrological model parameters in ungauged regions is performed with so-called regionalization approaches that exploit and transfer the hydrological information from gauged catchments (see reviews in Blöschl et al. (2013); Samaniego et al. (2010); Hrachowitz et al. (2013); Beck et al. (2020)) and additionally from various descriptors of catchments physical properties. The most widely used approach in early studies for regionalization involved independent calibrations catchment-by-catchment, 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 have interrelations based, for instance, on spatial proximity (Widén-Nilsson et al., 2007; Oudin et al., 2008), and physical or climatic similarity (Oudin et al., 2010; Beck et al., 2016). Statistical learning methods have also been applied in post-regionalization to explore the relationships between physical descriptors and calibrated parameter sets at gauged locations (e.g., Saadi et al. (2019); Wang et al. (2023)). However, post-regionalization approaches are limited to lumped parameters by catchment, thus ignoring within-catchment variabilities (see reviews in Samaniego et al. (2010); Razavi and Coulibaly (2013)), except if post-regionalizing, for spatially distributed models with physically interpretable parameters, the tuning factors of parameters maps obtained through pedotransfer functions or hydraulic frictions correspondence tables for instance (e.g., Garambois et al. (2015)). Nevertheless, post-regionalization methods are generally faced with the issue of equifinal parameter sets and hence equifinal estimated transfer laws, while spatial proximity is more adapted to densely gauged river networks and regions (Oudin et al., 2008; Reichl et al., 2009). Moreover, incorporating a statistical learning process, especially unsupervised learning approaches, in the post-regionalization step can exacerbate the existing issues in our opinion. A regionalized calibration simultaneously exploiting the information of multiple gauges, within spatial clustering 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-catchments nor within regional clusters determined by physical similarity, which can have a major impact on the forecasting of extreme floods in particular (Garambois et al., 2015; Jay-Allemand, Demargne, et al., 2022).

The simultaneous regionalization approach involves optimizing a transfer function between physical descriptors and model parameters. We refer to this approach as "pre-regionalization" because, in this case and contrarily to post-regionalization (cf. Samaniego et al. (2010)), the descriptors-to-parameters mapping is the first optimizable operator of the forward hydrological model. It enables to overcome most of the aforementioned problems and has been applied in several 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 introduce monotony and Lipschitz condition into the optimization problem to constrain the inferred spatial fields; TOPMODEL in Bastola et al. (2008) who use an Artificial Neural Network (ANN)-based mapping between catchment descriptors and variance-covariance matrix of model parameters. A multiscale parameter regionalization (MPR) method, combining descriptors maps upscalings functions and pre-regionalization transfer functions in form of multi-variate mappings from descriptors, implemented within a spatially distributed multiscale hydrological model (mHm), has been proposed by Samaniego et al. (2010), and later applied to over 400 European catchments at 0.25° resolution in Rakovec et al. (2016). This approach imposes a spatial regularization effect that is needed when working with spatially distributed hydrological models and spatially sparse discharge data. The MPR method from Samaniego et al. (2010) has also been used with other gridded hydrological models in large sample applications. For example, Mizukami et al. (2017) calibrate the VIC model at a resolution of 0.125° over 531 headwater catchments (area $< 2,000$ km^2) in the continguous US area, using a lumped pre-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 in the generalized likelihood uncertainty estimation (GLUE) framework, with high-resolution HRU spatially disaggregated model (DECIPHeR framework) at daily time resolution over a large sample of 437 catchments in the UK. However, the routing module in this study is calibrated separately with a simple random sampling approach. In Mizukami et al. (2017), the runoff routing model is a gamma distribution function with two parameters that were "directly calibrated for each basin". Therefore, those regionalization approaches essentially concern the runoff production, at a daily time step for mostly headwater catchments whose characteristic response/concentration time scale might be shorter. Same remark can be made for Beck et al. (2020) who work at daily time step on headwater catchments, simply without modeling routing. In all the above studies, state of the art optimization algorithms are used, especially Shuffle Complex Evolution algorithm (SCE) (Duan et al., 1992) in Mizukami et al. (2017), or Distributed Evolutionary Algorithms (DEAP) (Fortin et al., 2012) in Beck et al. (2020), or a classical uncertainty estimation framework with random sampling approach in Lane et al. (2021). Moreover, those optimization algorithms are limited to low-dimensional controls, which limits the affordable number of descriptors as well as the affordable spatialization of regional transfer parameters which are lumped in all methods above. Adressing these issues is required in view to maximize the information extraction, from larger sets of physical descriptors and hydrological response observations, with adequate spatial flexibility of regional controls regarding multi-scale variabilities of physical system, of data and modeling uncertainties.

A novel approach called HVDA-PR (Hybrid Variational Data Assimilation Parameter Regionalization) is presented in this article. HVDA-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 regarding spatial equifinality issue. It is designed to exploit the informative content of massive heterogeneous datasets over large spatio-temporal computational domains, and is therefore adapted for 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 pre-regionalization functions enabling the introduction of an explicit tunable mapping between heterogeneous input physical descriptors and spatially distributed unknown conceptual parameters into the direct hydrological model. This mapping allows for the estimation of 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 enable learning complex non linear descriptors-parameters mapping for optimizing regional spatially distributed hydrological model.
- A differentiable spatially distributed hydrological model into which the pre-regionalization operators have been implemented. It facilitates the computation of accurate spatially distributed gradients of the calibration objective function, with respect to the sought pre-regionalization parameters which can be of high dimension and are needed by optimization algorithms.

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

The remaining sections of this paper are organized as follows: Section 2 describes the HVDA-PR algorithms and the SMASH spatially distributed hydrological assimilation platform into which they have been implemented. In Section 3, we present the numerical experiments on two study areas and analyze the performance of HVDA-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 HVDA-PR method. An algorithm flowchart is provided in Figure 1 to help in global understanding.

First is presented the differentiable forward model consisting in: (i) a parsimonious and robust GR-like conceptual hydrological model structure that is spatially distributed and

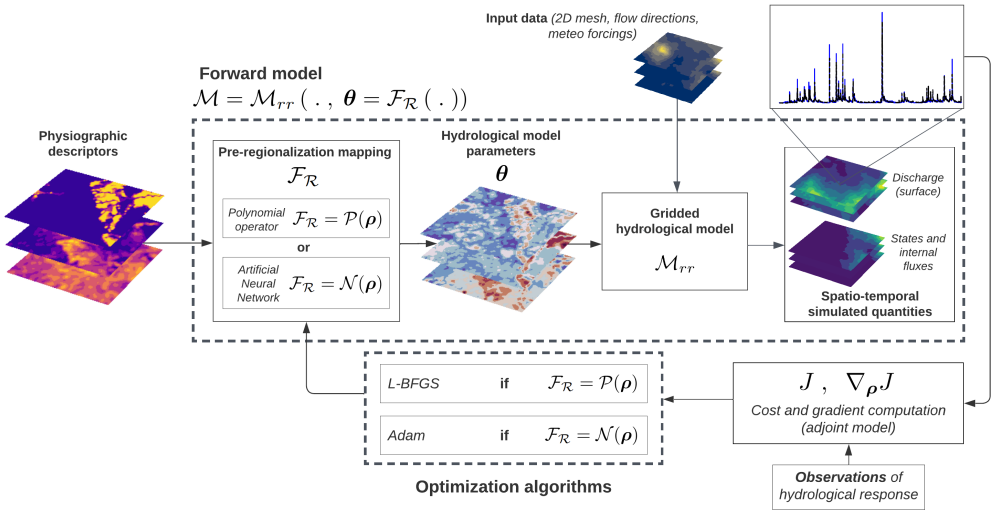

Figure 1. Flowchart of the forward-inverse algorithm used in HVDA-PR. The forward hydrological model is a gridded model (spatio-temporal regular grid at 1 km^2 and 1 h) using GR operators.

differentiable (Jay-Allemand et al., 2020); and (ii) pre-regionalization operators, consisting in either multivariate polynomial regressions or neural networks, for mapping descriptors onto hydrological model parameters.

Next is defined the calibration cost function enabling to work with multi-site (and potentially multi-source) observations.

Then are detailed the inverse optimization algorithms, that use the spatially distributed gradients of the cost function with respect to model parameters, and are capable of dealing with high-dimensional inverse problems such as encountered with tunable parameters of pre-regionalization descriptor-parameter mappings.

The core idea of HVDA-PR is the use of differentiable descriptors-to-parameters transfer functions, especially in form of neural networks, and the capability to automatically get accurate cost gradients. The latter enables to employ gradient-based variational optimization algorithms with high-dimensional regional parameter vectors. The method is applicable to any differentiable forward model as well as to multi-source heterogeneous datasets, hence being a powerful data assimilation framework.

2.1 Forward Model with Pre-Regionalization

First, let us define a 2D spatial domain $\Omega \subset \mathbb{R}^2$ 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 2D spatial coordinates over a domain Ω are denoted x. The number of active cells within the spatial domain Ω is denoted N_x. A 2D flow directions map $\mathcal{D}_\Omega$ is obtained from terrain elevation processing, with the only condition that a unique point in $\mathcal{T}_\Omega$ has the highest drainage area, and will be used for runoff routing.

Consider observed discharge time series $Q_g^*(t)$ at N_G observation cells of coordinates $x_g \in \Omega$, $g = 1..N_G$ ($N_G \geq 1$). For each observation cell, the corresponding gauged upstream sub-catchment is denoted Ω_g so that $\Omega_{ung} = \Omega \setminus \left(\cup_{g=1}^{N_G} \Omega_g\right)$ is the remaining ungauged part of the whole spatial domain Ω. 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 pre-regionalization operator $\mathcal{F}_R$ to estimate hydrological parameters $\boldsymbol{\theta}$ such that:

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

Now let us introduce and detail the hydrological model and the pre-regionalization operator along with their input variables.

The rainfall and potential evapotranspiration fields are respectively denoted as $\boldsymbol{P}(x,t)$ 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 the discharge field $Q(x,t)$ and states fields $\boldsymbol{h}(x,t)$ such that:

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

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 GR-like conceptual structure, which is the "gr-b" structure presented in Colleoni et al. (2023). The hydrological parameters vector is:

$$\boldsymbol{\theta}(x) = (c_p(x), c_{ft}(x), k_{exc}(x), l_r(x))^T,\ \forall x \in \Omega \tag{3}$$

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

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

$$\boldsymbol{\theta}(x) = \mathcal{F}_R(\boldsymbol{D}(x), \boldsymbol{\rho}),\ \forall x \in \Omega \tag{4}$$

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.

Now, we present two types of pre-regionalization operators used in HVDA-PR.

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

$$\begin{aligned} &\boldsymbol{\theta}(x, \boldsymbol{D}, \boldsymbol{\rho}) \coloneqq \mathcal{P}(\boldsymbol{D}(x), \boldsymbol{\rho}) \equiv \left[(\theta_k(x, \boldsymbol{D}, \rho_k))_{k=1}^{N_\theta}\right]^T,\ \forall x \in \Omega; \\ &\theta_k(x, \boldsymbol{D}, \rho_k) \coloneqq s_k\left(\alpha_{k,0} + \sum_{d=1}^{N_D} \alpha_{k,d} D_d^{\beta_{k,d}}(x)\right),\ \forall k \in [1..N_\theta] \end{aligned} \tag{5}$$

with $s_k(z) = l_k + (u_k - l_k)/(1 + e^{-z})$, $\forall z \in \mathbb{R}$, a transformation based on a Sigmoid function with values in $]l_k, u_k[$, thus imposing bound constraints in the direct hydrological model such that $l_k < \theta_k(x) < u_k$, $\forall x \in \Omega$. The lower and upper bounds l_k and u_k, assumed spatially uniform for simplicity here, associated to each parameter field θ_k of the hydrological model (Equation 2) are given. Each bound has the unit of its associated parameter (given in text after Equation 3). The regional control vector in this case is:

$$\boldsymbol{\rho} \equiv \left[(\rho_k)_{k=1}^{N_\theta}\right]^T \equiv \left[\left(\alpha_{k,0}, (\alpha_{k,d}, \beta_{k,d})_{d=1}^{N_D}\right)_{k=1}^{N_\theta}\right]^T \tag{6}$$

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

$$\boldsymbol{\theta}(x, \boldsymbol{D}, \boldsymbol{\rho}) \coloneqq \mathcal{N}\left(\boldsymbol{D}(x), \boldsymbol{W}, \boldsymbol{b}\right), \forall x \in \Omega \tag{7}$$

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 will be detailed in Appendix B and Equation B2. Note that an output layer consisting in a scaling transformation based on the Sigmoid function (cf. Equation B1) enables to impose $l_k < \theta_k(x) < u_k$, $\forall x \in \Omega$, i.e., bound constraints on the k^{th}-hydrological parameters. The regional control vector in this case is:

$$\boldsymbol{\rho} \equiv [\boldsymbol{W}, \boldsymbol{b}]^T \equiv \left[(W_j, b_j)_{j=1}^{N_L}\right]^T \tag{8}$$

For each pre-regionalization operator (Equation 5 or 7), the regional calibration problem consists in optimizing (in a sense defined below) the regionalization control $\boldsymbol{\rho}$ (Equation 6 or 8) 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 after. Importantly, note that by definition of the mathematical model and given the numerical implementation rules followed, the forward model is differentiable. This is a necessary condition for computing cost gradients with respect to spatially distributed hydrological parameters and obtain those of regional controls, as needed for solving the optimization problem. This is a key idea and property of our proposed algorithms.

The forward hydrological model is solved on a regular lattice $\mathcal{T}_\Omega$ composed of squares and continuously covering Ω. The spatial step is constant and denoted dx, the fixed temporal step is dt. The cell-to-cell flow routing is performed using a 2D flow direction map $\mathcal{D}_\Omega$ obtained from terrain elevation processing, with the only condition that a unique point in $\mathcal{T}_\Omega$ has the highest drainage area, on top of the routing scheme. All physical descriptors are mapped onto model grid for simplicity here.

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

2.2 Calibration Cost Function

A calibration cost function is defined in order to measure the misfit between simulated and observed discharge time series, respectively denoted $Q_g(t)$ and $Q_g^*(t)$, for $g \in 1..N_G$ gauged cells. A convex differentiable objective function is classically defined as follows:

$$J = J_{obs} + \gamma J_{reg} \tag{9}$$

with J_{obs} the observation term that measures the difference between observed and simulated quantities and J_{reg} a regularization term weighted by $\gamma > 0$.

In order to be able to employ information from several observation sites, the observation term in this study is expressed as follows:

$$J_{obs} = \sum_{g=1}^{N_G} w_g J_g^* \tag{10}$$

with w_g a weighting function explained afterwards, J_g^* a local quadratic metric "at the station", here $1-NSE$ or $1-KGE_2$ (cf. expressions in Appendix A) involving the response of the direct model. Thus the observation term J_{obs} depends on the control vector $\boldsymbol{\rho}$ through the direct model $\mathcal{M}$ (Equation 1) composed of the pre-regionalization operator $\mathcal{F}_R$ (Equation 4) and the direct hydrological model $\mathcal{M}_{rr}$ (Equation 2).

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

2.3 Regional Calibration Algorithms

The regional calibration aims to (i) reduce the misfit between observed and simulated discharges at spatially sparse gauging stations evaluated with Equation 9, while (ii) determining meaningful hydrological parameter maps $\boldsymbol{\theta}(x)$, using the spatial constraint introduced in the forward model (Equation 1) via the descriptors-to-parameters mapping $\mathcal{F}_R$. The inverse problem is written as the following convex optimization problem of the regional control vector $\boldsymbol{\rho}$:

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} J(\boldsymbol{\rho}) \tag{11}$$

The multi-site observation cost function defined in Equation 10 depends on the regional control vector $\boldsymbol{\rho}$ through the forward model (Equation 1). It is composed of the hydrological model $\mathcal{M}_{rr}$ (Equation 2) and of the regionalization operator $\mathcal{F}_R$. This regionalization operator can be expressed as either (i) a multi-polynomial mapping $\mathcal{F}_R \equiv \mathcal{P}$ (Equation 5), or (ii) an artificial neural network $\mathcal{F}_R \equiv \mathcal{N}$ (Equation 7). In both cases, the regional control vector $\boldsymbol{\rho}$ to optimize is large, and gradient based optimization methods adapted to high-dimensional inverse problems are employed.

2.3.1 Optimization Algorithm for Polynomial Pre-Regionalization

In this case, the forward model includes the polynomial descriptors-to-parameters mapping (Equation 5), i.e., $\mathcal{F}_R \equiv \mathcal{P}$ and the regional control vector is:

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

The optimization problem, represented in Equation 11, is solved using the L-BFGS-B algorithm (limited-memory Broyden–Fletcher–Goldfarb–Shanno bound-constrained) (Zhu et al., 1997). This algorithm is specially adapted to high-dimensional parameter spaces, and in our 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 solving the adjoint model, which is obtained by automatic differentiation using the Tapenade engine (Hascoet & Pascual, 2013). The entire process is implemented in the SMASH Fortran source code, where the full forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(.,\mathcal{P}(.))$ 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 global optimum $\bar{\boldsymbol{\theta}}^*$, which is obtained by a simple global optimization algorithm (Michel, 1989) of the inverse problem (Equation 11) where $\mathcal{M} \equiv \mathcal{M}_{rr}$ and $\boldsymbol{\rho} \coloneqq \bar{\boldsymbol{\theta}}$, 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 convergence criterion is determined based on the satisfaction of at least one of the following criteria:

- Maximum number of iterations;
- Cost function criterion: $\frac{J^{(i)}-J^{(i+1)}}{\max\left(|J^{(i)}|,|J^{(i+1)}|,1\right)} \leq \epsilon.10^{6}$;
- 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 Pre-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 11) can be solved, typically using Adam optimization algorithm (Kingma & Ba, 2014), that is an efficient stochastic gradient descent algorithm, capable to adapt the learning rate based upon the first and the approximation of the second moments of the gradients for fast convergence, and only requiring first order gradients of the cost function. In the present case, the cost function writes as:

$$J(\boldsymbol{\rho}) = J\big(Q^*, \mathcal{M}_{rr}(.\ ,\ \boldsymbol{\theta} = \mathcal{N}(\boldsymbol{D}, \boldsymbol{\rho}))\big) \tag{12}$$

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 $\boldsymbol{\theta}$ served as input for (ii) the hydrological model $\mathcal{M}_{rr}$ implemented in Fortran. In order to optimize J, we need its gradients with respect to $\boldsymbol{\rho}$. The main technical difficulty here is achieving a "seamless flow of gradients" through back-propagation. To overcome this, we divide the gradients into two parts and we 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 applied to the Fortran code corresponding to $\mathcal{M}_{rr}$. Then, $\nabla_{\boldsymbol{\rho}} \boldsymbol{\theta}$ is simply obtained by analytical calculus applicable given the explicit architecture of the ANN, consisting of a multilayer perceptron. The convergence criterion is determined by a specified number of iterations in the optimization algorithm. A detailed explanation of the network architecture, backward propagation, and the optimization process can be found in Appendix B.

3 Data and Numerical Experiment

3.1 Study Area and Experimental Design

The performance of HVDA-PR using various regional optimization algorithms are evaluated on high-resolution regional modeling of two flash flood prone study areas located in the South of France, in the Ardeche and Provence regions (Figure 2). They are characterized by contrasted physical properties and catchments behaviours. The modeling approach is applied to each regional window that contains multiple gauges downstream of nested and independent catchments, used at the same time in optimization through the cost function. This is a very challenging optimization context that is well representative of hydrological regionalization problem.

In order to examine the spatio-temporal extrapolation capabilities of HVDA-PR, a priori partition of the available discharge stations is made into calibration sites and pseudo-ungauged catchments for validation. We selected as gauged stations for regionalization those with good local model performances (i.e., "donor" catchments with potentially lower modeling error). Discharge time series at gauged sites are also split into a calibration and a validation period. A set of 7 physical descriptors (Table 1) available over the whole French territory is used following Odry (2017) and Jay-Allemand, Demargne, et al. (2022). Note that this is a sufficient design to show the regionalization property of our algorithms and

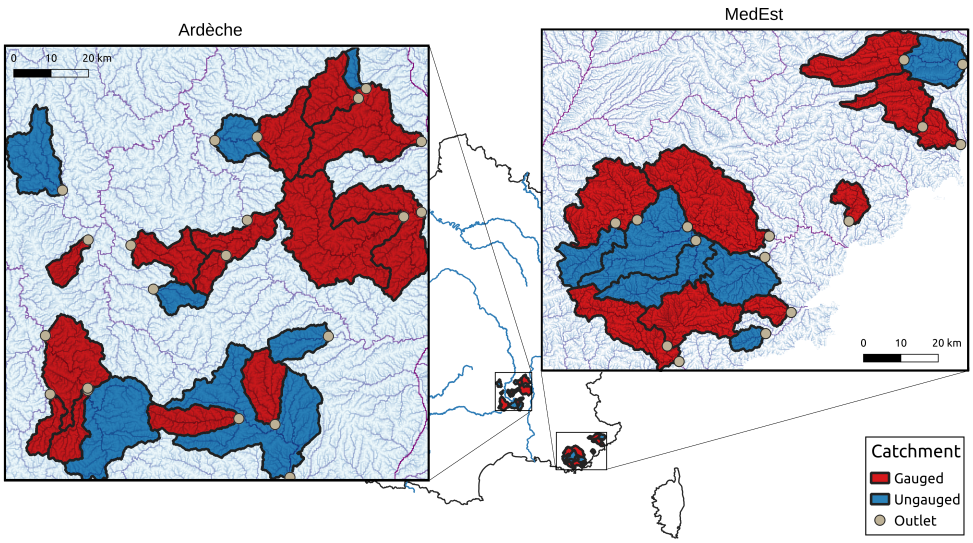

Figure 2. Two study areas on the map on France: Ardeche (left) and MedEst (right). For each area, the gauged catchments (red) are used for multi-site calibration while the pseudo-ungauged catchments (blue) are used for spatial validation.

keep the present article concise, while the issue of information selection (multi-source observations of hydrological responses and physical descriptors) is intentionally left for further research with several upgrades required for the algorithms (dedicated observation operators, descriptors selection layers, etc.).

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

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

The first zone studied is located in Provence region and is called "MedEst", 9 gauged catchments are used for calibration while 6 others are considered as pseudo-ungauged for validation. The second zone is located in the Ardeche region and called "Ardeche", 14 catchments are considered as gauged while 7 are considered as ungauged. Both study zones are challenging cases, with contrasted hydrological properties including steep topography and very heterogeneous soils and bedrock (e.g., Garambois et al. (2015)). They are affected by intense rainfall that trigger non linear flash flood responses. Moreover, the MedEst area is the most difficult one to model because of the significant proportion of karstic zones. The selection of catchments is based on the availability of long time series with high quality of observed flow and limited anthropogenic impacts. The SMASH model is run on a $dx = 1$ km spatial grid at $dt = 1$ h time step. It is forced by: (i) observed rainfall grids based on hourly ANTILOPE J+1 radar-gauge rainfall reanalysis from Météo-France (Champeaux et al.,

2009); (ii) potential evapotranspiration (PET) estimated using the formula of Oudin et al. (2005); and (iii) temperature data from SAFRAN reanalysis produced by Météo-France on a 8x8 km^2 spatial grid (Quintana-Seguí et al., 2008).

For each study area, we perform multi-site regional calibration methods using gauged catchments. Namely, the calibration metric chosen is NSE, which is computed using data from multiple gauges over a ten-year period (2006-2016). In what follows, the calibration methods compared consist in:

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

And multi-gauge regional calibrations with:

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

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

3.2 Regional Learning Performance

Here, we first assess four multi-gauge regional calibration methods: multi-linear regression (Multi-linear), multi-polynomial regression (Multi-polynomial), multilayer perceptron (ANN), and lumped model parameters (Uniform), with the latter serving as the baseline regionalized model. Our primary focus is on MedEst, which represents the most challenging study area.

Given the complexity and heterogeneity of the region, it is unsurprising that the global optimization approach with spatially uniform hydrological parameters $\bar{\boldsymbol{\theta}}$ is unable to accurately reproduce contrasted hydrodrological responses. Figure 3 (results for Ardeche given in Figure C3) conveys a typical result that the "Uniform" approach generally leads to poor performance in simulating discharges, whereas the other three regional learning methods result in noticeably similar signals with remarkably improved performance, both in gauged and pseudo-ungauged catchments.

These interpretations are further supported by the results presented in Figure 4 (results for Ardeche given in C1). It demonstrates that pre-regionalization methods, which incorporate information from physical descriptors, can lead to superior performance when compared to the classical calibration method using uniform mapping.

Specifically, the ANN-based mapping exhibits strong interpolative capabilities, achieving best results in gauged catchments. The efficiencies (NSE) over the nine calibrated catchments of MedEst area range between 0.76 and 0.93. In addition, the ANN yields high scores (ranging between 0.4 and 0.85) compared to the other three methods in 5 out of 6 pseudo-ungauged catchments for spatial validation. However, the complex heterogeneity in some areas can pose challenges for learning regionalization, as was observed in catchments

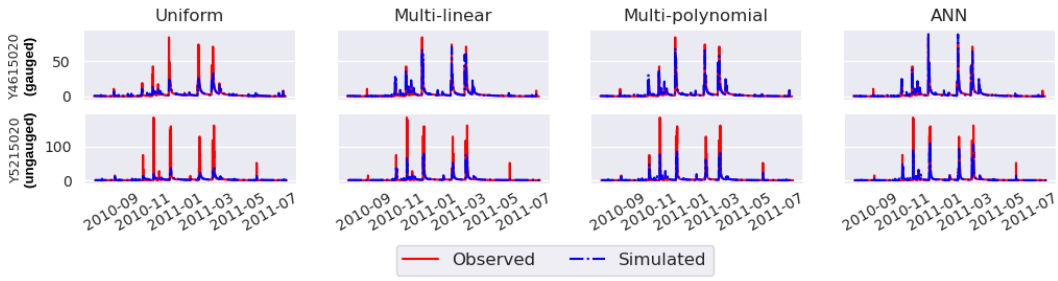

Figure 3. Study area: MedEst. Observed and simulated discharges (in m^3/s) in hourly time step at gauged catchment (Y4615020) and pseudo-ungauged catchment (Y5215020).

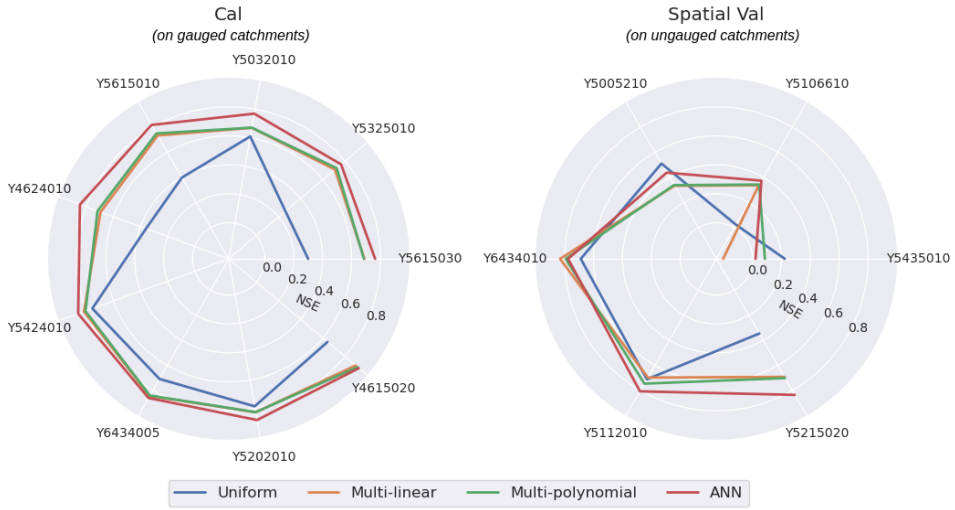

Figure 4. Study area: MedEst. Radial plots of the NSE (optimal value = 1) in gauged catchments (left) and pseudo-ungauged catchments (right) during the calibration period (2006-2016) for four calibration methods.

Y5435010 and Y5005210. The scores of local calibration methods (spatially uniform and spatially distributed calibrations) and the four regionalization methods are summarized in Figure 5 for both study areas, with a significant disparity in calibration and validation results observed among the regionalization methods in the MedEst area. In this case, the mulvariate regressions and ANN demonstrate impressive results, achieving a median NSE greater than 0.6 in both temporal and spatio-remporal validation. Notably, the ANN attains a performance level comparable to the reference benchmarks, reaching a median NSE score of approximately 0.9 in temporal validation for MedEst area.

Finally, to obtain a more robust evaluation criterion adapted to flood modeling, we consider validation in terms of hydrological signatures for numerous flood events, which are computed via an automated segmentation algorithm proposed by Huynh et al. (2023). Once again, the three pre-regionalization methods, with a particular focus on the ANN, demonstrate their ability to outperform the global optimization method, in both temporal and spatio-temporal validation, using multiple validation metrics based on flood event hydrological signatures, as shown in Figure 6.

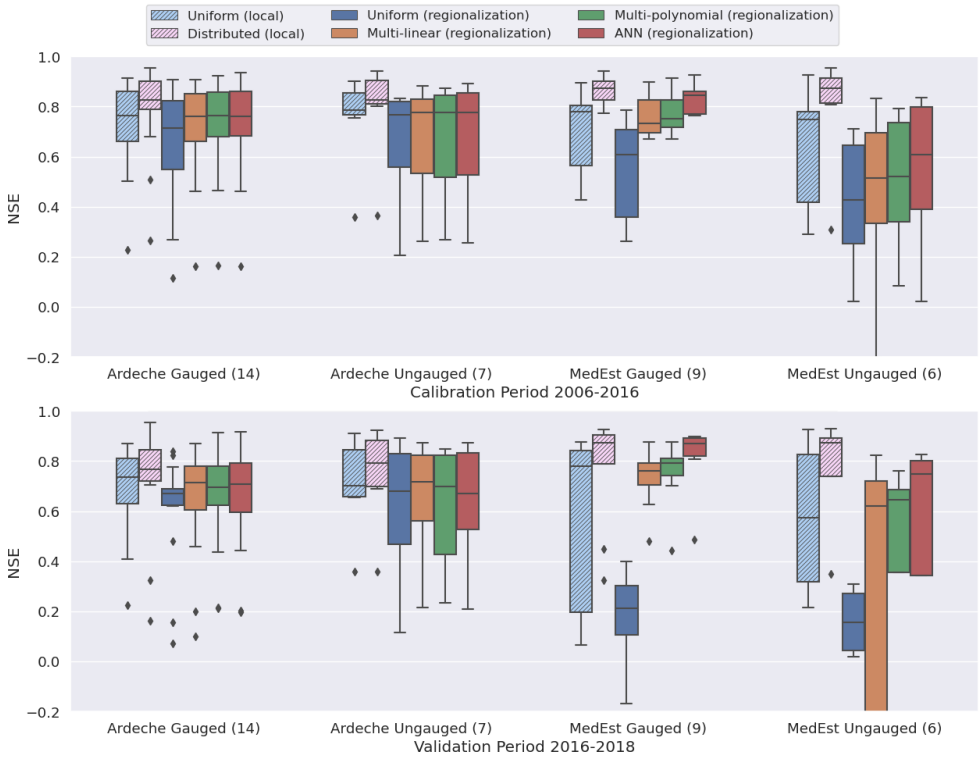

Figure 5. Comparison of NSE scores across gauged and pseudo-ungauged catchments during the calibration (upper sub-figure) and validation (lower sub-figure) periods for two local calibration methods and four regionalization methods in both study areas. The numbers in parentheses indicate the number of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

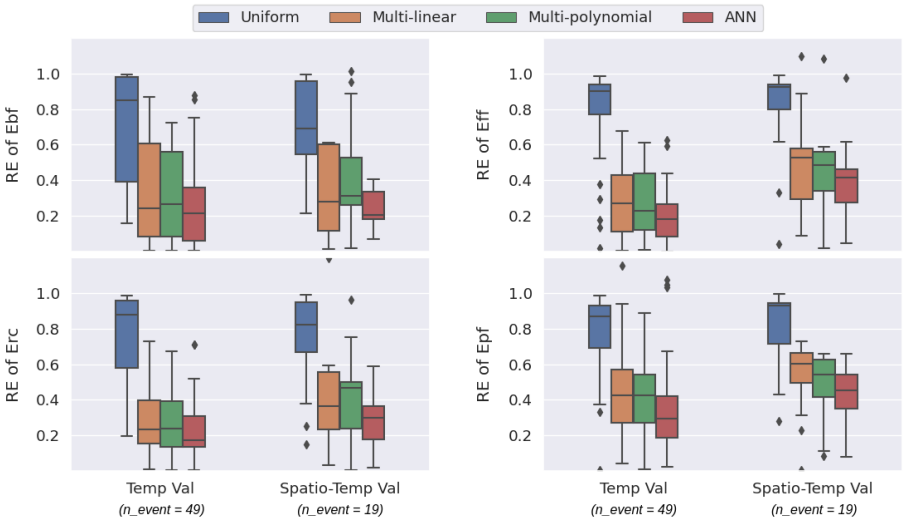

Figure 6. Study area: MedEst. Relative error (optimal value = 0) of four flood event signatures (Ebf - base flow, Eff - flood flow, Erc - runoff coefficient, Epf - peak flow) evaluated using 49 flood events ($n\_event = 49$) at gauged catchments for temporal validation (Temp Val) and 19 flood events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

The relative errors (median over flood events) of simulated signatures using these three pre-regionalization methods are around 0.2 compared to more than 0.8 for uniform calibration in temporal validation, and between 0.2 and 0.6 compared to more than 0.7 in spatio-temporal validation. It is noteworthy that these metrics were not included in the cost function during the model calibration process, which further supports the robustness and power of our pre-regionalization methods, particularly with ANN-based mapping, for seamless regionalization in hydrology in general, and flood modeling in particular.

4 Discussions

Learning the spatial variability of conceptual hydrological parameters is generally difficult to achieve with simple regression methods. However, complex regional methods, including ANNs, can result in a potential trade-off between regional model complexity and physically explainable results. In other words, complex regional mappings can reduce the misfit between observed and simulated hydrological responses, while their ability to produce physically interpretable results may be questioned. This section presents compelling findings and insightful observations obtained from the calibration of HVDA-PR, while providing a side-by-side comparison between the two study areas, with a particular focus on the learning process using ANN.

When calibrating a (hydrological) model with gradient-based optimization algorithms, it is also important to discuss on the descent of the cost function. This analysis enables us to understand how optimization algorithms converge towards the global or local minimum of the cost function, and identify potential trade-offs between model complexity and over-parameterization, in addition to validation results. We observed similar analysis results in both study areas, including the descent of the cost function J and its projected gradient $\nabla_{\boldsymbol{\rho}} J$, which are typically represented in Figure 7.

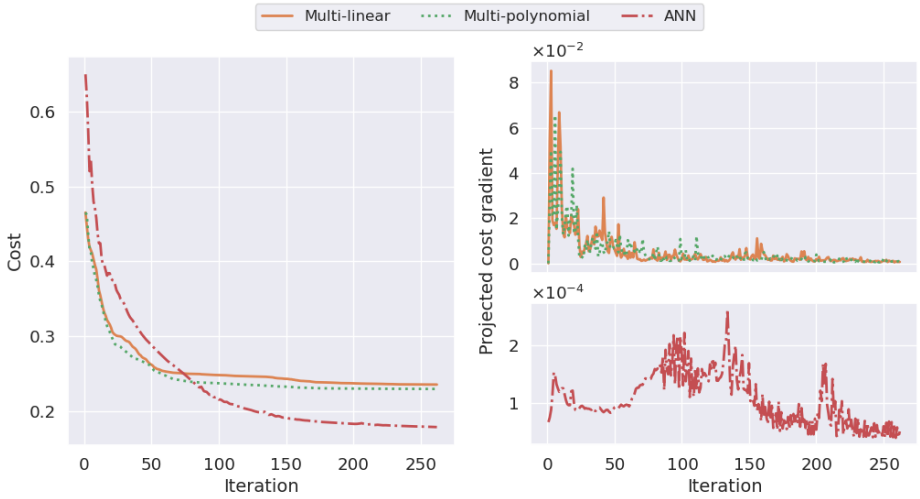

Figure 7. Study area: MedEst. Cost descent $J = 1 - NSE$ (left) and projected gradient descent $||\nabla_{\boldsymbol{\rho}} J||_\infty$ (right) of three regional calibration methods.

It is apparent that the cost functions in the two regression methods start from a more optimal point (approximately 0.46) than the ANN since they use a uniform background solution obtained by a global optimization method, as mentioned in 2.3.1. Furthermore, they converge after around 200 iterations and remain monotonous throughout the optimization process. The polynomial regression approach achieves a slightly lower cost than the linear

approach, likely due to its nature and the fact that it has nearly twice the number of parameters (Table 2), while the ANN, with a significantly larger number of parameters, can achieve a lower cost despite starting from a higher cost.

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

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

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

This is possibly due to the Adam optimizer used to optimize the ANN, which is more flexible in escaping local minima. Moreover, the variability of the cost function in this case can be changed throughout training, as evidenced by several instances within the first 100 iterations, further supporting this argument. Specifically, the projected gradient provides a clearer view that the solutions of the ANN can explore different paths and avoid getting stuck in local minima in certain situations. As shown in the bottom left sub-figure of Figure 7, there are four instances (around iterations 5, 90, 130, and 210) where significant changes occur in the control vector space (biases and weights) to update the optimum. This property could be essential in tackling equifinality and reach robust global optimum even with different starting points, which is tested with a specific stochastic initialization method, as will be discussed later. Another point to interpret is the gradient values in the Multi-linear case that are marginally higher than those in the Multi-polynomial case, but significantly higher than in the ANN case. This discrepancy can be attributed to the complexity of the regionalization mapping in each case, with the ANN case being the most complex. To alleviate the vanishing gradient problem inherent in the ANN case, we employed several techniques. First, we applied Xavier initialization to the weights, maintaining a reasonable magnitude of the gradients. Second, we utilized the ReLU activation function and 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 complexity and exacerbation of the vanishing gradient problem. Ultimately, we employed a relatively high initial learning rate (e.g., 0.005) to prevent the gradients from shrinking excessively during training.

Despite reasonably similar simulated discharges across the three regionalization methods for MedEst (Figure 3), the methods resulted in different validation performances (as shown in Figures 4 and 6). Now, we observe that the spatial variability of each hydrological model parameter differs notably across these three methods (Figure 8b). In contrast, for Ardeche, all regionalization methods result in similar distributed parameter maps with significantly reduced spatial variability (Figure C4b).

At this stage, one can inquire about the extraction of information from physical descriptors to hydrological parameters via the pre-regionalization mapping. For instance, how can we ensure that HVDA-PR maintains the desired physical properties of the original descriptors? Additionally, we may question the use of complex pre-regionalization mappings such as ANNs, which may result in hydrological parameters that are vastly different from the input descriptors. To address these questions, it is worth noting that the "safest"

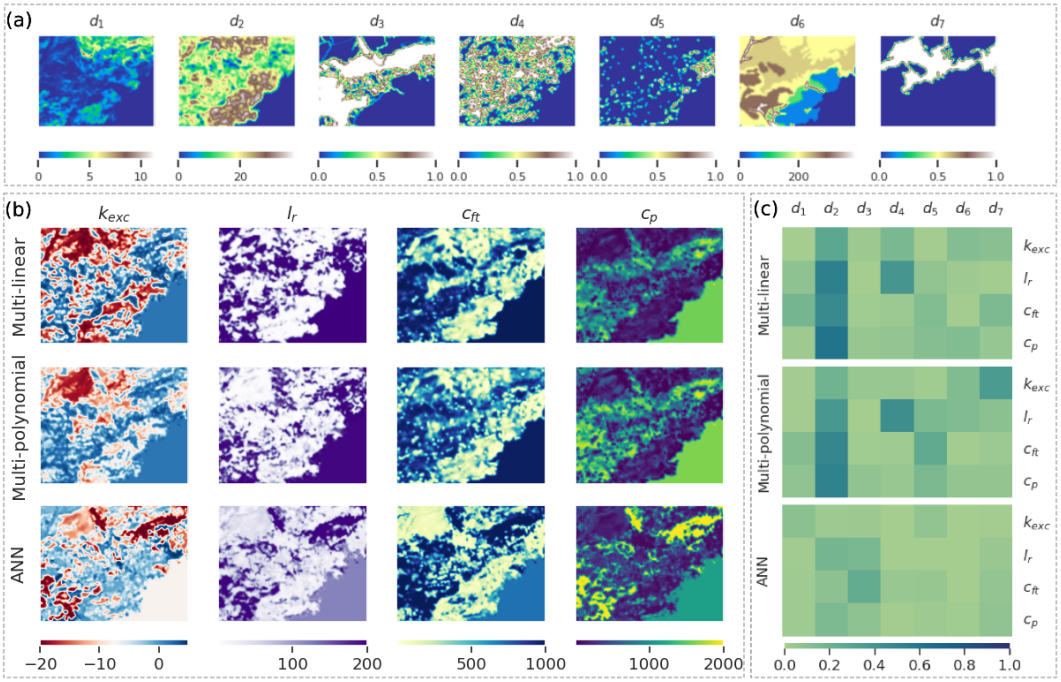

Figure 8. Study area: MedEst. Sub-figure a: The maps of input descriptors (d_1-d_7), whose information is provided in Table 1. Sub-figure b: Calibrated hydrological parameters ($k_{exc}, l_r, c_{ft}, c_p$) for three regionalization methods. Sub-figure c: Linear covariance between descriptor and parameter for three regionalization methods.

approach is to use multi-linear regression, which relies solely on the discernible correlation between descriptors and model parameters. Such correlation can be quantified using one-to-one parameter-descriptor linear correlation matrices (Figure 8c). In the case of multi-polynomial regression, the risk of losing physical properties may arise when the polynomial degree is unbounded. Optimization algorithms could employ super higher-order correlations between descriptors and parameters that do not exist physically. To mitigate this risk, our study imposed a boundary condition on the polynomial degree, $0.5 \leq \beta_{k,d} \leq 2$, as mentioned earlier in 2.3.1. The ANNs, however, pose the most complicated case, where the control vector (that is, the weights and biases) consists of numerous parameters that are difficult to physically constrain. Our hands-on experience indicates that a multilayer perceptron with two or three hidden layers is sufficient for learning the parameters of a parismonious conceptual distributed hydrological model without under- or over-extracting the physical information of the input descriptors. Note that the number of neurons in each layer must be reasonable, which should not exceed $\sqrt{N_D.N_x}$ based on our experiments. In our study, we can still observe physical properties in the resulting maps, primarily through visualization (Figure 8b), although the descriptor-parameter correlation is less evident compared to the two regression cases (Figure 8c). In general, the physical properties of the drainage density (d_2) are easily observable in the distributed parameter maps, as evidenced by the high linear correlation between this descriptor and the four model parameters. These correlations are not limited to linear and polynomial forms and can be explored further with ANN. Therefore, it is expected that the ANN would explore more complex correlations beyond linear and polynomial ones. Furthermore, it is noteworthy that the ANN can identify stronger linear correlations (compared to both regression cases) of the routing parameter (l_r) and the transfer parameter (c_{ft}) with the karst index (d_3), which is typically sophisticated in

hydrological regionalization. Compared to the results in Ardeche, where the problem is less challenging, we observe that the ANN can still employ linear correlations between descriptor and parameter, which are similar to those found in both regression cases (see Figure C4c). This finding demonstrates the robustness and flexibility of our ANN, which proves to be efficient not only in complex study areas but also in simpler ones. The validation results in Ardeche further support this, as the ANN performs well even under spatio-temporal validation using multiple hydrological signatures (see Figure C2).

Regarding regionalization over larger areas, such as for large basins or at country scales, for dealing with significant physical heterogeneity, an increased flexibility in the regional mapping might be needed. This can be achieved through the use of spatialized regional controls, for example as done in regional calibration for catchment clusters determined with a similarity measure (Huang et al., 2019). In our proposed HVDA-PR framework, the definition of transfer functions allows us to consider flexible mappings, as well as spatialized regional controls through masked descriptor maps, for each hydrological parameter independently or jointly. This enables to explore transfer functions at pixel scale or on a clustering of the spatial domain, for example into sub-regions or hydrological response units (HRU). In this work, the capabilities of HVDA-PR have been successfully demonstrated in a high-dimensional and challenging high-resolution flash-flood modeling context. Determining effective physical descriptor sets from large databases as well as finding optimal spatial flexibility represent interesting research avenues for further investigation of optimal regional data assimilation approaches.

Last but not least, we want to highlight that the proposed ANN scheme in the HVDA-PR algorithms has been constructed incrementally and that it largely relies on the mathematical and numerical understanding and know-how built with the regression methods studied in a known VDA context.

5 Conclusion

A Hybrid Variational Data Assimilation Parameter Regionalization (HVDA-PR) approach, extending the MPR technique, has been introduced in this study. We investigated the potential of incorporating learnable pre-regionalization mappings, including multivariate polynomial regressions and neural networks, into a high-resolution spatially distributed hydrological VDA framework. To the best of our knowledge, we present the first implementation of ANNs within this context, enabling seamless regionalization in hydrology. Effective optimization algorithms capable to perform high-dimensional optimizations from multi-source data have been obtained with:

- effective regional transfer functions of adaptable complexity, enabling to exploit information from heterogeneous data sources, with flexible formulations and spatial rigidity;
- a differentiable forward hydrological model, including the regional mappings, that enables accurate computation of spatially distributed gradients of multi-gauge cost - which is crucially needed in context of sparse observations (i.e., cost evaluation locations), and relatively small gradient values and spatial variability;
- optimization algorithms, adapted to high dimensional problems, with seamless flow of cost gradients enabling to obtain physically explainable results, especially when combined with physical descriptors and spatial gradients, which efficiently enhance the transferability of geophysical properties from gauged to ungauged locations.

HVDA-PR has been tested on two very challenging regional optimization cases from multi-gauge discharge and descriptors maps, with a high-resolution conceptual hydrological modeling at 1 km^2 and 1 h of two flash flood prone areas characterized by heterogeneous physical properties and non-linear hydrological behaviours. The results obtained on both zones, and especially on MedEst being the most challenging case, highlight the effectiveness

of HVDA-PR that utilizes physical descriptors, surpassing the performance of the global optimization method with lumped model parameters. Notably, the ANN exhibited superior performance, even comparable to the reference benchmarks, thereby establishing its remarkable capability in challenging modeling scenarios as well as on less challenging ones. The median NSE scores of all HVDA-PR methods (including multivariate regressions and ANN) are greater than 0.75 in calibration in both study areas, whereas the scores in spatio-temporal validation surpass 0.6 in MedEst (in comparison to the distributed calibration benchmark of approximately 0.85) and exceed 0.65 in Ardeche (in contrast to the distributed calibration benchmark of around 0.8). Various flood event signatures computed from Huynh et al. (2023) are also used as validation metrics to demonstrate the robustness of HVDA-PR, where the three regionalization methods using descriptors outperform the lumped model. For instance, when considering spatio-temporal validation across 19 flood events in MedEst, these three methods yield median relative errors of the peak flow below 0.6, whereas the remaining method yields errors exceeding 0.9. Several interesting findings and results have also been discussed in Section 4. These findings highlighted the cost descent behaviour and the interrelation between input descriptors and output parameter maps, demonstrating the effectiveness of HVDA-PR in extracting information from physical descriptors while producing physically explainable results.

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

A Metrics

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

- Nash-Sutcliffe observation cost function:

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

- Kling-Gupta observation cost function:

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

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

B Architecture and Forward/Backward Propagation of ANN-based Regionalization

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).

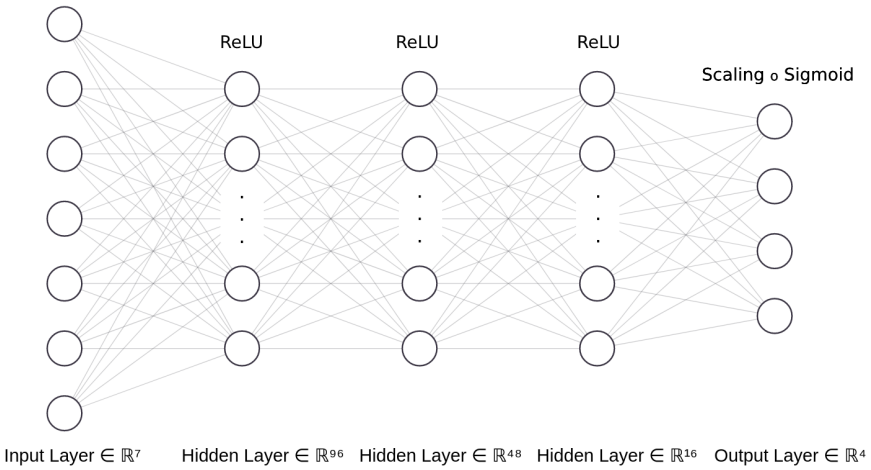

Figure B1. The architecture of our ANN consists of three hidden layers followed by the ReLU activation function and an output layer that uses the Sigmoid activation function in combination with a scaling function. In this particular case, we have $N_D = 7$, $N_L = 4$ and $N_\theta = 4$. The calculation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

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 W_j and bias b_j. Then, an input I of each layer is mapped to the input of the next layer by a linear function $\phi_j(I) = W_j I + b_j$, 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 $O_x = \sigma \circ \phi_{N_L}(., x) \in [0, 1]^{N_\theta}$ of the last layer is mapped to the range of the hydrological model parameters by a differentiable scaling function s:

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

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 matricial product.

Once denote $\Psi_j \equiv \begin{cases} \delta \circ \phi_j,\ j = 1..N_L - 1 \\ \sigma \circ \phi_j,\ j = N_L \end{cases}$, 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 11 optimized in this case by the regional control vector $\boldsymbol{\rho} \coloneqq [\boldsymbol{W}, \boldsymbol{b}]$, using a cost function as Equation 12. 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 $\frac{\partial J}{\partial \rho_j}$ for each layer, where $\rho_j \coloneqq [W_j, b_j]$. Since the forward model $\mathcal{M} \equiv \mathcal{M}_{rr}(., \mathcal{N}(.))$ with $\boldsymbol{\theta}$ being an output variable of the mapping function $\mathcal{N}$ and an input of $\mathcal{M}_{rr}$, we can write $\frac{\partial J}{\partial \rho_j} = \frac{\partial J}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \rho_j}$. Then these two gradients are obtained as follows:

- The gradients of the cost function with respect to the hydrological model parameters $\frac{\partial J}{\partial \boldsymbol{\theta}}$, computed by solving the adjoint model $\nabla_{\boldsymbol{\theta}} J$ of $\mathcal{M}_{rr}$;
- The gradients of the network output with respect to the weight and bias $\frac{\partial \boldsymbol{\theta}}{\partial \rho_j}$, 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.

Algorithm 1 The proposed back-propagation using Adam optimizer

▷ Randomly initialized weights and biases $\boldsymbol{\rho}^{(0)} = \left(\rho_1^{(0)}, ..., \rho_{N_L}^{(0)}\right)$

▷ Number of training iterations N_{ite}

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

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

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

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

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

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

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

$\beta_1 = 0.9$ 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

C Calibration results in case of Ardeche

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

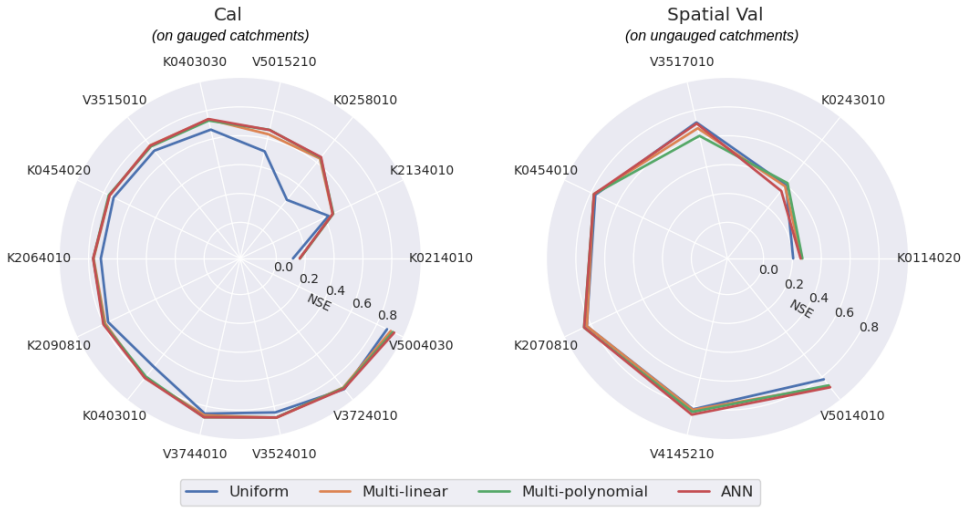

Figure C1. Study area: Ardeche. Radial plots of the NSE (optimal value = 1) in gauged catchments (left) and pseudo-ungauged catchments (right) during the calibration period (2006-2016) for four calibration methods.

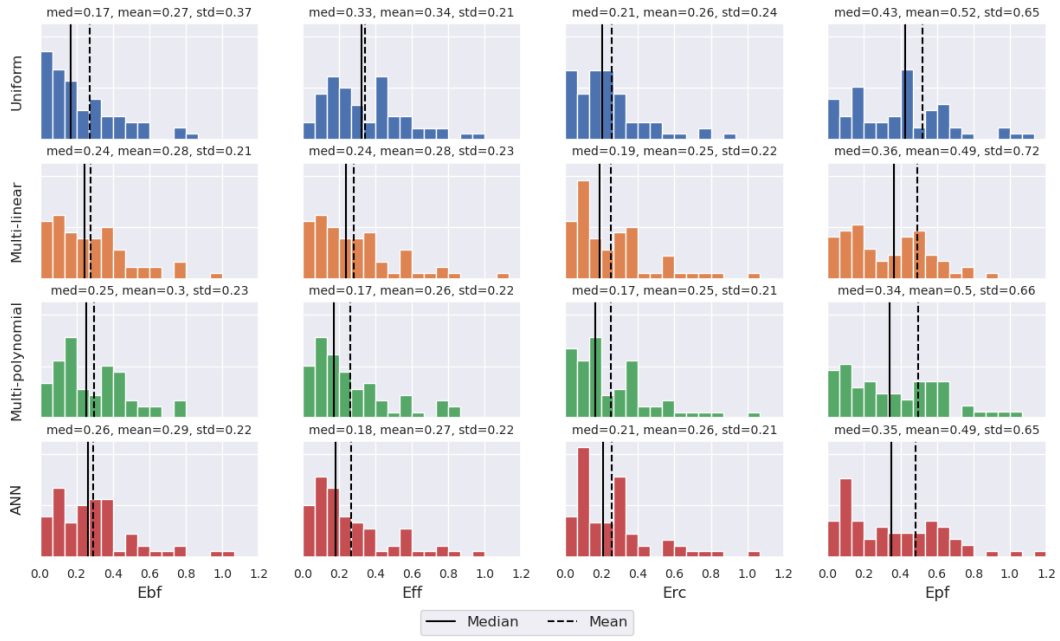

Figure C2. Study area: Ardeche. Distribution over 67 flood events of relative error (optimal value = 0) of four flood event signatures (Ebf - base flow, Eff - flood flow, Erc - runoff coefficient, Epf - peak flow) evaluated in pseudo-ungauged catchments during the validation period 2016-2018.

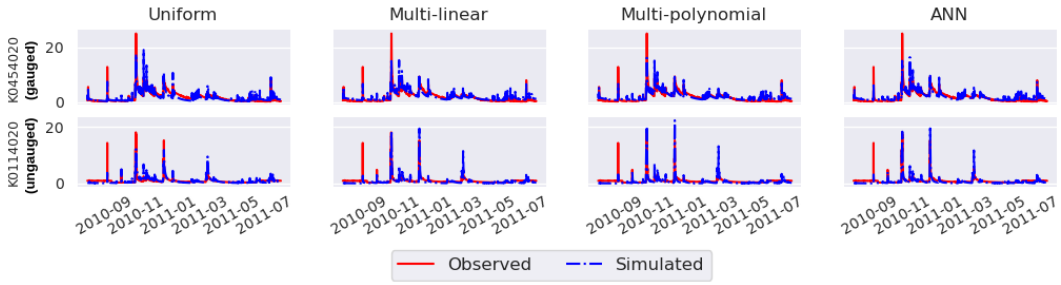

Figure C3. Study area: Ardeche. Observed and simulated discharges (in m^3/s) in hourly time step at gauged catchment (K0454020) and pseudo-ungauged catchment (K0114020).

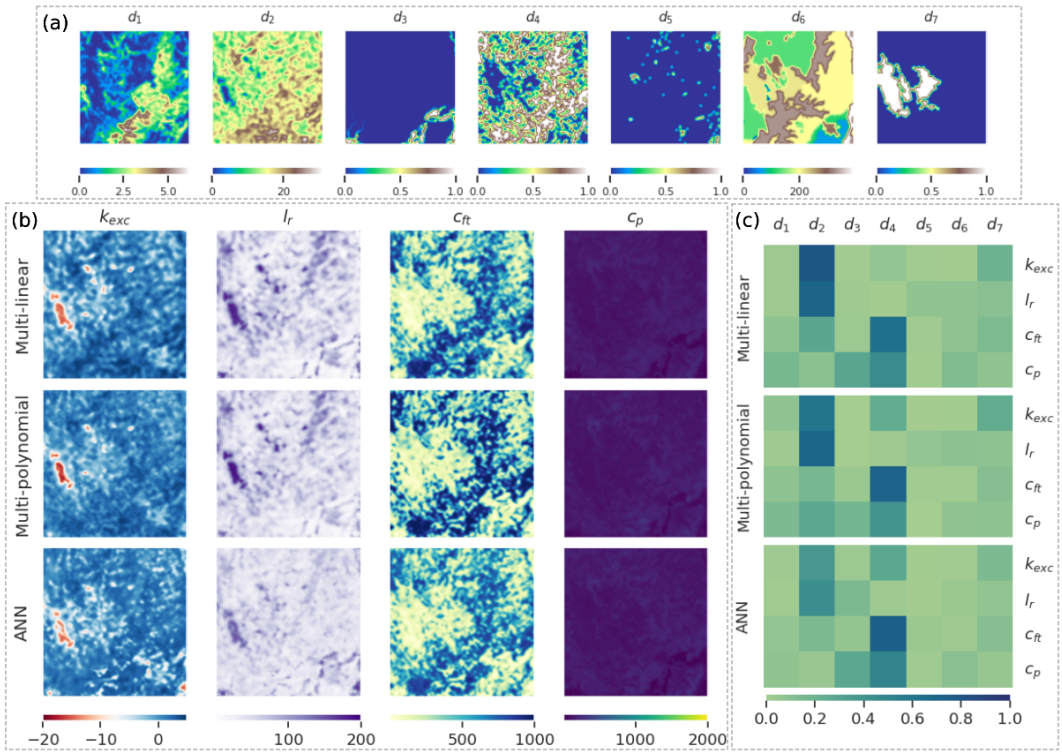

Figure C4. Study area: Ardeche. Sub-figure a: The maps of input descriptors (d_1-d_7), whose information is provided in Table 1. Sub-figure b: Calibrated hydrological parameters ($k_{exc}, l_r, c_{ft}, c_p$) for three regionalization methods. Sub-figure c: Linear covariance between descriptor and parameter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

Data and Software Availability Statements

The proposed algorithms were implemented into the SMASH platform, which is interfaced in Python (Jay-Allemand, Colleoni, et al., 2022). The SMASH source code can be accessed at: `https://github.com/DassHydro-dev/smash`. Additionally, the scripts used for conducting the numerical experiments and analysis in this study are provided

at: https://github.com/nghi-truyen/Regionalization-Learning. The data supporting this study will be made available by the corresponding author upon reasonable request.

Acknowledgments
The authors greatly acknowledge SCHAPI-DGPR and Météo-France for providing data used in this study; Killian Pujol-Nicolas for preliminary work on pre-regionalization with SMASH during his master thesis internship; 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.

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169408002187 doi: https://doi.org/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), e2019JD031485.

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. Retrieved from https://hess.copernicus.org/articles/5/1/2001/ 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.

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.*

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.

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

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015-1031. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/91WR02985 doi: https://doi.org/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), e2021MS002572.

Fekete, B. M., & Vörösmarty, C. J. (2007). The current status of global river discharge

monitoring and potential new technologies complementing traditional discharge measurements. *IAHS publ*, *309*, 129–136.

Fortin, F.-A., Rainville, F.-M. D., Gardner, M.-A., Parizeau, M., & Gagné, C. (2012). Deap: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, *13*(70), 2171–2175. Retrieved from http://jmlr.org/papers/v13/fortin12a.html

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169419311448 doi: https://doi.org/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.

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*.

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.

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.

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. Retrieved from https://doi.org/10.1080/02626667.2013.803183 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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0022169419302495 doi: https://doi.org/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., Garambois, P.-A., Colleoni, F., & Javelle, P. (2023). Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to mediterranean floods. *arXiv preprint arXiv:2305.19307*. Retrieved from https://arxiv.org/abs/2305.19307

Jay-Allemand, M., Colleoni, F., Garambois, P.-A., Javelle, P., & Julie, D. (2022). Smash - spatially distributed modelling and assimilation for hydrology: Python wrapping towards enhanced research-to-operations transfer. *IAHS*. Retrieved from https://hal.archives-ouvertes.fr/hal-03683657

Jay-Allemand, M., Demargne, J., Garambois, P.-A., Javelle, P., Gejadze, I., Colleoni, F., ... Fouchier, C. (2022). *Spatially distributed calibration of a hydrological model with variational optimization constrained by physiographic maps for flash flood forecasting in france* (Tech. Rep.). Copernicus Meetings. Retrieved from https://doi.org/10.5194/iahs2022-166 doi: 10.5194/iahs2022-166

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.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. Retrieved from https://arxiv.org/abs/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). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005WR004362 doi: https://doi.org/10.1029/2005WR004362

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), e2020WR028393.

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.

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0309170816302214 doi: https://doi.org/10.1016/j.advwatres.2016.07.005

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.

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. Retrieved from https://gmd.copernicus.org/articles/15/6085/2022/ 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. Retrieved from https://journals.ametsoc.org/view/journals/hydr/17/1/jhm-d-15-0054_1.xml doi: https://doi.org/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). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008WR007248 doi: https://doi.org/10.1029/2008WR007248

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).

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. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.5155 doi: https://doi.org/10.1002/hyp.5155

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. Retrieved from http://dblp.uni-trier.de/db/journals/toms/toms23.html#ZhuBLN97