Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to Mediterranean floods

Ngo Nghi Truyen Huynh, Pierre-André Garambois, François Colleoni, Pierre Javelle

Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to Mediterranean floods

Ngo Nghi Truyen Huynh[a], Pierre-André Garambois[a], François Colleoni[a] and Pierre Javelle[a]

[a]*INRAE, Aix Marseille Université, RECOVER, 3275 Route Cézanne, Aix-en-Provence 13182, France*

ARTICLE INFO

Keywords:
Hydrological modeling
Hydrological signatures
Variational data assimilation
Multi-objective optimization
Pareto-optimal solution
Variance-based sensitivity analysis

ABSTRACT

Classical calibration methods in hydrology typically rely on a single cost function computed on long-term streamflow series. Even when hydrological models achieve acceptable scores in NSE and KGE, imbalances can still arise between overall model performance and its ability to simulate flood events, particularly flash floods. Enhancing multi-criteria calibration methods with multi-scale signatures to improve distributed flood modeling remains a challenge. In this study, the potential of hydrological signatures computed continuously and at the scale of flood events, are employed within various multi-criteria calibration approaches to attain a more efficient hydrological model. We present a novel sensitivity and signatures-based calibration framework, implemented in the variational data assimilation algorithm of SMASH platform, which we apply to 141 catchments mostly located in the French Mediterranean region. Our approach involves computing several signatures, including flood event signatures, using an automated flood segmentation algorithm. We select suitable signatures for constraining the model based on their global sensitivity with the input parameters. We then perform two multi-criteria calibration strategies using the selected signatures, including a single-objective optimization approach, which transforms the multi-criteria problem into a single-objective function, and a multi-objective optimization approach, which uses a simple additive weighting method to select an optimal solution from a set of non-inferior solutions. Our results show significant improvements in both calibration and temporal validation metrics, especially for flood signatures, demonstrating the robustness and delicacy of our signatures-based calibration framework for enhancing flash flood forecasting systems.

1. Introduction

Numerical hydrological models are used extensively to simulate catchments responses to atmospheric signals and are a key component of floods forecasting systems where accuracy in terms of peak location, amplitude and timing is crucial. As a matter of facts, hydrological models, whatever their complexity and spatialization, consist in more or less empirical representations of flows through watersheds compartments and contain parameters that cannot be inferred directly from the available observations but can only be meaningfully estimated through a calibration procedure (e.g. Gupta et al. (2006); Vrugt et al. (2008)). Such procedures aim to improve the model capability in reproducing the available observations of hydrological responses dynamics by optimizing model parameters.

Nevertheless, the whole construction process of a hydrological model is faced with the issue of equifinality: distinct model structures and/or parameter sets can lead to similar (in a sense to be defined) simulations. The equifinality concept has been popularized in hydrology by Beven (1993) while the issues of uncertainty in determining

ngo-nghi-truyen.huynh@inrae.fr (N.N.T. Huynh); pierre-andre.Garambois@inrae.fr (P. Garambois); francois.colleoni@inrae.fr (F. Colleoni); pierre.javelle@inrae.fr (P. Javelle)
ORCID(s): 0000-0001-5078-3865 (N.N.T. Huynh)

environmental model structures and estimating their parameters were known (e.g. Beck (1987); Yeh (1986)). For a given hydrological model structure, the calibration of its parameters is in general an ill-posed inverse problem with non unique solutions and the definition of an optimization algorithm and of a calibration metric is an essential modeling decision. Indeed, it determines how hydrological information is seen and learn in the calibration process and it can substantially affect the quality and consistency of model simulations.

In hydrology, most calibration approaches attempt to optimize input parameters of a model such that they result in a minimal misfit between simulated and observed discharge. Nevertheless, because no single metric can exhaustively represent this misfit, the calibration of a hydrological model is "inherently multi-objective" as remarked by Gupta et al. (1998). Several performance metrics have been proposed over the past decades in the literature for hydrological modeling. The classical quadratic Nash-Sutcliffe efficiency (NSE) Nash and Sutcliffe (1970) (cf. Appendix A.1) has been used for long time. The Kling–Gupta (KGE) (cf. Appendix A.2) proposed in Gupta et al. (2009) and based on a decomposition of the NSE has also become widely used. Other metrics, in form of signature measures (see review in McMillan (2021)), have been proposed in the literature for model evaluation (e.g. Yilmaz et al. (2008)) and used in model optimization (e.g. Roux et al. (2011); Shafii and Tolson (2015); Mostafaie et al. (2018); Sahraei et al. (2020); Wu et al. (2021) and references therein). Hydrological signatures can be used to derive application-specific metrics such as for high flows in Mizukami et al. (2019) or Roux et al. (2011). Moreover, hydrological signatures are a useful tool to effectively evaluate models and diagnose the role of their components in explaining the discrepancy between the simulated and observed behavior (Gupta et al., 2009), especially when combined with global sensitivity analysis (Horner, 2020). Nonetheless, there is still a need for automated methods capable of computing signatures on observed and modeled hydrological responses, at multiple time scales with the underlying difficulties of consistent segmentation of flood events as highlighted in Tarasova et al. (2018), of computing global sensitivity analysis of simulated signatures with respect to the model parameters (Horner, 2020), and finally of performing signature based parametric optimization.

Although the concept of flood event is widely used in hydrology, there is no clear consensus on approaches for flood detection from continuous streamflow time series, as pointed out in Tarasova et al. (2018). Several studies have suggested segmentation algorithms for detecting flood events (refer to the references in Tarasova et al. (2018)). For instance, Li et al. (2022); Astagneau et al. (2021) used simple segmentation methods respectively involving fixed time windows before and after rainfall events or discharge thresholds to detect events. Meanwhile, Tarasova et al. (2018) developed an algorithm incorporating, baseflow separation technique (see also Pelletier and Andréassian (2020)), rainfall attribution methods and an iterative procedure to identify single-peak components of multiple-peak events. In this study, we propose an automated segmentation algorithm, consisting of, peak detection in discharge series, catchment rainfall time series analysis through a combination of rainfall gradients and rainfall energy criterion, which enables a robust determination of flood start time on contrasted catchment-floods, and a classical baseflow separation for determining the end of an event.

Hydrological calibration problems that incorporate multiple metrics, including multi-scale signatures, can be considered as multi-criteria optimization problems. Generally, three categories of methods are employed for solving multi-criteria optimization problems in various domains: (i) transforming the multi-criteria problem into a single-objective optimization problem (Ross et al., 2015; El-Ghandour and Elbeltagi, 2014; Veluscek et al., 2015); (ii) obtaining a non-inferior solution set (Pareto front) by solving the multi-objective optimization problem (Khorram et al., 2014; Tavakkoli-Moghaddam et al., 2011; Torres-Treviño et al., 2011); (iii) selecting a unique solution after obtaining the Pareto optimal solution set by adding constraints based on specific preferences (Chibeles-Martins et al., 2016; Wu et al., 2015). The state-of-the-art in multi-criteria optimization in hydrology is commonly accomplished through the first two approaches mentioned earlier. For instance, a simple approach on the choice of calibration metrics for flood modeling, including NSE, weighted KGEs, and annual peak flow biais, has been proposed for daily mHm and VIC models on 492 US catchments by Mizukami et al. (2019). For event-based flash flood modeling at high resolution, a metric that accounts for the shape of flash flood hydrographs, particularly their timing and maximum peak flow, has been studied in Roux et al. (2011). Nevertheless, generalizing these methods for multi-scale signatures and integrating them into variational data assimilation algorithms remain significant challenges. Research on calibration with multi-objective functions to generate a set of non-dominated solutions has also been conducted, as seen in studies by Yapo et al. (1998); Guo et al. (2014); Oliveira et al. (2021); Mostafaie et al. (2018). However, the selection of an optimal solution from the non-dominated set has not received much attention. Our goal in this work is to comprehensively investigate all feasible multi-criteria optimization methods using a more general approach. This research will address 4 aspects that have received relatively little attention in prior studies: (i) the need for an automated segmentation method applicable to large contrasted catchment-floods samples and capable to capture hydrological information at the scale of flash flood events; (ii) a global analysis of simulated errors across various hydrological signatures and their sensitivity with the model parameters; (iii) the need for a more intelligent approach to select the Pareto optimal solution in the case of optimization with multi-objective functions; and (iv) the computation of the cost function based on signatures within variational data assimilation algorithms.

In this work, we focus on multi-criteria calibration metrics with single-objective functions and with multi-objective functions for uniform parameters, then a multi-criteria calibration with single-objective function for distributed parameters of a distributed model aimed at flood modeling. Note that multi-objective optimization is a widely used method for multi-criteria calibration, which attempts to simultaneously minimize multi-objective functions to obtain a set of optimal solutions (also called Pareto solution) rather than a single solution. To address such a optimization problem in the present context, where the optimization of multiple conflicting objectives is often encountered, multi-objective genetic algorithms (MOGA) have been shown to be effective (Murata et al., 1995). Compared to multiple gradient descent algorithms (Désidéri, 2012; Mercier et al., 2018), MOGAs do not require gradient information and are therefore suitable for a wide range of objective functions. Instead, they rely on crossover and mutation operators, making them effective regardless of the nature of the problem functions. Non-dominated sorting genetic algorithm (NSGA), suggested by Deb et al. (2002), is a well known MOGA for solving multi/many-objective optimization problems, including fast and elitist approach (Deb et al., 2002). Namely, a fast sorting algorithm helps optimizing the computational complexity (even with a large population size) arising from the non-dominated sorting procedure in every generation. Into the bargain, NSGA possesses a diversity preservation property, based on a sharing function method, that prevents the loss of good solutions involved in the mating process. Recently, NSGA has also been implemented in the *pymoo* Python library (Blank and Deb, 2020), that is used in the present study thanks to the Python interface of our SMASH platform.

This study proposes an improved signature-based calibration approach for hydrological models. The approach employs hydrological signatures computed at the scale of flood events to enhance multi-criteria calibration. The proposed algorithm originally combines automated segmentation of flood events and signatures computation within a variational data assimilation (VDA) algorithm from Jay-Allemand et al. (2020) enabling high dimensional spatially distributed calibration, now with multi-criteria metrics adapted to floods. Classical global calibration algorithms have also been upgraded that way. These upgrades, including new cost functions and adjoint model update, have been implemented into the SMASH platform, which solvers are differentiable. Using the proposed algorithms, we investigate over a quite large dataset of Mediterranean flash floods the parametric sensitivity of a parsimonious distributed hydrological model for a large array of signatures from the literature, as well as the benefit of using a signature-based flood specific metric in calibration, and especially in performing variational spatially distributed optimization which has seldom been done to our best knowledge.

The remaining sections of this paper are organized as follows: section 2 describes our methodology for computing various hydrological signatures and our multi-criteria calibration algorithms, along with an overview of the SMASH forward model. In section 3, we present and analyze our results on signatures and calibration, including a summary of the data and numerical experiments. Finally, in section 4, we conclude our work and outline potential future directions.

2. Methodology

We propose a novel calibration strategy that leverages hydrological signatures and their sensitivity analysis in combination with the optimization algorithms discussed above. Our approach is illustrated in Fig. 1 and addresses the challenges of model calibration in the presence of multiple objectives and complex hydrological processes.

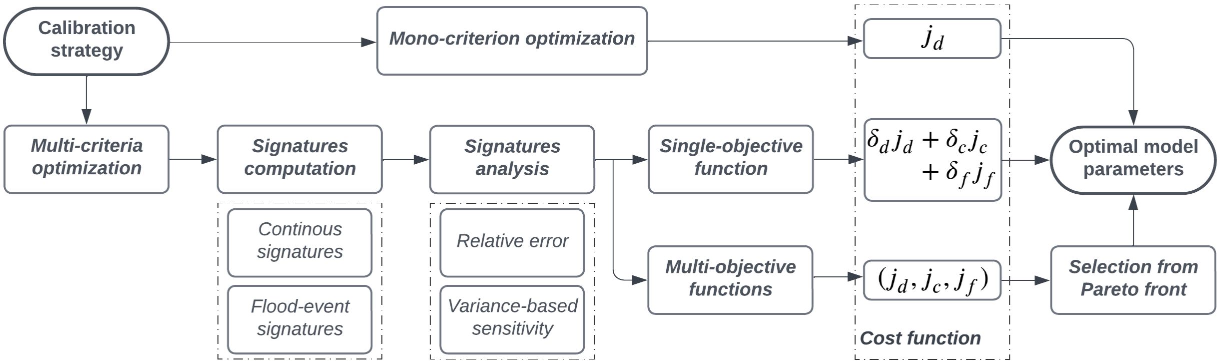

Figure 1: Flowchart of the multi-criteria calibration process using hydrological signatures. The different cost functions are denoted by j_d, j_c and j_f, while the corresponding optimal weights are denoted by δ_d, δ_c and δ_f. The notations used in the cost function will be explained in 2.4.

The computations of the signatures are first performed to quantify their sensitivities with the model parameters following Horner (2020). These computations involve performing both whole-period-based analysis to obtain continuous signatures and event-based analysis to capture the most significant events (flood event signatures). Through this analysis, we gain a more meaningful understanding of the parametric sensitivity, not just for discharge but also for other factors that need to be considered as part of our minimization criterion. Furthermore, we evaluate the sensitivity of signature error using variance-based sensitivity analysis (Sobol indices) to determine the most appropriate signatures for multi-criteria optimization. Based on these results, we conduct a multi-criteria optimization with single-objective or multi-objective functions, utilizing suitable hydrological signatures to improve the simulation performance.

The numerical algorithms proposed here are implemented in Python, on top of SMASH Fortran platform that is interfaced in Python (Jay-Allemand et al., 2022a) making accessible its forward-inverse algorithms (forward hydrological models, SBS and VDA Jay-Allemand et al. (2020) calibration algorithms) and internal variables.

The following subsections of this section detail the different elements of our methodology: 2.1 defines the hydrological model structure, the objective function and the proposed calibration algorithms; 2.2 explains which signatures are computed and how, including a description of the proposed hydrograph segmentation algorithm; 2.3 describes the method for computing global sensitivities of simulated hydrological signatures; 2.4 details the formulation of the multi-criteria cost functions including multi-scale signatures and the multi-objective optimization problems.

2.1. SMASH: An overview of the forward model and calibration algorithms

SMASH is a computational software framework dedicated to *Spatially distributed Modelling and ASsimilation for Hydrology*. It aims to tackle flexible spatially distributed hydrological modeling, signatures and sensitivity analysis, as well as high dimensional inverse problems using multi-source observations. This model is designed to simulate discharge hydrographs and hydrological states at any spatial location within a basin and reproduce the hydrological response of contrasted catchments, especially aiming at floods and low-flows modeling, by taking advantage of spatially distributed meteorological forcings, physiographic data and hydrometric observations.

First, the forward spatially distributed hydrological modeling problem is formulated as follows. Let $\Omega \subset \mathbb{R}^2$ be a 2D spatial domain (catchment) and $t > 0$ be the physical time. A regular lattice $\mathcal{R}_\Omega$ covers Ω and $D(x)$ is the drainage plan obtained from terrain elevation processing. The number of active cells within a catchment Ω is denoted N_x. Then the hydrological model is a dynamic operator $\mathcal{M}$ mapping observed input fields of rainfall and evapotranspiration $\boldsymbol{P}\left(x,t'\right),\ \boldsymbol{E}\left(x,t'\right), \forall\left(x,t'\right) \in \Omega \times [0,t]$ onto discharge field $Q(x,t)$ such that:

$$Q\left(x,t\right) = \mathcal{M}\left[\boldsymbol{P}\left(x,t'\right), \boldsymbol{E}\left(x,t'\right), \boldsymbol{h}\left(x,0\right), \boldsymbol{\theta}\left(x\right), t\right], \forall x \in \Omega, t' \in [0,t] \tag{1}$$

with $\boldsymbol{h}\left(x,t\right)$ the N_s-dimensional vector of model states 2D fields and $\boldsymbol{\theta}$ the N_p-dimensional vector of model parameters 2D fields. In the following, $\boldsymbol{\theta}$ is also called control vector in optimization context.

Then the forward hydrological model structure is defined as follows. In this study, a parsimonious 6 parameters model structure from Colleoni et al. (2022) is used (Fig. 2). For a given cell i of coordinates $x \in \Omega$, in the proposed model S6, four reservoirs $\mathcal{I}$, $\mathcal{P}$, $\mathcal{T}_r$ and $\mathcal{T}_l$ of respective capacity c_i, c_p, c_{tr} and c_{tl}, are considered for simulating, respectively, the interception, the production of runoff and its transfer within a cell. Their state vector is denoted $\boldsymbol{h}(x,t) \equiv \left(\boldsymbol{h}_i(x,t), \boldsymbol{h}_p(x,t), \boldsymbol{h}_{tr}(x,t), \boldsymbol{h}_r(x,t), \boldsymbol{h}_{tl}(x,t)\right)^T$, and the parameter vector of SMASH model structure S6 is $\boldsymbol{\theta}(x) \equiv \left(c_i(x), c_p(x), c_{tr}(x), c_r(x), ml(x), c_{tl}(x)\right)^T$. Hence the size of state vector is $N_s \times N_x$ and the size of parameter vector that is optimized in the following is $N_p \times N_x$. Considering tens of cells or more over a simulated catchments domain Ω, the calibration of $\boldsymbol{\theta}$ is a high dimensional inverse problem. All details related to hydrological model operator and model description are explained in Colleoni et al. (2022). The numerical model operates at hourly time step $dt = 1h$ and on a regular grid at $dx = 1km$.

Now, the calibration problem and optimization algorithms are presented, starting by the definition of a multi-criteria objective function enabling to account for multi-scale signatures.

In order to calibrate the hydrological model based on the simulated and observed discharge at gauged cells $x_k \in \Omega$, $k \in 1,..,Ng$, denoted as $Q_k(t)$ and $Q_k^*(t)$, respectively, we define the objective convex function as shown in Eq. 2.

$$J(\boldsymbol{\theta}) = J_{obs}(\boldsymbol{\theta}) + \alpha J_{reg}(\boldsymbol{\theta}) \tag{2}$$

where the observation cost function $J_{obs} = \frac{1}{N_g}\sum_{k=1}^{N_g} j_k^*$ measuring the misfit, via several adapted metrics that can include signatures as detailed later, between simulated and observed discharge. In this study, $N_g = 1$, that is for single gauge optimization. Note that simulated discharge $Q_k(t) = \mathcal{M}\left[\boldsymbol{P}\left(x,t'\right), \boldsymbol{E}\left(x,t'\right), \boldsymbol{h}\left(x,0\right), \boldsymbol{\theta}\left(x\right)\right], \forall x \in \Omega_k, t' \in [0,t]$ with $\Omega_k \subset \Omega$ denoting the spatial domain including all upstream cells of a gauge at x_k, depends on the control

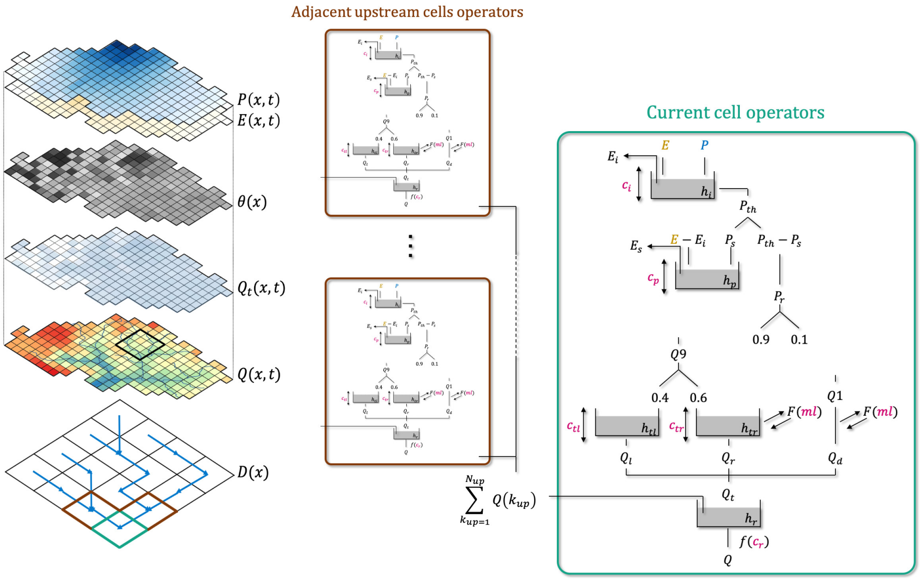

Figure 2: Distributed hydrological modeling with SMASH platform. Model fields from top to bottom: meteorological inputs, parameters, internal and output flux maps (left). Pixel scale and pixel-to pixel flow operators of SMASH model structure S6 studied (right). *Source:* Colleoni et al. (2022).

vector $\boldsymbol{\theta}$ via the hydrological model $\mathcal{M}$ (Eq. 1). The second term in Eq. (2) is weighted by α and set as a classical Thikhonov regularization $J_{reg} = \left\| \boldsymbol{B}^{-1/2} \left(\boldsymbol{\theta} - \boldsymbol{\theta}^* \right) \right\|_{L^2}^2$ with $\boldsymbol{B}$ the background error covariance, and $\boldsymbol{\theta}^*$ the first guess/background on $\boldsymbol{\theta}$. We set $\alpha = 10^{-4}$ for the spatially distributed optimizations presented in this study, $\alpha = 0$ otherwise if $\boldsymbol{\theta} \equiv \overline{\boldsymbol{\theta}}$, and $\boldsymbol{B}$ is simply defined from $\boldsymbol{\sigma}_{\theta}$ the vector of mean deviations of $\boldsymbol{\theta}$, as done in Jay-Allemand et al. (2020). The optimal estimate $\hat{\boldsymbol{\theta}}$ of the model parameter set can be obtained by minimizing the objective function $\boldsymbol{J}$ in Eq. 1, subject to an additional bound constrain on the model parameters, which can be expressed as Eq. 3.

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}_{min} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{max}}{\arg\min} \; \boldsymbol{J}(\boldsymbol{\theta}) \tag{3}$$

This inverse problem 3 is tackled with different global optimization algorithms considering a spatially uniform control, that is low dimensional optimization problems. For instance, optimization algorithms such as: Step-By-Step (SBS) (steepest descent algorithm summarized in Edijatno (1991)), Nelder–Mead and Genetic Algorithms (GA) can be applied in this scenario. Next, a spatially distributed control vector is sought with a VDA algorithm (Jay-Allemand et al., 2020) adapted to such high dimensional hydrological optimization problems. Considering a spatially distributed control vector $\boldsymbol{\theta}(x)$, its optimization is performed with the L-BFGS-B algorithm (limited-memory Broyden–Fletcher–Goldfarb–Shanno bound-constrained (Zhu et al., 1997)) adapted to high dimension. This algorithm requires the gradient of the cost function with respect to the sought parameters $\nabla_{\boldsymbol{\theta}} \boldsymbol{J}$, that is obtained by solving the adjoint model. This numerical adjoint model has been generated with the automatic differentiation engine TAPENADE

(Hascoet and Pascual, 2013) applied to the SMASH source code including the new models structures and validated with standard gradient test. The background value θ^*, used as a starting point for the optimization problem and in the regularization term, is set as in Jay-Allemand et al. (2020), i.e. as $\bar{\theta}$, a spatially uniform global optimum determined with a simple global-minimization algorithm from a uniform first guess $\bar{\theta}^*$. Given mildly non linear hydrological models as those considered in this study, this calibration approach is pertinent and sensitivity to priors is limited as shown in Jay-Allemand et al. (2020).

2.2. Signatures computation

Several signatures describing and quantifying properties of discharge time series are introduced in view to analyze and calibrate hydrological models (an exhaustive list is given in Appendix B). Signatures are denoted as S_i, $i \in 1..N_{crit}$, with N_{crit} being the number of different signature types considered. These signatures allow for the description of various aspects of the rainfall-runoff behavior, such as flow distribution (e.g. based on flow percentiles), flow dynamics (Le Mesnil, 2021), flow separation (Nathan and McMahon, 1990; Lyne and Hollick, 1979), and flow timing, among others. A so-called continuous signature is a signature that can be computed over the entire study period. Flood event signatures on the other hand focus on the behavior of the high-flows that are observed in flash flood events (Fig. 3). In this way, event segmentation algorithm is crucially needed before computing the flood event signatures.

We propose here an automated segmentation algorithm for detecting flood events with the aid of the rainfall gradient, rainfall energy and baseflow separation (Algorithm 1). First, we identify event peak discharges using a peak detection algorithm, which allows for several parameters to be set, such as minimum peak height (mph) or minimum distance between two successive peaks (mpd), among others (Duarte and Watanabe, 2021). For instance, we consider events that exceed the 0.995-quantile of the discharge as important events (mph criterion), and events are considered to be distinct if they are separated by at least 12 hours (mpd criterion). Subsequently, we determine the starting and ending dates for each event. The starting date of the event is considered to be the moment when the rain starts to increase dramatically, which is sometime 72 hours before the peak discharge. To calculate this, we compute the gradient of the rainfall and choose the peaks of rainfall gradient that exceed the 0.8-quantile. These peaks correspond to the moments when there is a sharp increase in rainfall. However, we also require an additional criterion called the "energy criterion", which takes into account the "rainfall energy" for a more robust detection of flood start time. The rainfall energy is computed as the sum of squares of the rainfall observed in a 24-hour period, counted from 1 hour before the peak of rainfall gradient. The starting date is the first moment when the rainfall energy exceeds 0.2 of the maximal rainfall energy observed in the 72-hour period before the peak discharge, based on the gradient criterion. Finally, we aim to find the ending date by using baseflow separation. We compute the difference between the discharge and its baseflow from the peak discharge until the end of study period (which lasts for 10 days from the starting date of the event). The ending date is the moment when the difference between the discharge and its baseflow is minimal in a 48-hour period, counted from 1 hour before this moment. Note that these values are adapted to the basins and flood scales studied.

Algorithm 1 Hydrograph segmentation algorithm

For each catchment, considering 2 time series (T, Q) and (T, P) where:
$T = (t_1, ..., t_n)$ is time (by hour),
$Q = (q_1, ..., q_n)$ is the discharge and
$P = (p_1, ..., p_n)$ is the rainfall.

1. Detecting peaks that exceed the 0.995-quantile of the discharge, that can be considered as important events:
   $E = (t_i)_{1 \leq i \leq n}$ s.t. $q_i > Quant_{0.995}(Q)$
2. For each event $t_j \in E$:
   (a) Determining a starting date based on the "rainfall gradient criterion" and the "rainfall energy criterion":
      i. Selecting rainfalls gradient those exceed its 0.8-quantile, considered as the "rainfall events":
         $RE = (t_k)_{t_k \in (t_j - 72, t_j)}$ s.t. $\nabla P(t_k) > Quant_{0.8}(\nabla P([t_j - 72, t_j]))$
      ii. Defining the rainfall energy function:
         $f(t_x) = ||(p_x - 1, ..., p_x + 23)||_2$
         then the starting date is the first moment the rainfall energy exceeds 0.2 of the maximal rainfall energy:
         $sd = \min(t_s)_{t_s \in RE}$ s.t. $f(t_s) > 0.2||(f(t_j - 72), ..., f(t_j))||_\infty$
   (b) Determining an ending date based on discharge baseflow $Qb = Baseflow(Q)$:
      $ed = \arg\min_{t_e} \sum_{t=t_e-1}^{t_e+47} |(Q - Qb)(t)|$ s.t. $t_j \leq t_e \leq sd + 10 \times 24$

Remark. If there exists $m + 1$ $(m > 0)$ consecutive events $(sd_u, ed_u), ..., (sd_{u+m}, ed_{u+m})$ occurring "nearly simultaneously", that means all of these events occur in no more than 10 days: $ed_{u+m} < sd_u + 10 \times 24$, then we merge these $m + 1$ events into a single event (sd_u, ed_{u+m}).

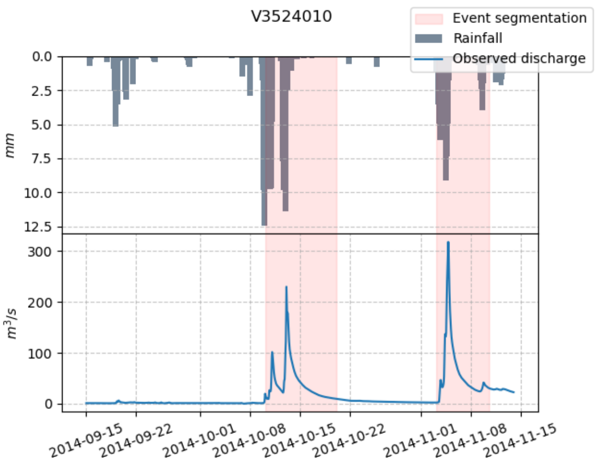

Figure 3: Example of flood events detected from hydrograph using the segmentation algorithm.

2.3. Signatures sensitivity

To perform a calibration process with hydrological signatures, it is important to investigate the sensitivity of simulated signatures with the model parameters, to guide the potential selection of the signatures which should be used to calibrate the model. The sensitivity analysis enables us to examine how the variation of a given output/signature can be apportioned to a variation in model inputs (Saltelli, 2002). If some signatures are not sensitive with the model parameters, then it may not have any significant impact to optimize an objective function based on these signatures. In this context, we consider a hydrological model $\mathcal{M}$ with m spatially uniform parameters $\bar{\theta} \equiv (\theta_1, ..., \theta_m)$. Then the simulated value of a signature S_i, calculated from the simulated discharges via a discharge-to-signature mapping f_i, is represented as $S_i^s \equiv f_i \circ \mathcal{M}\left(\boldsymbol{P}, \boldsymbol{E}, \boldsymbol{h}, \bar{\theta}\right)$. We are interested in Sobol indices called first-order and total-order. The first- (depending on θ_j) and total- (depending on $\theta_{\sim j}$, i.e. all parameters except θ_j) Sobol indices of the simulated signature S_i^s are respectively defined as follows:

$$s_i^{(1j)} = \frac{\mathbb{V}[\mathbb{E}[S_i^s|\theta_j]]}{\mathbb{V}[S_i^s]} = \frac{V_j}{V} \quad \text{and} \quad s_i^{(Tj)} = \frac{\mathbb{E}[\mathbb{V}[S_i^s|\theta_{\sim j}]]}{\mathbb{V}[S_i^s]} = 1 - \frac{\mathbb{V}[\mathbb{E}[S_i^s|\theta_{\sim j}]]}{\mathbb{V}[S_i^s]} = 1 - \frac{V_{\sim j}}{V}$$

where V_j (respectively, $V_{\sim j}$) is the variance of the expectation of output signature S_i^s conditioned by the input parameter θ_j (respectively, $\theta_{\sim j}$, i.e. all sampled inputs except θ_j). To estimate these indices, Azzini et al. (2021) proposed a method based on the Saltelli generator (Saltelli, 2002), which is implemented in the *SALib* Python library (Iwanaga et al., 2022; Herman and Usher, 2017). This method, that is shown to be relatively accurate in a recent benchmark (Puy et al., 2022), allows us to estimate the first-, second- and total-order variance-based sensitivity indices using Monte Carlo simulations. However, in our specific application with high-dimensional parameter spaces, we have encountered significant challenges in estimating the second-order variance-based sensitivity indices due to their computationally intensive nature (Saltelli, 2002; Campolongo et al., 2011). To achieve accurate results, a large number of Monte Carlo simulations are required, which can be time-consuming and computationally demanding. Therefore, for the purpose of this study, we focus on estimating the first- and total-order Sobol indices, which provide a sufficiently efficient means of capturing information about interaction effects while retaining an acceptable computational cost.

2.4. Multi-criteria calibration using hydrological signatures

This section defines the calibration objective functions and how they account for the multi-scale signatures that are provided by the segmentation algorithm detailed previously.

First, we define cost function parts corresponding respectively to classical metrics, continuous signatures and event based signatures. Let us consider a classical objective function j_d, which is the dominant criterion (or the most constrained criterion) in case of multi-criteria optimization, an objective function j_c combining continuous-signatures-based cost functions, and j_f combining flood-event-signatures-based cost functions. Then, the cost function to be minimized, denoted J, can be defined as Eq. 4.

$$J \equiv \begin{cases} \delta_d j_d + \delta_c j_c + \delta_f j_f \text{ for single-objective optimization,} \\ (j_d, j_c, j_f) \text{ for multi-objective optimization} \end{cases} \tag{4}$$

where $\delta_d, \delta_c, \delta_f$ are the corresponding optimization weights in the first case. Keep in mind that we take into account the use of signatures in both cases but the first case is a single-objective optimization while the second is a multi-objective optimization.

Then we detail how each cost function part is computed from signatures. For each signature S_i, denote by S_i^o and S_i^s the observation and simulation respectively. The set of continuous and flood event signatures denoted N_c and N_f respectively. Then, the components j_d, j_c and j_f can be defined as follows:

- $j_d \equiv 1 - NSE$ or $1 - KGE_{(\alpha,\beta,\gamma)}$ with varying weights α, β, γ (see Appendix A.2). This metric j_d is considered as a constraining objective function for selecting an optimal solution from non-inferior solutions in case of multi-objective optimization (see Appendix C.3).
- $j_c \equiv \begin{cases} \sum_{S_i \in N_c} \sigma_{S_i} j_c^{S_i}, \text{ for single-objective multi-criteria optimization;} \\ (\{j_c^{S_i}\}_{S_i \in N_c}), \text{ for multi-objective optimization} \end{cases}$

  where $j_c^{S_i} \equiv \left| \frac{S_i^s}{S_i^o} - 1 \right|$ is the objective function based on continuous signature S_i and σ_{S_i} is the corresponding optimization weight of S_i in case of single-objective function.
- $j_f \equiv \begin{cases} \sum_{S_i \in N_f} \sigma_{S_i} j_f^{S_i}, \text{ for single-objective multi-criteria optimization;} \\ (\{j_f^{S_i}\}_{S_i \in N_f}), \text{ for multi-objective optimization.} \end{cases}$

  In this case, and in the context of global optimization in time, $j_f^{S_i} \equiv \frac{1}{N_E} \sum_{e=1}^{N_E} \left| \frac{S_{i,e}^s}{S_{i,e}^o} - 1 \right|$ defines the scalar objective function related to flood signature $S_i \in N_f$ over the N_E events selected with the segmentation method described in Algorithm 1. Otherwise, to perform a season-based optimization on flood event signatures, we can compute for the events occurring in the selected season. For example, for a Spring-based optimization:

$$j_{f,spring}^{S_i} \equiv \frac{1}{\dim \mathcal{SP}} \sum_{e \in \mathcal{SP}} \left| \frac{S_{i,e}^s}{S_{i,e}^o} - 1 \right| \text{ s.t. } \forall e \in \mathcal{SP} \subset \{1, ..., N_E\}, S_{i,e} \text{ occurs in Spring.}$$

Finally, these cost functions enable to formulate, after the single objective calibration problem 3, the following multi-objectives calibration problems. The optimization problems taking into account signatures via the cost function defined in Eq. 4 can be developed as Eq. 5 for a single-objective optimization, and as Eq. 6 for a multi-objective optimization.

$$\min_{\theta \in \mathcal{O} \subset \mathbb{R}^n} \delta_d j_d(\theta) + \delta_c \sum_{S_i \in N_c} \sigma_{S_i} \left| \frac{S_i^s(\theta)}{S_i^o(\theta)} - 1 \right| + \delta_f \sum_{S_i \in N_f} \sigma_{S_i} \frac{1}{N_E} \sum_{e=1}^{N_E} \left| \frac{S_{i,e}^s(\theta)}{S_{i,e}^o(\theta)} - 1 \right| \quad (5)$$

$$\min_{\theta\in\mathcal{O}\subset\mathbb{R}^n}\left(j_d(\theta),\left\{\left|\frac{S_i^s(\theta)}{S_i^o(\theta)}-1\right|\right\}_{S_i\in N_c},\left\{\frac{1}{N_E}\sum_{e=1}^{N_E}\left|\frac{S_{i,e}^s(\theta)}{S_{i,e}^o(\theta)}-1\right|\right\}_{S_i\in N_f}\right) \tag{6}$$

While the minimization problem with single-objective function 5 is accessible for both global and distributed calibration methods, performing a multi-objective optimization as problem 6 is sophisticated for distributed calibration considering a spatially distributed control vector adapted to a high dimensional hydrological optimization problems, and requiring a lot of cost gradient information. In global calibration with multi-objective optimization approaches, a set of feasible solutions can be found instead of a unique optimal solution in single-objective optimization (Appendix C). In such a way, a so-called Pareto front contains non-inferior solutions (Appendix C.1) and thus a method is proposed for selecting an optimal solution from the Pareto as depicted in Appendix C.3.

Note that the objective functions j_c and j_f related to continuous and flood signatures have also been implemented in Fortran. This implementation imposes strict positivity of their components ($j_c^{S_i}$ and $j_f^{S_i}$) numerically to ensure that the total cost J remains convex and differentiable. The numerical adjoint model has been also re-derived as needed by the variational calibration algorithm (refer to section 2.1). The cost function based flood event signatures j_f can be computed thanks to a temporal mask of corresponding flood events selected by the segmentation algorithm, implemented in the Python, and passed to the Fortran via the wrapped interface.

3. Data and numerical results analysis

This section first presents the catchment-flood dataset used in this study. Next, flow signatures are analyzed via the comparison of observed and simulated signatures, in terms sensitivity with parameters, and finally some are selected for signature-based model calibration. The last part analyzes the performances of model calibration with classical and signature-based metrics.

3.1. Catchment information and data sources

A relatively large dataset of catchment-floods mostly located in the French mediterranean region is used. This dataset stems from Jay-Allemand (2020) and contains time series of hydro-meteorological variables and time invariant catchment attributes for four high rainfall-flow areas in France, identified as study areas of the PICS research project[1]. It encompasses 141 catchments including 23 outlet gauges, which are mostly located in the French Mediterranean region (Fig. 4). This is a subset of a larger dataset of 4,190 French catchments from INRAE-HYCAR research unit (Brigode et al., 2020; Delaigue et al., 2020). The hydrological model inputs consist of observation data, covering a period of about 13 years (2006 to 2019), that includes hourly distributed discharge and rainfall. Discharge data are collected by

[1] https://pics.ifsttar.fr

the French Ministry of Environment covering the period of the forcing data and have been extracted from the (Hydro) platform[2]. The rainfall grids are the radar observation reanalysis ANTILOPE J+1 provided by Météo-France at a grid resolution of 1 km^2 (Champeaux et al., 2009). The potential evapotranspiration (PET) is obtained by applying a simple formula (Oudin et al., 2005) to SAFRAN[3] (Quintana-Seguí et al., 2008) temperature grids at 8 km resolution an empirically desaggregated at hourly time step and 1 km spatial resolution, i.e. at the same spatio-temporal resolution than rainfall. Note that observation data, rainfall grids and discharge time series, over the selected catchments have few missing data as detailed in Table 1, so that it can be neglected when performing the computations and analysis in this study. Table 1 contains catchment information such as the river name, surface, code, number of upstream gauges, and missing rates in the outlet gauges. Raster maps, at 1 km resolution, of upstream drained area and D8 flow directions have been obtained by processing fine DEM provided by IGN (Institut Geographique National).

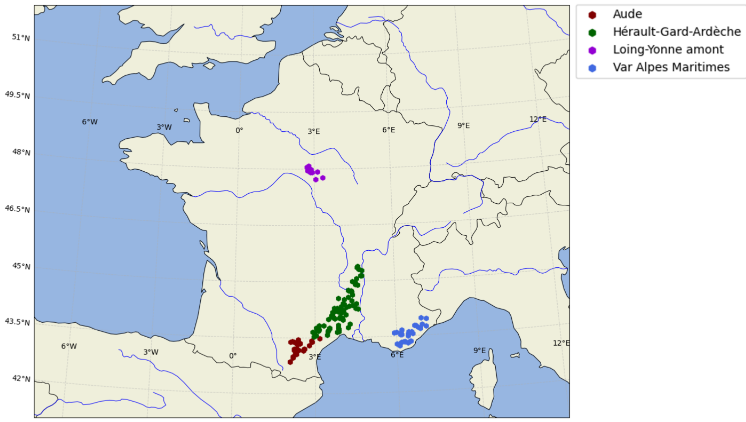

Figure 4: Spatial distribution of 141 catchments of the PICS dataset, consisting of 23 outlet gauges and 118 upstream gauges on the map of France with four regions denoted by different colors.

3.2. Sensitivity analysis and selection of signatures for model calibration

To start with, the relative error is analyzed between observed signatures and simulated ones with a model calibrated using SBS algorithm and spatially uniform parameters. Table 2 shows that some hydrological signatures with a significant simulation error such as: $Cfp2$, $Cfp10$, $Cfp50$, Elt and Epf that could be better constrained with a signature-based calibration process as investigated in next subsection (a list of all studied signatures with corresponding notations is presented in Appendix B).

Next, we survey the global sensitivity of these signatures with the model parameters. We considered over 10,000 spatially uniform sets of the 6 model parameters, sampled using Saltelli generator (Saltelli, 2002) to estimate the total-order Sobol indices across 23 gauged catchments (catchments downstream outlets of the dataset). Based on the

[2] http://www.hydro.eaufrance.fr/
[3] "Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige" in French

Table 1
General information about 23 outlet gauges of the PICS data. Code, river name, surface, missing rate of rainfall (respectively, discharge) in outlet gauge during the period 2006-2019, and number of upstream are represented by the columns from left to right.

Code	River name	Surface (km^2)	Missing rates (%)	Total upstream gauges
H3201010	Le Loing	2302	0.14 (3.68)	8
V3524010	La Cance	381	0.14 (4.31)	3
V3744010	Le Doux	621	0.14 (4.02)	2
V4154010	L'Eyrieux	649	0.14 (7.38)	3
V5064010	L'Ardèche	2264	0.14 (4.22)	9
V5474015	La Cèze	1112	0.14 (3.76)	6
V7164015	Le Gardon	1093	0.14 (16.62)	10
Y1232010	L'Aude	1828	0.14 (3.74)	11
Y1364010	Le Fresquel	935	0.14 (3.74)	4
Y1415020	L'Orbiel	242	0.14 (3.74)	2
Y1564010	L'Orbieu	589	0.14 (3.77)	3
Y1605050	La Cesse	251	0.14 (4.64)	1
Y2332015	L'Hérault	2208	0.14 (7.22)	12
Y2584010	L'Orb	1336	0.14 (4.04)	11
Y3204040	Le Lez	168	0.14 (15.55)	3
Y3444020	Le Vidourle	503	0.14 (7.97)	4
Y3534010	Le Vistre	496	0.14 (4.42)	1
Y4624010	Le Gapeau	535	0.14 (3.79)	6
Y5312010	L'Argens	2512	0.14 (5.08)	10
Y5444010	La Giscle	201	0.14 (9.96)	2
Y5534030	La Siagne	492	0.14 (5.30)	5
Y5615030	Le Loup	289	0.14 (3.79)	1
Y6434010	L'Estéron	442	0.14 (7.70)	1

results presented in Table 3, it can be observed that the non conservative water exchange parameter ml and the transfer parameter c_{tr} exhibit the highest sensitivities to the studied signatures, both in terms of first-order and total-order. Our analysis suggests that these two parameters have the most significant impact on the output signatures as a result of their interactions with other inputs. This is in coherence with highest sensitivities found for soil depth and subsurface flow parameters of an event flash flood model found in Garambois et al. (2013, 2015) on some catchments of the present set. Conversely, we found that parameters such as the interception c_i and the production of runoff c_p have little to no impact on the simulated signatures. We also observed that continuous signatures exhibit lower sensitivities than flood-event signatures in both first-order and total-order effects. Furthermore, constraining hydrological model by flood event signatures along with a classical calibration metric (e.g. $1 - NSE$ or $1 - KGE$), which is based primarily on continuous records of streamflow, is ideal to balance the model between the global score and the performance on flood events. We select for example the peak flow, denoted as Epf, which is one of flood event signatures having both significant relative error and high sensitivity, to perform multi-critera calibration methods. Note that multi-criteria optimization methods with multiple signatures are absolutely reachable but will not be shown in this study for sake of brevity and simplify results analysis.

Table 2
Relative error between simulated and observed signatures of the same model structure calibrated either with $1 - NSE$ or $1 - KGE$ by SBS algorithm for global optimization. The values (in the form of . [., .]) in each case represent respectively the median, mean and standard deviation of a signature over gauged catchments.

Notation	Signature type	Relative error on simulated signature	
		Cal. with j_d^{NSE}	Cal. with j_d^{KGE}
Crc	Continuous runoff coefficients	0.14 [0.28, 0.38]	0.16 [0.3, 0.46]
Crchf		0.24 [0.35, 0.35]	0.26 [0.4, 0.45]
Crclf		0.15 [0.3, 0.44]	0.15 [0.33, 0.54]
Crch2r		0.23 [0.4, 0.68]	0.22 [0.38, 0.69]
Cfp2	Flow percentiles	0.72 [3.99, 21.14]	0.76 [5.99, 29.98]
Cfp10		0.52 [2.64, 8.8]	0.52 [2.87, 9.42]
Cfp50		0.29 [0.49, 0.85]	0.2 [0.52, 0.99]
Cfp90		0.21 [0.37, 0.96]	0.18 [0.38, 0.99]
Eff	Flood flow	0.23 [0.32, 0.31]	0.19 [0.31, 0.37]
Ebf	Base flow	0.22 [0.33, 0.39]	0.22 [0.33, 0.41]
Erc	Flood event runoff coefficients	0.2 [0.28, 0.26]	0.18 [0.27, 0.26]
Erchf		0.23 [0.32, 0.31]	0.19 [0.31, 0.37]
Erclf		0.22 [0.33, 0.39]	0.22 [0.33, 0.41]
Erch2r		0.12 [0.19, 0.2]	0.13 [0.2, 0.24]
Elt	Lag time	0.48 [0.96, 1.25]	0.46 [0.82, 1.1]
Epf	Peak flow	0.28 [0.38, 0.35]	0.25 [0.36, 0.41]

Note also that global sensitivity analysis can be performed with local derivatives based approaches. A link between global Sobol indices and local derivatives has been proposed by Sobol and Kucherenko (2010) (refer also to Lamboni et al. (2013)). Global sensitivity matrices in three dimensions (sample size, parameters number, time) and sensitivity statistics, based on local derivatives computed by finite differences have been proposed in Gupta and Razavi (2018); Razavi and Gupta (2019) for geophysical models and applied to HBV-SASK lumped hydrologic model. Note that the variational data assimilation algorithm upgraded in the present work uses accurate local (in parameter space) cost function gradients, global in time and spatially distributed, computed with the adjoint method. Such method enables to compute accurate spatial sensitivity maps even for high dimensional parameter spaces (e.g. Monnier et al. (2016)) and deepening sensitivity analysis with our differentiable and spatially distributed hydrological model, along with accounting for sensitivity indices into the VDA algorithm, is a very interesting direction intentionally left for further research.

3.3. Performance comparison of classical and signature-based calibration metrics

In this section, we compare the performance of different calibrated models obtained using spatially uniform and distributed optimization methods with different calibration metrics including signature based ones. For spatially uniform calibration methods, we aim to compare different calibration metrics including classical single-objective optimization (CSOO), signature-based single-objective optimization (SSOO) and signature-based multi-objective optimization (SMOO). For spatially distributed calibration methods, two strategies selected for comparison are CSOO and SSOO. In both spatially uniform or distributed calibration scenarios, the models are calibrated on 23 outlet gauges

Table 3
Median across gauged catchments of first- (respectively, total-) order variance-based sensitivity indices of the studied signatures to the model parameters.

Signature	Model parameter					
	c_i	c_p	c_{tr}	c_{tl}	c_r	ml
Crc	-0.0 (0.0001)	-0.0004 (0.0004)	0.1336 (1.2998)	0.0006 (0.0002)	-0.0 (0.0)	0.1167 (1.3778)
Crchf	0.0038 (0.0103)	0.0268 (0.1155)	0.3739 (0.8506)	0.0153 (0.0123)	0.0367 (0.1513)	0.2245 (0.7919)
Crclf	-0.0 (0.0001)	-0.0004 (0.0003)	0.1299 (1.3018)	0.0006 (0.0001)	-0.0 (0.0)	0.1142 (1.3844)
Crch2r	-0.0004 (0.0017)	0.0193 (0.0255)	0.1014 (0.1426)	0.1099 (0.2055)	0.1833 (0.2449)	0.3984 (0.5481)
Cfp2	0.0014 (0.056)	0.0002 (0.0024)	0.1283 (1.6008)	0.0 (0.0)	0.0 (0.001)	-0.0026 (1.2871)
Cfp10	-0.0 (0.0001)	-0.0002 (0.0001)	0.128 (1.3353)	0.0002 (0.0)	0.0 (0.0)	0.092 (1.3922)
Cfp50	-0.0001 (0.0001)	-0.0001 (0.0001)	0.1267 (1.315)	0.0005 (0.0001)	0.0 (0.0)	0.1043 (1.3933)
Cfp90	-0.0002 (0.0001)	-0.0006 (0.0015)	0.1329 (1.2483)	0.001 (0.0006)	-0.0001 (0.0002)	0.1512 (1.3817)
Eff	0.0002 (0.0059)	0.0699 (0.1939)	0.306 (0.7807)	0.0242 (0.022)	0.0321 (0.1389)	0.1872 (0.7303)
Ebf	-0.0001 (0.001)	0.0014 (0.0159)	0.144 (1.1914)	0.0018 (0.0019)	-0.0002 (0.0056)	0.162 (1.3146)
Erc	-0.0001 (0.0015)	0.0076 (0.0314)	0.18 (1.1633)	0.0028 (0.0031)	-0.0001 (0.0011)	0.1705 (1.2433)
Erchf	0.0002 (0.0059)	0.0699 (0.1939)	0.306 (0.7807)	0.0242 (0.022)	0.0321 (0.1389)	0.1872 (0.7303)
Erclf	-0.0001 (0.001)	0.0014 (0.0159)	0.144 (1.1914)	0.0018 (0.0019)	-0.0002 (0.0056)	0.162 (1.3146)
Erch2r	0.0057 (0.0099)	0.0123 (0.0426)	0.0873 (0.2124)	0.0256 (0.0552)	0.4387 (0.5797)	0.1171 (0.2255)
Elt	-0.0002 (0.0116)	-0.0004 (0.0293)	0.0043 (0.087)	0.0009 (0.0048)	0.8832 (0.953)	0.0127 (0.0568)
Epf	-0.0008 (0.0026)	0.0357 (0.1235)	0.2505 (0.9199)	0.0081 (0.0074)	0.1099 (0.2632)	0.1257 (0.8049)

of the PICS data on the calibration period 2006-2013. The validation of calibrated models performances is done in space and time following the three setups:

- on 23 outlet gauges on the validation period 2013-2019 (temporal validation, T_Val),
- on 118 upstream gauges on the calibration period 2006-2013 (spatial validation, S_Val),
- on 118 upstream gauges on the validation period 2013-2019 (spatio-temporal validation, S-T_Val).

3.3.1. Spatially uniform calibrations with NSGA

We first perform global calibrations using NSGA for two single-objective-function-based approaches and one multi-objective-function-based approach. Table 4 displays the mean of different objective functions for calibration and validation (with 3 validation metrics), and for 3 optimization criteria (CSOO, SSOO and SMOO) with various cost functions. In CSOO, we interpret that the model calibrated with $j_d^{KGE} = 1 - KGE$ produces a better result on the peak flow j_f^{Epf}, compared to the one calibrated with $j_d^{NSE} = 1 - NSE$. This explains why KGE criterion is more robust than NSE for constraining a hydrological model, since it is built on the decomposition of NSE (Gupta et al., 2009), which emphasizes relative importance of several hydrological features. This finding is consistent with that of Mizukami et al. (2019). The authors calibrated daily models over numerous US catchments with multiple metrics, including NSE, weighted KGEs, annual peak flow bias (APFB), and they found that KGE resulted in better estimates of annual peak flows than NSE. Additionally, the best reproduction of annual peak flows was achieved with APFB, but this was at the expense of other high flow metrics.

Table 4

Calibration, temporal, spatial and spatio-temporal validation metrics with spatially uniform calibrations with three strategies (CSOO, SSOO, SMOO) and global algorithms (SBS or NSGA), (optimal fit for $cost = 0$). The mean of calibration and validation cost values for different objective functions are displayed for each calibration metric - mean is computed over the catchment set: over the 23 outlet gauges for Cal and T_Val, over the remaining 118 upstream gauges for S_Val and S-T_Val.

Method	Calibration metric	$\overline{j_d^{NSE}}$				$\overline{j_d^{KGE}}$				$\overline{j_f^{Epf}}$			
		Cal	T_Val	S_Val	S-T_Val	Cal	T_Val	S_Val	S-T_Val	Cal	T_Val	S_Val	S-T_Val
CSOO	j_d^{NSE}	0.274	0.277	0.901	0.616	0.239	0.369	0.687	0.736	0.279	0.324	0.387	0.357
	j_d^{KGE}	0.352	0.330	1.048	0.795	0.183	0.323	0.665	0.721	0.267	0.280	0.379	0.344
SSOO	$j_d^{NSE}/2 + j_f^{Epf}/2$	0.447	0.418	1.056	0.889	0.377	0.476	0.759	0.853	0.014	0.189	0.346	0.372
	$j_d^{KGE}/2 + j_f^{Epf}/2$	0.551	0.431	1.259	0.956	0.335	0.443	0.777	0.833	0.017	0.209	0.337	0.358
SMOO	$\{j_d^{NSE}, j_f^{Epf}\}$	0.341	0.351	1.020	0.845	0.271	0.420	0.703	0.803	0.087	0.215	0.336	0.391
	$\{j_d^{KGE}, j_f^{Epf}\}$	0.456	0.409	1.163	0.821	0.243	0.368	0.683	0.724	0.048	0.182	0.316	0.389

Using event signatures in addition to classical continuous metrics in SSOO, we found that simulated peak flow is highly improved in terms of relative error (about 15-18 times and 1.4-1.7 times on average, respectively, for calibration and temporal validation) while classical calibrated metrics are significantly deteriorated (about 1.4-1.6 times and 1.4-1.5 times on average, respectively, for calibration and temporal validation). This may arise from imbalances between global score and performance in simulating flood event signature. To address this issue, careful consideration of the optimization weights assigned to objective functions is necessary in order to achieve a balance between model performance on short and long-term series. It should be noted that this approach can be time-consuming, as it requires numerous simulations to determine the appropriate optimization weights for the objective functions, typically using a L-curve approach. Alternatively, the use of global calibration algorithms, which do not require gradient information and can be solved using lower-dimensional optimization problems, can also address these imbalances through the application of a multi-objective optimization approach. This approach offers the advantage of keeping acceptable levels of deterioration of NSE and KGE while significantly improving the simulation of peak flow as shown by multi-objective SMOO results in Fig. 5 and 6.

However, this global multi-objective optimization algorithm is not capable to deal with high dimensional control vectors and the spatially uniform parameter setup here (under-parameterization) led to unsatisfactory results in spatial and spatio-temporal validation metrics. Therefore, a distributed calibration approach, such as using our VDA algorithm accounting for signatures, could improve the model performances. This approach maintains the same optimization weights as described above, and its performance will be evaluated in the subsequent section.

As shown in Fig. 7, the corresponding optimal parameters obtained using various optimization strategies are presented. Based on our preliminary analysis, it is evident that the distribution over studied catchments of c_r has an important difference when performing traditional calibration (CSOO) and multi-criteria calibration methods (SSOO and SMOO). We recall that c_r is the routing parameter in our conceptual design (Fig. 2), so it has a crucial role in

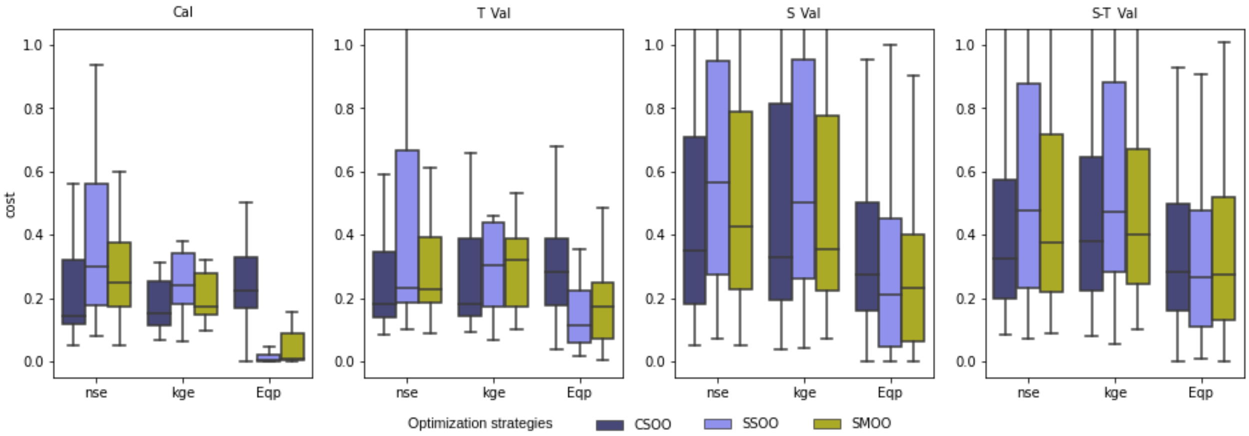

Figure 5: Comparison, with spatially uniform parameters, of calibration and validation metrics (optimal fit for $cost = 0$) for three optimization approaches (CSOO, SSOO, SMOO) by constraining $1 - NSE$ in case of global algorithms (SBS or NSGA). From left to right: calibration (Cal), temporal validation (T_Val), spatial validation (S_Val) and spatio-temporal validation (S-T_Val).

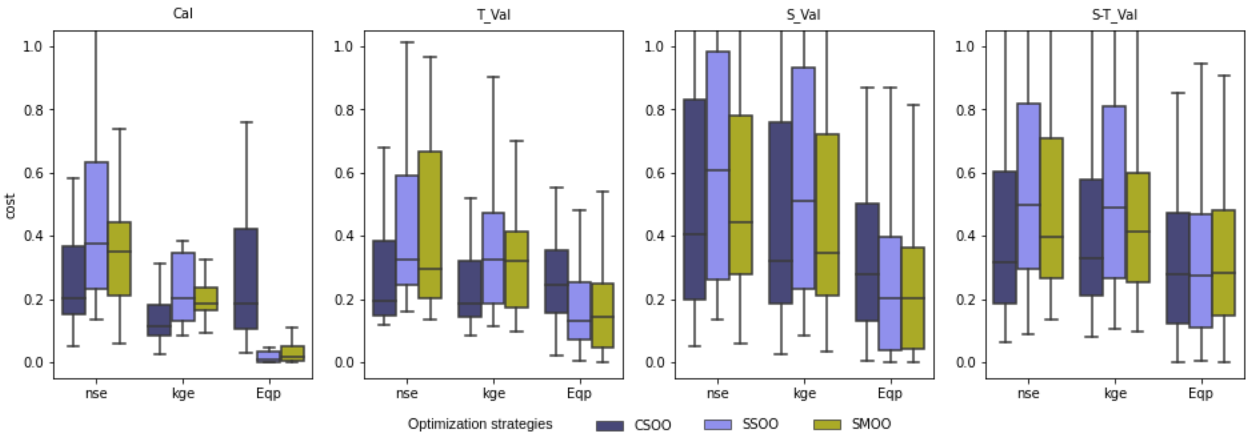

Figure 6: Comparison, with spatially uniform parameters, of calibration and validation metrics (optimal fit $cost = 0$) for three optimization approaches (CSOO, SSOO, SMOO) by constraining $1 - KGE$ in case of global algorithms (SBS or NSGA). From left to right: calibration (Cal), temporal validation (T_Val), spatial validation (S_Val) and spatio-temporal validation (S-T_Val).

producing the peak flow Epf. Additionally, the sensitivity analysis in Table 3 has indicated that c_r is one of the three parameters explaining most of the sensitivity of the peak flow.

The above result on the importance of lateral flow components in a flood hydrological model is in coherence with existing works, for example as shown in Garambois et al. (2013) on few catchments-flood events used in the present study, in addition to high sensitivity to subsurface flow parameter (see also Douinot et al. (2018)) the temporal sensitivity of kinematic wave compound friction parameters in a distributed flash flood model increases with flood magnitude. Improving hydraulic meaningfulness of hydrological models is an important topic since it can improve floods discharge modeling in high resolution catchment-flood models (e.g. Bout and Jetten (2018); Li et al. (2021);

Kirstetter et al. (2021) with shallow water models and simplifications) but also improve internal state-flux coherence and realism as required for instance to assimilate remote sensing observables of river suface such as height and width (e.g. Paiva et al. (2011); Pujol et al. (2020, 2022)).

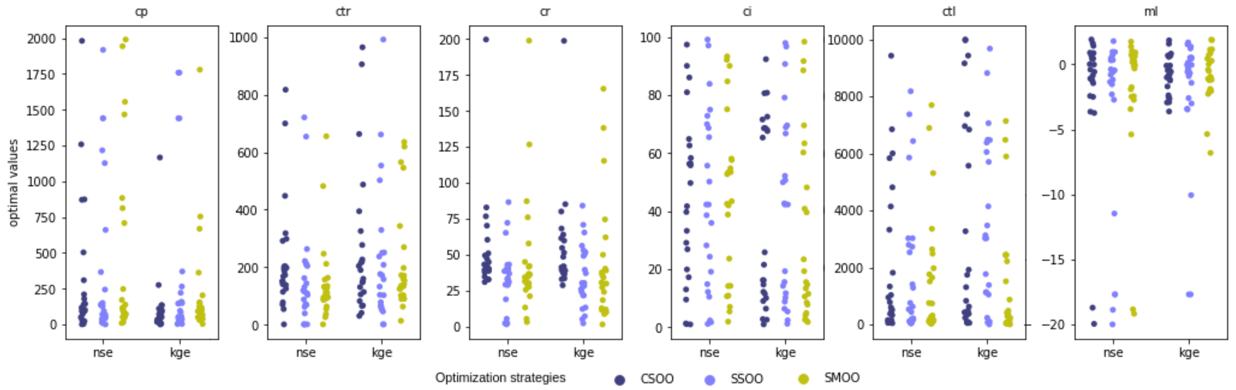

Figure 7: Analysis of spatially uniform calibrated parameters over the whole catchment sample. In each scatterplot, the first column present the distribution of a parameter for 3 optimization strategies (CSOO, SSOO, SMOO) using j_d^{NSE}, whereas the strategies in the second column use j_d^{KGE} as the dominant (or constrained) objective function. The boundary conditions of the model parameters are given in Appendix D.

3.3.2. Spatially distributed calibrations with VDA algorithm

Now, spatially distributed calibrations with the VDA algorithm using multi-criteria cost function including signatures are performed. We employ SSOO technique for a distributed calibration using L-BFGS-B algorithm provided a first guess by SBS algorithm. In overall, all of obtained scores in Table 5 are significantly enhanced compared to the uniform calibration method, thanks to spatially distributed control vectors granting more flexibility to reproduce observed discharge. Instead of a sharp decline of j_f^{Epf} as above, this relative error slightly decreases about 1.5 times (from about 0.25 down to 0.16) in calibration and from about 0.32 down to 0.28 in temporal validation, but instead, the scores (NSE and KGE) are slightly reduced in calibration and have an inappreciable deterioration in temporal validation. So in this case, we do not have imbalances between the model performances on short and long-term series when employing SSOO. We observe clearly in Fig. 8 and 9 that the error of simulated pick flow is significantly reduced while the deterioration level of the scores remains tolerable, particularly in calibration and temporal validation.

Ultimately, the scoring metrics are computed on 111 flood events picked from 23 outlet gauges (by segmentation method depicted in Algorithm 1) on the calibration period. The results plotted in Fig. 10 show that, in distributed calibration, the score of constrained calibration metric is not decreased but even improved from 0.80 (respectively, 0.71) up to 0.83 (respectively, 0.78) in median for NSE (respectively, KGE). It indicates that the optimum of the model parameters has moved to another location that produces a better performance in simulating flood events by slightly reducing the scores in simulating the low-flow.

Table 5
Calibration, temporal, spatial and spatio-temporal validation metrics with spatially distributed control vectors. The mean of calibration and validation cost values for different objective functions are displayed for each calibration metric.

Method	Calibration metric	$\overline{j_d^{NSE}}$				$\overline{j_d^{KGE}}$				$\overline{j_f^{Epf}}$			
		Cal	T_Val	S_Val	S-T_Val	Cal	T_Val	S_Val	S-T_Val	Cal	T_Val	S_Val	S-T_Val
CSOO	j_d^{NSE}	0.221	0.244	0.655	0.596	0.233	0.355	0.553	0.597	0.274	0.334	0.381	0.376
	j_d^{KGE}	0.239	0.231	0.802	0.702	0.140	0.292	0.617	0.701	0.226	0.295	0.365	0.364
SSOO	$j_d^{NSE}/2 + j_f^{Epf}/2$	0.251	0.241	0.831	0.639	0.231	0.305	0.586	0.612	0.183	0.298	0.392	0.383
	$j_d^{KGE}/2 + j_f^{Epf}/2$	0.297	0.245	0.964	0.671	0.190	0.300	0.617	0.647	0.152	0.271	0.376	0.387

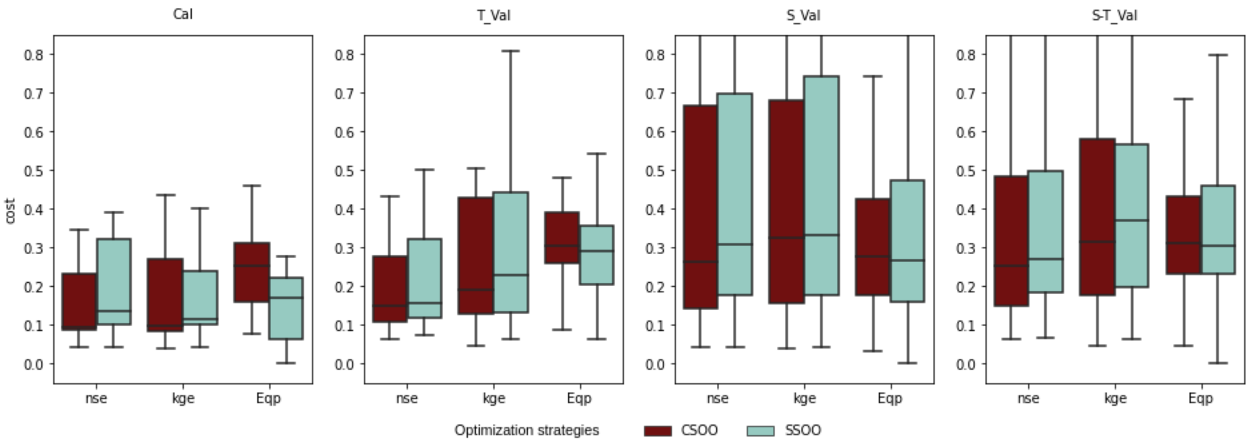

Figure 8: Comparison, with spatially distributed parameters, of calibration and validation metrics (optimal fit $cost = 0$) for two optimization approaches (CSOO, SSOO) by constraining $1 - NSE$ in case of distributed calibration. From left to right: calibration (Cal), temporal validation (T_Val), spatial validation (S_Val) and spatio-temporal validation (S-T_Val).

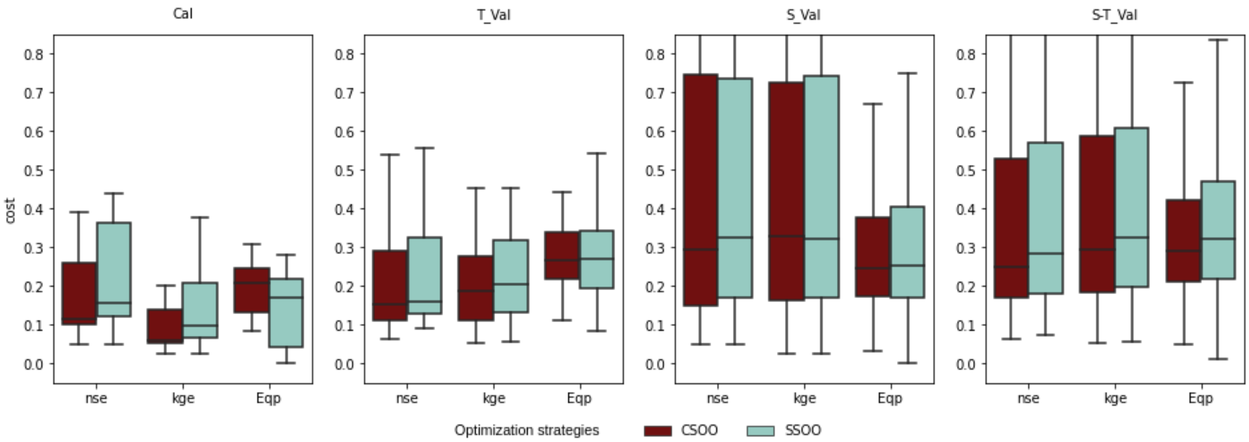

Figure 9: Comparison, with spatially distributed parameters, of calibration and validation metrics for two optimization approaches (CSOO, SSOO) by constraining $1 - KGE$ in case of distributed calibration. From left to right: calibration (Cal), temporal validation (T_Val), spatial validation (S_Val) and spatio-temporal validation (S-T_Val).

Regarding to the parameter space, Table 6 presents statistical analysis of the spatially uniform parameter sets obtained using 4 calibration metrics for the studied catchments. Comparing to the spatially distributed optimal

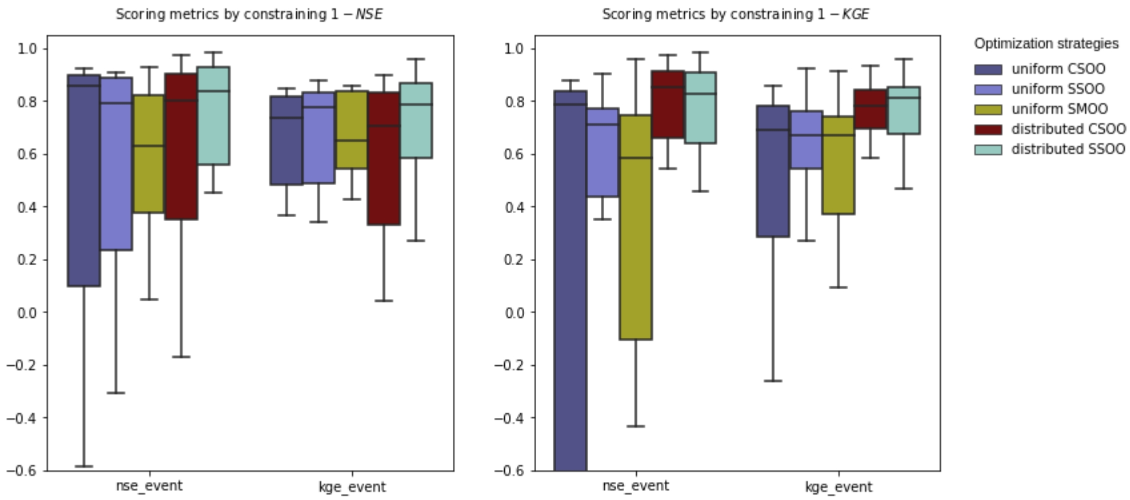

Figure 10: Comparison of scoring metrics computed on 111 events picked from 23 outlet gauges on the calibration period 2006-2013 for the five optimization strategies, by constraining $1 - NSE$ (left) and $1 - KGE$ (right).

parameters in Table 7, we interpret that the mean of distributed parameters over all catchments in the 4 cases (corresponding to 4 calibration metrics) is globally coherent to the distribution of the first guess. Several parameters are almost spatially uniform (e.g. the non conservative water exchange parameter *ml* has a small distributed deviation in median (respectively, in average) over all catchments 0.01 (respectively, 0.05) (calibrated with j_d^{NSE}) compared to its distributed mean in median (respectively, in average) −0.59 (respectively, −4.98)). Conversely, the transfer parameter c_{tl} has a great distributed deviation (in median over all catchments) 193.72 compared to its distributed average 347.87, that also has a massive difference to its distributed median 114.79. Fig. 11 illustrates the spatially distributed optimal parameters at the largest catchment (the Argens River), for a distributed calibration with $j_d^{KGE}/2 + j_f^{Epf}/2$.

Reducing the over-parameterization in distributed hydrological models calibration problems through spatial constrains while enhancing regional parameters consistency remains a key issue, especially for flash flood prediction at ungauged locations (e.g. classical post-regionalization in Garambois et al. (2015) on French Mediterranean flash floods). This issue can be tackled with calibration approaches accounting for physiographic descriptors through regularizations (e.g. De Lavenne et al. (2019); Jay-Allemand et al. (2022b) in multi-gauges calibration problems) or through pre-regionalization mappings, such as the multi-scale parameter regionalization approach (MPR) from Samaniego et al. (2010), used for example in Mizukami et al. (2017). In addition to exploiting the information of multi-scale signatures in calibration with the present VDA algorithm, the use of a pre-regionalization scheme, i.e. "strong constrains" in the forward model in form of a mapping between physiographic covariables and conceptual hydrological parameter fields, represent an interesting perspective for future research.

Table 6
Uniform optimal parameters calibrated by SBS algorithm with 4 calibration metrics for each catchment on its outlet gauge. The values (in the form of . [., .]) in each case represent respectively the median, mean and standard deviation of the optimal parameter values over all catchments of the dataset.

Parameter	Calibration metric			
	j_d^{NSE}	$j_d^{NSE}/2 + j_f^{Epf}/2$	j_d^{KGE}	$j_d^{KGE}/2 + j_f^{Epf}/2$
c_i	14.71 [20.3, 26.22]	16.93 [20.83, 26.48]	17.6 [27.15, 33.07]	17.27 [30.17, 35.83]
c_p	169.99 [291.17, 434.58]	146.04 [310.14, 505.68]	151.87 [286.79, 483.33]	141.56 [289.69, 466.99]
c_{tr}	171.76 [286.6, 269.5]	162.66 [313.49, 304.83]	266.32 [431.2, 355.04]	267.21 [436.83, 360.62]
c_{tl}	347.87 [812.15, 1274.12]	250.42 [1366.73, 2789.22]	383.51 [1413.93, 2749.35]	262.89 [1337.96, 2795.16]
c_r	41.32 [52.63, 34.05]	40.94 [50.97, 30.97]	41.33 [51.2, 30.29]	40.24 [50.2, 27.58]
ml	-0.59 [-4.98, 8.21]	0.0 [-3.81, 7.34]	-0.0 [-3.62, 7.31]	-0.0 [-3.28, 6.39]

Table 7
Analysis of spatially distributed parameter sets of the models corresponding to 4 calibration metrics. First, spatial median $(\tilde{\cdot})$, average $(\bar{\cdot})$ and standard deviation $(\sigma_{\cdot})$ for each parameter field are calculated for each catchment, then their median, mean and standard deviation over all catchments are represented in the form of . [., .].

Parameter	Calibration metric			
	j_d^{NSE}	$j_d^{NSE}/2 + j_f^{Epf}/2$	j_d^{KGE}	$j_d^{KGE}/2 + j_f^{Epf}/2$
$\tilde{c_i}$	15.45 [20.22, 26.34]	10.91 [20.14, 26.75]	17.6 [27.19, 33.16]	17.3 [30.32, 36.05]
$\overline{c_i}$	14.71 [20.3, 26.22]	16.93 [20.83, 26.48]	17.6 [27.15, 33.07]	17.27 [30.17, 35.83]
σ_{c_i}	0.22 [0.91, 1.46]	0.07 [1.13, 4.16]	0.13 [0.57, 1.28]	0.05 [0.82, 3.09]
$\tilde{c_p}$	161.81 [286.05, 435.48]	145.79 [314.19, 518.0]	156.65 [288.75, 476.53]	148.27 [296.94, 485.79]
$\overline{c_p}$	169.99 [291.17, 434.58]	146.04 [310.14, 505.68]	151.87 [286.79, 483.33]	141.56 [289.69, 466.99]
σ_{c_p}	38.52 [60.58, 59.64]	8.95 [37.57, 44.15]	31.08 [53.49, 57.35]	12.03 [46.82, 102.47]
$\tilde{c_{tr}}$	174.6 [287.44, 270.08]	158.48 [317.5, 311.03]	266.09 [429.0, 353.73]	267.12 [447.27, 372.32]
$\overline{c_{tr}}$	171.76 [286.6, 269.5]	162.66 [313.49, 304.83]	266.32 [431.2, 355.04]	267.21 [436.83, 360.62]
$\sigma_{c_{tr}}$	13.88 [28.84, 35.82]	3.23 [25.3, 57.2]	5.68 [24.28, 60.74]	1.45 [24.39, 74.56]
$\tilde{c_{tl}}$	114.79 [675.7, 1276.32]	127.09 [1322.22, 2806.59]	180.63 [1332.45, 2784.42]	146.96 [1322.17, 2803.08]
$\overline{c_{tl}}$	347.87 [812.15, 1274.12]	250.42 [1366.73, 2789.22]	383.51 [1413.93, 2749.35]	262.89 [1337.96, 2795.16]
$\sigma_{c_{tl}}$	193.72 [355.92, 433.62]	34.67 [139.21, 252.41]	69.91 [222.54, 388.17]	31.82 [61.87, 81.06]
$\tilde{c_r}$	41.37 [52.08, 34.42]	41.37 [52.08, 34.42]	41.37 [52.08, 34.42]	41.37 [52.08, 34.42]
$\overline{c_r}$	41.32 [52.63, 34.05]	40.94 [50.97, 30.97]	41.33 [51.2, 30.29]	40.24 [50.2, 27.58]
σ_{c_r}	4.66 [6.01, 5.04]	1.45 [5.17, 10.17]	3.01 [5.31, 10.29]	1.34 [5.08, 13.82]
$\tilde{ml}$	-0.59 [-4.98, 8.21]	0.0 [-3.72, 7.42]	0.0 [-3.61, 7.31]	-0.0 [-3.27, 6.39]
$\overline{ml}$	-0.59 [-4.98, 8.21]	0.0 [-3.81, 7.34]	-0.0 [-3.62, 7.31]	-0.0 [-3.28, 6.39]
σ_{ml}	0.01 [0.05, 0.09]	0.0 [0.17, 0.73]	0.02 [0.05, 0.09]	0.0 [0.08, 0.34]

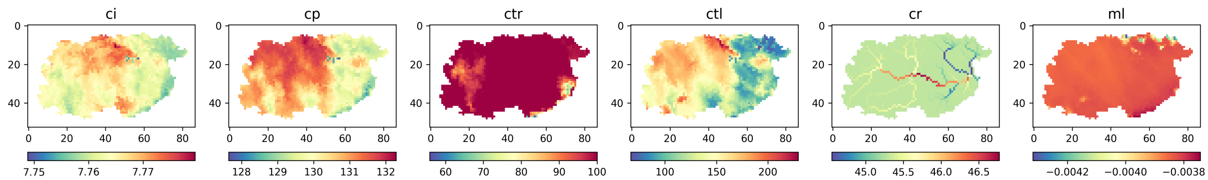

Figure 11: Spatially distributed optimal parameters $(\hat{\theta}(x) \equiv (c_i(x), c_p(x), c_{tr}(x), c_r(x), ml(x), c_{tl}(x))^T)$ for the Argens River basin, obtained by minimizing $j_d^{KGE}/2 + j_f^{Epf}/2$.

4. Conclusion

In this study, we enhanced the calibration process of the conceptual distributed hydrological model SMASH for Mediterranean floods by incorporating hydrological signatures and various multi-criteria optimization strategies. Firstly, we computed and analyzed both continuous signatures and flood event signatures. Subsequently, we used sensitivity analysis to select appropriate signatures for constraining the model. Finally, we performed signatures-based multi-criteria optimization approaches, which demonstrated their robustness and reliability in improving simulated peak flood events without significantly compromising the NSE and KGE. Notably, for distributed calibration, the model constrained by the signature performed better in simulating flood events and achieved higher NSE and KGE scores compared to the model calibrated without using signatures. These results highlight the superiority of signature-based calibration approaches, particularly in flash flood prediction. Furthermore, we compared the parameter spaces of different models to provide insights into the optimal transition from traditional calibration approaches to signature-based calibration methods.

Our proposed calibration strategy addresses the need for an intelligent approach to model calibration in the presence of multiple objectives and complex hydrological processes. This approach offers a new perspective on calibration that accounts for not only classical discharge metrics but also multi-scale hydrological signatures that can provide a more comprehensive assessment of model performance. This approach could be reinforced via the use of multi-source information such as from remotely sensed data products and of multi-gauge streamflow series in regionalization problems. The segmentation algorithm could be tested on larger flood samples, also including catchment rainfall moments (Zoccatelli et al., 2011; Emmanuel et al., 2015) describing rainfall paterns for floods analysis (e.g. Garambois et al. (2014); Saharia et al. (2021)) and in order to prepare learning sets for training hybrid flood modeling-correction approaches.

Future work will aim to (i) upgrade the variational calibration algorithm with Bayesian approach in context of equifinality and with penalization based on global sensitivity metrics that could be derived from local derivatives, including for spatially distributed controls using adjoint model, and also with improved spatial constrains through physiographic descriptors-to-parameters fields mappings (pre-regionalization); as well as to (ii) perform extended tests and analysis on larger samples of catchments and signatures, with complementary data using various model structures.

A. Classical calibration metrics in hydrology

A.1. Nash–Sutcliffe efficiency (NSE)

$$NSE = 1 - \frac{\sum_{t=0}^{T}(Q(t) - Q^*(t))^2}{\sum_{t=0}^{T}(Q^*(t) - \overline{Q^*})^2}$$

where $Q(t)$ is the simulated discharge at time t, $Q^*(t)$ is the observed discharge at time t and $\overline{Q^*}$ is the mean observed discharge.

A.2. Kling–Gupta efficiency (KGE)

$$KGE = 1 - \sqrt{\alpha(r-1)^2 + \beta(\frac{\sigma}{\sigma^*} - 1)^2 + \gamma(\frac{\mu}{\mu^*} - 1)^2}$$

where r is the linear correlation between observations and simulations, σ and σ^* are the standard deviation in simulations and observations, respectively, μ and μ^* are the mean discharge in simulations and observations, respectively, and α, β, γ are the optimization weight parameters.

B. List of studied signatures

Denote $P(t)$ and $Q(t)$ are the rainfall and runoff at time $t \in \mathbf{U}$, where $\mathbf{U}$ is the study period. Then $Qb(t)$ and $Qq(t)$ are the baseflow and quickflow computed using a classical technique for streamflow separation (please refer to Lyne and Hollick (1979) and Nathan and McMahon (1990) for more details).

B.1. Continuous signatures

The continuous signatures are calculated over the entire study period as Table 8.

B.2. Flood event signatures

For an event occurring in $\mathbf{E} \subset \mathbf{U}$, the flood event signatures are calculated as Table 9.

C. Multi-objective optimization with spatially uniform control vectors

We look into multi-objective optimization for a global calibration of spatially uniform parameters of the distributed model $\mathcal{M}$, i.e. a low dimensional control $\overline{\theta}$. The multi-objective calibration is simply defined as the optimization

Table 8
List of all studied continuous signatures.

Notation	Signature	Description	Formula	Unit
Crc	Runoff coefficients	Coefficient relating the amount of runoff to the amount of precipitation received	$\frac{\int^{t\in\mathbf{U}} Q(t)dt}{\int^{t\in\mathbf{U}} P(t)dt}$	-
Crchf		Coefficient relating the amount of high-flow to the amount of precipitation received	$\frac{\int^{t\in\mathbf{U}} Qq(t)dt}{\int^{t\in\mathbf{U}} P(t)dt}$	-
Crclf		Coefficient relating the amount of low-flow to the amount of precipitation received	$\frac{\int^{t\in\mathbf{U}} Qb(t)dt}{\int^{t\in\mathbf{U}} P(t)dt}$	-
Crch2r		Coefficient relating the amount of high-flow to the amount of runoff	$\frac{\int^{t\in\mathbf{U}} Qq(t)dt}{\int^{t\in\mathbf{U}} Q(t)dt}$	-
Cfp2	Flow percentiles	2%, 10%, 50% and 90%-quantiles from flow duration curve	$quantile(Q(t), 0.02)$	mm
Cfp10			$quantile(Q(t), 0.1)$	
Cfp50			$quantile(Q(t), 0.5)$	
Cfp90			$quantile(Q(t), 0.9)$	

Table 9
List of all studied flood event signatures.

Notation	Signature	Description	Formula	Unit
Eff	Flood flow	Amount of quickflow in flood event	$\int^{t\in\mathbf{E}} Qq(t)dt$	mm
Ebf	Base flow	Amount of baseflow in flood event	$\int^{t\in\mathbf{E}} Qb(t)dt$	mm
Erc	Runoff coefficients	Coefficient relating the amount of runoff to the amount of precipitation received	$\frac{\int^{t\in\mathbf{E}} Q(t)dt}{\int^{t\in\mathbf{E}} P(t)dt}$	-
Erchf		Coefficient relating the amount of high-flow to the amount of precipitation received	$\frac{\int^{t\in\mathbf{E}} Qq(t)dt}{\int^{t\in\mathbf{E}} P(t)dt}$	-
Erclf		Coefficient relating the amount of low-flow to the amount of precipitation received	$\frac{\int^{t\in\mathbf{E}} Qb(t)dt}{\int^{t\in\mathbf{E}} P(t)dt}$	-
Erch2r		Coefficient relating the amount of high-flow to the amount of runoff	$\frac{\int^{t\in\mathbf{E}} Qq(t)dt}{\int^{t\in\mathbf{E}} Q(t)dt}$	-
Elt	Lag time	Difference time between the peak runoff and the peak rainfall	$\arg\max_{t\in\mathbf{E}} Q(t) - \arg\max_{t\in\mathbf{E}} P(t)$	h
Epf	Peak flow	Peak runoff in flood event	$\max_{t\in\mathbf{E}} Q(t)$	mm

problem:

$$\min_{\theta\in\mathcal{O}\subset\mathbb{R}^n} (j_1(\boldsymbol{\theta}), ..., j_m(\boldsymbol{\theta})) \tag{7}$$

where $\boldsymbol{\theta}$ is the n-dimensional vector of model parameters in the feasible space $\mathcal{O} \subset \mathbb{R}^n$ and $j_1, ..., j_m$ are the m single-objective functions to be simultaneously minimized.

C.1. Pareto front

In single-objective optimization, the Pareto optimal solution is unique (in terms of objective space) but in multi-objective problem, it common to have several solutions that can not be defined which one is the best. If the optimization problem is non-dominated, or non-inferior (each objective function is its own entity, so no individual can be better off without making at least one individual worse off), then we call that Pareto optimality, or Pareto efficiency. A Pareto front (in terms of parameter space) is a set of all Pareto efficient solutions that need to be estimated. Let us consider two feasible solutions: $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{O}$. Then, $\boldsymbol{\theta}_1$ is said to Pareto dominate $\boldsymbol{\theta}_2$ if the following properties hold:

1. $\forall i \in \{1, ..., m\}, j_i(\theta_1) \leq j_i(\theta_2)$;
2. $\exists i \in \{1, ..., m\}, j_i(\theta_1) < j_i(\theta_2)$.

We call $\mathcal{P}$ the Pareto set representing all of Pareto solutions. By definition, a Pareto solution $\theta^* \in \mathcal{P}$ of problem 7 must fill the two following conditions:

1. $\nexists \theta' \in \mathcal{O} \setminus \mathcal{P}, \exists i \in \{1, ..., m\}, j_i(\theta') < j_i(\theta^*)$;
2. $\nexists \theta'' \in \mathcal{P}, \theta''$ dominates θ^*.

The first statement indicates that there does not exist other point in the feasible space that reduces at least one objective function while keeping others unchanged, so the Pareto set is the optimal set. The second says that, no other point exists in the Pareto set that decreases one objective function without increasing another one, so it is impossible to distinguish any solution as being better than the other in the Pareto set. Fig. 12 illustrates this for a simple problem where we have 2-objective functions j_1, j_2. The Pareto front (in terms of objective space) represents all of non-dominated optimal solutions. It implies that, it is impossible to move from any point in the feasible space and simultaneously decrease the two objective functions without violating a constraint.

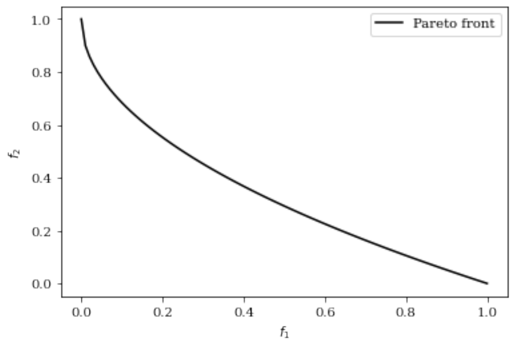

Figure 12: Illustration of Pareto front in terms of objective space.

C.2. Overview of GA

GA is a "heuristic algorithm" (or search heuristic) in optimization, inspired by the Theory of Natural Evolution, whose selection operators include "crossover" and "mutation". Basically, the process of a GA and a MOGA consist of the following 3 phases:

1. *Population initialization.* The population is randomly initialized based on the problem range and constraint. The size of the population determines also the number of solutions, called "pop-size".
2. *Parents selection (sorting).* A fitness function is defined to calculate the fitness score (also called Pareto ranking in multi-objective optimization) that determines how fit an individual is to the problem. Then, the fitness score decides the probability of selecting an individual as a parent to reproduce offspring population.

3. *Mating.* For each pair of parent to be mated, new offspring are created by exchanging the genes of parents among themselves (crossover operator). To maintain the diversity within the population and prevent premature convergence, some of the bits in the gene of certain new offspring can be flipped with a low random probability (mutation operator). Offspring are created until their pop-size is equal to the pop-size of previous generation.

C.3. Selection of an optimal solution from Pareto front

We aim to select an optimal solution that is acceptable for every objective within a constraint on principal objective function. Many strategies can be chosen to perform such a selection (e.g. based on the sensitivity ratio that is the ratio of the average variabilities of a certain non-inferior solution to the corresponding value of the objective function in the Pareto front (Wang et al., 2017), or the Euclidean distance from the ideal solution (Wang and Rangaiah, 2017)). A simple additive weighting (SAW) method in Wang and Rangaiah (2017) can be used in our case by adding a normalization operator and assigned weightage for the objective functions.

Considering an objective matrix $(j_{ij})_{1\leq i\leq m, 1\leq j\leq n}$, where m is the number of non-dominated solutions, n is the number of objective functions. Then each row i represents the i^{th} solution set of the Pareto front and each column j represents all non-inferior solutions of the j^{th} objective function. Denote c be the index of the classical objective function (for example $1-NSE$ or $1-KGE$), which is the most constrained function to find a unique optimal solution from Pareto front. This algorithm is detailed in the following three phases:

1. Objective matrix normalization $(F_{ij})_{1\leq i\leq m, 1\leq j\leq n}$:

$$F_{ij} = \frac{f_j^+ - f_{ij}}{f_j^+ - f_j^-} \text{ where } f_j^+ = \max_{1\leq i\leq m} f_{ij} \text{ and } f_j^- = \min_{1\leq i\leq m} f_{ij}$$

2. Assigning weightage for normalized objective matrix $(G_{ij})_{1\leq i\leq m, 1\leq j\leq n}$:

$$G_{ij} = w_j \times F_{ij} \text{ where } w_j = \begin{cases} e^d, \text{ if } j = c \\ e - e^d, \text{ otherwise.} \end{cases} \quad \text{and } d = f_c^+ - f_c^-$$

3. Finding optimal solution $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = (f_{k1}, ..., f_{kn}) \text{ where } k = \arg \max_{1\leq i\leq m} \left(\sum_{j=1}^{n} G_{ij} \right).$$

D. Calibration bounds

The parameter vector of SMASH model structure S6 is $\boldsymbol{\theta}(x) \equiv \left(c_i(x), c_p(x), c_{tr}(x), c_r(x), ml(x), c_{tl}(x)\right)^T$ and bound constrains used in optimization (Eq. 3) are set with values given in Table 10.

Table 10
Boundary conditions of SMASH 6-parameters model.

	c_i	c_p	c_{tr}	c_r	ml	c_{tl}
Lower boundary	1	1	1	1	-20	1
Upper boundary	100	2000	1000	200	5	10000

Acknowledgments

The authors greatly acknowledge SCHAPI-DGPR and METEO France for providing data used in this study. The first author was supported by funding from SCHAPI-DGPR and ANR grant ANR-21-CE04-0021-01 (MUFFINS project, "MUltiscale Flood Forecasting with INnovating Solutions").

CRediT authorship contribution statement

Ngo Nghi Truyen Huynh: research plan, conceptualization, software and numerical result, analysis, draft preparation, final redaction. **Pierre-André Garambois:** research methodology, conceptualization, analysis, draft preparation, final redaction, supervision, project administration, funding aquisition. **François Colleoni:** conceptualization, software and numerical results, analysis, final redaction. **Pierre Javelle:** research methodology, conceptualization, analysis, final redaction, funding aquisition.

References

Astagneau, P.C., Bourgin, F., Andréassian, V., Perrin, C., 2021. When does a parsimonious model fail to simulate floods? learning from the seasonality of model bias. Hydrological Sciences Journal 66, 1288–1305. URL: https://doi.org/10.1080/02626667.2021.1923720, doi:10.1080/02626667.2021.1923720, arXiv:https://doi.org/10.1080/02626667.2021.1923720.

Azzini, I., Mara, T.A., Rosati, R., 2021. Comparison of two sets of monte carlo estimators of sobol'indices. Environmental Modelling & Software 144, 105167.

Beck, M.B., 1987. Water quality modeling: a review of the analysis of uncertainty. Water resources research 23, 1393–1442.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in water resources 16, 41–51.

Blank, J., Deb, K., 2020. pymoo: Multi-objective optimization in python. IEEE Access 8, 89497–89509.

Bout, V., Jetten, V., 2018. The validity of flow approximations when simulating catchment-integrated flash floods. Journal of hydrology 556, 674–688.

Brigode, P., Génot, B., Lobligeois, F., Delaigue, O., 2020. Summary sheets of watershed-scale hydroclimatic observed data for france. URL: https://doi.org/10.15454/UV01P1, doi:10.15454/UV01P1.

Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. a unified approach. Computer physics communications 182, 978–988.

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 , 28–34.

Chibeles-Martins, N., Pinto-Varela, T., Barbosa-Póvoa, A.P., Novais, A.Q., 2016. A multi-objective meta-heuristic approach for the design and planning of green supply chains-mbsa. Expert Systems with Applications 47, 71–84.

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. submitted .

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

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE transactions on evolutionary computation 6, 182–197.

Delaigue, O., Génot, B., Lebecherel, L., Brigode, P., Bourgin, P.Y., 2020. Database of watershed-scale hydroclimatic observations in france. URL: https://webgr.inrae.fr/base-de-donnees.

Désidéri, J.A., 2012. Multiple-gradient descent algorithm (mgda) for multiobjective optimization. Comptes Rendus Mathematique 350, 313–318.

Douinot, A., Roux, H., Garambois, P.A., Dartus, D., 2018. Using a multi-hypothesis framework to improve the understanding of flow dynamics during flash floods. Hydrology and Earth System Sciences 22, 5317–5340.

Duarte, M., Watanabe, R.N., 2021. Notes on Scientific Computing for Biomechanics and Motor Control. URL: https://doi.org/10.5281/zenodo.4599319, doi:10.5281/zenodo.4599319.

Edijatno, 1991. Mise au point d’un modele elementaire pluie-debit au pas de temps journalier. Ph.D. thesis. Universite Louis Pasteur, ENGEES, Cemagref Antony.

El-Ghandour, H.A., Elbeltagi, E., 2014. Optimal groundwater management using multiobjective particle swarm with a new evolution strategy. Journal of Hydrologic Engineering 19, 1141–1149.

Emmanuel, I., Andrieu, H., Leblois, E., Janey, N., Payrastre, O., 2015. Influence of rainfall spatial variability on rainfall–runoff modelling: benefit of a simulation approach? Journal of hydrology 531, 337–348.

Garambois, P.A., Larnier, K., Roux, H., Labat, D., Dartus, D., 2014. Analysis of flash flood-triggering rainfall for a process-oriented hydrological model. Atmospheric research 137, 14–24.

Garambois, P.A., Roux, H., Larnier, K., Castaings, W., Dartus, D., 2013. Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in mediterranean catchments. Hydrology and Earth System Sciences .

Garambois, P.A., Roux, H., Larnier, K., Labat, D., Dartus, D., 2015. Characterization of catchment behaviour and rainfall selection for flash flood hydrological model calibration: catchments of the eastern pyrenees. Hydrological sciences journal 60, 424–447.

Guo, J., Zhou, J., Lu, J., Zou, Q., Zhang, H., Bi, S., 2014. Multi-objective optimization of empirical hydrological model for streamflow prediction. Journal of Hydrology 511, 242–253.

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

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of hydrology 377, 80–91.

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

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research 34, 751–763.

Hascoet, L., Pascual, V., 2013. The tapenade automatic differentiation tool: Principles, model, and specification. ACM Trans. Math. Softw. 39. URL: https://doi.org/10.1145/2450153.2450158, doi:10.1145/2450153.2450158.

Herman, J., Usher, W., 2017. SALib: An open-source python library for sensitivity analysis. The Journal of Open Source Software 2. URL: https://doi.org/10.21105/joss.00097, doi:10.21105/joss.00097.

Horner, I., 2020. Design and Evaluation of hydrological signatures for the diagnosis and improvement of a process-based distributed hydrological model. Ph.D. thesis. Université Grenoble Alpes.

Iwanaga, T., Usher, W., Herman, J., 2022. Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses. Socio-Environmental Systems Modelling 4, 18155. URL: https://sesmo.org/article/view/18155, doi:10.18174/sesmo.18155.

Jay-Allemand, M., 2020. Estimation variationnelle des paramètres d'un modèle hydrologique distribué. Ph.D. thesis. Aix-Marseille.

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

Jay-Allemand, M., Demargne, J., Garambois, P.A., Javelle, P., Gejadze, I., Colleoni, F., Organde, D., Arnaud, P., Fouchier, C., 2022b. Spatially distributed calibration of a hydrological model with variational optimization constrained by physiographic maps for flash flood forecasting in France. Technical Report. Copernicus Meetings.

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, 5519–5538.

Khorram, E., Khaledian, K., Khaledyan, M., 2014. A numerical method for constructing the pareto front of multi-objective optimization problems. Journal of Computational and Applied Mathematics 261, 158–171.

Kirstetter, G., Delestre, O., Lagrée, P.Y., Popinet, S., Josserand, C., 2021. B-flood 1.0: an open-source saint-venant model for flash-flood simulation using adaptive refinement. Geoscientific Model Development 14, 7117–7132.

Lamboni, M., Iooss, B., Popelin, A.L., Gamboa, F., 2013. Derivative-based global sensitivity measures: General links with sobol'indices and numerical tests. Mathematics and Computers in Simulation 87, 45–54.

Le Mesnil, M., 2021. Signatures Hydrologiques des Bassins Karstiques. Theses. Montpellier SupAgro. URL: https://tel.archives-ouvertes.fr/tel-03578569.

Li, K., Huang, G., Wang, S., Razavi, S., Zhang, X., 2022. Development of a joint probabilistic rainfall-runoff model for high-to-extreme flow projections under changing climatic conditions. Water Resources Research 58, e2021WR031557.

Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J.J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., et al., 2021. Crest-imap v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environmental Modelling & Software 141, 105051.

Lyne, V., Hollick, M., 1979. Stochastic time-variable rainfall-runoff modelling, in: Institute of Engineers Australia National Conference, Institute of Engineers Australia Barton, Australia. pp. 89–93.

McMillan, H.K., 2021. A review of hydrologic signatures and their applications. Wiley Interdisciplinary Reviews: Water 8, e1499.

Mercier, Q., Poirion, F., Désidéri, J.A., 2018. A stochastic multiple gradient descent algorithm. European Journal of Operational Research 271, 808–817.

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

Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H.V., Kumar, R., 2019. On the choice of calibration metrics for "high-flow" estimation using hydrologic models. Hydrology and Earth System Sciences 23, 2601–2614.

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.

Mostafaie, A., Forootan, E., Safari, A., Schumacher, M., 2018. Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily grace data. Computational Geosciences 22, 789–814.

Murata, T., Ishibuchi, H., et al., 1995. Moga: multi-objective genetic algorithms, in: IEEE international conference on evolutionary computation, IEEE Piscataway, NJ, USA. pp. 289–294.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part i—a discussion of principles. Journal of hydrology 10, 282–290.

Nathan, R.J., McMahon, T.A., 1990. Evaluation of automated techniques for base flow and recession analyses. Water resources research 26, 1465–1473.

Oliveira, A.M., Fleischmann, A., Paiva, R., 2021. On the contribution of remote sensing-based calibration to model hydrological and hydraulic processes in tropical regions. Journal of Hydrology 597, 126184.

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, 290–306.

Paiva, R.C., Collischonn, W., Tucci, C.E., 2011. Large scale hydrologic and hydrodynamic modeling using limited data and a gis based approach. Journal of Hydrology 406, 170–181.

Pelletier, A., Andréassian, V., 2020. Hydrograph separation: an impartial parametrisation for an imperfect method. Hydrology and earth system sciences 24, 1171–1187.

Pujol, L., Garambois, P.A., Finaud-Guyot, P., Monnier, J., Larnier, K., Mose, R., Biancamaria, S., Yesou, H., Moreira, D., Paris, A., et al., 2020. Estimation of multiple inflows and effective channel by assimilation of multi-satellite hydraulic signatures: The ungauged anabranching negro river. Journal of Hydrology 591, 125331.

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, 6085–6113.

Puy, A., Becker, W., Piano, S.L., Saltelli, A., 2022. A comprehensive comparison of total-order estimators for global sensitivity analysis. International Journal for Uncertainty Quantification 12.

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

Razavi, S., Gupta, H.V., 2019. A multi-method generalized global sensitivity matrix approach to accounting for the dynamical nature of earth and environmental systems models. Environmental modelling & software 114, 1–11.

Ross, M., Abbey, C., Bouffard, F., Jos, G., 2015. Multiobjective optimization dispatch for microgrids with a high penetration of renewable generation. IEEE Transactions on Sustainable Energy 6, 1306–1314.

Roux, H., Labat, D., Garambois, P.A., Maubourguet, M.M., Chorda, J., Dartus, D., 2011. A physically-based parsimonious hydrological model for flash floods in mediterranean catchments. Natural Hazards and Earth System Sciences 11, 2567–2582.

Saharia, M., Kirstetter, P.E., Vergara, H., Gourley, J.J., Emmanuel, I., Andrieu, H., 2021. On the impact of rainfall spatial variability, geomorphology, and climatology on flash floods. Water Resources Research 57, e2020WR029124.

Sahraei, S., Asadzadeh, M., Unduche, F., 2020. Signature-based multi-modelling and multi-objective calibration of hydrologic models: Application in flood forecasting for canadian prairies. Journal of Hydrology 588, 125095.

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. Computer physics communications 145, 280–297.

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

Shafii, M., Tolson, B.A., 2015. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. Water Resources Research 51, 3796–3814. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016520, doi:https://doi.org/10.1002/2014WR016520, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR016520.

Sobol, I., Kucherenko, S., 2010. Derivative based global sensitivity measures. Procedia-Social and Behavioral Sciences 2, 7745–7746.

Tarasova, L., Basso, S., Zink, M., Merz, R., 2018. Exploring controls on rainfall-runoff events: 1. time series-based event separation and temporal dynamics of event runoff response in germany. Water Resources Research 54, 7711–7732.

Tavakkoli-Moghaddam, R., Azarkish, M., Sadeghnejad-Barkousaraie, A., 2011. A new hybrid multi-objective pareto archive pso algorithm for a bi-objective job shop scheduling problem. Expert Systems with Applications 38, 10812–10821.

Torres-Treviño, L.M., Reyes-Valdes, F.A., López, V., Praga-Alejo, R., 2011. Multi-objective optimization of a welding process by the estimation of the pareto optimal set. Expert systems with applications 38, 8045–8053.

Veluscek, M., Kalganova, T., Broomhead, P., Grichnik, A., 2015. Composite goal methods for transportation network optimization. Expert Systems with Applications 42, 3852–3867.

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.

Wang, N., Zhao, W.j., Wu, N., Wu, D., 2017. Multi-objective optimization: a method for selecting the optimal solution from pareto non-inferior solutions. Expert Systems with Applications 74, 96–104.

Wang, Z., Rangaiah, G.P., 2017. Application and analysis of methods for selecting an optimal solution from the pareto-optimal front obtained by multiobjective optimization. Industrial & Engineering Chemistry Research 56, 560–574.

Wu, H., Chen, B., Ye, X., Guo, H., Meng, X., Zhang, B., 2021. An improved calibration and uncertainty analysis approach using a multicriteria sequential algorithm for hydrological modeling. Scientific Reports 11, 1–14.

Wu, Y.K., Liu, C.C., Lur, Y.Y., 2015. Pareto-optimal solution for multiple objective linear programming problems with fuzzy goals. Fuzzy Optimization and Decision Making 14, 43–55.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of hydrology 204, 83–97.

Yeh, W.W.G., 1986. Review of parameter identification procedures in groundwater hydrology: The inverse problem. Water resources research 22, 95–108.

Yilmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the nws distributed hydrologic model. Water Resources Research 44. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007WR006716, doi:https://doi.org/10.1029/2007WR006716, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007WR006716.

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, 550–560. URL: http://dblp.uni-trier.de/db/journals/toms/toms23.html#ZhuBLN97.

Zoccatelli, D., Borga, M., Viglione, A., Chirico, G., Blöschl, G., 2011. Spatial moments of catchment rainfall: rainfall spatial organisation, basin morphology, and flood response. Hydrology and Earth System Sciences 15, 3767–3783.