https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Adjoint-based spatially distributed calibration of a grid GR-based parsimonious hydrological model over 312 French catchments with SMASH platform

François Colleoni[1], Pierre-André Garambois[1], Pierre Javelle[1], Maxime Jay-Allemand[2], and Patrick Arnaud[1]

[1]INRAE, Aix Marseille Université, RECOVER, 3275 Route Cézanne, Aix-en-Provence, 13182, France
[2]Hydris Hydrologie corp., Montpellier, France

Correspondence: pierre-andre.garambois@inrae.fr

Abstract. Reducing uncertainty and improving robustness and spatio-temporal extrapolation capabilities remain key challenges in hydrological modeling especially for flood forecasting over large areas. Parsimonious model structures and effective optimization strategies are crucially needed to tackle the difficult issue of distributed hydrological model calibration from sparse integrative discharge data, that is in general high dimensional inverse problems. This contribution presents the first evaluation of Variational Data Assimilation (VDA), very well suited to this context but still rarely employed in hydrology because of high technicality, and successfully applied here to the spatially distributed calibration of a newly taylored grid-based parsimonious model structure and corresponding adjoint. It is based on the Variational Data Assimilation (VDA) framework of SMASH (*Spatially distributed Modelling and ASsimilation for Hydrology*) platform, underlying the French national flash flood forecasting system Vigicrues Flash. It proposes an upgraded distributed hourly rainfall-runoff model structure employing GR-based operators, including a non-conservative flux, and its adjoint obtained by automatic differentiation for VDA. The performances of the approach are assessed over annual, seasonal and floods timescales via standard performance metrics and in spatio-temporal validation. The gain of using the proposed non-conservative 6-parameters model structure is highlighted in terms of performance and robustness, compared to a simpler 3-parameters structure. Spatially distributed calibrations lead to a significant gain in terms of reaching high performances in calibration and temporal validation on the catchments sample, with median efficiencies respectively of NSE = 0.88 (resp. 0.85) and NSE = 0.8 (resp. 0.79) over the total time window on period $p2$ (resp. $p1$). Simulated signatures in temporal validation over 1443 (resp. 1522) flood events on period $p2$ (resp. $p1$) are quite good with median flood (NSE; KGE) of (0.63; 0.59) (resp. (0.55; 0.53)). Spatio-temporal validations, i.e. on pseudo ungauged cases, lead to encouraging performances also. Moreover, the influence of certain catchment characteristics on model performance and parametric sensitivity is analyzed. Best performances are obtained for Oceanic and Mediterranean basins whereas it performs less well over Uniform basins with significant influence of multi-frequency hydrogeological processes. Interestingly, regional sensitivity analysis revealed that the non conservative water exchange parameter and the production parameter, impacting the simulated runoff amount, are the most sensitive parameters along with the routing parameter especially for faster responding catchments. This study is a first step in the construction of a flexibe and versatile multi-model and optimization framework with hydbrid methods for regional hydrological modeling with multi-source data assimilation.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Keywords: hydrological modeling, distributed, parsimonious, GR, high dimension, calibration, variational data assimilation, large sample, accuracy, robustness

1 Introduction

In the context of climate change and potential intensification of the hydrological cycle (e.g. Shukla et al., 2019; Allan et al., 2020), advanced high resolution modeling tools are needed to perform accurate and reliable local forecasts, in terms of location, magnitude and timing of hydrological responses. A core ingredient of a hydrometeorological forecasting system is hydrological modeling (Pagano et al., 2014) which yet remains a hard task tinged with uncertainty due to the high heterogeneity and variability of physical processes and their limited observability (e.g. Beven, 1989; Milly, 1994; Blöschl and Sivapalan, 1995; Refsgaard, 1997; Vereecken et al., 2019). Reducing uncertainty and improving robustness and spatio-temporal extrapolation capabilities remain key challenges in hydrological modeling (Blöschl and al., 2019) especially over large catchment samples with contrasting physiographic properties and hydrological functioning (e.g. Lane et al., 2019; Beck et al., 2020). A trade off has to be made between modeling goals (quantities of interest, resolution, etc) and model complexity given data availability.

The "resolution–complexity continuum" (Clark et al., 2017) has been investigated over the past 5 decades by many studies with various modeling approaches, ranging from point-scale processes numerically integrated at larger scales (e.g. catchment) to spatially lumped representation of the system response (Hrachowitz and Clark, 2017). Among the variety of existing hydrological models, and hypothesis they rely on, their components generally describe water storage and transfer (e.g. Fenicia et al. (2011)) via various combinations and parameterizations of vertical and lateral storage-flux operators. Several model comparison experiments, aiming at analyzing the differences between various modeling approaches, tested the performances in terms of streamflow modeling (Perrin et al., 2001; Reed et al., 2004; Duan et al., 2006; Orth et al., 2015) but also in terms of internal state such as soil moisture (Orth et al., 2015; Bouaziz et al., 2021). Orth et al. (2015) concluded that "added complexity does not necessarily lead to improved performance of hydrological models (...)". Note that, parsimonious conceptual models, lumped or semi lumped, performed efficiently in large sample studies (e.g. GR model in Perrin et al. (2001), GRSD model in De Lavenne et al. (2019), GR and MORDOR models in Mathevet et al. (2020), FUSE models in Lane et al. (2019) and references therein). Such large sample studies, enabling more general and statistically sound analysis of model performances (Andréassian et al., 2009; Gupta et al., 2014), are useful to address large scale challenges with consistent methodologies across scales and contrasted conditions and hydrological functioning. Consequently, national-scale hydrological modeling studies are increasingly applied and constitute appropriate benchmarks for model development as well as for identifying where models fail over a range of catchment characteristics and functioning (Lane et al., 2019). This lattest study points the potential need to account for groundwater exchanges as done in GR models for instance (cf. (Perrin et al., 2001; Le Moine, 2008)).

All hydrological models are to some degree conceptual and due to limitations and uncertainties (in their structure, parameters representativity, data availability, and even initial and boundary conditions), calibration/learning is generally required. Whatever their status and complexity, hydrological models are most often calibrated and validated using discharge time series at the catchment outlet(s) (Sebben et al., 2013). However, calibrating hydrological model parameters from sparse and

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

integrative discharge data is a difficult inverse problem faced with equifinality issues (Bertalanffy, 1968; Beven, 1993, 2001), especially for distributed models with larger number of cells and parameters. Calibration approaches in hydrology generally involve regularization strategies, such as imposing spatial patterns via descriptor-to-parameter mapping laws with few tunable hyperparameters (cf. Pokhrel and Gupta (2010) and references therein, used in (Samaniego et al., 2010) along with descriptor upscaling laws), relaxation strategies for semi-lumped parameter fields in multi-gauge calibration problem (e.g. (De Lavenne et al., 2019)). Calibration from spatially sparse discharge data of spatially distributed hydrological model parameters, assuming no a priori spatial structure, is a high dimensional ill-posed inverse problems and the variational data assimilation (VDA) approach proposed in Jay-Allemand et al. (2020) is well adapted to this context but still rarely employed in hydrology due to high technicality related to the obtention of an adjoint model (see Monnier et al. (2016); Larnier et al. (2020); Jay-Allemand et al. (2020) and references therein). This approach has been tested with a simple 3-parameters continuous distributed hydrological model, on a single flash flood prone catchment (the Gardon d'Anduze in south of France) with promising results (Jay-Allemand et al., 2020).

This contribution presents the first evaluation over a large sample, of Variational Data Assimilation (VDA) for the spatially distributed calibration of a newly taylored parsimonious distributed hydrological model structure and corresponding adjoint. An upgraded grid based model structure is proposed in the variational calibration framework proposed by Jay-Allemand et al. (2020) and its potential is demonstrated over a large sample. Compared to this initial study Jay-Allemand et al. (2020), the following new insights have been introduced. First, instead of one catchment from the South of France, results are here evaluated over 312 catchments across France, representative of various hydro-climatological French conditions. To our knowledge, this is the first evaluation of a spatially distributed variational calibration algorithm over a large sample. Second, the present work also proposes a more complex distributed structure (i.e. 6 parameters per cell including a non conservative "water exchange" parameter), obtained after extended testing and differentiated to obtain its adjoint, and compares the results with the initial version (i.e. 3 parameters per cell). Both versions are "GR-based", that is to say they are inspired by the conceptual "Génie Rural" models (lumped or semi-distributed) developed at INRAE-Antony this last 40 years. Those parsimonious and "easy-to-apply" models are now recognized in various international studies to perform well over large samples outside France (see for instance Coron et al., 2012; de Boer-Euser et al., 2017; Bouaziz et al., 2021; Astagneau et al., 2021b; Mathevet et al., 2020). Lastly, a global sensitivity analysis of the proposed model structure to its parameters is performed considering hydrological regimes, as well as a signature analysis over flood events over the whole sample.

The specific research questions we study are as follows:

1. How well does a parsimonious distributed model structure perform across France when assessed over annual, seasonal and floods timescales via standard performance metrics and in spatio-temporal validation?

2. What is the gain of using a complexified model structure including non conservative exchange, in terms of performances and robustness?

3. What is the gain of distributed parameters calibration with a variational data assimilation algorithm in terms of performance and robustness, compared with an uniform calibration?

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

4. What is the influence of certain catchment characteristics on model performance and parametric sensitivity?

The article is organized as follows. Section 2 details the methodology, models and calibration algorithm. The dataset and numerical experimental design is presented in Section 3. Results are analyzed and discussed in Section 4. Conclusions and perspectives are formulated in Section 5.

2 Methods

This section first presents the proposed distributed hydrological model along with the variational calibration algorithm.

2.1 Parsimonious distributed hydrological models with SMASH

SMASH is a computational software framework dedicated to *Spatially distributed Modeling and data ASsimilation for Hydrology*, enabling to tackle high dimensional inverse problems with adjoint-based variational method.This modular platform contains conceptual representations and numerical approximations of dominant hydrological processes while aiming to maintain a relative parsimony. SMASH enables to work at multiple spatio-temporal resolutions, and to compare different modelling operators and structures. All are defined on a regular grid and run continuously. For each time step they take as inputs rainfall and potential evapotranspiration grids, and outputs a field of routed discharges following a drainage grid.

Two model structures are tested in this work, including an upgraded parsimonious one that stems from numerous numerical experiments that consisted in testing structures of increasing complexity. The first structure, denoted S3 in the following, has 3 components per cell: the production store and the transfer store of the GR4J model described in Perrin et al. (2003), and a simple cell to cell routing scheme (linear reservoir) to convey the discharge downstream. This structure is equivalent to those from Jay-Allemand et al. (2020). The second structure, denoted S6 in the following, is described Fig. 1. It is obtained by adding to the first structure the following components: an interception store introduced by Ficchì et al. (2019) in a lumped GR model, a second transfer store coupled with a direct runoff branch, and the water exchange function described in Perrin et al. (2003). This last function enables to simulate losses or gains of water, which can be required in cases of non-conservative catchments (groundwater exchange) and/or data uncertainties.

In this study, both structures are running at a 1-km spatial grid resolution and an hourly time step.

2.2 Mathematical formulation

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, with the only condition that a unique point in Ω has the highest drainage area; the number of active cells within a catchment Ω is denoted N_x.

The hydrological model is a dynamic operator $\mathcal{M}$ mapping observed input fields of rainfall and evapotranspiration $\boldsymbol{P}(x,t')$, $\boldsymbol{E}(x,t')$, $\forall (x,t') \in \Omega \times (0,t)$ onto discharge field $Q(x,t)$ such that:

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

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

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

In the proposed S6 model described Fig. 1, 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 given cell i of coordinates $x \in \Omega$. Their stages are respectively denoted h_i, h_p h_{tr} and h_{tl}. First, part of the rainfall fills the linear interception reservoir determining the remaining throughfall P_{th}. Then, the partition of P_{th} between an infiltration part P_s filling the production reservoir (soil moisture accounting), and an effective rainfall $P_r = P_{th} - P_s$ filling the transfer reservoirs is done with a production operator. The interception and production reservoirs are then emptied respectively from E_i and $E_p - E_i$, with E_p the actual evaporation calculated with a GR evaporation operator. The interception and production reservoirs are then emptied respectively from E_i and E_s. The effective rainfall P_r is divided into 10% of direct runoff $Q1$ and 90% of runoff $Q9$ inflowing transfer part. Then, $Q9$ inflows, with a fixed split (40%; 60%) inspired from GR6J routing part (Pushpalatha, 2013) and corroborated by testing various couples, two non linear reservoirs to represent transfer within a cell. A water exchange term F depending on ml, that is a non conservative operator, is applied to direct runoff component and to one transfer reservoir. The total runoff amount of a cell $Q_t = Q_l + Q_r + Q_d$ is then routed between cells. Given known flow directions $D(x)$, classically obtained from DEM, the cell to cell routing is done with a linear routing reservoir $\mathcal{R}$ of stage h_r whose emptying is parameterized by c_r.

Note that the S3 model reduces to considering only 3 reservoirs for production, transfer and routing as done in Jay-Allemand et al. (2020) yet with slightly different but as parsimonious and comparable routing operator.

We detail in Appendix A the flow operators, mostly consisting in first order ordinary differential equations which are analytically integrated in time, enabling simple computation of the forward model. Note that the forward hydrological model (Eq. (1)) combining those operators is differentiable with explicit dependency to parameters, which is necessary for deriving the numerical adjoint model used in the calibration algorithm.

The state vector of SMASH model structure S6 (resp. S3) is $\boldsymbol{h}(x,t) \equiv (h_i(x,t), h_p(x,t), h_{tr}(x,t), h_r(x,t), h_{tl}(x,t))^T$ (resp. $\boldsymbol{h}(x,t) \equiv (h_p(x,t), h_{tr}(x,t), h_r(x,t))^T$).

The parameter vector of SMASH model structure S6 (resp. S3) is $\boldsymbol{\theta}(x) \equiv (c_i(x), c_p(x), c_{tr}(x), c_r(x), ml(x), c_{tl}(x))^T$ (resp. $\boldsymbol{\theta}(x) \equiv (c_p(x), c_{tr}(x), c_r(x))^T$). 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.

2.3 Variational calibration algorithm

The adjoint-based variational calibration algorithm proposed and detailled in Jay-Allemand et al. (2020), adapted to high dimensional hydrological inverse problems, is used in this study and summarized in the flowchart Fig. 2.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

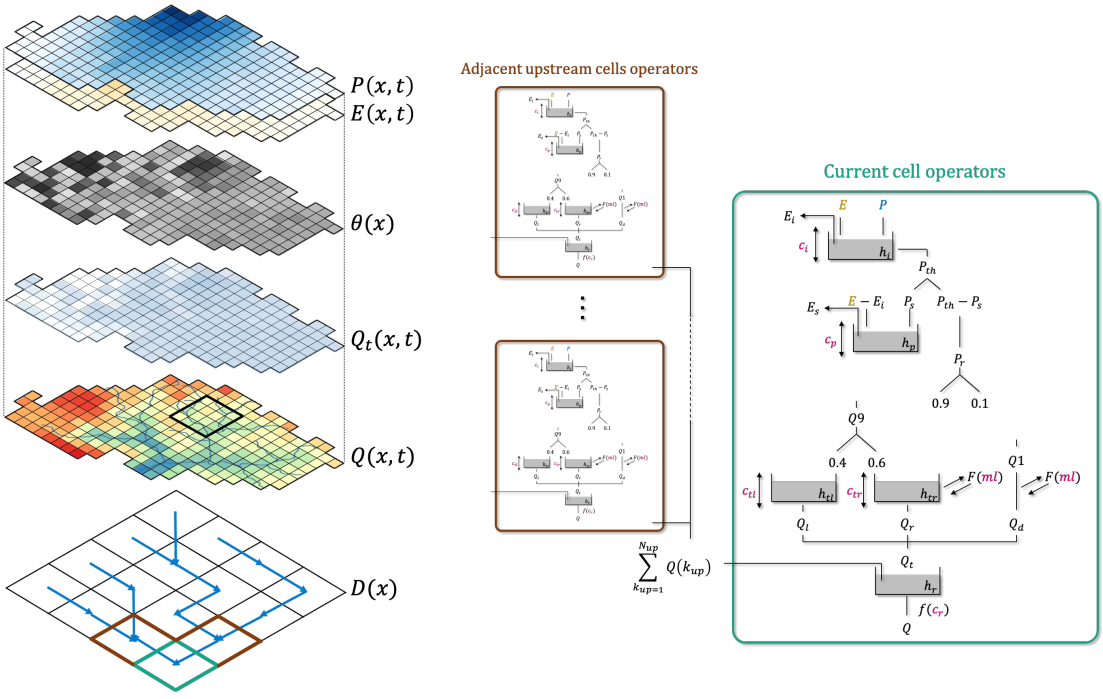

Figure 1. Distributed hydrological modeling with SMASH platform: (left) model fields from top to bottom: meteorological inputs, parameters, internal and output flux maps; (right) pixel scale and pixel-to pixel flow operators of the 6 parameters model S6.

Given simulated and observed discharge at Ng gauged cells $x_k \in \Omega$, $k \in 1,..,Ng$, respectively denoted $Q_k(t)$ and $Q_k^*(t)$ we define the objective function as:

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

where the observation cost function is $J_{obs} = \frac{1}{N_g}\sum_{k=1}^{N_g} j_k^*$. In this study, $N_g = 1$, that is for single gauge optimization, and j_k^* is classically set as $1 - NSE$, with NSE the Nash-Sutcliffe efficiency (this quadratic metric is recalled in Appendix Eq. (A20)), and measures the misfit between simulated and observed discharge. Note that simulated discharge depends on the control vector $\boldsymbol{\theta}$ via the hydrological model (Eq. (1)) such that $Q_k(t) = \mathcal{M}\left[\boldsymbol{P}(x,t'), \boldsymbol{E}(x,t'), \boldsymbol{h}(x,0), \boldsymbol{\theta}(x), t\right], \forall x \in \Omega_k, t' \in [0,t]$. $\Omega_k \subset \Omega$ denotes the spatial domain including all upstream cells of a gauge at x_k. 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 B the background error covariance, and $\boldsymbol{\theta}^*$ the first guess on $\boldsymbol{\theta}$. For the optimizations presented in this study, we set $\alpha = 10^{-4}$ and $\boldsymbol{B}$ is simply defined from $\boldsymbol{\sigma}_\theta$ the vector of mean deviations of θ, as done in Jay-Allemand et al. (2020).

An additional constraint is introduced for parameter optimization such that:

$$\boldsymbol{\theta}_{min} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{max} \tag{3}$$

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Given the hydrological model (Eq. (1)) and the constraint (Eq. (3)), an optimal estimate $\hat{\boldsymbol{\theta}}$ of model parameter set is obtained from the condition:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \tag{4}$$

The minimisation 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, $\nabla_{\boldsymbol{\theta}} J$, that is obtained by solving the adjoint model. This model has been generated with the the automatic differentiation engine TAPENADE (Hascoet and Pascual, 2013) applied to the SMASH source code including the new model structures and validated with standard gradient test. The background value $\boldsymbol{\theta}^*$, used as a starting point for the optmization and in the regularization term, is set as in Jay-Allemand et al. (2020), i.e. as $\bar{\boldsymbol{\theta}}$, a spatially uniform global optimum determined with a simple global-minimization algorithm from a uniform first guess $\bar{\boldsymbol{\theta}}^*$, the steepest descent method summarized in (Edijatno, 1991). Given mildly non linear hydrological models as those considered in this study, this calibration approach is pertinent and the sensitivity to uniform prior is relatively limited as shown in Jay-Allemand et al. (2020).

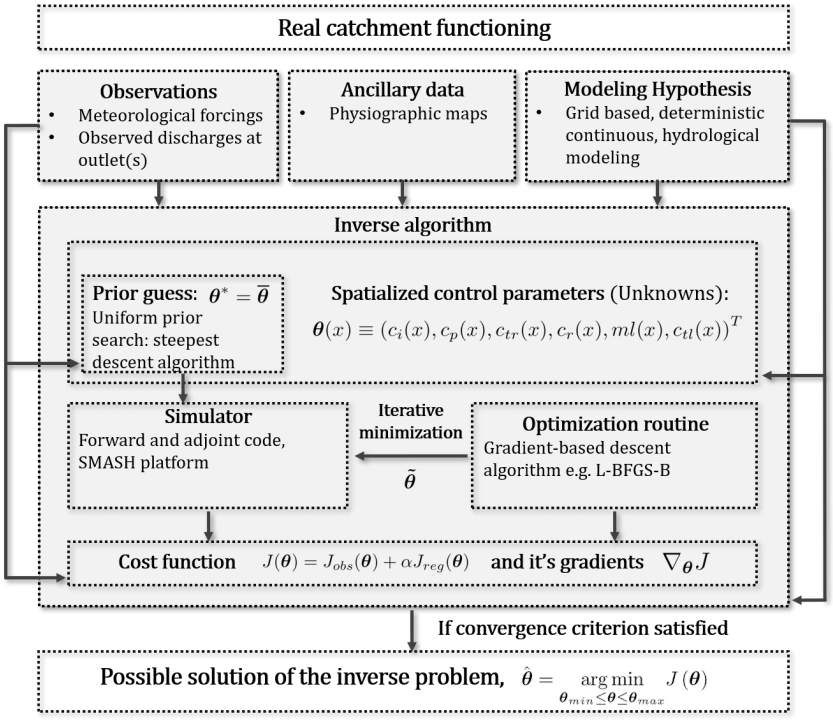

Figure 2. Distributed hydrological calibration with SMASH data assimilation platform: inverse algorithm flowchart. We denote by $\tilde{\boldsymbol{\theta}}$ a parameter vector estimate during the optimization process.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

3 Data and numerical experiments

The following section presents the large sample of 312 "natural" gauged catchments across contiguous France territory and the numerical experiments design.

3.1 Data

3.1.1 Catchment selection and hydroclimatic data

Our catchment sample includes 312 catchments which is a subset from Hashemi et al. (2021), subset of a larger dataset of 4190 French catchments from INRAE-HYCAR research unit (Delaigue et al., 2020; Brigode et al., 2020). Based on a sample of 740 catchments, catchments were removed if (i) regime is Nival and Nival-Pluvial, (ii) or if area is greater than 3000 km^2, (iii) or if more than 20% gaps exist in discharge data on one of the two periods ($p1$, $p2$) defined after.

Concerning the hydrological model input, 14 years of hourly distributed rainfall, PET and locally observed discharges, covering the period 1 August 2006 to 31 July 2020, were used:

- Discharge data are collected by the French Ministry of Environment covering the period of the forcing data and have been extracted from the (Hydro) platform (http://www.hydro.eaufrance.fr/).
- Rainfall: We use rainfall data from the radar observation reanalysis ANTILOPE J+1, which merges radar and insitu gauge observations. This data is provided by Météo-France at a grid resolution of 1 km^2 exactly corresponding to rasters defining model grids.
- Potential evapotranspiration (PET): The interannual temperature data is provided by the SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) reanalysis that is run by Météo France at a resolution of 8×8 [km2] (Quintana-Seguí et al., 2008; Vidal et al., 2010) reanalysis and then used to calculate the potential evapotranspiration using the Oudin formula (Oudin et al., 2005). PET are at the same resolution as the rainfall data. This data is specific to the continuous models (GR5H and SMASH).

3.1.2 Catchment classification

The classification used here is derived from Hashemi et al. (2021) inspired by Blache (1933) and Sauquet (2006) and is based on two interannual monthly variables namely runoff (QM [mm per year]) and total areal precipitation (PM [mm per year]). From these variables, the two classification indices IQ [-] and IP [-] are defined as follows for each catchment. $IQ = (QM_{max} - QM_{min})/\overline{QM}$, with $\overline{QM}$ the temporal average of monthly discharges. $IP = (PM_{max} - PM_{min})/\overline{PM}$, with $\overline{PM}$ the temporal average of monthly areal rainfall. In this definition, the positive IQ and IP indices give information on runoff variability and precipitation variability over the year, respectively. Low values of IQ and IP indicate a uniform distribution of them over the year, while a high value indicates the presence of contrasted dry and humid seasons. A low IQ can also imply the presence of ground water effects or reservoirs (natural or artificial) tending to attenuate runoff fluctuations

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

at catchment outlet. Based on the defined indices (IQ, IP), the following classification criteria are applied to all catchments in the sample to determine their hydrological regime:

- Mediterranean: $IP > 1$
- Uniform: $IP \leq 1$ and $IQ < 1$
- Oceanic: $IP \leq 1$ and $IQ \geq 1$

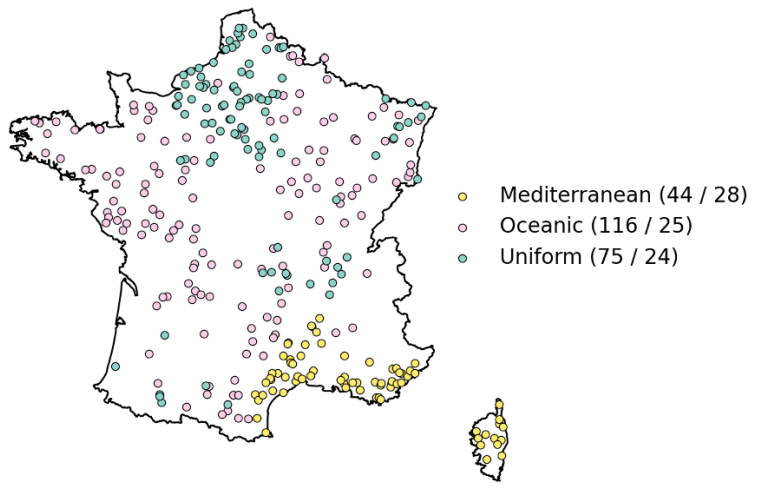

Figure 3. Catchments sample distribution across France. 312 gauges including 235 downstream gauges and 77 upstream gauges. The values in brackets represent the number of catchments in each regime for the downstream gauges on the left and upstream gauges on the right

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Variable	Mediterranean	Oceanic	Uniform
Catchment area (km^2)	406 [609]	480 [603]	442 [561]
Baseflow index (-)	0.42 [0.116]	0.482 [0.117]	0.754 [0.118]
Runoff coefficient (-)	0.468 [0.228]	0.394 [0.283]	0.289 [0.126]
Wetness index (-)	1.422 [0.525]	1.352 [0.293]	1.254 [0.289]
Median elevation (m)	586 [342]	307 [245]	223 [180]

Variable	Mediterranean	Oceanic	Uniform
Catchment area (km^2)	412 [413]	335 [410]	185 [182]
Baseflow index (-)	0.474 [0.106]	0.504 [0.078]	0.764 [0.102]
Runoff coefficient (-)	0.498 [0.25]	0.422 [0.227]	0.342 [0.189]
Wetness index (-)	1.544 [0.525]	1.415 [0.203]	1.389 [0.381]
Median elevation (m)	581 [314]	476 [382]	284 [226]

Table 1. Catchments characteristics for downstream gauges (top) and upstream gauges (bottom). Each value represents the mean with the standard deviation in square brackets.

3.2 Numerical experiments design

In order to test the proposed grid based hydrological model S6, with the variational calibration algorithm, the following experimental design is proposed on the large dataset presented above. The predictive performances of calibrated-models are evaluated on data not used in calibration.

- The models considered are (i) the S3 and S6 structures in SMASH (Section 2.2) with the variational calibration algorithm (Section 2.3) with parameter bounds from Table 2, and (ii) the lumped GR5H model with its calibration algorithm, used as reference given some of its operators are applied at cell scale in SMASH. Recall SMASH S3 is equivalent to the structure in Jay-Allemand et al. (2020) with slightly simplified routing - found not sensitive with the considered model-inversion settings.

- A split sample spatio-temporal validation procedure (Klemeš, 1983) is used to assess the predictive performance of the calibrated models. The time window covered by hydrometric data is split into two complementary subsets over sub-periods of 7 years: $p1$ (from 1 August 2006 to 31 July 2013) and $p2$ (from 1 August 2013 to 31 July 2020) are both used for calibration and validation. For all models, the NSE is used as calibration cost function J_{obs}, not accounting for discharge data over the first year of a calibration period that is left as "warm-up" for a forward model not to be too sensitive to its initial condition $\boldsymbol{h}(x,0)$ (as recommended for GR4 in (Perrin et al., 2003)). Indeed 1 year is supposedly greater than the characteristic relaxation time of the model.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

- Calibrations of the considered S3 and S6 model structures are systematically performed on the 235 downstream gauges with the temporal split sample procedure. In the following, the suffix -U (resp. -D) denotes spatially uniform (resp. distributed) calibration with the algorithm presented in Section 2.3.

- Model performance is also depicted at the scale of flood events, in terms of peak discharge shape, magnitude and timing restitution. This analysis relies on an automatic hydrograph segmentation algorithm (Huynh and Garambois, 2022) applied to simulated and observed discharge time series over the whole dataset.

- A spatio-temporal validation procedure, aiming to assess the capabilities at ungauged locations of downstream calibrated SMASH, consists in applying it at 77 upstream gauges not seen in calibration. Local calibrations are also performed at those upstream gauges to assess the loss of efficiency.

- A global parametric sensitivity analysis over the whole catchment sample is performed to better understand model response surface. A classical regional sensitivity analysis method is used Saltelli et al. (2008).

Parameter name	Parameter description	Lower bound $\boldsymbol{\theta}_{min}$	Upper bound $\boldsymbol{\theta}_{max}$	Uniform first guess $\bar{\boldsymbol{\theta}}^*$
c_i	Interception parameter (mm)	1	100	1
c_p	Production parameter (mm)	1	2000	200
c_{tr}	First transfer parameter (mm)	1	1000	100
c_{tl}	Second transfer parameter (mm)	1	10000	100
ml	Water exchange parameter (mm/h)	-20	5	0
c_r	Linear routing parameter (min)	1	200	5

Table 2. Parameters, ranges and uniform first guess used for calibration and sensitivity analysis for the model SMASH.

4 Results and discussion

4.1 Downstream gauges: global performances

Our first goal was to assess how well a parsimonious distributed model structure performs across France when assessed over annual timescales via standard performance metrics. The second goal was to study the benefits of performing distributed calibration with the variational algorithm. The distribution of model performance obtained with the split sample procedure across all 235 downstream catchments, in terms of calibration and validation, is presented for period $p2$ in Fig. 4 and for period $p1$ in Fig. B1.

First, concerning spatially uniform calibrations, the proposed model S6-U clearly outperforms the simple S3-U configuration with median NSE of 0.85 (resp. 0.81) compared to 0.73 (resp. 0.67) in calibration on period $p2$ (resp. $p1$) as shown on the left panel of Fig. 4 (resp. B1). In temporal validation, S6-U also outperforms S3-U with median NSE of 0.78 (resp. 0.76)

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

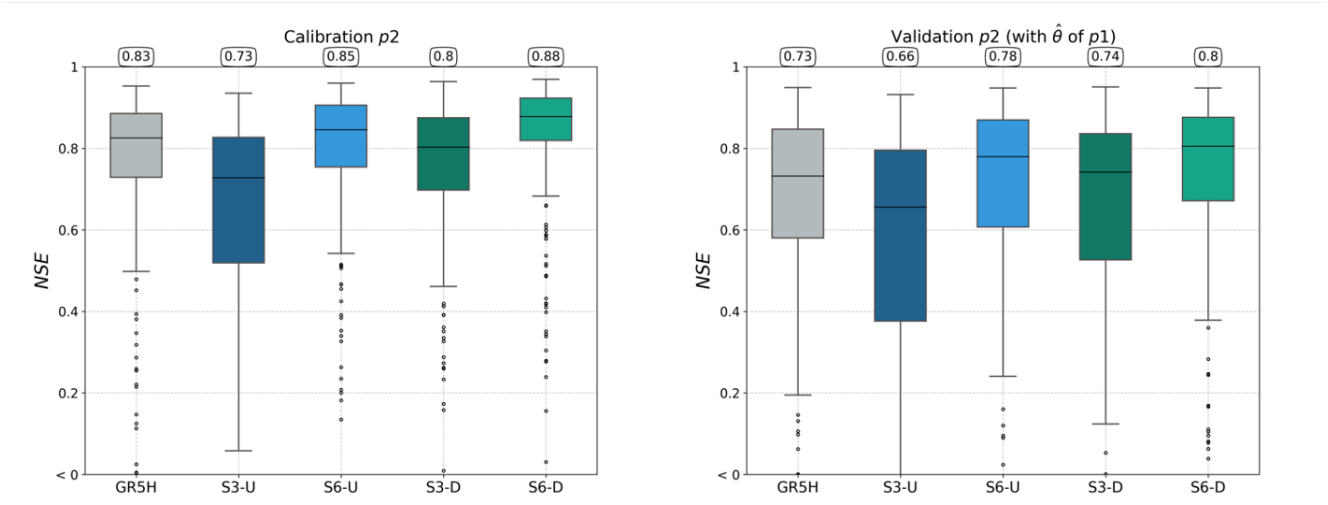

Figure 4. Performances comparison between the two SMASH distributed model structures (S3 and S6) for each calibration settings (U and D). GR5H performances are plotted as a reference. Calibration and validation on $p2$ (see Fig. B1 for performances on $p1$).

compared to 0.66 (resp. 0.63) in temporal validation on period $p2$ (resp. $p1$) as shown on the right panel of Fig. 4 (resp. B1). This shows a clear improvement of performances over a large sample obtained by increasing the model complexity in SMASH, from the 3 parameters structure of Jay-Allemand et al. (2020) to the present one with 6 parameters, with good robustness in time. Note the slightly better performances of S6-U compared to the 5 parameters GR5H model only shown here as a reference - benchmark is not the scope of this work.

Next, concerning spatially distributed calibration, similar trends as in uniform calibration are observed but with higher performances. The proposed model S6-D clearly outperforms the simple S3-D configuration with median NSE of 0.88 (resp. 0.85) compared to 0.80 (resp. 0.77) in calibration on period $p2$ (resp. $p1$) as shown on the left panel of Fig. 4 (resp. B1). In temporal validation, S6-D also outperforms S3-D with median NSE of 0.8 (resp. 0.79) compared to 0.74 (resp. 0.7) in temporal validation on period $p2$ (resp. $p1$) as shown on the right panel of Fig. 4 (resp. B1). A relatively narrower interquantile range is obtained with S6-D both in calibration and validation, with better performances than the 5 parameters GR5H model only shown here as a reference - recall benchmark is not the scope of this work. This will be further discussed after.

Those good NSE performances show an improved capability of the proposed 6 parameters SMASH structure, which was expected with this increased model complexity compared to a 3 parameters model, to fit observed streamflows in calibration and a good temporal robustness on this dataset. Using distributed calibration increases the model performances in terms of NSE for calibration period and even for validation period, with higher performances and temporal robustness for S6-D. The rest of the paper therefore focuses on the analysis of the SMASH structure S6.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Our NSE in validation $p2$ (S6-U and S6-D) are represented on a map on Fig. 5 which points low performances especially localized in northern basins. These basins are mainly characterized as Uniform (Fig. 3), that is a hydrological regime mainly influenced by groundwater flows (high BFI value in Table 1). This type of catchment, even by considering a non-conservative exchange function F (Perrin et al., 2003) in the model structure, is not properly simulated.

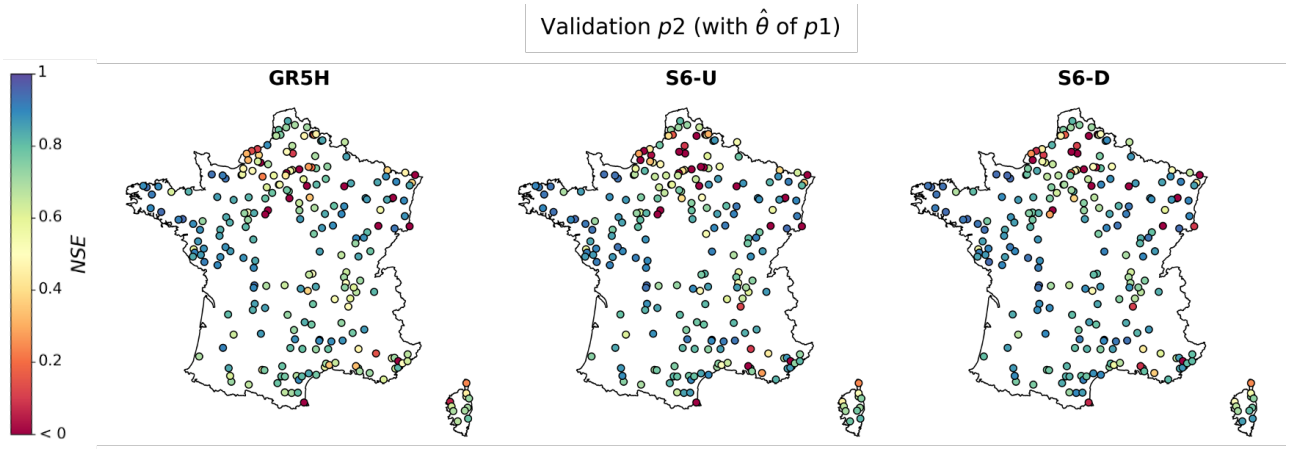

Figure 5. Performances maps in temporal validation on period $p2$ (see in Fig. B2 for $p1$) in terms of NSE criterion for (left) lumped GR5H, spatially distributed 6 parameters SMASH model in (middle) spatially uniform S6-U and (right) distributed S6-D calibration.

To deepen the analysis by catchment regime, the model performance are represented in function of the regime (Section 3.1.2) in Fig. 6 for calibration and validation on $p2$ (Fig. B3 for $p1$). Similar trends as before in terms of performances and robustness on the whole sample are obtained for each of the three hydrological regimes. The 6 parameters model with spatially distributed calibration S6-D is the best, compared to uniform calibration S6-U, both in calibration and validation for the three groups. The best performances with S6-D, in calibration and validation on $p2$ Fig. 6 (resp. $p1$ Fig. B3), are obtained for Oceanic basins: median NSE of 0.92 and 0.87 (resp. 0.89 and 0.84). High performances are also obtained on Mediterranean basins: median NSE of 0.89 and 0.78 (resp. 0.82 and 0.68). Lower performances are obtained on Uniform basins as concluded on the maps above Fig. 5 with median NSE of 0.79 and 0.67 (resp. 0.76 and 0.57).

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

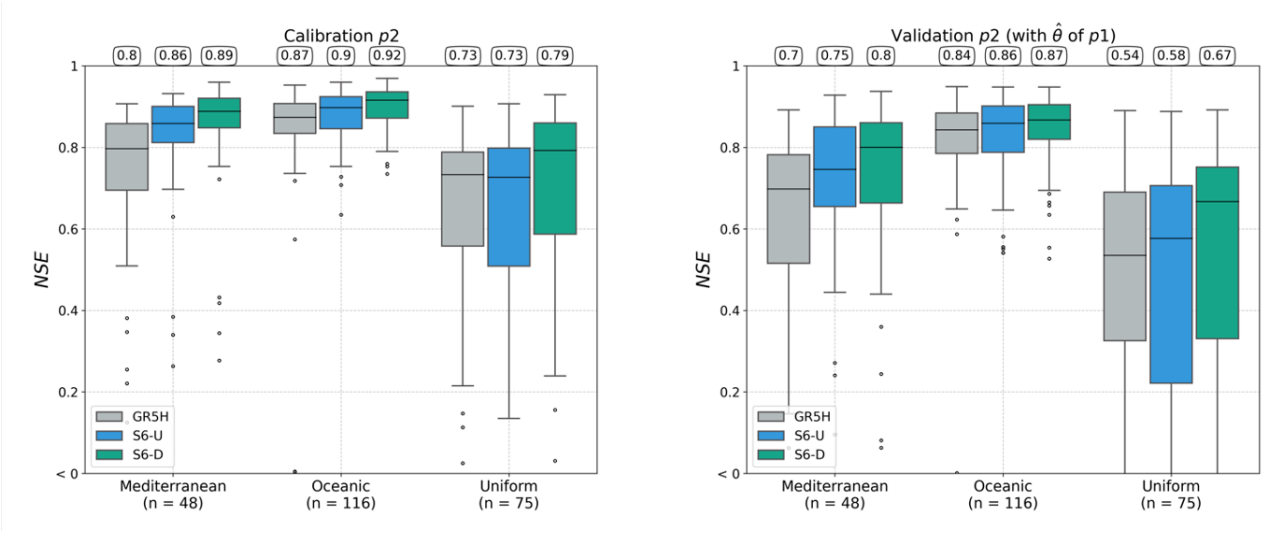

Figure 6. Performances comparison between the two SMASH distributed model structures (S3 and S6) for each calibration settings (U and D) and catchment regime (Mediterranean, Oceanic and Uniform). GR5H performances are plotted as a reference. Calibration and validation on $p2$ (see Fig. B3 for performances on $p1$)

The difference of hydrological information, contained in observed signal for each downstream gauge, and learnt in calibration, is briefly analyzed in Fig. 7. It shows the difference between $p1$ and $p2$ of runoff coefficients (Eq. (A22)) against the wetness index (Eq. (A23)), computed for each period and calibration gauge. The most important spread for each group is on the wetness index difference between periods. The highest spread on runoff coefficient is for Mediterranean basins, which can be apportioned to more inter period variability especially in relation to strong flood occurrence on such basins. Therefore, in addition to input and modeling errors, this disparity of learnt hydrological signals can partly explain the slight variability of model performance in split sample, for all regimes and in particular Mediterranean regime. This is corroborated by the reference GR5H lumped model, well-known to be robust, but affected the same way in split-sample test on the present dataset.

4.2 Downstream gauges: flood signatures

In complement to the previous analysis of model performance on multi-years time periods, a detailed analysis at the scale of flood events is presented to provide an assessment of model performance and signatures at shorter time scale and in a flood forecasting perspective.

For each catchment $b \in [1, N_B]$, are extracted $N_{E,b}$ flood events by applying an automatic segmentation algorithm (Huynh and Garambois, 2022) to discharge time series at downstream gages. Considering peaks that exceed the 0.999-quantile of the sorted hourly discharge values, starting dates are determined based on a "rainfall gradient criterion" and "rainfall energy criterion", ending dates based on discharge base flow obtained by a classical baseflow separation method (see details Huynh and Garambois, 2022). The segmentation applied on period $p2$ (resp. $p1$) to time series of 235 downstream gauges results in a

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

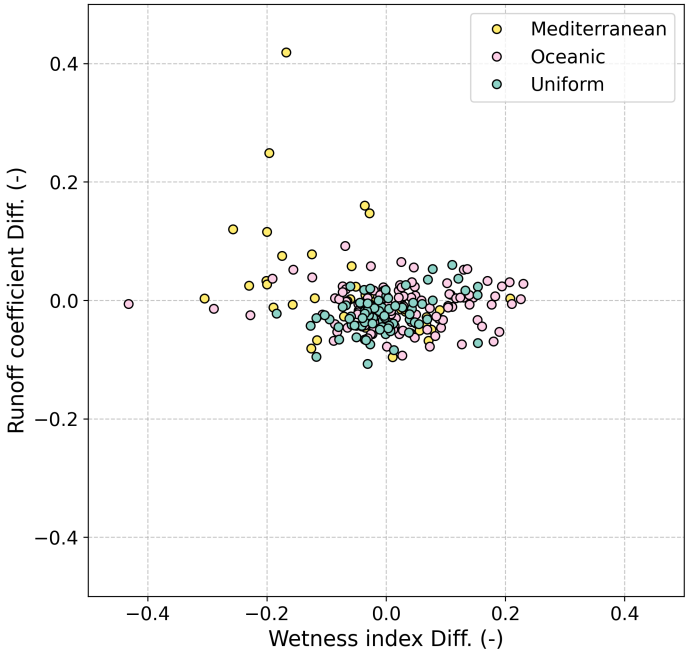

Figure 7. Representing for the 235 downstream gauges the difference between $p1$ and $p2$ of the wetness index and the runoff coefficient for each catchment regime.

total sample of respectively 1443 (resp. 1522) flood events as summarized in Table 3. Flood events repartition between $p1$ and $p2$ is rather balanced, as well as between hydrological groups.

Variable	Total	Mediterranean	Oceanic	Uniform
Number of events	1443 (1522)	363 (356)	572 (613)	508 (553)
Median specific discharge (mm/h)	0.625 (0.513)	2.982 (2.138)	0.660 (0.564)	0.160 (0.134)

Table 3. Flood events characteristics in terms of number of events and median specific discharge for $p2$ ($p1$ between brackets).

For each flood events of the list $\mathcal{F}_{E,b} = \{e \in [1, N_{E,b}]\}$ obtained with the segmentation algorithm for a given catchment $b \in [1, N_B]$ the following metrics are computed:

- The classical Nash-Sutcliff NSE and Kling-Gupta KGE metrics (Eq. (A20), (A21))
- The relative error on the peak discharge $RPD = \frac{Q_{\max}}{Q^*_{\max}} - 1$, where $Q_{\max} = Q(t_{\max})$ and $Q^*_{\max} = Q^*(t^*_{\max})$ are the predicted and the observed maximum discharges, respectively. These maximum values are achieved at different time instants $t_{\max}$ and $t^*_{\max}$, within the chosen time period.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

– The synchronous relative error on the peak discharge $SRPD = \frac{Q_{\max}(t^*_{\max})}{Q^*_{\max}} - 1$

Those metrics are computed on temporal validation runs for the 1443 (resp. 1522) flood events of period $p2$ (resp. $p1$) at the 235 downstream gauges and plotted in Fig. 8 (resp. Fig. B4). First, the median flood (NSE; KGE) over all events and catchment groups is of (0.55; 0.54) (resp. (0.41; 0.46)) for uniform calibration S6-U and is better with a (0.63; 0.59) (resp. (0.55, 0.53)) for spatially distributed calibration S6-D which is also true for floods in each catchment groups. Best performances are obtained with S6-D for Oceanic (0.76; 0.71) (resp. (0.67; 0.62)) and Mediterranean basins (0.71; 0.63) (resp. (0.65; 0.53)), while lowest performances (0.31; 0.43) (resp. (0.31; 0.45)) are again for Uniform catchment where modeling but also events segmentation is the most difficult. Regarding peak discharge estimation for S6-D, the median relative error on peak flows RPD is of -0.23 (resp. -0.3), that is a slight global underestimation of Q_p in validation, flood timing is quite accurate as shown by $SRPD$ compared to RPD bias. Note that this slight bais is found both in calibration and validation over each time period which could be improved via a calibration strategy giving more weight to flood events.

4.3 Model performances at upstream un-calibrated gauges

An important feature sought in distributed hydrological modeling is consistent spatialized modeling of discharge, especially at ungauged locations. Un-shown results indicates that the 6 parameters structure clearly improves the results of the 3 parameters structure also at upstream gauges, when calibration is done only on the downstream station, confirming the robustness of the new structure. For the sake of clarity the results at upstream gauges are detailed only for the 6 parameters structure.

Spatial validation of SMASH S6 is shown for the 77 upstream gauges not used in calibration in Fig. 9 for period $p2$ (resp. Fig. B5 for period $p1$). The spatial validations ('sp v', in solid color) on calibration period (left panel) results for spatially distributed and uniform calibrations, in median NSE respectively of 0.73 and 0.72 (resp. 0.67 and 0.62). This represents good performances compared to local calibrations ('c', in hatched color) performed at those upstream gauges just for comparison: decrease on median NSE of 0.16 (resp. 0.17) in distributed calibration (S6-D 'c' to 'sp v') and 0.14 (resp. 0.16) in uniform calibration (S6-U 'c' to 'sp v'). One can remark that part of these upstream information has been seen in calibration since it is performed on downstream gauges.

The right panel of Fig. 9 (resp. Fig. B5 shows the even more challenging spatio-temporal validation ('sp-tp v', in solid color) that is validation on upstream gauges and on validation period, that is totally unseen data. Spatially distributed and uniform parameter sets in spatio-temporal validation gives median NSE respectively of 0.7 and 0.62 (resp. 0.62 and 0.6). This represents good performances compared to temporal validations of local calibrations ('tp v', in hatched color) performed at those upstream gauges just for comparison: decrease on median NSE of 0.1 (resp. 0.15) in distributed calibration (S6-D 'sp-tp v' to 'tp v') and 0.17 (resp. 0.14) (S6-U 'sp-tp v' to 'tp v') in uniform calibration.

Remark that the performance decrease is on the same order in temporal or spatial validation which might stem from the relatively short calibration periods (7 years) considered in split sample.

This spatial and spatio-temporal validations on upstream gauges are depicted for the three catchment groups in function of hydrological regime in Fig. 10 for $p2$ (resp. Fig. B6). Similar trends as above are found for spatial 'sp v' and spatio-temporal

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

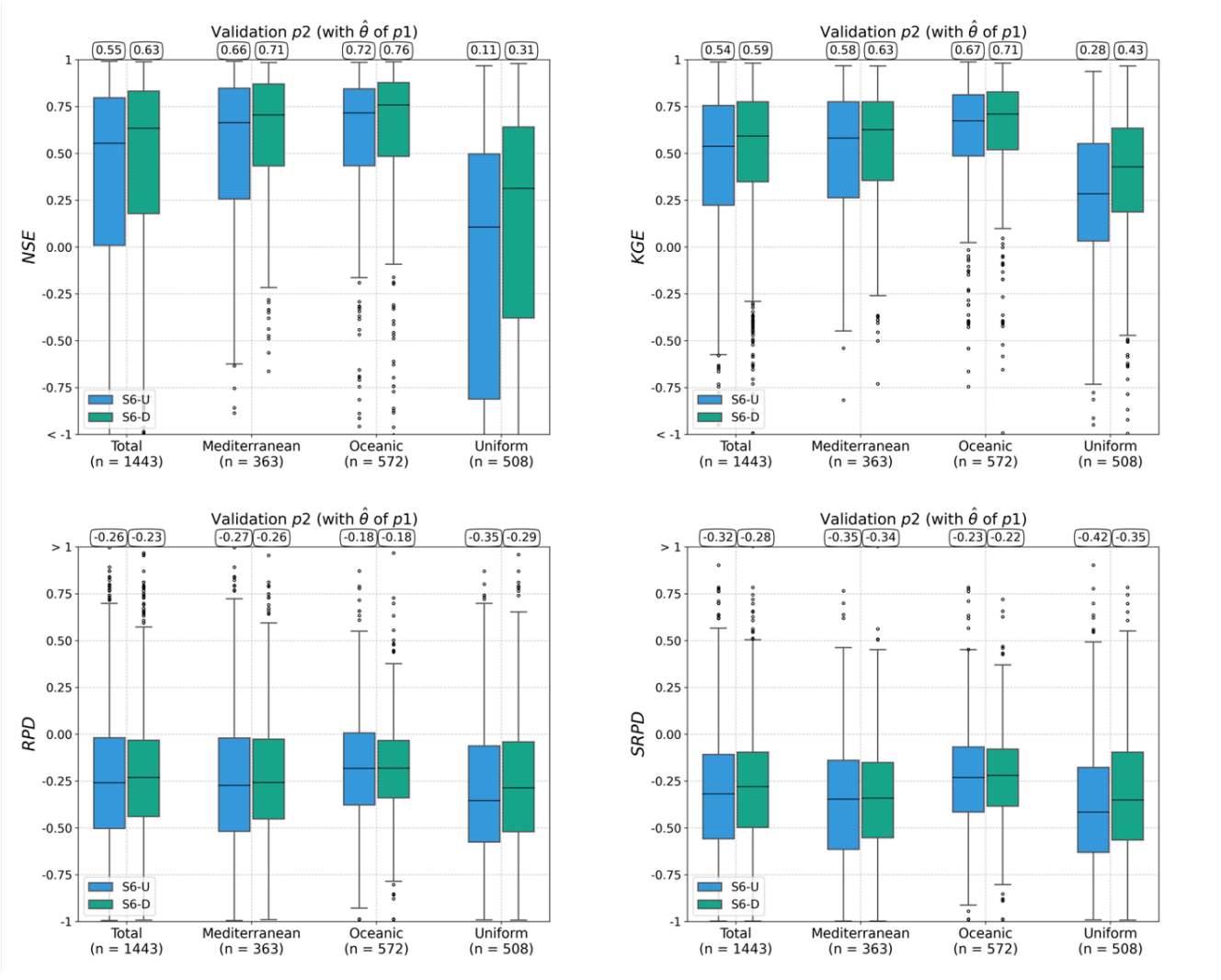

Figure 8. Boxplot of flood event performances and signatures on 235 downstream gauges on validation period $p2$, i.e. 1443 flood events. (See period $p1$ in Fig B4).

'sp-st v' validations, with again the best performances for Oceanic followed by Mediterranean catchments. Unsurprisingly, lower performances are obtained on Uniform catchments. The introduction of spatial regularizations could improve the spatial constrain on the sought parameter fields and hopefully model performance in regionalization which is a research subject in itself that is left for further work. Nevertheless a rigorous analysis of calibrated parameter fields is provided after.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

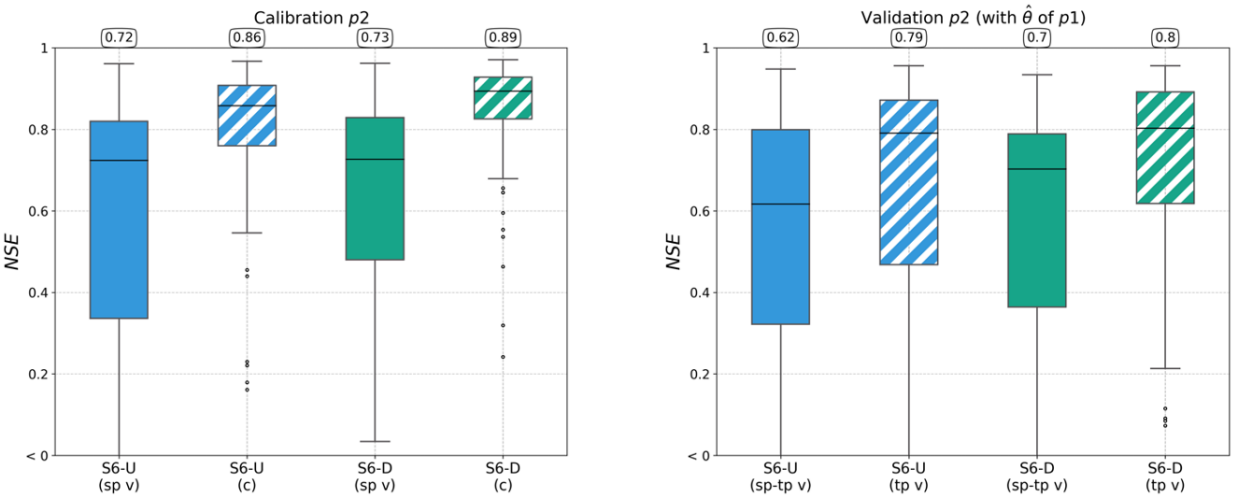

Figure 9. Boxplot of NSE performances on 77 upstream gauges on period $p2$. Spatial validation 'sp v' (left) and spatio-temporal validation 'sp-st v' (right) ; label 'c' and hatched boxes stand for local calibrations performed at those upstream gauges for sake of comparison. (see Fig. B5 in appendix for $p1$).

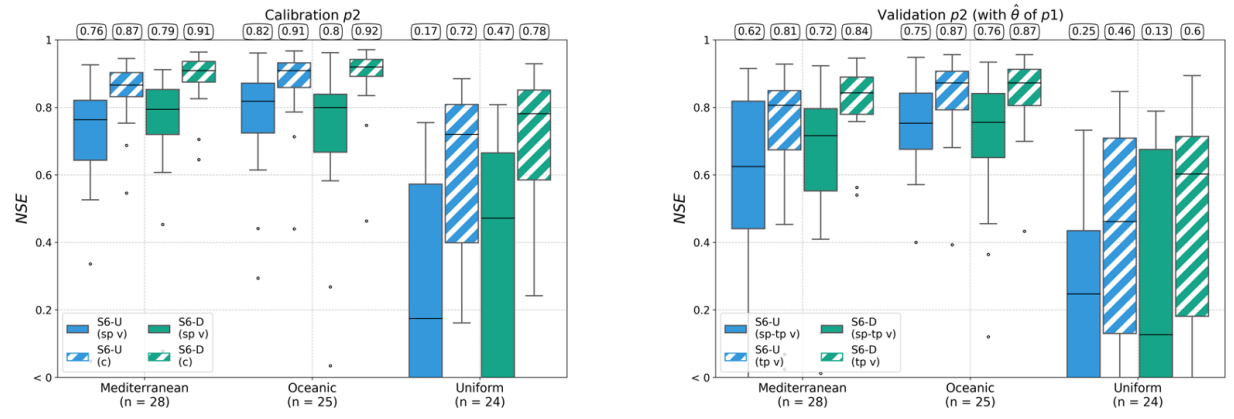

Figure 10. Boxplot of NSE performances on 77 upstream gauges on period $p2$. Spatial validation (left) and spatio-temporal validation (right). (see Fig. B6 in appendix for $p1$).

4.4 Model parameters analysis

The parameter sets calibrated for SMASH S6, over the large sample of French catchments, are summarized for spatially uniform setup S6-U in terms of average and standard deviation over the dataset in Table 4 and for spatially distributed calibration setup S6-D in Table 5

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

First concerning spatially uniform calibrations S6-U, it is interesting to see similar trends, in terms of parameter means and standard deviations for each regime and between regimes over the two periods. For each regime, parameter values are rather stable between the two periods with higher differences for Mediterranean catchments, especially for second transfer parameter c_{tl} and water exchange parameter ml; a lower negative ml being obtained on $p1$. This can arise from (i) hydrometric data errors (both forcings and discharge, especially for high intensity events), (ii) modeling errors, (iii) hydrological information contained in calibration data (e.g. analysis in last section Section 4.1 and scatter plot in Fig. 7).

The variability of calibrated parameters over the dataset has some "physical" coherence: Median routing parameter c_r is smallest, that is more rapid routing, on fastly responding Mediterranean catchments; median second (slow) transfer parameter c_{tl} and its dispersion are highest on Uniform catchments. For those Uniform catchments, highest mean values and dispersion of reservoir capacities c_p, c_{tr}, c_{tl} and routing parameter c_r (whose high value induces a slower emptying of the reservoir) are found for both periods in calibration. Reaching such values and dispersion may testify of the relative difficulty that a parsimonious hydrological model structure has to reproduce such hydrological behaviours including complex multi-frequency and non linear hydrogeological processes (Schuite et al., 2019, and references therein).

Higher production capacity c_p with higher dispersion are obtained on Mediterranean and Uniform regimes. Again, the difference of hydrological information between calibration periods pointed out in Section 4.1 and in Fig. 7 can partly explain the differences obtained in calibrated parameter values, especially on the exchange parameter ml on Mediterranean catchments. This could also be linked to the complexity of physical processes integrated in catchment responses. Indeed, even for high frequency processes such as floods, few studies showed some relation of drainage-storage capacity with bedrock type as in Vannier et al. (2014); Garambois et al. (2015) in the context of distributed modeling of French Mediterranean catchments and more recently in Le Mesnil et al. (2021) focusing on flood signatures in several French karstic areas.

Regarding spatially distributed calibration S6-D, spatial averages of each calibrated parameter field is computed for each catchment and period, next statistics are computed over the dataset for each parameter and presented in Table 5. Note that these simple spatial statistics are computed on catchments of various surfaces. Analogously to uniform calibrations one can observe similar trends over the catchment set, in terms of calibrated parameter means and standard deviations, between the two periods, per regime and between regimes. In terms of within catchment spatial variability, interestingly, the lower ones are those of the routing parameter c_r and of the interception capacity c_i. This indicates those parameters are relatively well constrained in the current model-optimization setup, which corroborates this finding for routing parameter with the present optimization algorithm and a 3 parameters version on one catchment in Jay-Allemand et al. (2020) (see also therein an analysis of sensitivity to first guess). For the other four parameters, controlling runoff production and transfer within each pixel, higher but relatively moderate spatial variability within catchments are found, which indicates there is room for progress in terms of spatial constrains. Again, the introduction of spatially distributed information and spatial regularizations on the sought parameter fields is a research subject in itself that is left for further work. Nevertheless, recall that with the present setup, high performances are reached in calibration with good robustness in temporal and spatio-temporal validations as found and analyzed above on each period.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Parameter	Mediterranean	Oceanic	Uniform
c_i	24 [29]	21 [19]	10 [20]
c_p	345 [513]	200 [137]	828 [748]
c_{tr}	325 [352]	309 [334]	637 [428]
c_{tl}	715 [1809]	412 [1189]	4350 [4240]
c_r	39 [17]	95 [44]	117 [59]
ml	-1.949 [3.989]	-0.669 [2.52]	-3.472 [6.448]

Parameter	Mediterranean	Oceanic	Uniform
c_i	15 [24]	16 [15]	8 [18]
c_p	401 [625]	170 [110]	607 [650]
c_{tr}	281 [341]	340 [357]	484 [437]
c_{tl}	1208 [2716]	682 [1969]	4479 [3760]
c_r	51 [48]	93 [41]	114 [59]
ml	-4.510 [7.597]	-0.433 [2.184]	-3.920 [7.364]

Table 4. SU-3.6 mean and standard deviation between square brackets over the 235 downstream calibration catchments of spatially uniform parameter sets $\hat{\hat{\boldsymbol{\theta}}}$, for $p2$ (left) and $p1$ (right).

Parameter	Mediterranean	Oceanic	Uniform
$\overline{c_i}$	26 [28]	24 [18]	15 [20]
σc_i	3 [5]	4 [5]	7 [9]
$\overline{c_p}$	367 [515]	204 [132]	834 [722]
σc_p	95 [111]	59 [44]	116 [133]
$\overline{c_{tr}}$	330 [334]	311 [322]	655 [407]
σc_{tr}	54 [54]	35 [38]	43 [75]
$\overline{c_{tl}}$	748 [1697]	440 [1140]	4476 [3879]
σc_{tl}	282 [259]	220 [243]	619 [892]
$\overline{c_r}$	40 [18]	96 [44]	115 [54]
σc_r	9 [9]	7 [7]	16 [14]
$\overline{ml}$	-1.877 [3.953]	-0.631 [2.508]	-3.709 [6.318]
σ_{ml}	0.145 [0.215]	0.117 [0.137]	0.765 [1.347]

Parameter	Mediterranean	Oceanic	Uniform
$\overline{c_i}$	17 [24]	19 [15]	14 [19]
σc_i	3 [4]	4 [5]	8 [9]
$\overline{c_p}$	414 [604]	182 [103]	638 [633]
σc_p	86 [62]	64 [48]	111 [129]
$\overline{c_{tr}}$	305 [314]	342 [351]	511 [420]
σc_{tr}	69 [75]	27 [27]	42 [59]
$\overline{c_{tl}}$	1403 [2630]	699 [1832]	4387 [3487]
σc_{tl}	600 [897]	260 [327]	461 [585]
$\overline{c_r}$	52 [46]	94 [41]	112 [55]
σc_r	12 [11]	8 [8]	15 [14]
$\overline{ml}$	-4.361 [7.646]	-0.371 [2.181]	-4.133
σ_{ml}	0.200 [0.340]	0.118 [0.142]	0.688 [1.286]

Table 5. Analysis of spatially distributed parameter sets of S6 calibration by catchment: first, spatial average and standard deviation for each parameter field are calculated for each catchments, then their mean and standard deviation between brackets over the 235 downstream calibration catchments of areal statistics of spatially distributed parameter sets $\hat{\hat{\boldsymbol{\theta}}}$, $p2$ (left) and $p1$ (right)

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

4.5 Parameters sensitivity

The parametric sensitivity of model variant S6 is studied here based on a regional sensitivity analysis (RSA) method applied on each catchment (for details see Saltelli et al., 2008; Song et al., 2015, and the references therein).

RSA was performed for each catchment as follows. We considered 10000 spatially uniform sets of the 6 model parameters, randomly sampled using Sobol quasi-random sequences within calibration ranges (Table 2). Each parameter set is used to perform a discharge simulation with the forward model and compute the resulting NSE value on the 235 downstream gauges, over period $p2$ only since it provides sufficient information for a summary parametric sensitivity analysis over this large sample. For each basin, a threshold NSE is chosen such as obtaining a partition of model simulations into: 10% best ones called "behavioural" and 90% non behavioral. Next, the magnitude of the Kolmogorov-Smirnoff (KS) statistics D, representing the maximum difference between the cumulative density functions (CDF) of the two classes is used to rank the parameters based on their sensitivity. As noted by Beven (2011), the KS test can be very sensitive to small differences, and will thus report significant differences between the two classes.

The parametric sensitivities of SMASH model variant S6 are depicted by catchment group, in terms of density plot in Fig. 11, and in terms of rankings of parameters sensitivities in Table 6 and clearly show:

- For all regimes, the least sensitive parameters are in order the interception c_i and next transfer c_{tl}, c_{tr} parameters. Despite this low sensitivity to interception, this component is found to critically improve modeling performances. That is why interception parameter could be fixed a priori which would potentially reduce the complexity of calibration problem.

- For Uniform and Oceanic catchments, the most sensitive parameters are the non conservative water exchange parameter ml followed by the routing c_r or production c_p parameters respectively.

- For fast responding Mediterranean catchments, the routing parameter c_r is the most sensitive followed by the non conservative water exchange parameter ml. Note that the bimodal shape of density function of c_p on Mediterranean catchments may indicate one group very sensitive to this parameter while another one is less sensitive to this parameter. Besides, the spatial distribution of sensitivities to c_p on Mediterranean catchments looks correlated to soil and bedrock types (maps not shown), which has been highlighted in previous analysis in similar context (e.g. Vannier et al., 2014; Garambois et al., 2015).

Recall that such a simple sensitivity analysis is itself sensitive to parameter ranges, sampling method and performance threshold definition. Nevertheless, the highest sensitivities highlight that modeling efforts could focus on improving the representation of runoff production and the exchanges function.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

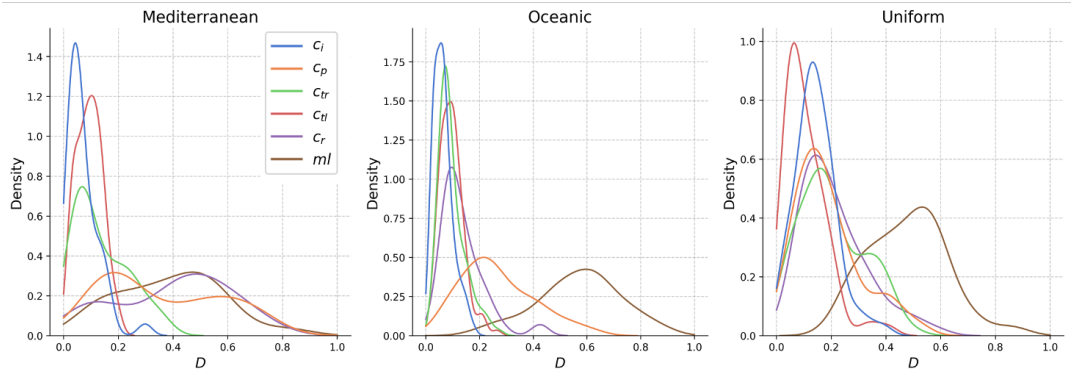

Figure 11. Parametric sensitivity of SMASH model variant S6 over period $p2$ with spatially uniform parameters sampled in previous calibration ranges, 10 000 parameter sets by catchment and NSE threshold determined by catchment such as partinoning simulations into 10 % and 90 % behavioural and non behavioural simulations denoted B and $\bar{B}$. Density of Kolmogorov-Smirnoff sensitivity (D) measure between the cumulative distribution curves of B and $\bar{B}$ of each catchment, over catchment groups.

Regime	c_i	c_p	c_{tr}	c_{tl}	c_r	ml
Mediterranean	6	3	4	5	1	2
Oceanic	6	2	5	4	3	1
Uniform	5	4	3	6	2	1

Table 6. Parameters sensitivity ranking for each catchment group for SMASH model variant S6 over period $p2$. For each parameters sensitivity distribution the rank has been calculated by sorting in ascending order the distribution mean value. Rank 1 is the most sensitive parameter and rank 6 the less in terms of distribution mean value.

5 Conclusions

This article has presented the first evaluation over a large sample of Variational Data Assimilation (VDA) applied to the spatially distributed calibration of a newly taylored parsimonious hydrological model. This improved parsimonious grid GR-based model has been implemented in SMASH platform and its adjoint generated as requeted by VDA algorithm. The spatially distributed calibration of this single model structure over a large sample of French catchments spanning different hydrological regimes results in high performances. From the obtained results and analysis at annual, seasonal and floods timescales and in spatio-temporal validation, as well as from parametric sensitivity analysis, the following conclusions can be raised:

- The complexified model structure with 6 parameters, including one non conservative component, leads to higher performances, both in calibration and spatio-temporal validation over this large sample, compared to the 3 parameters model structure and to the reference GR5H lumped model.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

- Spatially distributed calibration, despite inverse problem overparameterization, leads to a significant gain in terms of performances in calibration and temporal validation on this large sample with median efficiencies respectively of NSE = 0.88 (resp. 0.85) and NSE = 0.8 (resp. 0.79) over the total time window on period $p2$ (resp. $p1$).

- Simulated signatures over 1443 (resp. 1522) flood events on period $p2$ (resp. $p1$) are also quite good with median flood (NSE; KGE) of (0.63; 0.59) (resp. (0.55; 0.53)), peak flow restitution RPD = -0.23 (resp. -0.3) and timing $SRPD$ = -0.28 (resp. -0.35). This slight bias over flood peaks could be reduced by giving more weight to them via a dedicated event metric in addition to the global one over the relatively long calibration time window.

- Best modeling performances are obtained for Oceanic and Mediterranean basins with this 6 parameters model structure whereas it performs less well over Uniform basins where influence of groundwater processes can be significant - multi-frequency hydrogeological processes with stream aquifer coupling (e.g. Schuite et al., 2019; Flipo et al., 2014, 2022, and references therein).

- Those results obtained on temporal validation at downstream gauges are also confirmed in spatio-temporal validation, using un-calibrated upstream gauges.

- Finally, a sensitivity analysis revealed that the non conservative water exchange parameter and the production parameter to be the most sensitive parameters, along with the routing parameter especially for faster responding catchments. Hence, modeling efforts could focus on improving the representation of runoff production and water exchanges and ultimately model versatility. Interception is the least sensitive while improving significantly model performance which indicates it should be fixed.

Those encouraging performances obtained for a wide range of catchment and hydrological conditions over the french territory, with the proposed distributed model in the variational framework, represents a solid base for further research. Immediate to mid-term work perspectives are as follows.

- Forward model improvements: improve and test multiple hydrological model structures, in state-space formulation enabling adaptative time stepping and flexible model structure definition in the variational framework (with possibility to combine multiple operators within a catchment and enhance model versatility for various flow conditions and sub-catchments functionning).

- High dimensional inverse algorithm improvements: (i) multicriteria variational calibration strategy adapted to floods or low flows; (ii) descriptor upscaling and mapping into model parameter space for regionalization (Samaniego et al., 2010), along with taylored regularizations ; (iii) multi site and multi-source data constrains with "relaxation strategies"; (iv) global-local optimization strategies with high dimensional distributed controls in equifinality context.

- Coupling to hydraulic models of the DassFlow variational data assimilation platform (Monnier et al., 2016) also recently wrapped in python, adapted to use satellite datasets either with 1D network model (Brisset et al., 2018; Larnier

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

et al., 2020; Pujol et al., 2020; Malou et al., 2021) or a multi-D model with fine 2D zooms (Pujol et al., 2022) with regularizations taylored for hydrological-hydraulic inverse problems.

- Massive use of multi-source datasets including regional and worldwide datasets gained from satellite remote sensing.

Note that SMASH platform has recently been interfaced in python (Jay-Allemand et al., 2022) enabling to use powerful libraries such as for signal processing and machine learning for building hybrid deterministic-ML methods in the powerful VDA framework.

Author contributions. research plan: PAG, FC, PJ, PA. Code upgrade and data curation: FC and PAG. Runs and plots: FC. Manuscript preparation: PAG and FC. Analysis and final redaction: all. Funding management PA.

Acknowledgements. The authors greatly aknowledge (i) Etienne Leblois from INRAE Riverly (Lyon) lab for fine terrain elevation processing at multiple scales over the French territory, (ii) Ngo Nghi Truyen Huynh who implemented the segmentation algorithm used in this work, (iii) and also INRAE Hycar (Paris) lab, SCHAPI and METEO France for providing data used in this study. First author was partly funded by the Hydrodemo and EXPLORE2 projects.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Appendix A: SMASH Model operators

The state equations of SMASH models used in this study and summarized in section 2.2 are provided here. We use GR operators for interception (Ficchì et al., 2019), production (Edijatno and Michel, 1989; Edijatno, 1991) and transfer (Perrin et al., 2003). See also in Astagneau et al. (2021a) and references therein, details on the lumped model GR5H-I (Ficchì et al., 2019) that is used in the present study via airGR package (Coron et al., 2017, 2022). The operators of the computational model are obtained from analytical integration in time of state-space equations consisting mostly in first order ordinary differential equations (see details in (Perrin et al., 2003) and (Jay-Allemand et al., 2020)). Finally both SMASH model structures in this study include conceptual GR-based operators for runoff production and water exchange at pixel scale (Edijatno and Michel, 1989; Perrin et al., 2003; Ficchì et al., 2019) on top of a within cell transfer part and a cell to cell routing scheme.

For a given cell $x \in \Omega$ and time step $t > 0$, fluxes and states are computed as folows, with $P(t)$ and $E(t)$ denoting the local total rainfall and evapotranspiration. Free parameters are written in red.

A1 Interception (from Ficchì et al. (2019)

Evapotranspiration from the interception store :

$$E_i(t) = \min\left[E(t),\ P(t) + h_i(t-1)\right] \tag{A1}$$

Remaining rainfall after interception:

$$P_{th}(t) = \max\left[0,\ P(t) - (c_i - h_i(t-1)) - E_i(t)\right] \tag{A2}$$

Update of the interception reservoir state:

$$h_i(t) = h_i(t-1) + P(t) - E_i(t) - P_{th}(t) \tag{A3}$$

A2 Production (from Perrin et al. (2003))

Net rainfall filling the production reservoir:

$$P_s(t) = c_p\left(1 - \left(\frac{h_p(t-1)}{c_p}\right)^2\right)\frac{\tanh\left(\frac{P_{th}(t)}{c_p}\right)}{1 + \left(\frac{h_p(t-1)}{c_p}\right)\tanh\left(\frac{P_{th}(t)}{c_p}\right)} \tag{A4}$$

If $E(t) - E(i) > 0$ the evapotranspiration from the production reservoir is :

$$E_s(t) = h_p(t-1)\left(2 - \frac{h_p(t-1)}{c_p}\right)\frac{\tanh\left(\frac{E(t)-E_i(t)}{c_p}\right)}{1 + \left(1 - \frac{h_p(t-1)}{c_p}\right)\tanh\left(\frac{E(t)-E_i(t)}{c_p}\right)} \tag{A5}$$

Direct runoff amount:

$$P_r(t) = P_{th}(t) - P_s(t) \tag{A6}$$

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Update of the production reservoir state:

$$h_p(t) = h_p(t-1) + P_s(t) - E_s(t) \tag{A7}$$

A3 Transfer within pixel x (adapted from Perrin et al. (2003))

Let $\phi_1, \phi_2, \phi_3 \in [0,1]$ be flux partitioning weights such that $\phi_1 + \phi_2 + \phi_3 = 1$. For model structure S-3.6, $\phi_1 = 0.54$, $\phi_2 = 0.36$ and $\phi_3 = 0.1$

water exchange:

$$F(t) = ml\left(\frac{h_{tr}(t-1)}{c_{tr}}\right)^{7/2} \tag{A8}$$

Update of the "fast" transfer reservoir state $0 \leq \phi_1 \leq 1$:

$$h_{tr}(t') = \max\left[\epsilon,\ h_{tr}(t-1) + \phi_1 P_r(t) + F(t)\right] \tag{A9}$$

Outflow from the "fast" transfer reservoir:

$$Q_r(t') = h_{tr}(t') - \left[h_{tr}(t')^{-4} + {c_{tr}}^{-4}\right]^{-1/4} \tag{A10}$$

Update of the "fast" transfer reservoir state:

$$h_{tr}(t) = h_{tr}(t') - Q_r(t') \tag{A11}$$

Update of the "slow" transfer reservoir state $0 \leq \phi_2 \leq 1$: :

$$h_{tl}(t') = h_{tl}(t-1) + \phi_2 P_r(t) \tag{A12}$$

Outflow from the "slow" transfer reservoir:

$$Q_l(t') = h_{tl}(t') - \left[h_{tl}(t')^{-4} + {c_{tl}}^{-4}\right]^{-1/4} \tag{A13}$$

Update of the "fast" transfer reservoir state:

$$h_{tl}(t) = h_{tl}(t') - Q_l(t') \tag{A14}$$

Outflow from direct runoff with $0 \leq \phi_3 \leq 1$:

$$Q_d(t) = \max\left[0,\ \phi_3 P_r(t) + F(t)\right] \tag{A15}$$

Total cell outflow :

$$Q_t(t) = Q_r(t) + Q_l(t) + Q_d(t) \tag{A16}$$

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

A4 Routing between pixels

Given N_{up} adjacent upstream cells within Ω flowing into cell x as imposed by flow direction map, the upstream runoff is:

$$Q_{up}(x,t) = \sum_{k_{up}=1}^{N_{up}} Q(k_{up},t) \tag{A17}$$

with for a given upstream cell k_{up}:

$$Q(k_{up},t) = h_r(k_{up},t)\left(1-\exp\left(-\frac{\Delta t}{60\ c_r}\right)\right) + Q_t(k_{up},t) \tag{A18}$$

Update of local routing state:

$$h_r(t) = h_r(t-1) + Q_{up}(x,t) \tag{A19}$$

A5 Notations

– Inputs

P: Precipitation, mm/h

E: Potential evapotranspiration, mm/h

– States

h_i: Interception reservoir state, mm

h_p: Production reservoir state, mm

h_{tr}: First transfer reservoir state, mm

h_{tl}: Second transfer reservoir state, mm

h_r: Routing reservoir state, mm

– Parameters

c_i: Interception parameter, mm

c_p: Production parameter, mm

c_{tr}: First transfer parameter, mm

c_{tl}: Second transfer parameter, mm

ml: Water exchange parameter, mm/h

c_r: Linear routing parameter, min

– Fluxes

E_i: Evapotranspiration from the interception reservoir, mm/h

P_{th}: Remaining rainfall after interception, mm/h

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

P_s: Net rainfall filling the production reservoir, mm/h

E_s: Evapotranspiration from the production reservoir, mm/h

P_r: Direct runoff, mm/h

F: water exchange, mm/h

Q_r: Outflow from the first transfer reservoir, mm/h

Q_l: Outflow from the second transfer reservoir, mm/h

Q_d: Outflow from direct runoff, mm/h

Q_t: Total cell outflow, mm/h

Q_{up}: Upstream inflow, mm/h

Q_c: Total outflow, mm/h

– Other

ϕ: Direct runoff partitioning weights, with $\sum_{i=1}^{3} \phi_i = 1$

k_{up}: Index denoting the adjacent upstream cells of a given cell $x \in \Omega$, k_{up} =1,...,N_{up}

A6 Metrics

– NSE

$$j_k^* = \frac{\sum_{t=0}^{T} \left[Q_k(t) - Q_k^*(t)\right]^2}{\sum_{t=0}^{T} \left[Q_k^*(t) - \overline{Q_k^*}(t)\right]^2} \tag{A20}$$

– KGE

$$\begin{aligned} j_k^* &= \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \\ r &= \frac{\mathrm{Cov}[Q_k, Q_k^*]}{\sqrt{\mathrm{Var}[Q_k]}\ \sqrt{\mathrm{Var}[Q_k^*]}} \\ \alpha &= \frac{\overline{Q_k}}{\overline{Q_k^*}} \\ \beta &= \frac{\sqrt{\mathrm{Var}[Q_k]}}{\sqrt{\mathrm{Var}[Q_k^*]}} \end{aligned} \tag{A21}$$

where Q_k and Q_k^* are simulated and observed discharges at gauge $x_k \in \Omega$.

A7 Indices

$$RC = \frac{\sum_{t=0}^{T} Q\left(x_k, t\right)}{\sum_{t=0}^{T} \left(\frac{1}{N_x} \sum_{x \in \Omega} \boldsymbol{P}\left(x, t\right)\right)} \tag{A22}$$

$$WI = \frac{\sum_{t=0}^{T} \left(\sum_{x \in \Omega} \boldsymbol{P}\left(x, t\right)\right)}{\sum_{t=0}^{T} \left(\sum_{x \in \Omega} \boldsymbol{E}\left(x, t\right)\right)} \tag{A23}$$

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Appendix B: Results

B1 Downstream gauges: global performances

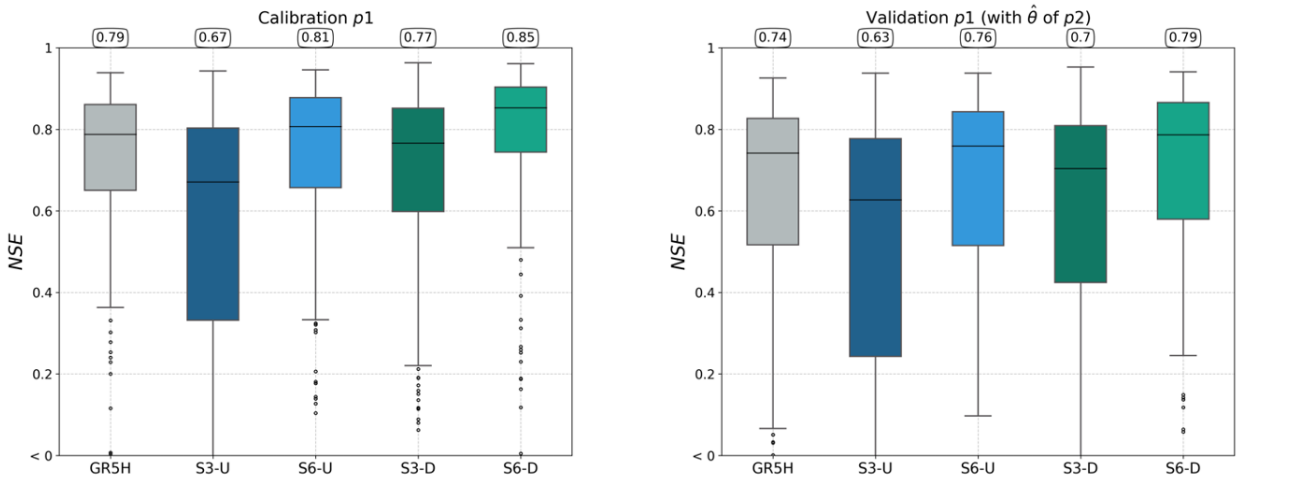

Figure B1. Performances comparison between the two SMASH distributed model structures (S3 and S6) for each calibration settings (U and D). GR5H performances are plotted as a reference. Calibration and validation on $p1$. (See in Fig. 4 for period $p2$).

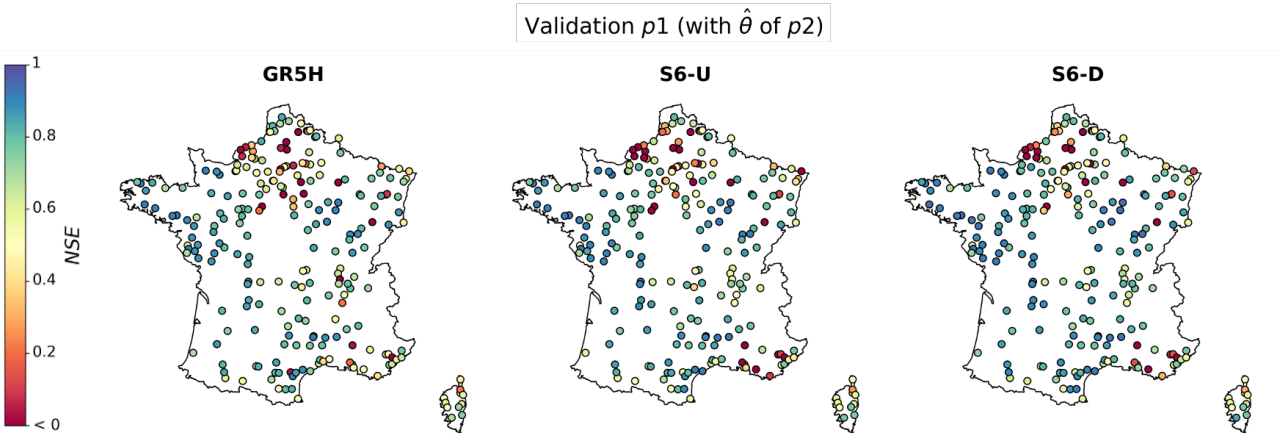

Figure B2. Performances maps in temporal validation on period $p1$ period in terms of NSE criterion for (left) lumped GR5H, spatially distributed 6 parameters SMASH model in (middle) spatially uniform S6-U and (right) distributed S6-D calibration. (See in Fig. 5 for period $p2$).

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

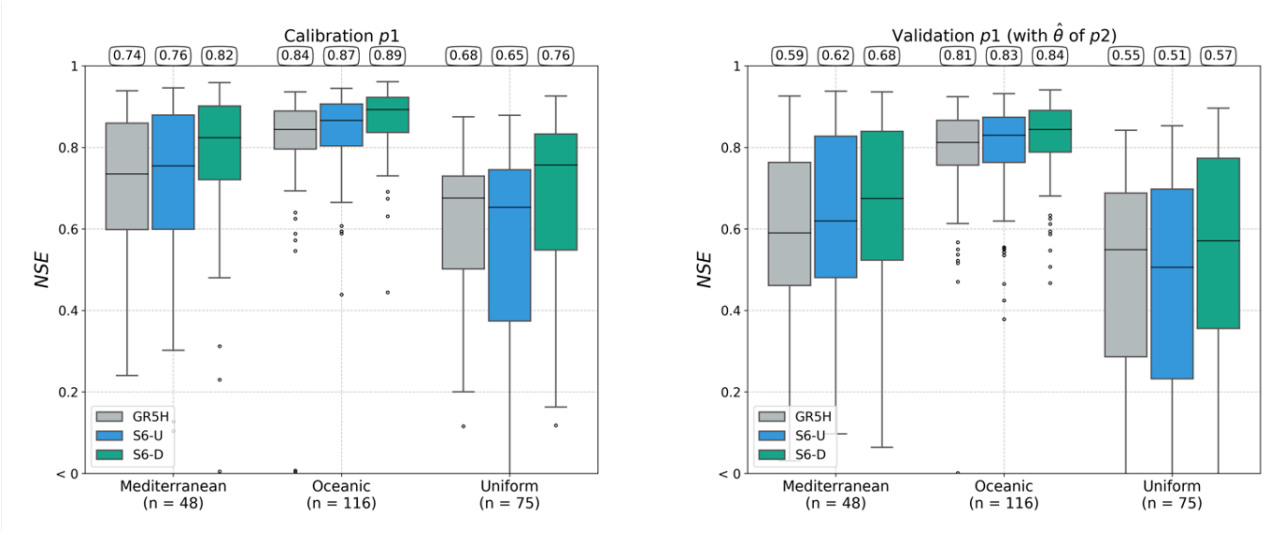

Figure B3. Boxplot by class NSE performances on 235 downstream gauges on period $p1$. (See in Fig. 6 for period $p2$).

B2 Downstream gauges: flood signatures

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

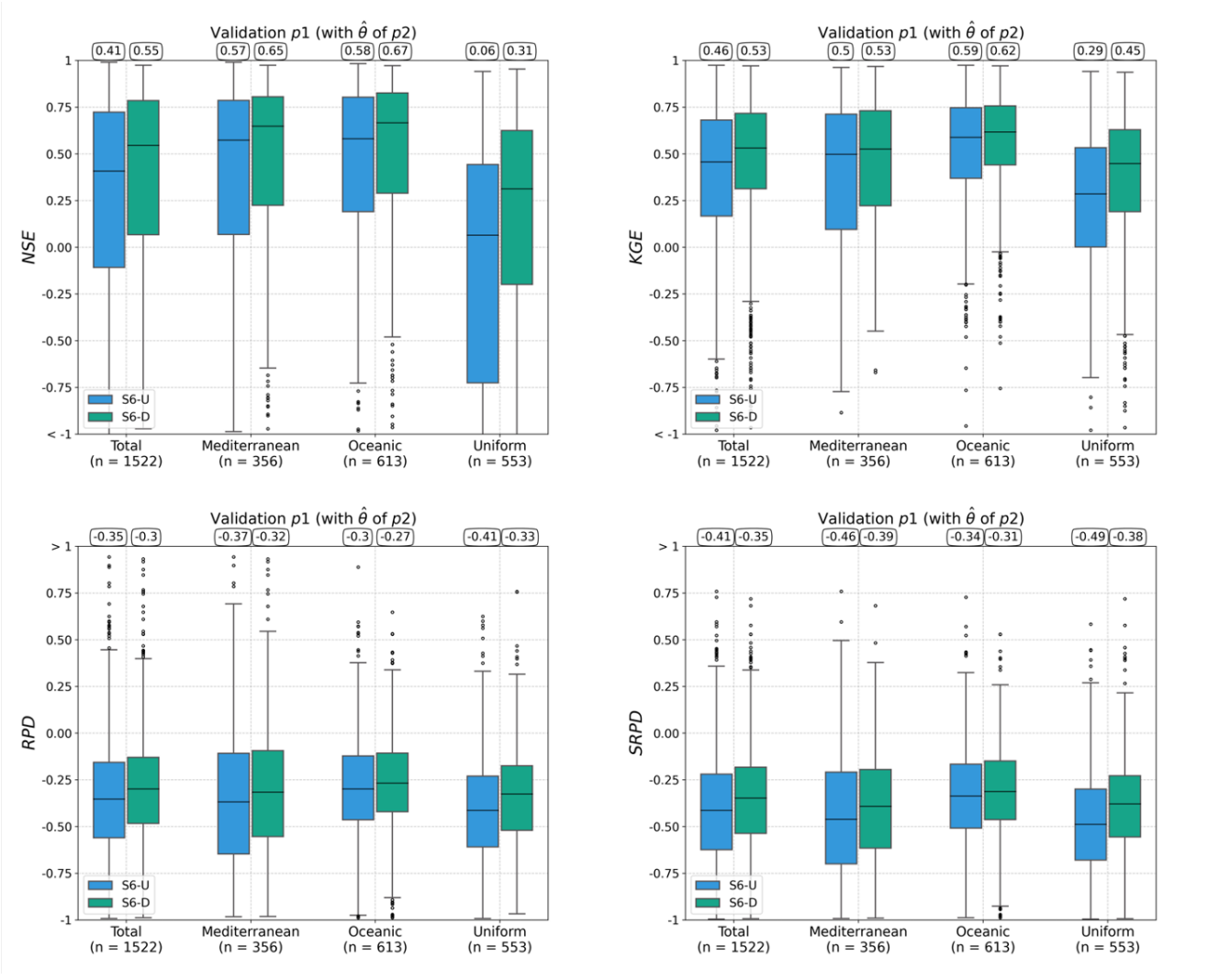

Figure B4. Boxplot of flood event performances and signatures on 235 downstream gauges on validation period $p1$, i.e. 1522 flood events. (See in Fig. 8 for period $p2$).

B3 Model performances at upstream un-calibrated gauges

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

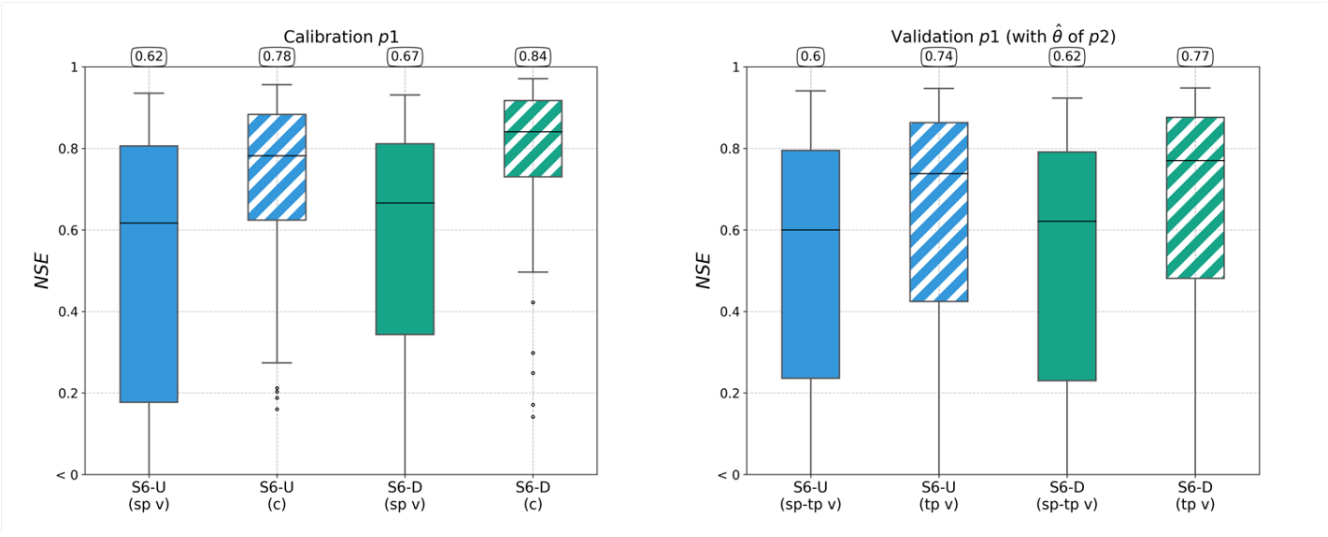

Figure B5. Boxplot of NSE performances on 77 upstream gauges on period $p1$. Spatial validation 'sp v' (left) and spatio-temporal validation 'sp-st v' (right) ; label 'c' and hatched boxes stand for local calibrations performed at those upstream gauges for sake of comparison. (See in Fig. 9 for period $p2$).

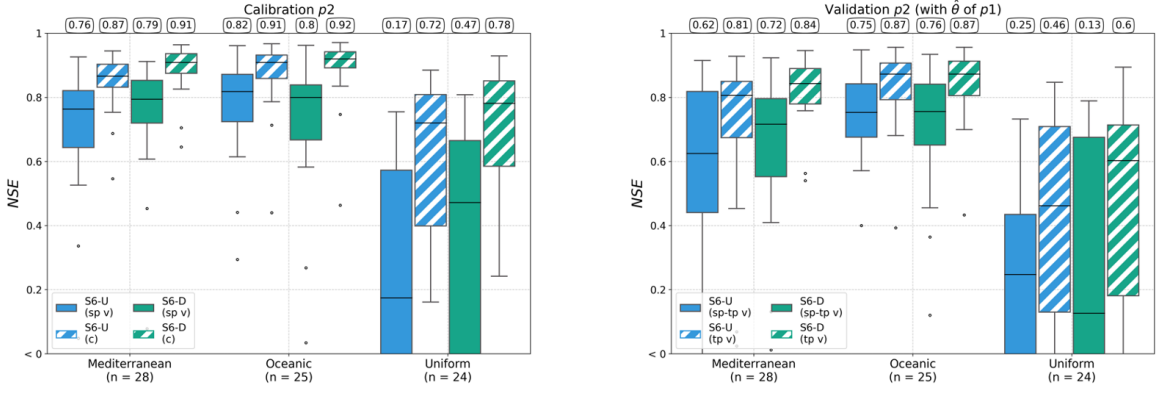

Figure B6. Boxplot of NSE performances on 77 upstream gauges on period $p1$. Spatial validation (left) and spatio-temporal validation (right). (See in Fig. 6 for period $p2$).

References

565 Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., Gan, T. Y., Pendergrass, A. G., Rosenfeld, D., Swann, A. L. S., Wilcox, L. J., and Zolina, O.: Advances in understanding large-scale responses of the water cycle to climate change, Annals of the New York Academy of Sciences, 1472, 49–75, https://doi.org/10.1111/nyas.14337, 2020.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrology and Earth System Sciences, 13, 1757–1764, https://doi.org/10.5194/hess-13-1757-2009, 2009.

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

Astagneau, P. C., Thirel, G., Delaigue, O., Guillaume, J. H. A., Parajka, J., Brauer, C. C., Viglione, A., Buytaert, W., and Beven, K. J.: Technical note: Hydrology modelling R packages – a unified analysis of models and practicalities from a user perspective, Hydrology and Earth System Sciences, 25, 3937–3973, https://doi.org/10.5194/hess-25-3937-2021, 2021b.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I. J. M., and Wood, E. F.: Global Fully Distributed Parameter Regionalization Based on Observed Streamflow From 4,229 Headwater Catchments, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 485, https://doi.org/10.1029/2019JD031485, e2019JD031485 10.1029/2019JD031485, 2020.

Bertalanffy, L. v.: General system theory: Foundations, development, applications, G. Braziller, 1968.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, Journal of Hydrology, 105, 157–172, https://doi.org/10.1016/0022-1694(89)90101-7, 1989.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling., Advances in Water Resources, 16, 41–51, https://doi.org/10.1016/0309-1708(93)90028-E, 1993.

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

Beven, K. J.: Rainfall - Runoff Modelling, The Primer, John Wiley and Sons, LTD, 2011.

Blache, J.: Pardé (Maurice). — Fleuves et rivières., Revue de Géographie Alpine, 21, 851–854, https://www.persee.fr/doc/rga_0035-1121_1933_num_21_4_5386_t1_0851_0000_1, 1933.

Blöschl, G. and al.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, Hydrological Sciences Journal, 64, 1141–1158, https://doi.org/10.1080/02626667.2019.1620507, 2019.

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrological Processes, 9, 251–290, https://doi.org/10.1002/hyp.3360090305, 1995.

Bouaziz, L. J. E., Fenicia, F., Thirel, G., de Boer-Euser, T., Buitink, J., Brauer, C. C., De Niel, J., Dewals, B. J., Drogue, G., Grelier, B., Melsen, L. A., Moustakas, S., Nossent, J., Pereira, F., Sprokkereef, E., Stam, J., Weerts, A. H., Willems, P., Savenije, H. H. G., and Hrachowitz, M.: Behind the scenes of streamflow model performance, Hydrology and Earth System Sciences, 25, 1069–1095, https://doi.org/10.5194/hess-25-1069-2021, 2021.

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

Brisset, P., Monnier, J., Garambois, P.-A., and Roux, H.: On the assimilation of altimetric data in 1D Saint-Venant river flow models, Advances in water resources, 119, 41–59, https://doi.org/10.1016/j.advwatres.2018.06.004, 2018.

Clark, M. P., Bierkens, M. F., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrology and Earth System Sciences, 21, 3427–3440, 2017.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resources Research, 48, https://doi.org/10.1029/2011WR011721, 2012.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The Suite of Lumped GR Hydrological Models in an R package, Environmental Modelling and Software, 94, 166–171, https://doi.org/10.1016/j.envsoft.2017.05.002, 2017.

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., and Michel, C.: airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling, https://doi.org/10.15454/EX11NA, r package version 1.7.0, 2022.

de Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G., and Willems, P.: Looking beyond general metrics for model comparison – lessons from an international model intercomparison study, Hydrology and Earth System Sciences, 21, 423–440, https://doi.org/10.5194/hess-21-423-2017, 2017.

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

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

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3–17, https://doi.org/10.1016/j.jhydrol.2005.07.031, the model parameter estimation experiment, 2006.

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

Edijatno, N. and Michel, C.: Un modèle pluie-débit journalier à trois paramètres, La Houille Blanche, 2, 113–121, 1989.

Fenicia, F., Kavetski, D., and Savenije, H. H.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, 2011.

Ficchì, A., Perrin, C., and Andréassian, V.: Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching, Journal of Hydrology, 575, 1308–1327, https://doi.org/10.1016/j.jhydrol.2019.05.084, 2019.

Flipo, N., Mouhri, A., Labarthe, B., Biancamaria, S., Rivière, A., and Weill, P.: Continental hydrosystem modelling: the concept of nested stream–aquifer interfaces, Hydrology and Earth System Sciences, 18, 3121–3149, https://doi.org/10.5194/hess-18-3121-2014, 2014.

Flipo, N., Gallois, N., and Schuite, J.: Regional coupled surface-subsurface hydrological model fitting based on a spatially distributed minimalist reduction of frequency-domain discharge data, Geoscientific Model Development Discussions, 2022, 1–40, https://doi.org/10.5194/gmd-2022-24, 2022.

Garambois, P.-A., Roux, H., Larnier, K., Labat, D., and Dartus, D.: Parameter regionalization for a process-oriented distributed model dedicated to flash floods, Journal of Hydrology, 525, 383 – 399, https://doi.org/https://doi.org/10.1016/j.jhydrol.2015.03.052, 2015.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrology and Earth System Sciences, 18, 463–477, https://doi.org/10.5194/hess-18-463-2014, 2014.

Hascoet, L. and Pascual, V.: The Tapenade Automatic Differentiation Tool: Principles, Model, and Specification, ACM Trans. Math. Softw., 39, https://doi.org/10.1145/2450153.2450158, 2013.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Hashemi, R., Brigode, P., Garambois, P.-A., and Javelle, P.: How can regime characteristics of catchments help in training of local and regional LSTM-based runoff models?, Hydrology and Earth System Sciences Discussions, 2021, 1–33, https://doi.org/10.5194/hess-2021-511, 2021.

Hrachowitz, M. and Clark, M. P.: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology, Hydrology and Earth System Sciences, 21, 3953–3973, 2017.

Huynh, N. N. T. and Garambois, P.-A.: Signatures-and-sensitivity-based multi-criteria optimization for constraining distributed conceptual hydrological models applied to Mediterranean floods, submitted, 2022.

Jay-Allemand, M., Javelle, P., Gejadze, I., Arnaud, P., Malaterre, P.-O., Fine, J.-A., and Organde, D.: On the potential of variational calibration for a fully distributed hydrological model: application on a Mediterranean catchment, Hydrology and Earth System Sciences, pp. 1–24, https://doi.org/10.5194/hess-24-5519-2020, 2020.

Jay-Allemand, M., Colleoni, F., Garambois, P.-A., Javelle, P., and Julie, D.: SMASH - Spatially distributed Modelling and ASsimilation for Hydrology: Python wrapping towards enhanced research-to-operations transfer, IAHS, https://hal.archives-ouvertes.fr/hal-03683657, 2022.

Klemeš, V.: Conceptualization and scale in hydrology, Journal of Hydrology, 65, 1–23, https://doi.org/10.1016/0022-1694(83)90208-1, scale Problems in Hydrology, 1983.

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J. A., and Reaney, S. M.: Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain, Hydrology and Earth System Sciences, 23, 4011–4032, https://doi.org/10.5194/hess-23-4011-2019, 2019.

Larnier, K., Monnier, J., Garambois, P.-A., and Verley, J.: River discharge and bathymetry estimation from SWOT altimetry measurements, Inverse Problems in Science and Engineering, pp. 1–31, https://doi.org/10.1080/17415977.2020.1803858, 2020.

Le Mesnil, M., Moussa, R., Charlier, J.-B., and Caballero, Y.: Impact of karst areas on runoff generation, lateral flow and interbasin groundwater flow at the storm-event timescale, Hydrology and Earth System Sciences, 25, 1259–1282, https://doi.org/10.5194/hess-25-1259-2021, 2021.

Le Moine, N.: Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?, Ph.D. thesis, UPMC, Cemagref Antony, 2008.

Malou, T., Garambois, P.-A., Paris, A., Monnier, J., and Larnier, K.: Generation and analysis of stage-fall-discharge laws from coupled hydrological-hydraulic river network model integrating sparse multi-satellite data, Journal of Hydrology, 603, 126 993, https://doi.org/10.1016/j.jhydrol.2021.126993, 2021.

Mathevet, T., Gupta, H., Perrin, C., Andréassian, V., and Le Moine, N.: Assessing the performance and robustness of two conceptual rainfall-runoff models on a worldwide sample of watersheds, Journal of Hydrology, 585, 124 698, https://doi.org/10.1016/j.jhydrol.2020.124698, 2020.

Milly, P.: Climate, interseasonal storage of soil water, and the annual water balance, Advances in Water Resources, 17, 19–24, https://doi.org/10.1016/0309-1708(94)90020-5, mIT Colloquium on Hydroclimatology and Global Hydrology, 1994.

Monnier, J., Couderc, F., Dartus, D., Larnier, K., Madec, R., and Vila, J.-P.: Inverse algorithms for 2D shallow water equations in presence of wet dry fronts: Application to flood plain dynamics, Advances in Water Resources, 97, 11–24, https://doi.org/10.1016/j.advwatres.2016.07.005, 2016.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, Journal of Hydrology, 523, 147–159, https://doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

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

Pagano, T. C., Wood, A. W., Ramos, M.-H., Cloke, H. L., Pappenberger, F., Clark, M. P., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., and Verkade, J. S.: Challenges of Operational River Forecasting, Journal of Hydrometeorology, 15, 1692 – 1707, https://doi.org/10.1175/JHM-D-13-0188.1, 2014.

Perrin, C., Michel, C., and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, Journal of Hydrology, 242, 275–301, https://doi.org/10.1016/S0022-1694(00)00393-0, 2001.

Perrin, C., Michel, C., and Andrèassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of hydrology, 279, 275–289, 2003.

Pokhrel, P. and Gupta, H. V.: On the use of spatial regularization strategies to improve calibration of distributed watershed models, Water Resources Research, 46, https://doi.org/https://doi.org/10.1029/2009WR008066, 2010.

Pujol, L., Garambois, P.-A., Finaud-Guyot, P., Monnier, J., Larnier, K., Mosé, R., Biancamaria, S., Yesou, H., Moreira, D., Paris, A., and Calmant, S.: Estimation of multiple inflows and effective channel by assimilation of multi-satellite hydraulic signatures: The ungauged anabranching Negro river, Journal of Hydrology, 591, 125 331, https://doi.org/10.1016/j.jhydrol.2020.125331, 2020.

Pujol, L., Garambois, P.-A., and Monnier, J.: Multi-dimensional hydrological-hydraulic model with variational data assimilation for river networks and floodplains, EGUsphere, 2022, 1–44, https://doi.org/10.5194/egusphere-2022-10, 2022.

Pushpalatha, R.: Low-flow simulation and forecasting on French river basins : A hydrological modelling approach, Theses, AgroParisTech, https://pastel.archives-ouvertes.fr/pastel-00912565, 2013.

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

Reed, S., Koren, V., Smith, Z., Moreda, Fekadu, Seo, Dong-Jun, Butts, M., and DMIP: Reed S., Koren V., Smith M. Zhang Z, Moreda F, Seo D-J and DMIP Participants (Butts M. B), 2004: Overall distributed model intercomparison project results, Journal of Hydrology, Volume 298, Issues 1-4, 1 October 2004, Pages 27-60., Journal of Hydrology, 298, 27–60, 2004.

Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological models, Journal of Hydrology, 198, 69–97, https://doi.org/10.1016/S0022-1694(96)03329-X, 1997.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/10.1029/2008WR007327, 2010.

Sauquet, E.: Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies, Journal of Hydrology, 331, p. 300 – p. 314, https://doi.org/10.1016/j.jhydrol.2006.05.018, 2006.

Schuite, J., Flipo, N., Massei, N., Rivière, A., and Baratelli, F.: Improving the Spectral Analysis of Hydrological Signals to Efficiently Constrain Watershed Properties, Water Resources Research, 55, 4043–4065, https://doi.org/10.1029/2018WR024579, 2019.

Sebben, M. L., Werner, A. D., Liggett, J. E., Partington, D., and Simmons, C. T.: On the testing of fully integrated surface subsurface hydrological models, Hydrological Processes, 27, 1276–1285, https://doi.org/10.1002/hyp.9630, publisher: John Wiley & Sons, Ltd, 2013.

https://doi.org/10.5194/egusphere-2022-506
Preprint. Discussion started: 14 July 2022

Shukla, P., Skea, J., Buendia, E. C., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Pereira, J. P., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J., and (eds.): Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems, In press, 2019.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, Journal of hydrology, 523, 739–757, 2015.

Vannier, O., Braud, I., and Anquetin, S.: Regional estimation of catchment-scale soil properties by means of streamflow recession analysis for use in distributed hydrological models, Hydrological Processes, 28, 6276–6291, https://doi.org/10.1002/hyp.10101, 2014.

Vereecken, H., Weihermüller, L., Assouline, S., Šimůnek, J., Verhoef, A., Herbst, M., Archer, N., Mohanty, B., Montzka, C., Vanderborght, J., Balsamo, G., Bechtold, M., Boone, A., Chadburn, S., Cuntz, M., Decharme, B., Ducharne, A., Ek, M., Garrigues, S., Goergen, K., Ingwersen, J., Kollet, S., Lawrence, D. M., Li, Q., Or, D., Swenson, S., de Vrese, P., Walko, R., Wu, Y., and Xue, Y.: Infiltration from the Pedon to Global Grid Scales: An Overview and Outlook for Land Surface Modeling, Vadose Zone Journal, 18, 180 191, https://doi.org/10.2136/vzj2018.10.0191, 2019.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, International Journal of Climatology, 30, 1627–1644, https://doi.org/10.1002/joc.2003, 2010.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization., ACM Trans. Math. Softw., 23, 550–560, http://dblp.uni-trier.de/db/journals/toms/toms23.html#ZhuBLN97, 1997.