# A Distributed Hybrid Physics-AI Framework for Learning Corrections of Internal Hydrological Fluxes and Enhancing High-Resolution Regionalized Flood Modeling

Ngo Nghi Truyen Huynh[1], Pierre-André Garambois[1], Benjamin Renard[1], François Colleoni[1], Jérôme Monnier[2], and Hélène Roux[3]

[1]INRAE, Aix-Marseille Université, RECOVER, 3275 Route Cézanne, 13182 Aix-en-Provence, France
[2]INSA, Institut de Mathématiques de Toulouse (IMT), Université de Toulouse, 31400 Toulouse, France
[3]Institut de Mécanique des Fluides de Toulouse (IMFT), Université de Toulouse, CNRS, 31400 Toulouse, France

**Correspondence:** Ngo Nghi Truyen Huynh (ngo-nghi-truyen.huynh@inrae.fr) and Pierre-André Garambois (pierre-andre.garambois@inrae.fr)

**Abstract.** To advance the discovery of scale-relevant hydrological laws while better exploiting massive multi-source data, merging artificial intelligence with process-based modeling has emerged as a compelling approach, as demonstrated in recent lumped hydrological modeling studies. This research proposes a general spatially distributed hybrid modeling framework that seamlessly combines differentiable process-based modeling with neural networks. We focus on hybridizing a differentiable hydrological model with neural networks, leveraging the temporal memory effect of the original model, on top of a differentiable kinematic wave routing over a flow direction grid. We evaluate flood modeling performance and analyze the interpretability of learned conceptual parameters and corrections of internal fluxes using two high-resolution data sets ($dx = 1\ \mathrm{km}$, $dt = 1\ \mathrm{h}$). The first data set involves 235 catchments in France, used for local calibration-validation and model structure comparisons between the classical GR-like model and the hybrid approach. The second dataset presents a challenging multi-catchment modeling setup in flash flood-prone areas to demonstrate the framework's regionalization learning capabilities. The results show that the hybrid models achieve superior accuracy and robustness compared to classical approaches in both spatial and temporal validation. Analysis of the spatially distributed parameters and internal fluxes reveals the hybrid models' nuanced behavior, their adaptability to diverse hydrological responses, and their potential for uncovering physical processes.

## 1 Introduction

Faced with the socio-economic challenges of floods and drought forecasting in a context of climate change, modeling approaches that make the most of the maximum amount of information available are needed to make accurate forecasts at high spatio-temporal resolution. Nevertheless, given the complexity and non-linearity of the coupled surface and subsurface physical processes involved, and their limited observability with respect to the number of parameters to estimate ("curse of dimensionality"), hydrological modeling remains a difficult task tinged with uncertainties (e.g., Liu and Gupta, 2007). Moreover, in the absence of directly exploitable first principles in hydrology (e.g., Dooge, 1986), as opposed to flow mechanistic equations in continuous media such as river hydraulics, meteorology or oceanography, and given the high heterogeneities of continental hydrosystems compartments as well as the lack of "scale-relevant theories" (Beven, 1987), process-based hydrological models generally include a certain amount of empiricism, which represents an avenue for the fusion of data assimilation (DA) and uncertainty quantification (UQ) with machine learning (ML) and deep learning (DL) techniques to better exploit the informative richness of multi-source data.

Pure ML applications in hydrology started decades ago (e.g., references in Maier and Dandy (2000) or Artigue et al. (2012) on flash floods). A recent explosion of artificial intelligence (AI) applications, stemming from the rise of big data, computational power, and their capabilities to extract multi-level information from large data sets (LeCun et al., 2015), has led to a bloom of studies, in particular in hydrology (e.g., reviews by Nearing et al. (2021); Shen and Lawson (2021)) and water related disciplines (e.g., Tripathy and Mishra, 2024). The potential of using long short-term memory (LSTM) network (Hochreiter and Schmidhuber, 1997), a recurrent neural network (RNN) adapted to long time series, for lumped continuous rainfall-runoff modeling, was introduced by Kratzert et al. (2018) and explored in hundreds of studies since (Shen and Lawson, 2021). In addition to the capability of LSTM to learn multi-frequential aspects, training these networks over large catchment samples using meteorological forcings time series and catchment physical descriptors within lumped models enhances performance in daily runoff prediction and in regionalization (Kratzert et al., 2018; Hashemi et al., 2022). A convolutional LSTM architecture, combining the strength of LSTM for capturing multiscale temporal dynamics and of convolutional layers for spatial patterns extraction, is found effective for spatio-temporal rainfall nowcasting (Shi et al., 2015) and for hydrological modeling (e.g., Xu et al., 2022; Chen et al., 2022). Nevertheless, pure ML/DL algorithms are hardly interpretable and do not use the effective physical models, solvers and DA techniques developed over the past century. Hybrid approaches, that leverage ML/DL in sequential combination with process-based numerical models via their inputs/outputs, have been explored recently and enable improving the accuracy of hydrologic predictions (e.g., Konapala et al. (2020), with DA in Roy et al. (2023) or with UQ in Tran et al. (2023)).

Merging process-based differential equations with ML can be very advantageous as recently shown with physics-informed neural networks (PINN) in Raissi et al. (2019), where the process-based model is used as a weak constrain in the training cost function and is well adapted to assimilate observations (e.g., He et al., 2020), or in universal differential equations that embed a universal approximator (Chen et al., 2019; Rackauckas et al., 2021; Yin et al., 2021). In hydrology, integrating ML into process-based models shows promise, as demonstrated in recent studies on daily lumped models (Kumanlioglu and Fistikoglu, 2019; Jiang et al., 2020; Höge et al., 2022; Feng et al., 2022). Kumanlioglu and Fistikoglu (2019) replaced the routing component of a lumped GR model (Perrin et al., 2003)–an algebraic model derived from temporally integrable ordinary differential equations (ODE)–with an artificial neural network (ANN), achieving superior performance compared to using the GR model or ANN alone on a single basin. Including an ANN for flux correction into a spatially lumped process-based hydrological ODE (Höge et al., 2022), and adding an ANN-based regionalization pipeline (Feng et al., 2022), or to a semi-lumped model (Li et al., 2024) has resulted in learnable lumped model structures that exhibit interesting performance and improved interpretability after training. However, these approaches do not natively account for spatially distributed information, such as detailed meteorological forcings and descriptors of basin physical properties, which is crucially needed for high-resolution hydrological modeling, particularly for extreme events with strong variability.

A spatially distributed hybrid approach, HDA-PR, which incorporates a regionalization neural network into the forward model, was recently proposed by Huynh et al. (2024b). This approach has enhanced regional flash flood modeling at a relatively high-resolution ($dx = 1$ km, $dt = 1$ h). Meanwhile, Wang et al. (2024) introduced a large-scale spatialized hybrid hydrological model that improves evapotranspiration modeling for the Amazon basin by incorporating a replacement neural network and a regionalization neural network. In their study, Wang et al. (2024) combine a conceptual hydrological model with a simple bucket-based routing structure, using a hybrid approach that predicts conceptual parameters and corrects evapotranspiration (adjusting a Penman-Monteith estimate via a 3D convolutional neural network (CNN)) without correcting other internal fluxes, and operates at a coarser resolution ($dx = 0.5°$, $dt = 1$ day). In addition, enhancing the physical modeling of runoff and flood propagation necessitates the use of hydraulic models that must be differentiable to facilitate gradient-based optimization, a requirement for effective hybrid modeling. This approach has been demonstrated for the optimization of large parameter sets from heterogeneous data at the river network scale using various approaches based on differentiable hydraulic models, including: a 2D or multi-dimensional complete shallow water model coupled with a differentiable semi-lumped GR model (Pujol et al., 2022); a 1D Saint-Venant river network model (e.g., Larnier et al. (2025), though without a differentiable hydrological model); and a kinematic wave model integrated into a grid differentiable spatialized hydrological model (Huynh et al., 2024a). These hydraulic models consist in partial differential equations (PDE). Therefore, hybrid hydrological models embedding neural networks should advance from lumped, semi-lumped or spatialized ODE-based models, to spatially distributed differentiable hybrid PDE-based hydrological-hydraulic models in view to improve modeling capabilities at basin scale and realism, at least for hydraulic processes which modeling is less uncertain than hydrological processes occurring in the earth critical zone (the near-surface environment where complex interactions between water, soil, rock, and living organisms regulate the Earth's surface dynamics).

To address the aforementioned challenges, this research proposes a spatially distributed hybrid modeling framework combining process-based modeling and neural networks, which is amenable to hydrological-hydraulic (H&H) models and other geophysical models. This study is based on a spatially distributed, parsimonious GR-like hydrological model structure that is well-suited for flood modeling and regionalization learning (Huynh et al., 2024b), coupled with a differentiable kinematic wave model for spatially distributed flow routing. The original version of this framework was first proposed in Huynh et al. (2024a), while the present article enhances the approach with a more comprehensive case study, testing a larger sample set and offering more detailed analyses. The research continues to focus on correcting internal fluxes via a simple neural network. A relatively "parsimonious hybridization" with a dense flux correction neural network is proposed, as it effectively captures non-linear flux corrections while leveraging the memory effect already embedded in the original hydrological model. This reduces the need for recurrent architectures such as LSTMs, which are typically designed for sequential data but introduce additional complexity that is unnecessary for this application. The approach is implemented in the open-source SMASH software (Spatially distributed Modeling and ASsimilation for Hydrology) that enables multiscale modeling, numerical adjoint model derivation via automatic differentiation and variational data assimilation (VDA). The performance, robustness and interpretability of the proposed approach are studied at a relatively high spatio-temporal resolution of 1 km and 1 h, over a large sample of French catchments and also over a challenging flash flood-prone area. This study aims to demonstrate the feasibility and advantages of distributed hybrid modeling for spatio-temporal learning of hydrological processes at basin and regional scales. It provides a general framework for PDE-based spatially distributed modeling, taking advantage of AI and big data.

## 2 Method

### 2.1 Forward differentiable spatially distributed model statement

The forward differentiable hybrid model $\mathcal{M}$ is obtained by partially composing (i) a dynamic process-based, differentiable, and spatially distributed rainfall-runoff model with simplified hydraulic routing $\mathcal{M}_{rr-hy}$ with (ii) a learnable (neural network-based), differentiable process-parameterization and regionalization operator $\phi$, resulting in Equation 1.

$$\mathcal{M} = \mathcal{M}_{rr-hy}\left(\,.\,;\,\phi(.)\right) \tag{1}$$

Let $\Omega \subset \mathbb{R}^2$ denote a 2D spatial domain with $x \in \Omega$ the spatial coordinate and $t \in \left]0, T\right]$ the physical time, $\mathcal{D}_\Omega$ a 8-direction ("D8") drainage plan. The spatially distributed rainfall-runoff model $\mathcal{M}_{rr-hy}$ is a dynamic operator projecting the input fields of atmospheric forcings $\boldsymbol{\mathcal{I}}$ onto the fields of surface discharge $Q$, internal states $\boldsymbol{h}$, and internal fluxes $\boldsymbol{q}$, as expressed in Equation 2.

$$\boldsymbol{U}(x,t) = \left[Q, \boldsymbol{h}, \boldsymbol{q}\right](x,t) = \mathcal{M}_{rr-hy}\left(\mathcal{D}_\Omega, \boldsymbol{\mathcal{I}}(x,t); \boldsymbol{f}_q(x,t), \left[\boldsymbol{\theta}, \boldsymbol{h}_0\right](x)\right) \tag{2}$$

with $\boldsymbol{U}(x,t)$ the modeled state-flux variables, $\boldsymbol{f}_q$ the vector of spatially distributed corrections of $\boldsymbol{q}$ the internal fluxes (which will be explained later), $\boldsymbol{\theta}$ and $\boldsymbol{h}_0$ the spatially distributed parameters and initial states of the hydrological model. Note that neural network-based estimation of initial states is also feasible, for instance, for short-range DA; however, this is beyond the scope of the current work.

A neural network-based estimator $\phi$, with trainable parameters $\boldsymbol{\rho}$, is embedded into the hydrological model $\mathcal{M}_{rr-hy}$ and forms part of the complete model $\mathcal{M}$. Its purpose is to predict corrections of internal fluxes $\boldsymbol{f}_q$ as well as to estimate parameters $\boldsymbol{\theta}$ regionally, based on various input data, including atmospheric forcings $\boldsymbol{\mathcal{I}}$ and spatialized physical descriptors $\boldsymbol{\mathcal{D}}$, as described in Equation 3.

$$\phi : \left(\boldsymbol{\mathcal{I}}, \boldsymbol{h}, \boldsymbol{\mathcal{D}}; \boldsymbol{\rho}\right) \rightarrow \left(\boldsymbol{f}_q, \boldsymbol{\theta}, \boldsymbol{h}_0\right) \tag{3}$$

By construction, the complete forward model $\mathcal{M}$ is learnable, through the neural network-based mapping $\phi$ embedded into $\mathcal{M}_{rr-hy}$. Moreover, if $\mathcal{M}_{rr-hy}$ and $\phi$ are differentiable, then $\mathcal{M}$ is differentiable which is required to obtain its output gradient derivatives with respect to the neural network parameters as needed for their optimization.

The differentiable spatially distributed hydrological model studied thereafter is based on ODEs for local runoff production coupled with neural networks. This runoff is then conveyed on the spatial grid with a PDE-based hydraulic model.

### 2.2 Case of differentiable spatially distributed GR-like and kinematic wave model

The spatially distributed differentiable, learnable, regionalizable model proposed and studied in this article is detailed in this section. Its hydrological component is based on GR4, that is a parsimonious, widely used and efficient lumped hydrological model (Perrin et al., 2003). Interception, production and fast and slow transfer branches structures are used at pixel scale, without unit hydrographs that are not needed for small scale pixels (refer to Huynh et al. (2024b) and references therein) and could be, if needed for example for a semi distributed model with larger sub-catchments, replaced by a nash cascade that is differentiable and quasi equivalent (cf. Santos et al. (2018)). Moreover, the selected hydrological operators simply consist in algebraic relations obtained from analytical integration in time of first order ODEs that describe state evolution in interception, production and transfer reservoirs plus closure laws. This simple hydrological model produces "runoff" at pixel scale that is then routed on the spatial grid with a kinematic wave model (Te Chow et al., 1988). These model operators and their numerical implementation are fully differentiable, a property further detailed in the following sections. Additionally, this dynamical hydrological model, implemented as recurrence relations, maintains a spatio-temporal memory via the reservoir states $\boldsymbol{h}(x,t)$. This property is used for defining the hybridization for internal flux corrections via a simple ANN based on the principle of parsimony. The ANN uses previous reservoir states and atmospheric forcings as inputs, which requires no additional information beyond the original model. The proposed model, which is spatially distributed and differentiable, is schematized in Figure 1. The following sections describe the complete forward model, including details on the neural networks used for flux corrections and parameter regionalization.

#### 2.2.1 "Runoff" production

For a given cell $x \in \Omega$ and time step $t > 0$, $P(x,t)$ and $E(x,t)$ represent the local precipitation and potential evapotranspiration. For simplicity, spatio-temporal dependencies are omitted, with flux corrections highlighted in brown and regionalized parameters in turquoise.

**Interception.** First, an interception reservoir of capacity $c_i$, automatically computed with flux matching technique (cf. Ficchì et al. (2019)), enables computing the neutralized rainfall $P_n$ and the neutralized evapotranspiration $E_n$.

**Production.** Then, the infiltration flux $\tilde{P}_s$ into the production reservoir is obtained by applying a correction term $f_{q,1}$, predicted by the neural network $\phi_1$, to the classical GR infiltration flux $P_s$ as follows:

$$\tilde{P}_s = \min(P_n, (1+f_{q,1})P_s) \text{ with } P_s = c_p \left(1 - \left(\frac{h_p^-}{c_p}\right)^2\right) \frac{\text{TanH}\left(\frac{P_n}{c_p}\right)}{1+\left(\frac{h_p^-}{c_p}\right)\text{TanH}\left(\frac{P_n}{c_p}\right)} \tag{4}$$

where $c_p$ represents the capacity of the production reservoir predicted by the neural network $\phi_2$, and $h_p^-$ is the production state at previous time step. The actual evapotranspiration flux $\tilde{E}_s$ substracted to the production reservoir is obtained by applying a correction term $f_{q,2}$, predicted by $\phi_1$, to the classical GR evapotranspiration flux $E_s$ as follows:

$$\tilde{E}_s = \min(E_n, (1+f_{q,2})E_s) \text{ with } E_s = h_p^- \left(2 - \frac{h_p^-}{c_p}\right) \frac{\text{TanH}\left(\frac{E_n}{c_p}\right)}{1+\left(1-\frac{h_p^-}{c_p}\right)\text{TanH}\left(\frac{E_n}{c_p}\right)} \tag{5}$$

**Transfer (subgrid).** A subgrid transfer with slow and fast "lateral" flow components is fed by a learnable partition of net rainfall using a correction term $f_{q,3}$, predicted by $\phi_1$, where the net rainfall $P_r = P_n - \tilde{P}_s$ is split into $\tilde{Q}_{dl}$ and $\tilde{Q}_{dr}$, representing

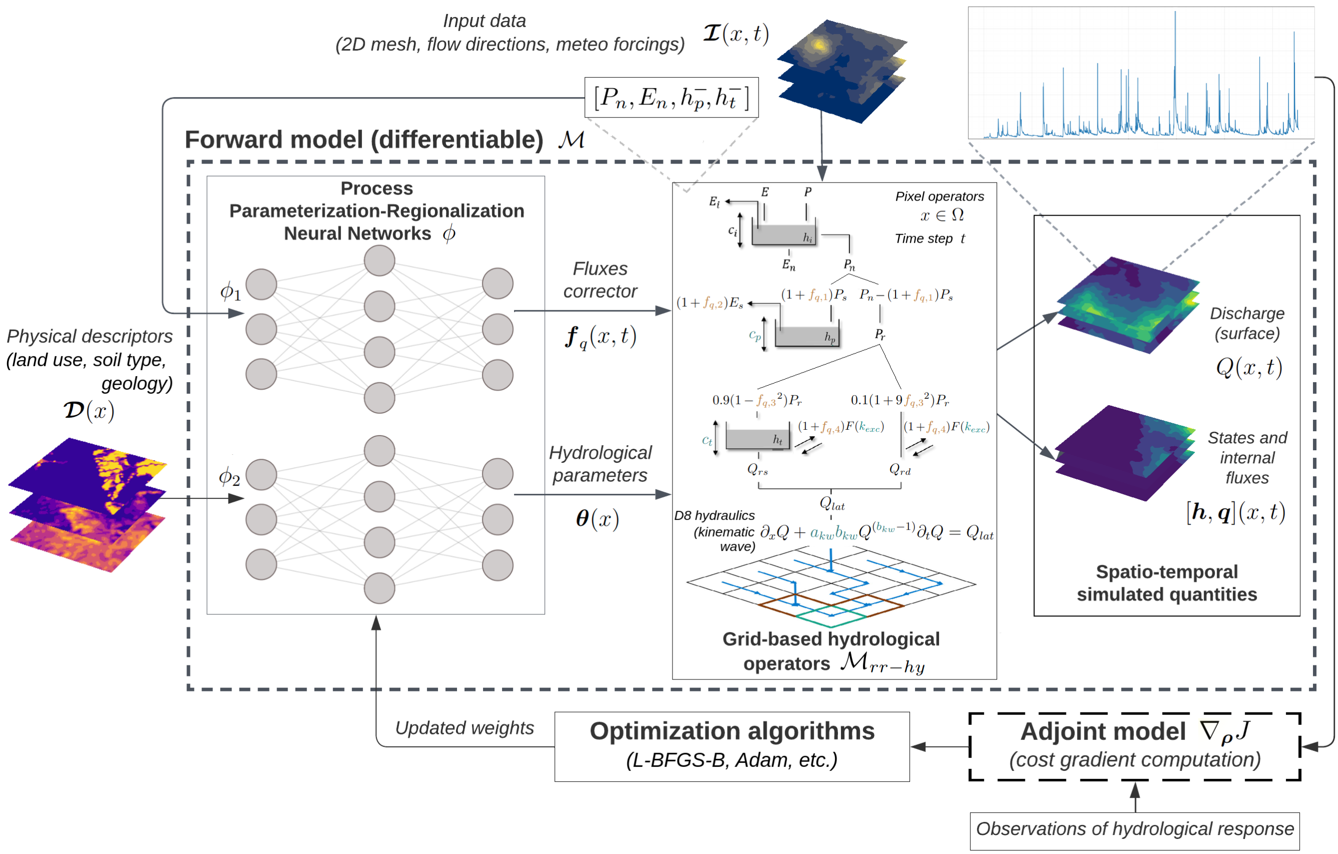


**Figure 1.** Hybrid physics-AI framework, applied to the spatially distributed GR-like and kinematic wave model, involving a pair of process-parameterization and regionalization neural networks. The pair of neural networks is used to (i) correct internal fluxes using "neutralized" (by interception reservoir, terminology of GR models from Perrin et al. (2003); Santos et al. (2018)) atmospheric data and (ii) estimate the model parameters using physical descriptors, with their weights optimized through high-dimensional optimization algorithms using an adjoint model to obtain accurate gradients of the cost function.

the fluxes that feed the delayed and direct transfer branches:

$$\tilde{Q}_{dl} = 0.9(1 - {f_{q,3}}^2)P_r \text{ and } \tilde{Q}_{dr} = 0.1(1 + 9{f_{q,3}}^2)P_r \quad (6)$$

A non conservative exchange flux $\tilde{F}$, applied to both the transfer reservoir and the fast transfer branch (direct runoff), is obtained by applying a correction term $f_{q,4}$, predicted by $\phi_1$, to the classical GR exchange flux $F$:

$$\tilde{F} = (1 + f_{q,4})F \text{ with } F = k_{exc} \times \left(\frac{h_t^-}{c_t}\right)^{7/2} \quad (7)$$

where $k_{exc}$ and $c_t$ represent respectively the exchange coefficient and the capacity of the transfer reservoir predicted by $\phi_2$, and $h_t^-$ the transfer reservoir state at previous time step. The outflow subtracted to the transfer reservoir is:

$$Q_{rs}(t) = h_t^- - \left((h_t^-)^{-4} + {c_t}^{-4}\right)^{-1/4} \tag{8}$$

The remaining net rainfall flux $\tilde{Q}_{dr}$ feeds the direct transfer branch where the exchange flux $\tilde{F}$ is also applied, and its outflow is $Q_{rd} = \tilde{Q}_{dr} + \tilde{F}$. The hydrological runoff flux produced at pixel scale is $Q_{lat} = Q_{rs} + Q_{rd}$ and is routed over a 2D mesh with a simple hydraulic routing module.

Note that, the values of $f_{q,i=1..4}$, which are the outputs of $\phi_1$, are bounded between -1 and 1 (section 2.2.3). Thus, by definition, the transformation functions applied to these internal flux corrections (i.e., $1+f_{q,1}$, $1-f_{q,3}^2$, etc.) enable the preservation of the original conceptual model structure when the neural network output is zero, as all transformations equal 1 in this case. These terms were defined based on the specific fluxes being corrected and mathematical constraints. For example, the correction $f_{q,3}$ is squared to ensure non-negativity, and its transformations ($1-f_{q,3}^2$ and $1+9f_{q,3}^2$) are specifically designed to preserve mass conservation in the transfer branch partitioning, as $0.9(1-f_{q,3}^2)+0.1(1+9f_{q,3}^2) = 1$ for any value of $f_{q,3}$, while allowing the model to learn and adjust the partition between delayed and direct transfer branches from their default values of 0.9 and 0.1 respectively.

### 2.2.2 Pixel-to-pixel flow routing of runoff with a partial differential equation

The routing module used here is based on a conceptual 1D kinematic wave model that is numerically solved with a linearized implicit numerical scheme (Te Chow et al., 1988). The discharge routing problem is classically reduced to a 1D problem by considering a "D8" drainage plan $\mathcal{D}_\Omega(x)$, obtained by terrain digital elevation model processing with the condition that a unique pixel has the highest drained area.

The kinematic wave model is a PDE obtained by simplifying the 1D Saint-Venant equations assuming that the momentum reduces to flow friction slope equal bottom slope. Using a conceptual parameterization of the momentum $A = a_{kw} Q^{b_{kw}}$, with $A$ the flow cross sectional area, $Q$ the discharge, $a_{kw}$ and $b_{kw}$ two parameters to be estimated, and injecting it into the mass equation $\partial_t A + \partial_x Q = Q_{lat}$, with $Q_{lat}$ the lateral discharge (total runoff produced at a pixel from GR operators presented above), a single-equation model is obtained. The model is discretized with a classical finite differences approach (cf. Te Chow et al. (1988)) resulting in the following expression for the discharge propagation model:

$$\partial_x Q + a_{kw} b_{kw} Q^{(b_{kw}-1)} \partial_t Q = Q_{lat} \tag{9}$$

### 2.2.3 Learnable mappings for spatialized GR-like model on top of kinematic wave routing

In this study, we use two multilayer perceptrons, the first one $\phi_1$ for spatio-temporal corrections of the model internal fluxes $\boldsymbol{f}_q(x,t)$ and the second $\phi_2$ for spatialized parameters $\boldsymbol{\theta}(x)$ regionalization as used in Huynh et al. (2024b). Namely, $\phi$ consists of a pair of neural networks designed to ingest (i) neutralized atmospheric inputs $\boldsymbol{\mathcal{I}}_n = (P_n, E_n)$ (using the wording of GR conceptual model (Perrin et al., 2003; Santos et al., 2018)), along with the model states at previous time step $\boldsymbol{h}(x,t-1)$, for correcting spatio-temporal internal fluxes $\boldsymbol{q}$ (process-parameterization pipeline) and (ii) physical descriptors $\boldsymbol{\mathcal{D}}$ (refer to Appendix A for information on the studied descriptors) for estimating spatialized hydrological parameters $\boldsymbol{\theta}$ (regionalization pipeline), as shown in Equation 10.

$$\phi : \begin{cases} \boldsymbol{f}_q(x,t) & = \phi_1\left(\boldsymbol{\mathcal{I}}_n(x,t), \boldsymbol{h}(x,t-1); \boldsymbol{\rho}_1\right) \\ \boldsymbol{\theta}(x) & = \phi_2\left(\boldsymbol{\mathcal{D}}(x); \boldsymbol{\rho}_2\right) \end{cases} \tag{10}$$

with $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ the vector of trainable parameters, invariant to the spatial coordinate $x$ over $\Omega$, of the (pair of) neural network(s). Note that more advanced neural networks, such as CNN, RNN, or LSTM, can be explored in future studies. For instance, applying a CNN to the regionalization neural network $\phi_2$ is possible and has been implemented into SMASH, but not investigated since it is out of scope of this paper.

Here, the first neural network $\phi_1$ has a single hidden layer with 16 neurons, followed by a Leaky ReLU activation function. The output layer uses a TanH activation function, which is bounded from -1 to 1. Then, the flux corrections $\boldsymbol{f}_q = (f_{q,i=1..4})^T$, predicted by $\phi_1$, are applied for each pixel $x$ and time $t$ to correct simultaneously the internal fluxes of the GR hydrological operators as described in section 2.2.1. The second network $\phi_2$ consists in 3 hidden layers with 96, 48 and 16 neurons. ReLU activation functions are used between hidden layers, while the Sigmoid function is applied in the output layer and followed by a scaling function to constrain the model parameters in accordance with their feasible bounds. The vector of conceptual spatialized parameters, mapped by $\phi_2$, is $\boldsymbol{\theta} = (c_p, c_t, k_{exc}, a_{kw}, b_{kw})^T$ composed of production and transfer reservoir capacities $c_p$ and $c_t$, exchange coefficient $k_{exc}$, kinematic wave parameters $a_{kw}$ and $b_{kw}$. Finally, the parameter control vector to optimize is $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$, i.e., the weight and biais of the process-parameterization and regionalization mappings.

### 2.3 Inverse problem and analysis of the hybrid physics-AI framework

Given observed and simulated discharge time series $\boldsymbol{Q}^* = (Q^*_{g=1..N_G})^T$ and $\boldsymbol{Q} = (Q_{g=1..N_G})^T$ with $N_G$ being the number of gauges over the study domain $\Omega$, the model misfit to multi-catchment observations is measured through a cost function $J$, as shown in Equation 11.

$$J(\boldsymbol{Q}^*, \boldsymbol{Q}) = \sum_{g=1}^{N_G} w_g j(Q^*_g, Q_g) \tag{11}$$

where $\sum_{g=1}^{N_G} w_g = 1$ (with $w_g = 1/N_G$ in this study), and $j(Q^*_g, Q_g) = 1 - NSE(Q^*_g, Q_g)$ at each gauge, with NSE being the quadratic Nash-Sutcliffe efficiency. Thus, $J$ is a convex and differentiable function, involving the response of the forward model $\mathcal{M}$ through its output $\boldsymbol{Q}$, and consequently depending on the model parameters $\boldsymbol{\theta}$ and the flux corrections $\boldsymbol{f}_q$, hence on the parameters $\boldsymbol{\rho}$ of the ANNs (cf. Equation 10). Accordingly, the VDA optimization problem is formulated as shown in Equation 12.

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} J\left(\boldsymbol{Q}^*, \mathcal{M}_{rr-hy}(., \phi(., \boldsymbol{\rho}))\right) \tag{12}$$

This high-dimensional inverse problem can be tackled through gradient-based optimization algorithms. A limited-memory quasi-Newton approach, such as L-BFGS-B (Zhu et al., 1997), is suitable for smooth objective functions, while an adaptive learning rate approach, exemplified by Adam (Kingma and Ba, 2014), is effective for non-smooth objective functions. These approaches necessitate obtaining the cost gradient with respect to the parameters sought $\nabla_{\boldsymbol{\rho}} J$, achieved through numerical code differentiability rules and automatic differentiation using the Tapenade engine (Hascoet and Pascual, 2013).

After optimization with the proposed approach, enabling to jointly learn physical processes parameterization and regionalization, a hybrid process-based spatially distributed calibrated hydrological model $\mathcal{M}_{\hat{\boldsymbol{\rho}}}$ is obtained and is therefore reusable for space-time extrapolation. Contrarily to PINNs where the physical model residual serves as a weak constrain in optimization, in our proposed conceptualization, the physics is used as a strong constrain. In this sense, the approach can be seen as a learnable spatialized physical model. Moreover, contrarily to PINNs and LSTM, which are composed of neural networks only, our hybrid model is physically interpretable through its conceptual parameters $\boldsymbol{\theta}(x)$, internal states $\boldsymbol{h}(x,t)$ and fluxes $\boldsymbol{q}(x,t)$. Moreover, the ANNs $\phi_1$ and $\phi_2$ coupled with the conceptual model $\mathcal{M}_{rr-hy}$ at the pixel scale for each time step, are capable of capturing non-linear and multi-resolution effects. The conceptualization, where the physics is used as a strong constrain in the forward model, enables using other differentiable hydrological and hydraulic models for example, on structured or unstructured meshes. Such an approach enables integrating data that are not directly usable nor explicitly represented in the model such as the physical descriptors for regionalization of conceptual parameters here.

## 3 Data and experimental design

We evaluated our method on two data sets (see Figure 2). The first data set includes 235 non-nested catchments selected from Hashemi et al. (2022), which is part of a larger data set containing 4,190 French catchments provided by the INRAE-HYCAR research unit (Delaigue et al., 2020; Brigode et al., 2020). The second data set consists of 21 catchments, a subset of the ArcMed region, taken from Huynh et al. (2024b).

The SMASH model is run on a spatial grid with a resolution of $dx = 1\ \mathrm{km}$ and a temporal step of $dt = 1\ \mathrm{h}$. It is forced by the following data:

– Discharge: Collected by the French Ministry of Environment, covering the period of the forcing data and extracted from the HydroPortail platform[1].

– Rainfall: We use rainfall data from the ANTILOPE J+1 radar observation reanalysis, which merges radar data with in-situ gauge observations. This data is provided by Météo-France at a grid resolution of $1\ \mathrm{km}^2$, matching the resolution of the model grid rasters.

– Potential Evapotranspiration (PET): Temperature data for calculating PET is sourced from the SAFRAN[2] reanalysis, provided by Météo-France at a resolution of $8 \times 8\ \mathrm{km}^2$ (Quintana-Seguí et al., 2008; Vidal et al., 2010). The PET is then computed using the Oudin formula (Oudin et al., 2005) and has the same resolution as the rainfall data.

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

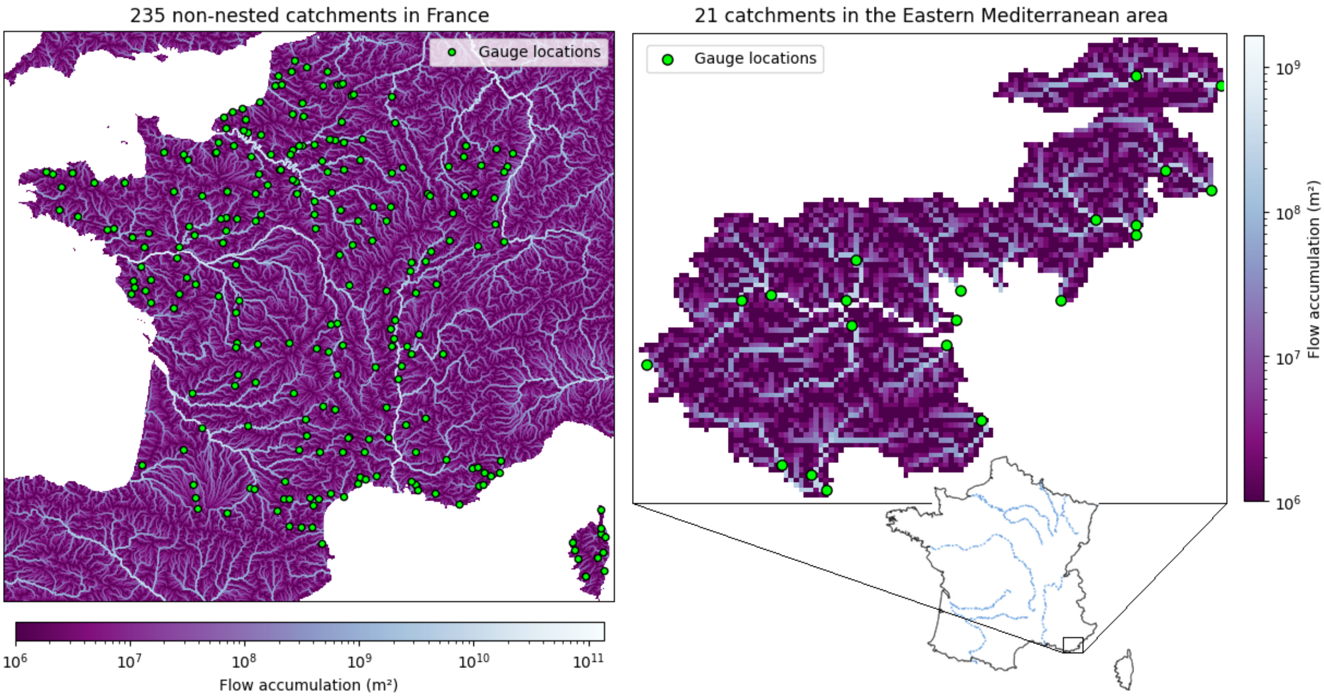


**Figure 2.** Study areas used for evaluation. The first area consists of 235 non-nested catchments in France, while the second area includes 21 catchments in a multiple-catchment setup in the Eastern Mediterranean region, representing contrasting hydrological conditions.

The first data set contains hourly time series over a 13-year period (August 2006 to July 2019) for downstream gauges only. It is used to evaluate single-gauge optimization (local calibration) without regionalization, solely focusing on the process-parameterization neural network $\phi_1$, which is the key novelty of this study. The 13-year period is divided into two segments: the calibration period covers the first 7 years (including a one-year warmup), and the remaining 6 years are used for temporal validation. Four methods are compared to evaluate the learning capacity of the neural network $\phi_1$:

– Two classic GR models with spatially uniform parameters (GR.U) and spatially distributed parameters (GR.D), which, in some cases, exhibit under- or over-parameterization issues in the spatially distributed hydrological model;

– Two hybrid GR models that integrate the neural network $\phi_1$ (called $\phi_1$-hybrid) with spatially uniform parameters (GRNN.U) and spatially distributed parameters (GRNN.D).

The second data set includes hourly time series over 7 years (August 2009 to July 2016) for both nested and independent catchments in the Eastern Mediterranean region (known as "MedEst"). This data set is used to assess the relevance of the learnable structure for simultaneous multi-gauge regionalization with physical descriptors. A set of seven descriptors (Table A1 and Figure A1 in Appendix A), with a spatial resolution of $0.01°$ in the WGS 84 projection, encompassing various types such as topography, morphology, land use, and hydrogeology, is used as inputs for the regionalization mapping $\phi_2$. The MedEst region presents a challenging case due to its contrasting hydrological properties, including steep topography and highly heterogeneous soils and bedrock (e.g., Garambois et al., 2015). This region is prone to intense rainfall events that trigger non-linear flash flood responses and contains a significant proportion of karstic zones. The first 4-year time series, including a one-year warmup period, is used for calibration, while the remaining 3 years are used for validation. Four methods are compared to evaluate the learning capacity of both the process-parameterization neural network $\phi_1$ and the regionalization neural network $\phi_2$:

- The classic GR model with regional, spatially uniform parameters (GR.U);
- The hybrid GR model integrating the neural network $\phi_1$ ($\phi_1$-hybrid) to correct internal fluxes in hydrological processes, with regional, spatially uniform parameters (GRNN.U);
- The classic GR model integrating the neural network $\phi_2$ ($\phi_2$-hybrid) to learn the mapping between physical descriptors and spatially distributed hydrological parameters (GR.NN);
- The hybrid GR model integrating both neural networks $\phi_1$ and $\phi_2$ (GRNN.NN), representing the fully integrated hybrid approach ($\phi$-hybrid) among the studied methods.

We recall that the design of the architecture and the hyper-parameters of both neural networks were determined as described in section 2.2.3.

# 4 Results and discussion

In this section, we first present the performance of the hybrid spatially distributed models tested on both data sets. Then, we will provide further interpretation and discussion of the learning process to analyze the proposed framework and enhance the understanding of hydrological behaviors in the process-based model through internal fluxes.

## 4.1 Model performance analysis

### 4.1.1 Local calibration over 235 French catchments

Figure 3 illustrates typical simulated streamflows from different methods for small, medium, and large catchments. Overall, we observe significant improvements in the hybrid methods compared to the classic models in simulating both peak flows and low flows. For example, in the case of the medium catchment, GRNN.U more accurately predicts the peak flows in January 2014 compared to GR.U, while also reliably reproducing the low flows, particularly during the period between 2018 and 2019.

Figure 4 shows a global comparison of performances in terms of Nash–Sutcliffe efficiency (NSE), Kling–Gupta efficiency (KGE), and root mean squared error (RMSE) across both calibration and validation periods for different methods. The results suggest that $\phi_1$-hybrid methods (GRNN.U and GRNN.D) consistently achieve superior efficiency scores and lower errors compared to the classic models (GR.U and GR.D) in various scenarios. In calibration, both hybrid models outperform the classic ones, with significantly higher median NSE scores (0.85 and 0.86 compared to 0.79 and 0.83), a narrower and higher interquartile range, and a shorter lower whisker.

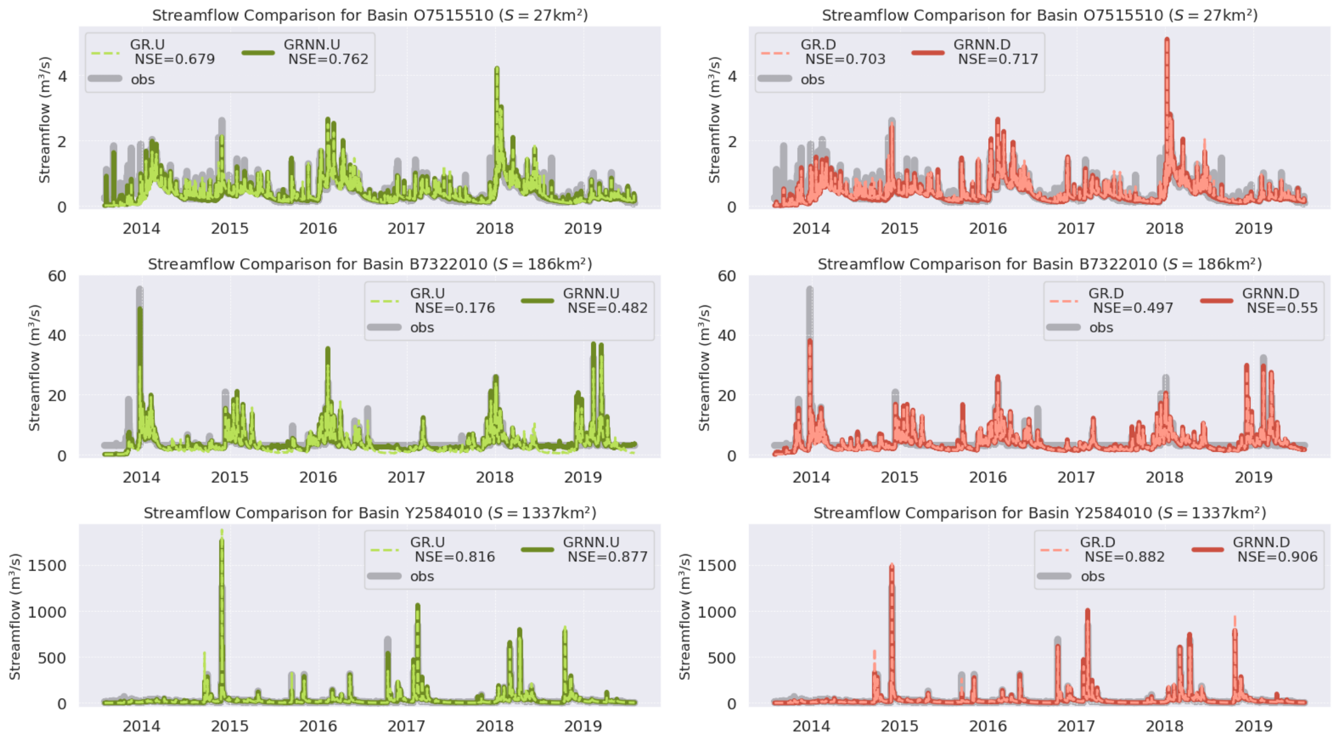


**Figure 3.** Comparison of streamflow simulation across representative small, medium, and large catchments randomly selected among the 235 catchments in France during the validation period.

In temporal validation, GRNN.U achieves a median NSE of 0.73 compared to 0.76 for GR.D, and both models reach a median KGE of 0.75, while GRNN.U shows a lower median RMSE of 1.38 compared to 1.42 for GR.D. Although median improvements may appear small, it is important to consider the entire distribution. In addition to the median values, we observe notable enhancements in other statistical measures, such as the interquartile range (0.25 and 0.75 quantiles) and whiskers in the boxplots. For catchments that already exhibit satisfactory performance, the effect of hybridization is relatively small, leading to similar median, 0.75 and 0.95 quantile values. However, for poorly performing basins, the hybrid models provide substantial improvements, as evidenced by enhanced performance in the lower quartiles. Notably, the hybrid model with spatially uniform hydrological parameters (GRNN.U) performs comparably to, and in some cases surpasses, the classic GR model with spatially distributed parameters (GR.D). This result is promising, as it demonstrates that while the original model with spatially uniform conceptual parameters (GR.U) inherently leads to under-parameterization compared to GR.D, this limitation can be compensated for by the spatially distributed flux correction in GRNN.U.

To evaluate flood simulation performance, Huynh et al. (2023) introduced a method to compute several flood event signatures using an automatic segmentation algorithm. These signatures help depict the model behavior during flood events. Relative error is used as the evaluation metric to quantify the difference between simulated and observed flood event signatures, including

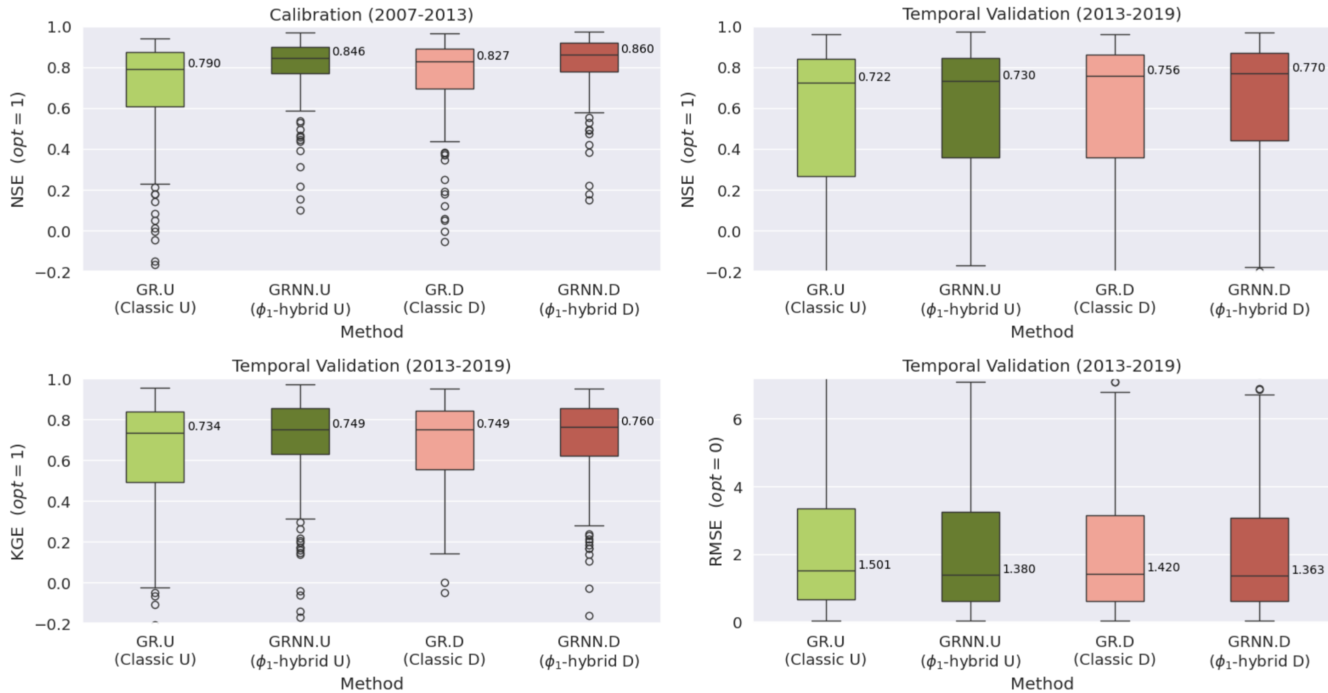


**Figure 4.** Model performance comparison of local calibration methods across 235 catchments in France. The two boxes in lighter colors represent the classical GR models (GR.U and GR.D), while the two boxes in darker colors represent the $\phi_1$-hybrid models (GRNN.U and GRNN.D). The evaluation is based on simulated discharges over the calibration and validation period, using NSE, KGE, and RMSE metrics.

peak flow, runoff coefficient, flood flow, and baseflow. Figure 5 shows the cumulative distribution function (CDF) of the relative error for these signatures, based on over 2,700 flood events that occurred during the 6-year validation period. The hybrid models achieve the best performance, outperforming the classic GR.U model, with their CDF lines consistently above. Notably, the hybrid model GRNN.U, using only spatially uniform parameters, attains performance comparable to or even better than the classic model with spatially distributed parameters (GR.D). GRNN.U shows similar performance to GR.D in reproducing peak flows (with the same median error of 0.32) and flood flow (both with a median error of 0.28), while performing slightly better in reproducing the runoff coefficient (0.20 compared to 0.22) and baseflow (0.21 compared to 0.22). This highlights the strength of the hybrid process-parameterization framework, particularly its relevance in improving flood modeling systems.

### 4.1.2 Multi-catchment regionalization over a flash-flood prone Mediterranean area

Here, we investigate how the hybrid models perform in the study of a multi-catchment regionalization setup. In calibration, it is evident from Figure 6 that the $\phi_2$-hybrid model (GR.NN) and the $\phi$-hybrid (GRNN.NN) model, both using the regionalization neural network $\phi_2$, outperform the models with lumped parameters (GR.U and GRNN.U). Notably, the fully integrated hybrid model GRNN.NN dominates the radar plot, with a large shape extending toward the outer edges, fully enveloping the other

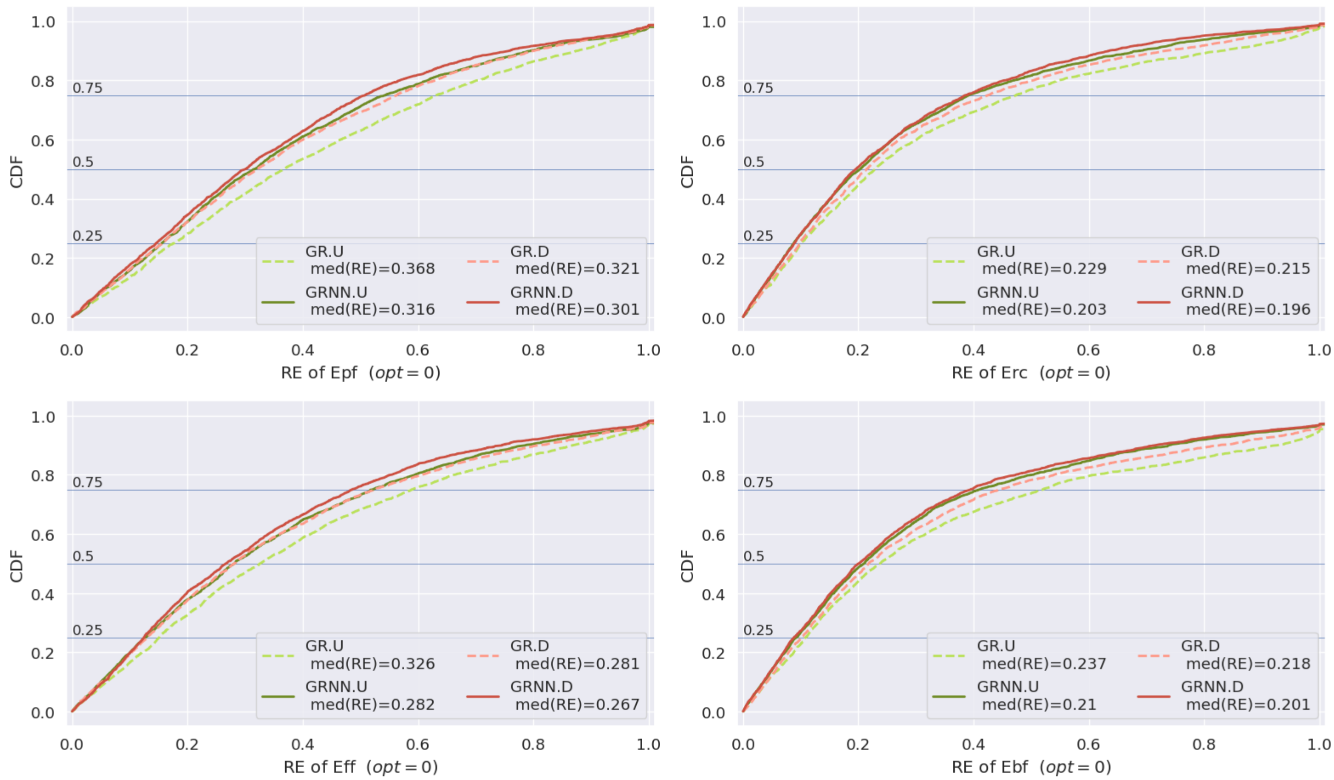


**Figure 5.** Comparison of model performance in simulating flood event signatures, presented as the cumulative distribution function (CDF) of the relative error (RE) between observed and simulated values for peak flow (Epf), runoff coefficient (Erc), flood flow (Eff), and baseflow (Ebf). The evaluation is based on 2,718 flood events across 235 catchments during the validation period (08/2013–07/2019).

methods. Although GRNN.U clearly falls short of the two regionalization-based models, it still shows a significant improvement over the classic GR.U model. Similar results can be observed in both temporal and spatio-temporal validation, as seen in the boxplots. This proves that using physical descriptors with a learnable mapping is an effective approach in this regionalization setup (multi-catchment in a large, flash flood-prone area with high spatio-temporal resolution data), compared to lumped models (without physical descriptors) or simpler regionalization methods (e.g., multi-linear or multi-polynomial mappings) as demonstrated in Huynh et al. (2024b). Interestingly, while the $\phi$-hybrid model GRNN.NN, which delivers the best overall performance, shows a moderate gain over GR.NN–with median NSE scores of 0.75 (compared to 0.72) and 0.51 (compared to 0.48) in temporal and spatio-temporal validation–the $\phi_1$-hybrid model GRNN.U, that uses lumped parameters without regionalization using physical descriptors, makes a dramatic improvement over the classic GR.U model (median NSE of 0.56 compared to 0.14, and 0.43 compared to 0.16). In this way, learning internal flux corrections has made it possible to improve the regionalizability of a distributed conceptual hydrological model even with spatially uniform conceptual parameters (i.e., without using physical descriptors). This may represent a compelling research direction for reducing structural uncertainty

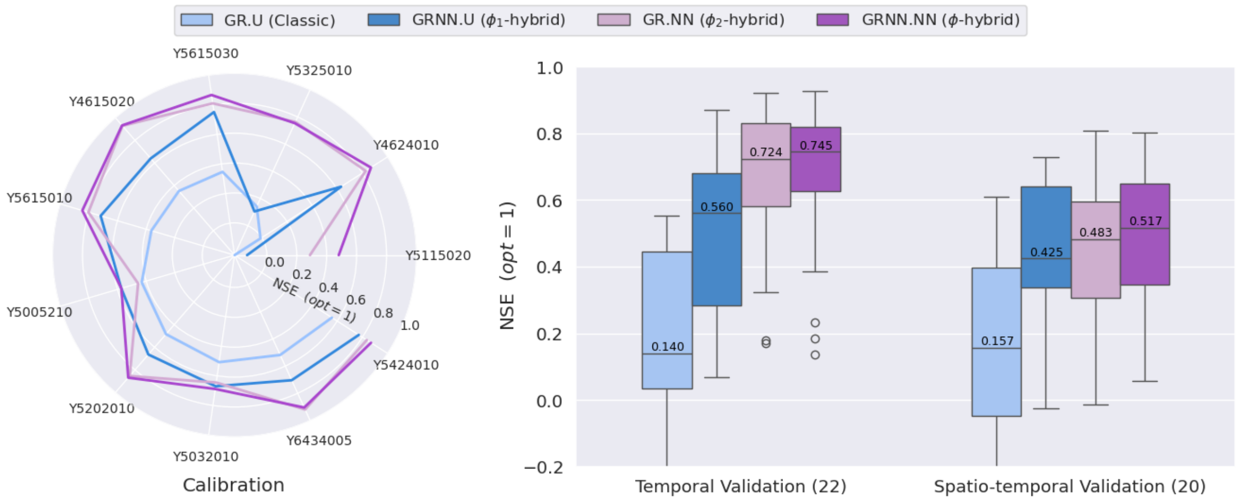


**Figure 6.** Comparison of multi-catchment regionalization performance for different methods. The evaluation is based on NSE scores computed over three periods: the 3-year calibration period (excluding a 1-year warmup), the first 18 months, and the last 18 months of the 3-year validation period for 11 calibration catchments and 10 validation catchments. The numbers in parentheses on the boxplots indicate the total number of samples evaluated, with the validation period split into two 18-month periods (sample counts are doubled for all catchments).

in modeling, using a minimum of data and enabling more efficient extraction of multi-scale information through hybrid flux correction using GRNN.U, with potential for flexible semi-spatializations of conceptual parameters and even proximity-based regionalizations for spatially dense gauging networks (cf. Oudin et al. (2008)).

In the context of flood prediction, the hybrid models (GRNN.U, GR.NN, and GRNN.NN) consistently yield superior performance compared to the classic GR.U model. Figure 7a presents the RMSE and NSE metrics computed using short time series from 143 flood events that occurred across the entire MedEst area during the validation period from August 2013 to July 2016 (similar graphs for additional evaluation metrics are shown in Figure B4 in Appendix B). Figure 7b shows typical streamflow simulations, demonstrating a significant enhancement for both hybrid process-parameterization and regionalization-based approaches, compared to classical methods, in simulating high flow characteristics and behaviors during flood events.

Although the NSE values computed for the 143 flood events are relatively low across all models, it is important to note that NSE for flood events–which are short time series with high values–is highly sensitive to small timing errors. Even slight discrepancies in peak timing can lead to substantial decreases in NSE. Additionally, data and modeling uncertainties may vary between events, making the accurate prediction of highly contrasted events particularly challenging. The models are calibrated on the entire time series, and we evaluate the validation results specifically for flood events, where classical approaches often struggle to accurately estimate water dynamics. This discrepancy highlights the difficulty in capturing the rapid and intense nature of flood events, even with advanced hybrid models. This underscores the need to investigate potential sources of error, including input data quality and model structural limitations, as well as the impact of using a calibration metric based solely on flood events. These factors could explain the overall challenges in flood event simulation.

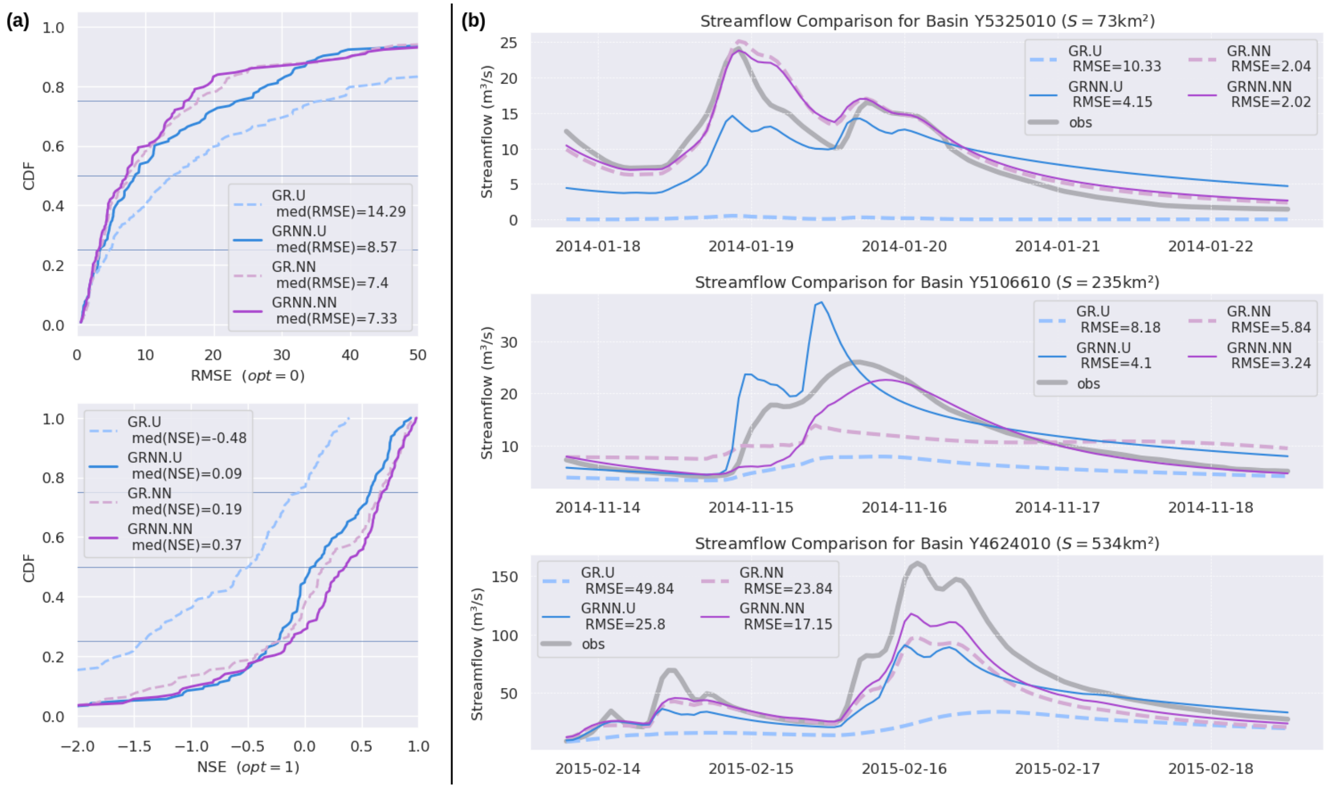


**Figure 7.** (a) Comparison of NSE and RMSE metrics computed for 143 flood events detected in the validation period across 21 catchments in the MedEst region. (b) Comparison of streamflow simulations for representative small, medium, and large catchments during several flood events selected in the validation period.

## 4.2 Towards learning hydrological behaviors

Here, we focus on uncovering the hydrological behaviors inferred with the hybrid approach consisting in neural networks embedded into a physical model for learnable correction of internal fluxes. In the studied hybrid structure GRNN, the learned correction of GR-like model consists in 4 flux corrections $\boldsymbol{f}_q(x,t) = (f_{q,i}(x,t))_{i=1..4} = \phi_1(P_n, E_n, h_p, h_t)(x,t)$ for each pixel and time step of the simulation domain from atmospheric forcings and previous model states. A positive (or negative) correction of $f_{q,i}$ (where $i = 1, 2, 4$), with values bounded in $]-1;1[$ due to the TanH activation function used in the output layer, results in an increase (or decrease) in the original fluxes $P_s$, $E_s$, and $|F|$–the absolute value of $F$ (cf. Equations 4, 5, 7), thereby influencing the simulated mass balance. Meanwhile, $f_{q,3}^2$, with values in $[0;1[$, produces a conservative re-repartition of net rainfall $P_r$ between direct and delayed transfer branches (cf. Equation 6), thereby affecting the subgrid transfer dynamics. The following quantitative analysis begins with the spatio-temporal averages of these flux corrections and proceeds to explore their variability across the 235 independently calibrated catchments, as well as in the regionalization test case.

### 4.2.1 Analysis of internal flux corrections

Figure 8 shows the maps of spatio-temporal average flux corrections (the corrections are first averaged spatially within each catchment, then temporally across the calibration period), obtained through local calibrations using GRNN.U (see Figure B1 in Appendix B for GRNN.D). Red and blue indicate positive and negative flux corrections in the spatio-temporal average,

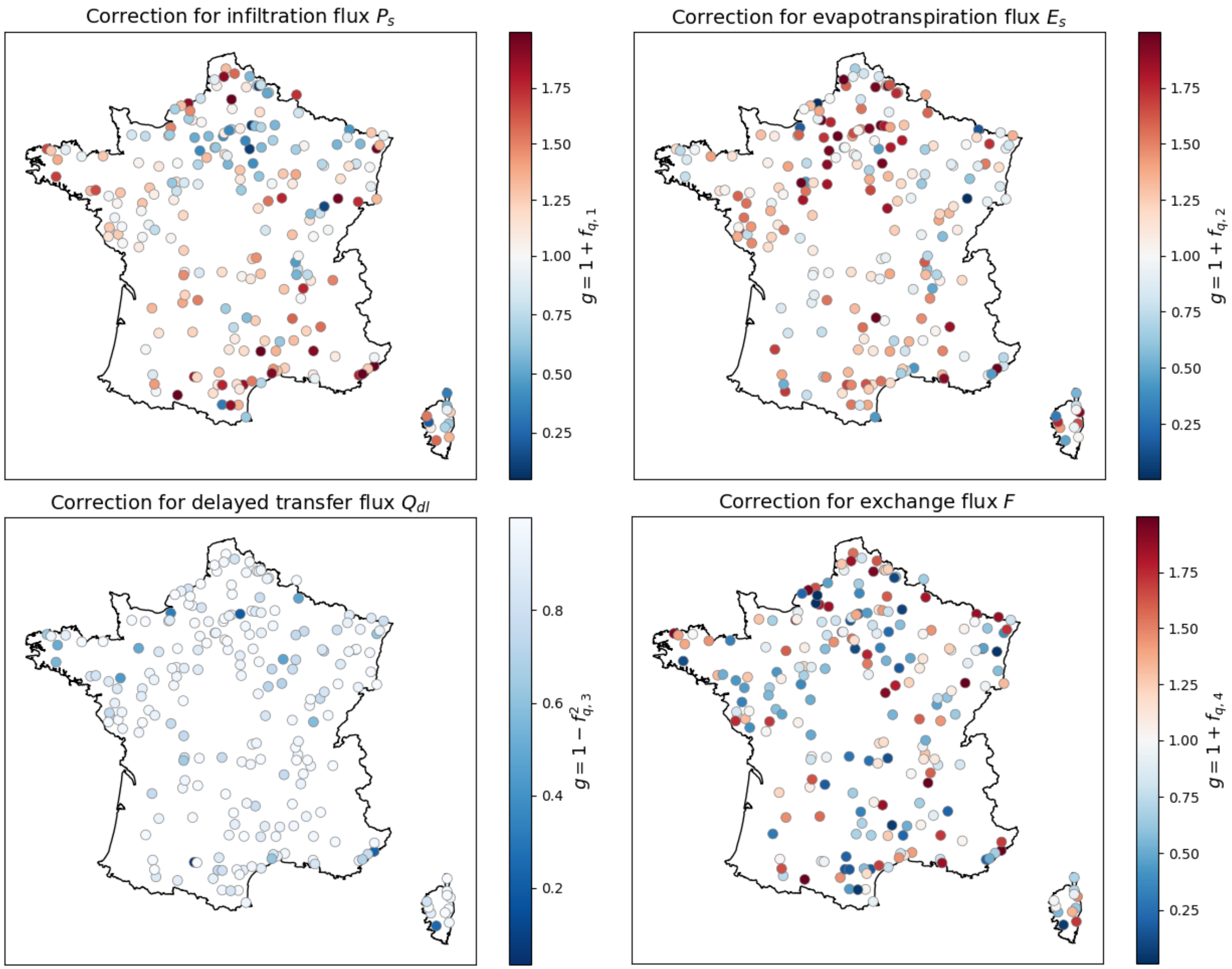


**Figure 8.** Maps of spatio-temporal average flux corrections $\overline{\overline{g\left(\boldsymbol{f}_q\right)}^{x\in\Omega_j}}^{t}$, where $j = 1..N_g$, for the $N_g = 235$ catchments, obtained through local calibrations of spatially uniform parameters with the hybrid model structure (GRNN.U). The function $g(.)$ represents the transformation applied to the neural network output $\boldsymbol{f}_q$, which may differ depending on the specific flux being corrected. Red indicates corrections that tend to increase the current flux, while blue indicates corrections that reduce it, with white representing minimal or no effective correction.

respectively. For fluxes affecting the production reservoir, namely infiltrating rainfall $P_s$ and evapotranspiration $E_s$, the average corrections show opposite signs for the majority of basins and the same sign for a minority. These maps reveal different trends of flux corrections across France. Several regions exhibit strong corrections (either negative or positive) for $P_s$ and $E_s$, while others show near-zero corrections (white points with values close to 1). However, the exchange flux is generally the most influenced by the corrections, as indicated by the dark colors across the maps, playing a crucial role in refining the model's state dynamic. Note that the transformation function $1 - f_{q,3}^2$ applied to correct the delayed transfer flux $Q_{dl}$ result in reduction of this flux and conservative augmentation of $Q_{dr}$ flux feeding the direct branch.

Now, we turn to the analysis of time series of spatially averaged flux corrections presented in Figure 9 (see Figure B2 for GRNN.D). Figure 9a illustrates these corrections over time for three randomly selected catchments that are representative of the 235 French catchments. We observe opposite signs in the corrections $f_{q,1}$ and $f_{q,2}$ for infiltration $P_s$ and evapotranspiration $E_s$

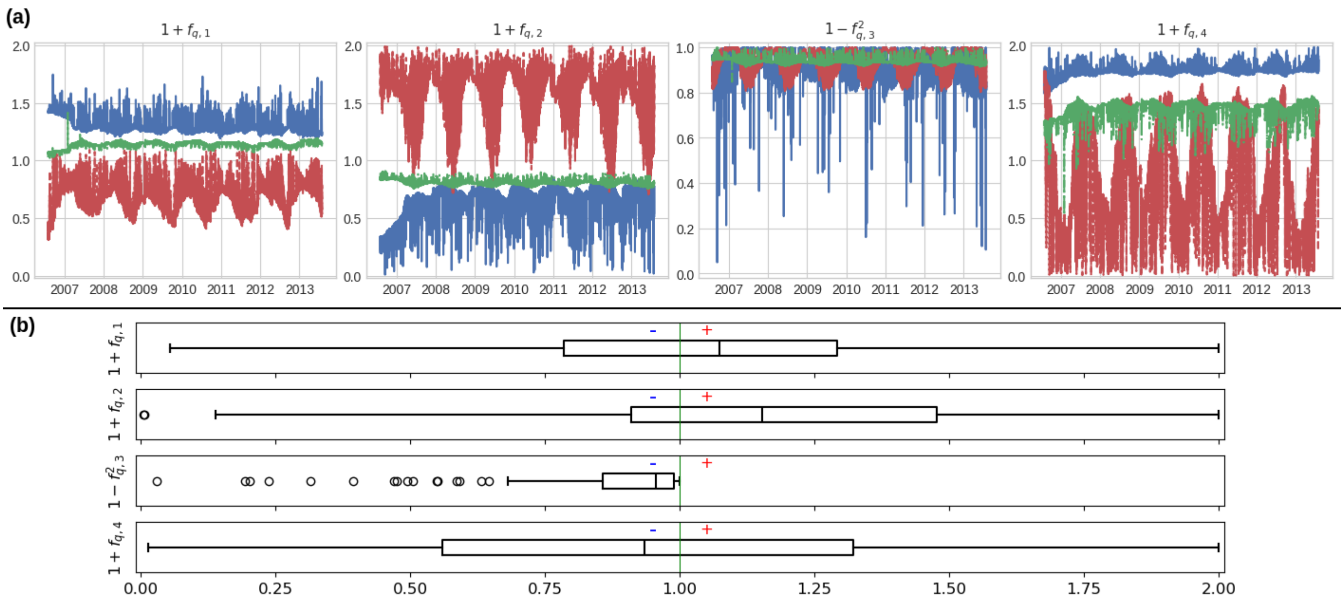


**Figure 9.** (a) Spatially averaged flux correction time series $\overline{g\left(\boldsymbol{f}_q(t)\right)}^{x\in\Omega_j}$ for three randomly selected catchments from the set of 235. These are obtained through local calibrations of spatially uniform parameters with the hybrid model structure (GRNN.U). (b) Boxplot of spatio-temporal average flux corrections $\overline{\overline{g\left(\boldsymbol{f}_q\right)}^{x\in\Omega_j}}^{t}$, where each boxplot represents 235 catchment-specific spatio-temporal averages.

from the production reservoir. Catchments exhibiting positive corrections to $P_s$ and negative corrections to $E_s$ suggest that more water is directed towards the production reservoir and less is lost by evapotranspiration, leading to increased moisture state. Conversely, in catchments with negative corrections for $P_s$ and positive corrections for $E_s$, reduced infiltration and increased evapotranspiration imply lower moisture states. Furthermore, periodic behaviors are observed over time for corrections of the four water fluxes, highlighting the temporal patterns of flux corrections. This pattern likely reflects the footprint of the annual periodicity of the production state $h_p$, which is an input to the neural network $\phi_1$. Overall, the corrected infiltrating rainfall $\tilde{P}_s$ is generally 10% higher than the original, as indicated by the median of $1 + f_{q,1}$ being approximately 1.10 in Figure 9b. This implies an increased water level in the production reservoir, and hence more water being directed there rather than feeding the transfer branch. This observation somehow explains why the production capacity $c_p$, calibrated for the hybrid models, is generally slightly higher than that of the classical models (see Table B1 in Appendix B). Additionally, fewer corrections are obtained for re-repartition of net rainfall flux into the direct and delayed transfer branches (i.e., the corrections show less variation and are closer to 1). Negative corrections that tend to reduce the flux magnitude are applied to the delayed transfer branch $\tilde{Q}_{dl}$, which implies positive corrections for the direct transfer branch $\tilde{Q}_{dr}$. In this case, the hybrid model suggests that more water reaches the outflow $Q_{rd}$ via the direct transfer branch. Both transfer branches are affected by the exchange flux $\tilde{F}$ for which, across most catchments, a reduction is obtained by flux correction (the median of $1+f_{q,4}$ is lower than 1).

In multi-gauge regionalization setup, distinct spatio-temporal patterns emerge over the MedEst area, as shown in Figure 10 for GRNN.U (corresponding results for GRNN.NN shown in Figure B5 in Appendix B). Figure 10a illustrates that the spatially averaged corrections for infiltrating rainfall $P_s$ show relatively high temporal variability; moreover, they still exhibit stable periodic patterns after the first-year warmup. The spatial average of the corrected flux $\tilde{P}_s$ tends to be lower during moderate-rain

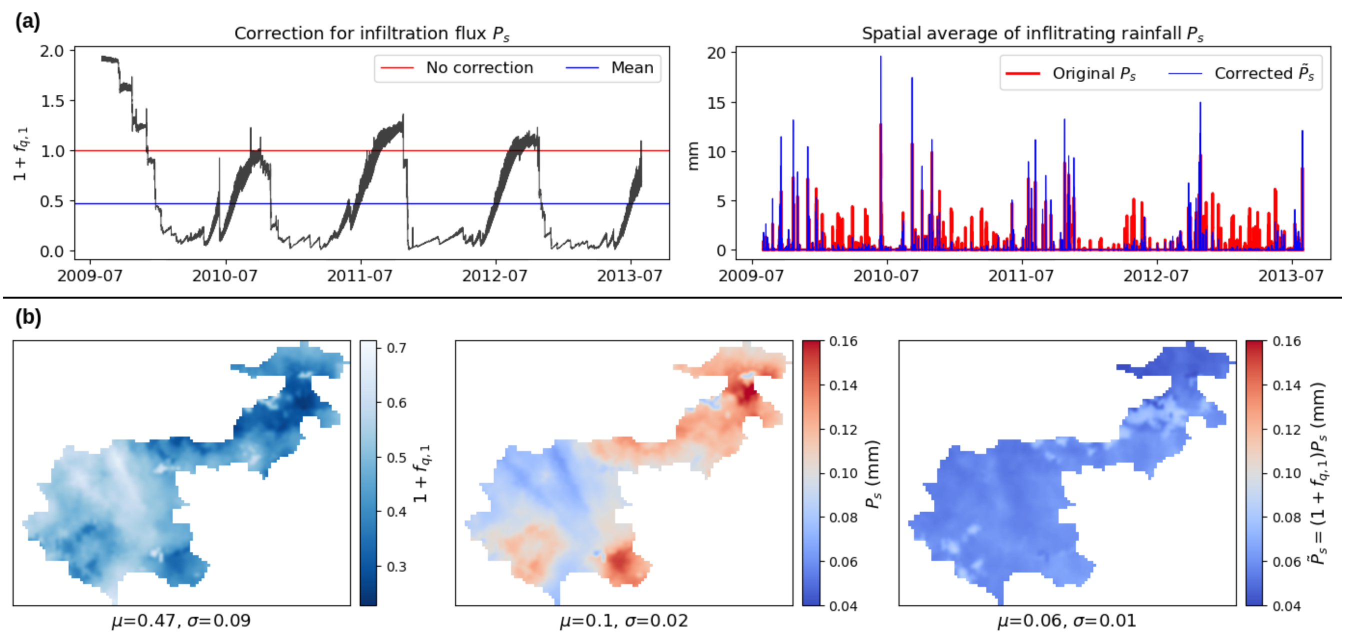


**Figure 10.** Visualization of flux corrections in the MedEst region obtained through regional calibration of spatially uniform parameters with the hybrid model (GRNN.U): (a) Spatial average of infiltrating flux correction $\overline{1+f_{q,1}(t)}^x$, original and corrected infiltrating rainfall $\overline{P_s(t)}^x$, $\overline{\tilde{P}_s(t)}^x$; (b) Maps of time-averaged infiltrating flux correction $\overline{1+f_{q,1}(x)}^t$, original and corrected infiltrating rainfall $\overline{P_s(x)}^t$, $\overline{\tilde{P}_s(x)}^t$, where $\mu$ and $\sigma$ represent the spatial average and standard deviation.

events, while it is higher during high-rain events compared to the original flux $P_s$. This suggests that the hybrid model directs more rainfall into the transfer branch during moderate-rain events (which may have longer duration), while the opposite behavior is observed for high-rain events (which can be shorter in duration). The spatial maps of time-averaged flux corrections in

Figure 10b further indicate that the hybrid model generally applies negative corrections, reducing the spatio-temporal mean of infiltrating rainfall from $0.1$ mm to $0.06$ mm. Interestingly, these maps also reveal spatial variability in internal flux corrections, which may explain the improved regionalizability of the hybrid GRNN models, as demonstrated by its performance in spatio-temporal validation, even with spatially uniform conceptual parameters (without regionalization using physical descriptors). While temporal patterns of flux corrections (e.g., periodic behaviors) emerge from the production and transfer states, similar to the case of the 235 catchments, spatial patterns are likely due to the spatial variability of atmospheric data.

#### 4.2.2 Hybridization effect on main mass fluxes involved in basin's water balance

This section examines the effect of $\phi_1$-hybridization on the primary mass fluxes involved in the hydrological mass balance, as simulated using the original GR-like spatially distributed model structure. For a given catchment domain $\Omega$, the annual catchment-scale flux $\Psi_{f,\mathcal{A}}$ of a state-flux $f(x,t)$–such as actual evapotranspiration, exchange flux, or runoff flux–simulated using either the classical model or the hybrid model (with flux corrections omitted for brevity) is computed as follows:

$$\Psi_{f,\mathcal{A}} = \frac{1}{|\Omega|} \int\limits_{t\in\mathcal{A}} \int\limits_{x\in\Omega} f(x,t)dxdt \tag{13}$$

where $\mathcal{A}$ denotes the annual period, $|\Omega|$ represents the drainage area.

A basin scale analysis is performed for each of the 235 French basins simulated, focusing on the flux of rainfall $P$ inflowing the model and three key fluxes affected by the $\phi_1$-hybridization: evapotranspiration $E_s$ from production store, exchange $F$, and pixel-scale discharge $Q_{lat}$ prior to routing. The annual average of each flux is calculated using Equation 13, and the interannual averages of these water gain or loss fluxes over the 6-year calibration period (2007–2013) are shown in Figure 11. This figure quantitatively illustrates the impact of $\phi_1$-hybridization on the classical GR-like model, with uniform conceptual parameters for each basin and each model structure. Over the variety of hydrological behaviors and annual rainfall regimes of this large catchment set, it is noteworthy that hybridization results show almost no change for nearly all basins in terms of interannual discharge runoff volume, with a median of 246.6 mm for GR.U and 251.3 mm for GRNN.U and similar quantiles, while dynamic changes have been obtained as suggested by improved NSE, flood signatures, and hydrographs (cf. performance analysis in section 4.1), as well as internal flux corrections (such as infiltration and repartition between direct and delayed lateral transfer branches).

The simulated water balance is influenced by the correction of $k_{exc}$, which represents the exchange flux and can result in either gains or losses relative to the original model (which is already non-conservative). In this case, the exchange flux is moderately affected by hybridization, with a median trend of reduced exchange from -23.5 mm to -13.2 mm. This reduction is compensated, in terms of water balance, by an increase in evaporation from the production reservoir (from 254.5 mm to 265.1 mm in median). Notably, both fluxes exhibit a larger interquartile range across basins compared to the classic model structure. Therefore, the proposed $\phi_1$-hybridization enables the learning of spatio-temporal corrections of internal model dynamics, resulting in physically interpretable fluxes that remain within imposed ranges and lead to overall model improvement.

Figure 12 depicts the versatile nature of the learnable hybrid model in comparison to classical conceptual models for correcting internal fluxes and vividly illustrates the learned non-linear relationship between the corrected net rainfall and neutralized

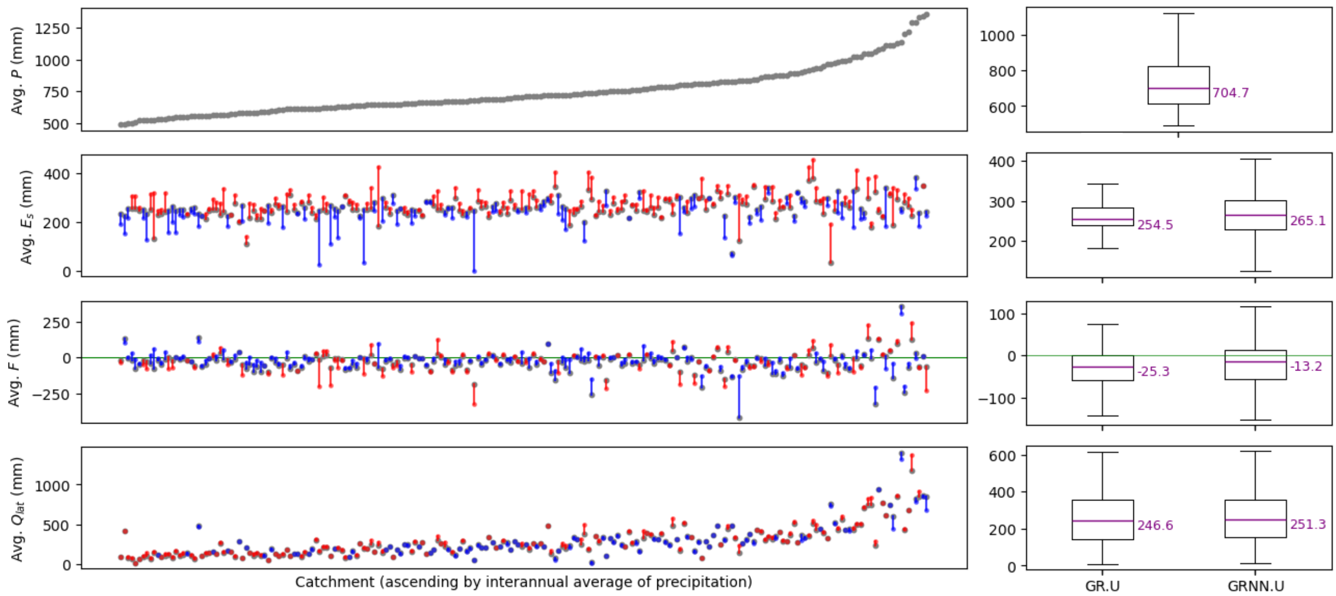


**Figure 11.** Comparison of mass fluxes affecting water balance in interannual mean at basin scale over the 235 catchment set, for the classical GR model (GR.U) and the $\phi_1$-hybrid model (GRNN.U) using uniform mapping "U" for conceptual parameters $\boldsymbol{\theta}$, in local calibration. The x-axis shows 235 catchments, sorted by their average precipitation. Grey dots represent the interannual averages of precipitation $P$, actual evapotranspiration $E_s$, exchange flux $F$, and lateral discharge $Q_{lat}$ over the calibration period for the classical model. Red dots and lines represent increases according to the hybrid model, while blue ones indicate decreases. For cases where $F < 0$, red indicates a larger magnitude in $F$ for the hybrid model (more negative), while blue indicates a lesser magnitude (closer to zero).

data, as well as internal states. The model response surface of the net rainfall $P_r = P_n - \tilde{P}_s$, obtained with the corrected infiltrating rainfall $\tilde{P}_s = (1 + f_{q,1})P_s$, is shown for different levels of the production state $h_p$ and neutralized rainfall $P_n$. Interestingly, this corrected net rainfall $P_n - \tilde{P}_s$, regardless of the level of production state (i.e., $h_p = 0.3, 6$, and $15$ mm), exhibits a non-monotonic behavior with respect to the intensity of neutralized rainfall $P_n$ (Figure 12a). However, this non-monotonic behavior becomes less pronounced as the production state $h_p$ approaches the production capacity $c_p$. Figure 12b further clarifies the non-linear response surface, showing that the corrected net rainfall undergoes two changes in monotonicity as the neutralized rainfall when the reservoir is less than half utilized ($h_p < c_p/2$). In contrast, when the reservoir is fully or nearly fully utilized ($h_p \approx c_p$), the corrected flux $P_n - \tilde{P}_s$ behaves similarly to the original flux $P_n - P_s$. Interestingly, a non-linear infiltration behavior is obtained after learning with the hybrid GRNN model structure, especially for drier conditions of the production reservoir where classical GR models are known to fail in flood generation (cf. Astagneau et al. (2021)). Further research could focus on deeper analysis of learned physical behaviors, for example by investigating the approximation of learned behaviors with known mathematical functions. One could also investigate how to impose physical a priori using other mathematical expressions directly into the forward model structure, for example to impose a explicit monotonicity or even a

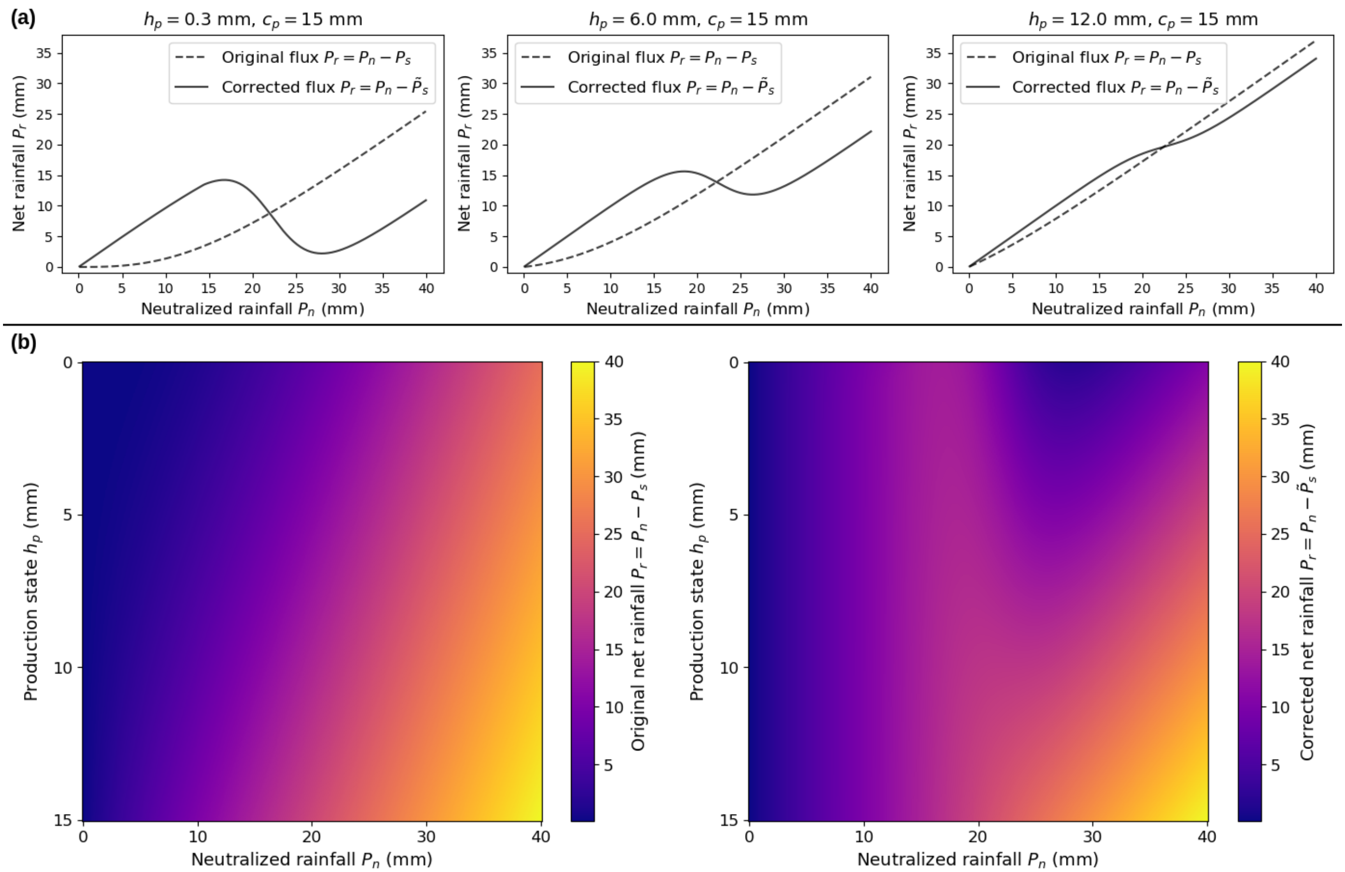


**Figure 12.** Hydrological interpretation of the non-linear response surface obtained using the learned flux correction neural network for infiltrating rainfall $\tilde{P}_s = (1 + f_{q,1})P_s$, plotted with a production capacity of $c_p = 15$ mm and neutralized evapotranspiration of $E_n = 0$ mm: (a) Original and corrected net rainfall $P_r$ for different levels of the production state $h_p$; (b) Response surface of original and corrected net rainfall $P_r$ as a function of both the production state $h_p$ and neutralized rainfall $P_n$.

shape of a dependency, such as the rainfall intensity related modifications of original lumped GR model in Astagneau et al. (2022).

### 4.3 Research perspectives and further discussion

This article proposed a spatially distributed hybrid GR-like model and a comprehensive analysis of a large catchment sample. Future research should concentrate on refining the model's hybridization strategy to enhance its applicability across even larger datasets (e.g., the CARAVAN database in Kratzert et al. (2023)) and to improve extrapolation capabilities for extreme hydro-meteorological events. This quest for generalized structures of spatially distributed hydrological models requires scalable hybrid solvers applicable over very large domains. Immediate work will focus on developing a SMASH version for parallel

GPU-based forward-inverse computation, and adapting the $\phi$-hybrid model to a state-space GR model (cf. Santos et al. (2018)), thereby enabling the investigation of additional non-linearities in hydrological model differential equations. In addition, improving the routing model may deliver a more realistic flood wave propagation. Such improvement could be based on the use of known hydraulic models (e.g., kinematic wave in Roux et al. (2011); Vergara et al. (2016), non-inertial 1D or 2D in Fleischmann et al. (2020), or full 1D Saint-Venant at network scale in Larnier et al. (2025)), and/or the use of fine topography data such as LiDAR made during low flows (capturing a significant part of river bathymetry), and/or the use of more observations of flow depth, extent, and velocity.

Although hybrid models have more degrees of freedom compared to classical GR models, it is important to note that the inputs and outputs of the flux correction model are physically consistent and of the same dimension as the original model. This design allows the hybrid model to learn nonlinearities in the internal flux laws, which we analyze thoroughly in the flux correction analysis in both time and space throughout the paper. The hybrid models do not necessarily have more conceptual parameters (maintaining the same number of reservoirs and connections here), they do introduce more nonlinearity in the internal flux laws corrections with neural network $\phi_1$. This added complexity effectively increases the model's degrees of freedom while maintaining robustness in both spatial and temporal validations, as demonstrated by numerical results.

It is worth noting that the two neural networks will not be extrapolated in the same way when the model is used in prediction. The regionalization neural network $\phi_2$ will not be extrapolated as long as the model is used in the region it has been calibrated. At the opposite, the flux correction neural network $\phi_1$ is bound to be extrapolated since its inputs $(P_n, E_n, h_p, h_t)$ are varying in time, so that the range observed in calibration will be exceeded sooner or later. This is particularly the case when the model is calibrated locally, as done in the first case study involving 235 French catchments. By contrast, a multi-catchment regionalization setup (Mediterranean case study) is advantageous since it offers more opportunity to expose $\phi_1$ to extreme values of its inputs. Quantifying the uncertainty affecting the estimated parameters of the neural networks would be useful to raise awareness of a likely loss of precision when $\phi_1$ is extrapolated, but this comes with many difficulties (e.g., Papamarkou et al., 2022). An alternative approach would be to look for parsimonious regressions that are able to adequately reproduce the behavior revealed by $\phi_1$, while being amenable to uncertainty quantification.

Finally, the proposed hybrid hydrological framework should be extended to other models structures, as other GR or VIC available in SMASH platform, but also using more complex physics based modeling approaches and hypothesis testing such as in Douinot et al. (2018) with various subsurface flow modelings. Note that the proposed physics-AI framework for spatially distributed modeling could help unifying top-down approaches such as GR or other data based conceptual models with bottom-up physics based hydrological models that suffer from (up)scaling problems of physical laws and parameterization. In the context of relatively sparse discharge data compared to model dimensionality, such a model discovery process could greatly benefit from the wealth of surface information provided by remote sensing. This includes data on terrain and vegetation properties, surface moisture, snow cover, surface temperature, and total water storage (Meyer Oliveira et al., 2021), along with river network data (e.g., river flow surface topography variability through altimetry and imagery), which necessitates a differentiable river network hydraulic model to achieve coherence with hydraulic observables while enabling the inference of complex and large spatio-temporal parameters from heterogeneous data (Larnier et al., 2025). Such a model would also support information feedback from these data to the hydrological model within a differentiable H&H coupling framework (Pujol et al., 2022).

## 5 Conclusions

This article introduces a hybrid physics-AI framework that integrates neural networks to infer spatio-temporal internal fluxes and spatially distributed conceptual parameters within a differentiable, gridded hydrological model, all encapsulated in a VDA algorithm. Numerical results from local calibration-validation across 235 French catchments and regionalization in a complex, flash flood-prone area demonstrate the superiority of the hybrid models. These models excel not only in performance scores during both calibration and validation but also in producing physically interpretable results, with improved representations of simulated hydrological behavior.

The proposed approach, relying on process-based equations hybridized with ANNs, allows obtaining interpretable spatially distributed hydrological models, contrarily to pure machine learning approaches, while taking advantage of non-linear and multi-resolution effects of neural networks. Accordingly, it is applicable to any other differentiable hydrological, hydraulic or geophysical model, on structured or unstructured meshes.

Future work aims to enhance the hybrid framework by: (i) studying the generalizability of structural corrections across larger data sets and diverse model structures; (ii) investigating more complex neural networks, including deeper ANNs to capture multi-scale information over larger data sets in global optimization, or at the opposite simpler tools that could reproduce the behavior revealed by the ANNs, while facilitating uncertainty quantification; (iii) exploring mathematical properties, such as equifinality issues between neural networks and conceptual parameters, and analyzing the response surfaces of universal differential equation sets for flexible hydrological modeling in time and space; and (iv) coupling with differentiable river network hydraulic models to improve 1D-2D hydrodynamic realism. This coupling will enable feedback by assimilating hydraulic observations into a differentiable H&H chain (Pujol et al., 2022), such as the unprecedented hydraulic visibility (Garambois et al., 2017) brought by SWOT (Surface Water and Ocean Topography) and multi-satellite data (e.g., with VDA in Pujol et al. (2020); Malou et al. (2021)). Such differentiable and learnable H&H modeling frameworks are expected to enhance the representation of basins internal state fluxes and enable the efficient fusion of machine learning with process-based modeling, advancing the discovery of scale-relevant hydrological laws through the maximal extraction of information from multi-source data.

*Code and data availability.* The data sets that support this study comprise preprocessed data sourced from SCHAPI-DGPR and Météo-France, and are available at https://doi.org/10.5281/zenodo.13826145 (Huynh, 2024). The proposed algorithms were implemented into the SMASH source code, Version 1.1-dev, which is preserved at https://doi.org/10.5281/zenodo.13696078 (Huynh and Colleoni, 2024), available via GNU-3 license and developed openly at https://github.com/DassHydro/smash.

## Appendix A: Input Physical Descriptors for Learning Regionalization

Table A1 and Figure A1 provide information on the physical descriptors used as input data for regionalization learning methods. Note that before the optimization process, all descriptors are standardized between 0 and 1 using min-max scaling.

**Table A1.** Descriptors used as input data for regionalization methods.

| Notation | Type | Description | Unit | Source |
|---|---|---|---|---|
| $d_1$ | Topography | Slope | ° | Odry (2017) |
| $d_2$ | Morphology | Drainage density | - | Organde et al. (2013) |
| $d_3$ | Influence | Percentage of basin area in karst zone | % | Caruso et al. (2013) |
| $d_4$ | Land use | Forest cover rate | % | Agency (2019) |
| $d_5$ | Land use | Urban cover rate (including artificial and non-vegetated areas) | % | Agency (2019) |
| $d_6$ | Hydrogeology | Potential available water reserve | mm | Poncelet (2016) |
| $d_7$ | Hydrogeology | High storage capacity basin rate | % | Finke et al. (1998) |

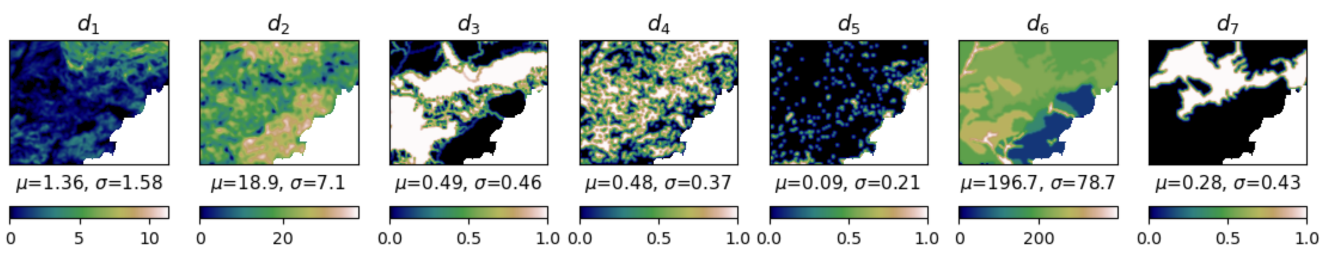


**Figure A1.** Maps of seven physical descriptors in the MedEst area at a resolution of 0.01° in the WGS 84 projection, where $\mu$ and $\sigma$ represent the spatial average and standard deviation for each descriptor.

## Appendix B: Further Results and Visualizations

Table B1 presents statistical quantities (mean, median, standard deviation) of the calibrated hydrological parameters across the 235 French catchments, obtained using different methods.

Figure B1, Figure B2, and Figure B3 present similar graphs to Figure 8, Figure 9, and Figure 11, but these are obtained using local calibrations of spatially distributed parameters with the hybrid model structure (GRNN.D).

Figure B4 shows graphs similar to those in Figure 7a but presents different evaluation metrics, including mean absolute error (MAE), percent bias (PBIAS), and peak flow ratio (PFR).

Figure B5 presents graphs similar to those in Figure 10, but these are obtained using the fully integrated hybrid model, which includes both the process-parameterization neural network and the regionalization neural network (GRNN.NN).

**Table B1.** Median (mean; standard deviation) of the calibrated hydrological parameters across the study area of 235 catchments.

| Method | $c_p$ | $c_t$ | $k_{exc}$ | $a_{kw}$ | $b_{kw}$ |
|---|---|---|---|---|---|
| GR.U | 281.2 (384.5; 283.8) | 155.4 (388.1; 758.6) | -0.58 (-2.15; 4.73) | 3.41 (6.5; 9.92) | 0.82 (0.77; 0.25) |
| GRNN.U | 344.8 (469.2; 398.9) | 150.7 (450.2; 1135) | -0.64 (-1.75; 5.37) | 4.8 (9.21; 12) | 0.82 (0.73; 0.29) |
| GR.D | 271.4 (379; 299.6) | 151.5 (461.5; 984.5) | -0.62 (-1.76; 3.67) | 4.6 (6.11; 6.46) | 0.74 (0.7; 0.19) |
| GRNN.D | 301.3 (411.1; 309) | 184.2 (524.8; 991.7) | -0.63 (-1.37; 3.4) | 4.62 (6.42; 6.9) | 0.74 (0.71; 0.19) |

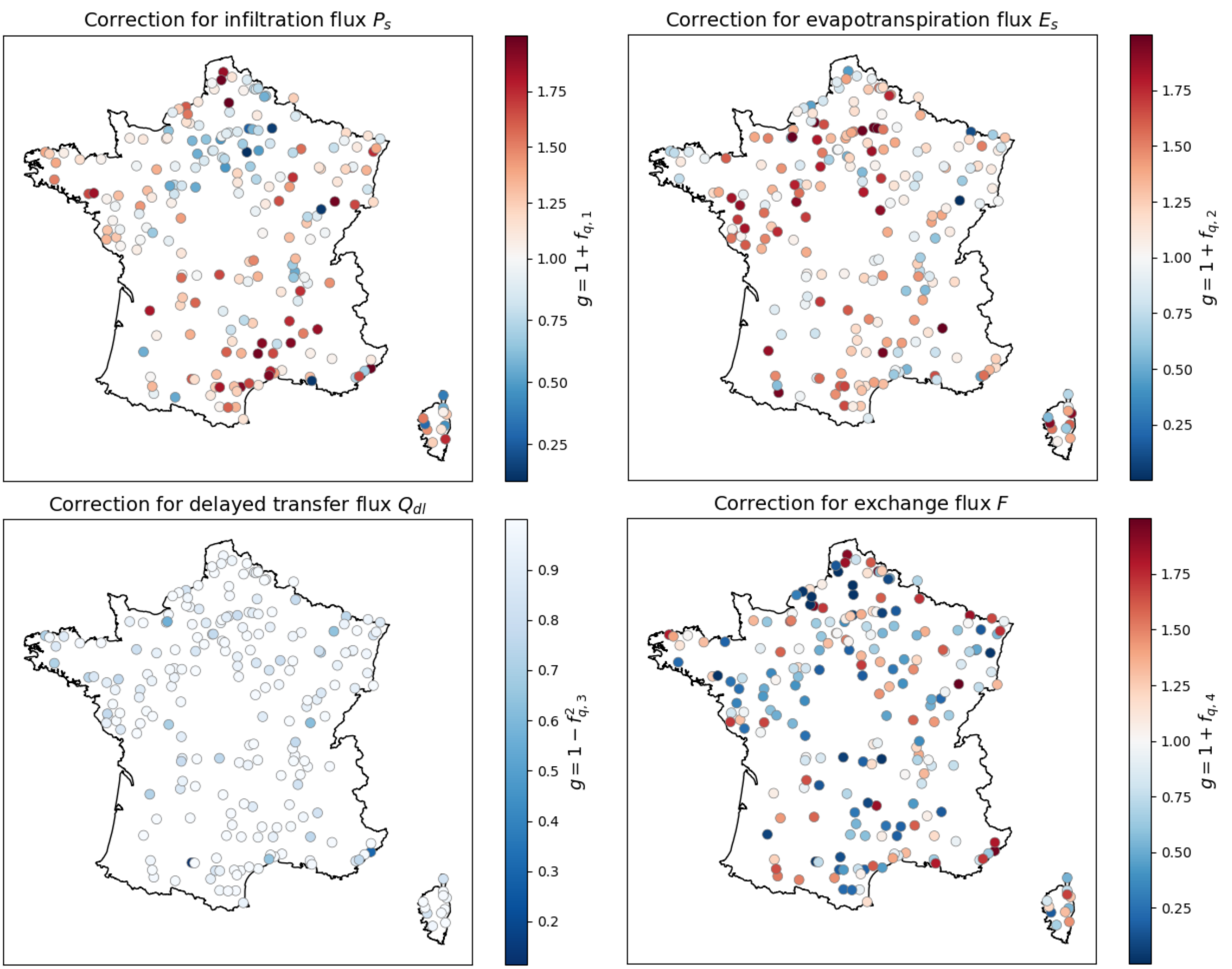


**Figure B1.** Maps of spatio-temporal average flux corrections $\overline{g\left(\boldsymbol{f}_q\right)}^{x\in\Omega_j}{}^{t}$, where $j = 1..N_g$, for the $N_g = 235$ catchments, obtained through local calibrations of spatially distributed parameters with the hybrid model structure (GRNN.D). The function $g(.)$ represents the transformation applied to the neural network output $\boldsymbol{f}_q$, which may differ depending on the specific flux being corrected. Red indicates corrections that tend to increase the current flux, while blue indicates corrections that reduce it, with white representing minimal or no effective correction.

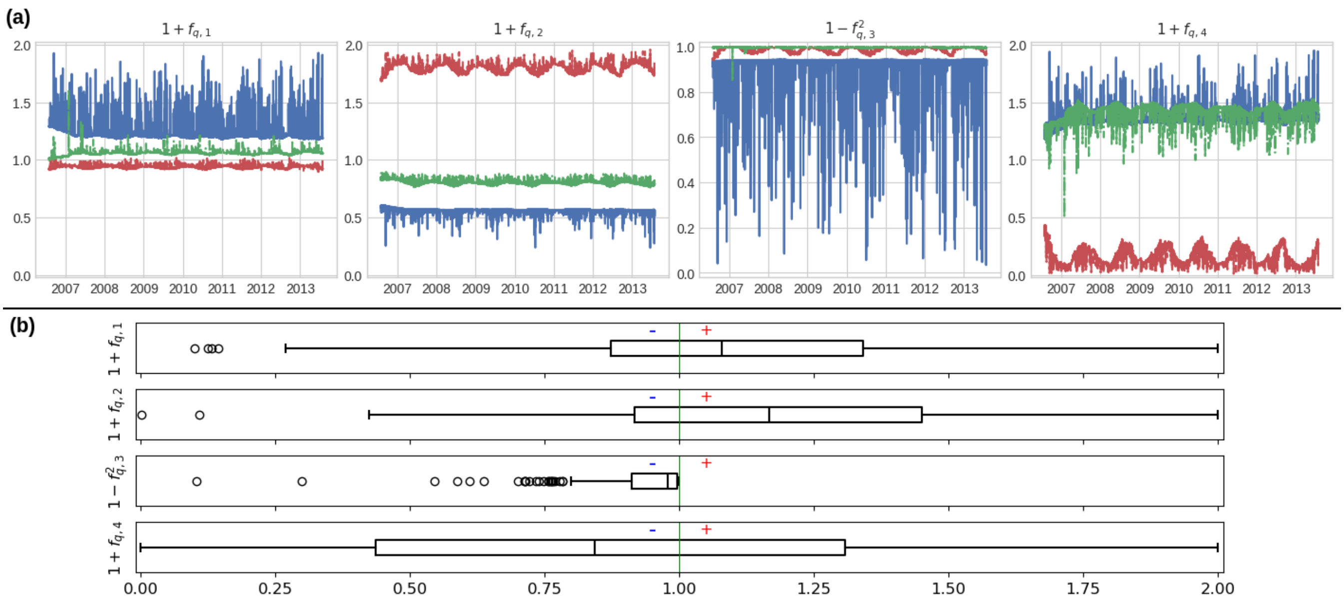


**Figure B2.** (a) Spatially averaged flux correction time series $\overline{g\left(\boldsymbol{f}_q(t)\right)}^{x\in\Omega_j}$ for the same three catchments selected in Figure 9a. These are obtained through local calibrations of spatially distributed parameters with the hybrid model structure (GRNN.D). (b) Boxplot of spatio-temporal average flux corrections $\overline{\overline{g\left(\boldsymbol{f}_q\right)}^{x\in\Omega_j}}^{t}$, where each boxplot represents 235 catchment-specific spatio-temporal averages.

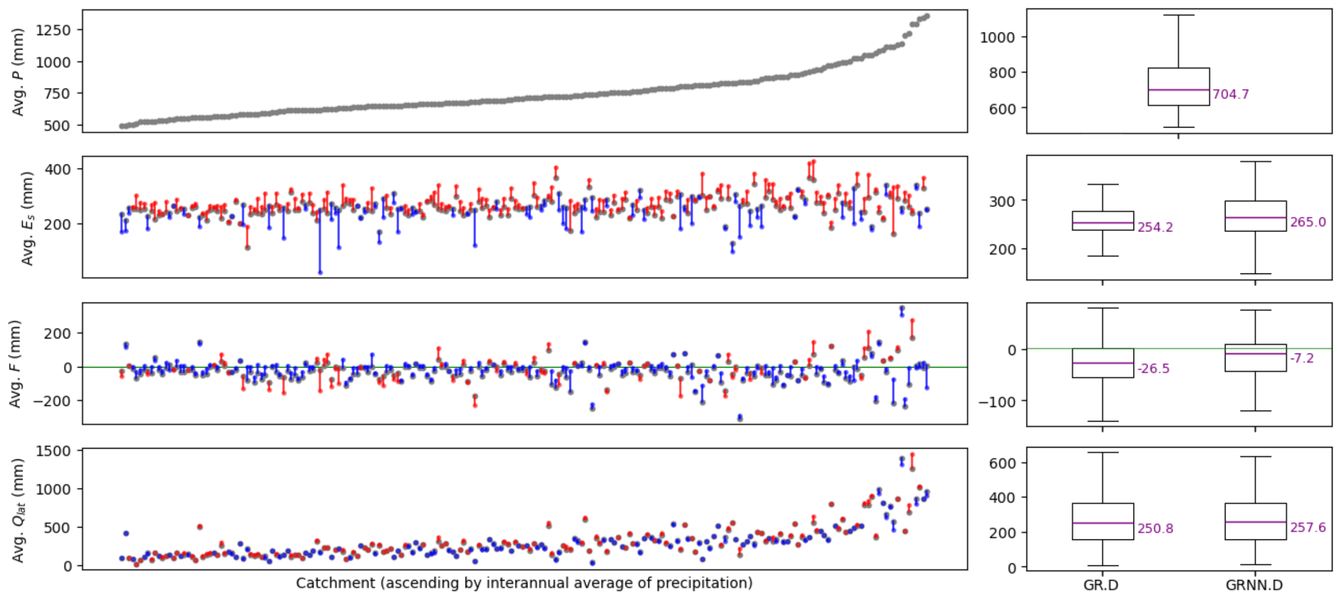


**Figure B3.** Comparison of mass fluxes affecting water balance in interannual mean at basin scale over the 235 catchment set, for the classical GR model (GR.D) and the $\phi_1$-hybrid model (GRNN.D) using distributed mapping "D" for conceptual parameters $\boldsymbol{\theta}$, in local calibration. The x-axis shows 235 catchments, sorted by their average precipitation. Grey dots represent the interannual averages of precipitation $P$, actual evapotranspiration $E_s$, exchange flux $F$, and lateral discharge $Q_{lat}$ over the calibration period for the classical model. Red dots and lines represent increases according to the hybrid model, while blue ones indicate decreases. For cases where $F < 0$, red indicates a larger magnitude in $F$ for the hybrid model (more negative), while blue indicates a lesser magnitude (closer to zero).

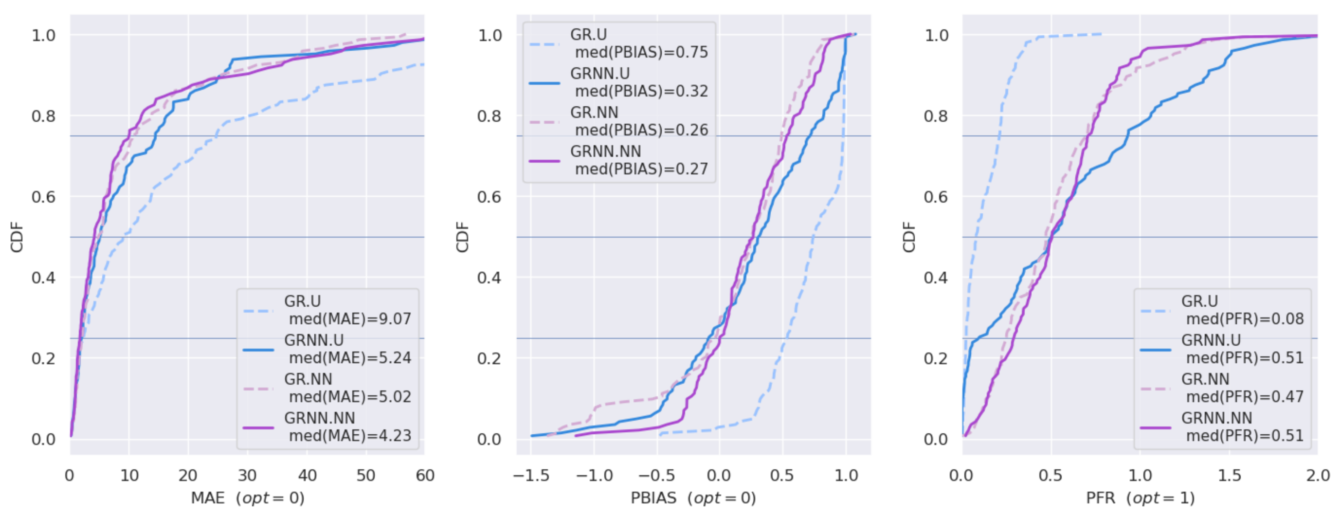


**Figure B4.** Comparison of mean absolute error (MAE), percent bias (PBIAS), and peak flow ratio (PFR) metrics computed for 143 flood events detected in the validation period across 21 catchments in the MedEst region.

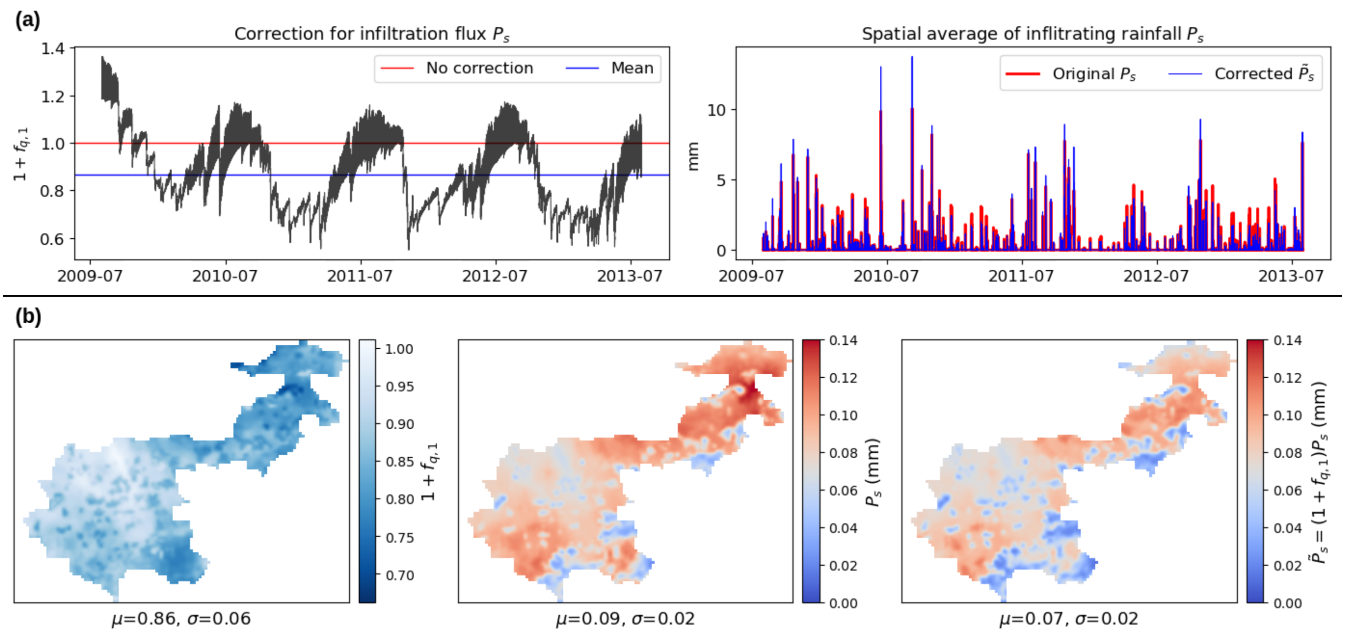


**Figure B5.** Visualization of flux corrections in the MedEst region obtained through regional calibration of spatially distributed parameters with the fully integrated hybrid model (GRNN.NN): (a) Spatial average of infiltrating flux correction $\overline{1 + f_{q,1}(t)}^x$, original and corrected infiltrating rainfall $\overline{P_s(t)}^x$, $\overline{\tilde{P}_s(t)}^x$; (b) Maps of time-averaged infiltrating flux correction $\overline{1 + f_{q,1}(x)}^t$, original and corrected infiltrating rainfall $\overline{P_s(x)}^t$, $\overline{\tilde{P}_s(x)}^t$, where $\mu$ and $\sigma$ represent the spatial average and standard deviation.

*Author contributions.* Conceptualization: NNTH and PAG; Software and data curation: NNTH, FC, and PAG; Experiment and computation: NNTH; Main writing: NNTH and PAG; Method and results analysis: NNTH, PAG, BR, JM, and HR; Final review: all; Funding and supervision: PAG.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* Pierre-André Garambois greatly acknowledges Pierre Brasseur from CNRS-IGE Grenoble for a fruitful discussion and a remark on “no directly exploitable first principles in hydrology and avenue for data assimilation and machine learning”. This work was supported by funding from SCHAPI-DGPR, ANR grant ANR-21-CE04-0021-01 (MUFFINS project, “MUltiscale Flood Forecasting with INnovating Solutions”, https://muffins-anr-project.hub.inrae.fr), and NEPTUNE European project DG-ECO.

## References


Agency, E. E.: CORINE Land Cover 2012 (raster 100 m), https://doi.org/10.2909/a84ae124-c5c5-4577-8e10-511bfe55cc0d, 2019.

Artigue, G., Johannet, A., Borrell, V., and Pistre, S.: Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin (southern France), Natural Hazards and Earth System Sciences, 12, 3307–3324, 2012.

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, 2021.

Astagneau, P. C., Bourgin, F., Andréassian, V., and Perrin, C.: Catchment response to intense rainfall: Evaluating modelling hypotheses, Hydrological Processes, 36, e14 676, 2022.

Beven, K.: Towards a new paradigm in hydrology, IN: Water for the Future: Hydrology in Perspective. IAHS Publication, 1987.

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.

Caruso, A., Guillot, A., and Arnaud, P.: Notice sur les indices de confiance de la méthode Shyreg-Débit–Définitions et Calculs, in: Aix en Provence: IRSTEA, Convention DGPR/SNRH, 2013.

Chen, C., Jiang, J., Liao, Z., Zhou, Y., Wang, H., and Pei, Q.: A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China, Journal of Hydrology, 607, 127 535, https://doi.org/10.1016/j.jhydrol.2022.127535, 2022.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D.: Neural Ordinary Differential Equations, https://arxiv.org/abs/1806.07366, 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.

Dooge, J. C. I.: Looking for hydrologic laws, Water Resources Research, 22, 46S–58S, https://doi.org/10.1029/WR022i09Sp0046S, 1986.

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

Feng, D., Liu, J., Lawson, K., and Shen, C.: Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy, Water Resources Research, 58, e2022WR032 404, https://doi.org/10.1029/2022WR032404, 2022.

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.

Finke, P., Hartwich, R., Dudal, R., Ibanez, J., Jamagne, M., King, D., Montanarella, L., and Yassoglou, N.: Geo-referenced soil database for Europe. Manual of procedures, version 1.0, European Communities, 1998.

Fleischmann, A. S., Paiva, R. C. D., Collischonn, W., Siqueira, V. A., Paris, A., Moreira, D. M., Papa, F., Bitar, A. A., Parrens, M., Aires, F., and Garambois, P. A.: Trade-Offs Between 1-D and 2-D Regional River Hydrodynamic Models, Water Resources Research, 56, e2019WR026 812, https://doi.org/10.1029/2019WR026812, 2020.

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

Garambois, P.-A., Calmant, S., Roux, H., Paris, A., Monnier, J., Finaud-Guyot, P., Samine Montazem, A., and Santos da Silva, J.: Hydraulic visibility: Using satellite altimetry to parameterize a hydraulic model of an ungauged reach of a braided river, Hydrological Processes, 31, 756–767, https://doi.org/10.1002/hyp.11033, 2017.

Hascoet, L. and Pascual, V.: The Tapenade automatic differentiation tool: principles, model, and specification, ACM Transactions on Mathematical Software (TOMS), 39, 1–43, 2013.

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

He, Q., Barajas-Solano, D., Tartakovsky, G., and Tartakovsky, A. M.: Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport, Advances in Water Resources, 141, 103 610, https://doi.org/10.1016/j.advwatres.2020.103610, 2020.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., and Fenicia, F.: Improving hydrologic models for predictions and process understanding using Neural ODEs, Hydrology and Earth System Sciences Discussions, pp. 1–29, 2022.

Huynh, N. N. T.: Medest Region and France 235-Catchments Data, https://doi.org/10.5281/zenodo.13826145, 2024.

Huynh, N. N. T. and Colleoni, F.: SMASH: Version 1.1-dev, https://doi.org/10.5281/zenodo.13696078, 2024.

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., and Javelle, P.: Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to Mediterranean floods, Journal of Hydrology, 625, 129 992, https://doi.org/10.1016/j.jhydrol.2023.129992, 2023.

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., Renard, B., Monnier, J., and Roux, H.: Multiscale Learnable Physical Modeling and Data Assimilation Framework: Application to High-Resolution Regionalized Hydrological Simulation of Flash Flood, Authorea Preprints, https://doi.org/10.22541/au.170709054.44271526/v2, 2024a.

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., Renard, B., Roux, H., Demargne, J., Jay-Allemand, M., and Javelle, P.: Learning Regionalization using Accurate Spatial Cost Gradients within a Differentiable High-Resolution Hydrological Model: Application to the French Mediterranean Region, Water Resources Research, 60, e2024WR037 544, https://doi.org/10.1029/2024WR037544, 2024b.

Jiang, S., Zheng, Y., and Solomatine, D.: Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning, Geophysical Research Letters, 47, e2020GL088 229, 2020.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Konapala, G., Kao, S.-C., Painter, S. L., and Lu, D.: Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US, Environmental Research Letters, 15, 104 022, https://doi.org/10.1088/1748-9326/aba927, 2020.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrology and Earth System Sciences, 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., et al.: Caravan-A global community dataset for large-sample hydrology, Scientific Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

Kumanlioglu, A. A. and Fistikoglu, O.: Performance Enhancement of a Conceptual Hydrological Model by Integrating Artificial Intelligence, Journal of Hydrologic Engineering, 24, 04019 047, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001850, 2019.

Larnier, K., Garambois, P.-A., Emery, C., Pujol, L., Monnier, J., Gal, L., Paris, A., Yesou, H., Ledauphin, T., and Calmant, S.: Estimating channel parameters and discharge at river network scale using hydrological-hydraulic models, SWOT and multi-satellite data, https://hal.inrae.fr/hal-04681079, hal open science, 2025.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436–444, https://doi.org/10.1038/nature14539, 2015.

Li, B., Sun, T., Tian, F., Tudaji, M., Qin, L., and Ni, G.: Hybrid hydrological modeling for large alpine basins: a semi-distributed approach, Hydrology and Earth System Sciences, 28, 4521–4538, https://doi.org/10.5194/hess-28-4521-2024, 2024.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water resources research, 43, 2007.

Maier, H. R. and Dandy, G. C.: Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, Environmental Modelling & Software, 15, 101–124, https://doi.org/10.1016/S1364-8152(99)00007-9, 2000.

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, 2021.

Meyer Oliveira, A., Fleischmann, A. S., and Paiva, R. C. D.: On the contribution of remote sensing-based calibration to model hydrological and hydraulic processes in tropical regions, Journal of hydrology (Amsterdam), 597, 126 184, 2021.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resources Research, 57, https://doi.org/10.1029/2020WR028091, 2021.

Odry, J.: Prédétermination des débits de crues extrêmes en sites non jaugés : régionalisation de la méthode par simulation SHYREG, Ph.D. thesis, http://www.theses.fr/2017AIXM0424, thèse de doctorat dirigée par Arnaud, Patrick Géosciences de l'environnement. Hydrologie Aix-Marseille 2017, 2017.

Organde, D., Arnaud, P., Fine, J.-A., Fouchier, C., Folton, N., and Lavabre, J.: Régionalisation d'une méthode de prédétermination de crue sur l'ensemble du territoire français: la méthode SHYREG, Revue des Sciences de l'Eau, 26, 65–78, 2013.

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

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resources Research, 44, 2008.

Papamarkou, T., Hinkle, J., Young, M. T., and Womble, D.: Challenges in Markov Chain Monte Carlo for Bayesian Neural Networks, Statistical Science, 37, https://doi.org/10.1214/21-STS840, 2022.

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

Poncelet, C.: Du bassin au paramètre : jusqu'où peut-on régionaliser un modèle hydrologique conceptuel ?, Ph.D. thesis, http://www.theses.fr/2016PA066550, thèse de doctorat dirigée par Andréassian, Vazken et Oudin, Ludovic Hydrologie Paris 6 2016, 2016.

Pujol, L., Garambois, P.-A., Finaud-Guyot, P., Monnier, J., Larnier, K., Mose, R., Biancamaria, S., Yesou, H., Moreira, D., Paris, A., et al.: 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, 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.

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.

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., and Edelman, A.: Universal Differential Equations for Scientific Machine Learning, 2021.

Raissi, M., Perdikaris, P., and Karniadakis, G.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, 378, 686–707, https://doi.org/10.1016/j.jcp.2018.10.045, 2019.

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

Roy, A., Kasiviswanathan, K. S., Patidar, S., Adeloye, A. J., Soundharajan, B.-S., and Ojha, C. S. P.: A Novel Physics-Aware Machine Learning-Based Dynamic Error Correction Model for Improving Streamflow Forecast Accuracy, Water Resources Research, 59, https://doi.org/10.1029/2022WR033318, 2023.

Santos, L., Thirel, G., and Perrin, C.: Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the GR4 model using state-space GR4 (version 1.0), Geoscientific Model Development, 11, 1591–1605, 2018.

Shen, C. and Lawson, K.: Applications of Deep Learning in Hydrology, chap. 19, pp. 283–297, John Wiley & Sons, Ltd, ISBN 9781119646181, https://doi.org/10.1002/9781119646181.ch19, 2021.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., and Woo, W.-c.: Convolutional LSTM Network: a machine learning approach for precipitation nowcasting, in: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS'15, p. 802–810, MIT Press, Cambridge, MA, USA, 2015.

Te Chow, V., Maidment, D. R., and Mays, L. W.: Applied Hydrology, McGraw-Hill Series in Water Resources and Environmental Engineering, 1988.

Tran, V. N., Ivanov, V. Y., Xu, D., and Kim, J.: Closing in on Hydrologic Predictive Accuracy: Combining the Strengths of High-Fidelity and Physics-Agnostic Models, Geophysical Research Letters, 50, https://doi.org/10.1029/2023GL104464, 2023.

Tripathy, K. P. and Mishra, A. K.: Deep learning in hydrology and water resources disciplines: concepts, methods, applications, and research directions, Journal of Hydrology, 628, 130 458, https://doi.org/10.1016/j.jhydrol.2023.130458, 2024.

Vergara, H., Kirstetter, P.-E., Gourley, J. J., Flamig, Z. L., Hong, Y., Arthur, A., and Kolar, R.: Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States, Journal of Hydrology, 541, 421–433, 2016.

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.

Wang, C., Jiang, S., Zheng, Y., Han, F., Kumar, R., Rakovec, O., and Li, S.: Distributed Hydrological Modeling With Physics-Encoded Deep Learning: A General Framework and Its Application in the Amazon, Water Resources Research, 60, e2023WR036 170, https://doi.org/10.1029/2023WR036170, 2024.

Xu, T., Longyang, Q., Tyson, C., Zeng, R., and Neilson, B. T.: Hybrid Physically Based and Deep Learning Modeling of a Snow Dominated, Mountainous, Karst Watershed, Water Resources Research, 58, https://doi.org/10.1029/2021WR030993, 2022.

Yin, Y., Guen, V. L., Dona, J., de Bézenac, E., Ayed, I., Thome, N., and Gallinari, P.: Augmenting physical models with deep networks for complex dynamics forecasting, Journal of Statistical Mechanics: Theory and Experiment, 2021, 124 012, https://doi.org/10.1088/1742-5468/ac3ae5, 2021.

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, 1997.