Supporting Information for "Multiscale Learnable Physical Modeling and Data Assimilation Framework: Application to High-Resolution Regionalized Hydrological Simulation of Flash Floods"

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

Contents of this file

Introduction

The supporting information includes mathematical details concerning the complete forward model $\mathcal{M} = \mathcal{M}_{rr}\left(\ .\ ,\ \phi\left(.\right)\right)$ presented in the main article (Text S1 to S3), information regarding the model setup and watershed study (Text S4, Figures S1 and S2), and supplementary results on the learning performance (Figures S3 to S6).

Text S1. Notations

• Surface discharge: $Q(x,t) \in \mathbb{R}^N$, where $N = N_x \times N_t$ with N_x the number of cells in the 2D spatial domain $\Omega \in \mathbb{R}^2$ and N_t the number of simulation time steps in $]0,T]$;

• Internal states: $\boldsymbol{h} = (h_1(x,t), ..., h_{N_h}(x,t)) \in \mathbb{R}^{N \times N_h}$ with N_h the number of distinct state variables;

• Internal fluxes: $\boldsymbol{q} = \big(q_1(x,t), ..., q_{N_q}(x,t)\big) \in \mathbb{R}^{N \times N_q}$ with N_q the number of distinct flux variables;

• Atmospheric forcings: $\boldsymbol{\mathcal{I}} = (I_1(x,t), ..., I_{N_I}(x,t)) \in \mathbb{R}^{N \times N_I}$ with N_I the number of distinct forcings;

• Physical descriptors: $\boldsymbol{\mathcal{D}} = (D_1(x), ..., D_{N_D}(x)) \in \mathbb{R}^{N_x \times N_D}$ with N_D the number of distinct descriptors;

• Model parameters: $\boldsymbol{\theta} = (\theta_1(x), ..., \theta_{N_\theta}(x)) \in \mathbb{R}^{N_x \times N_\theta}$ with N_θ the number of distinct model parameters;

• Model initial states: $\boldsymbol{h}_0 = \boldsymbol{h}(x, t = 0)$.

In this study, we investigate a hydrological model structure comprising the following model parameters and internal states:

• c_i: capacity of interception reservoir [mm];

• c_p: capacity of production reservoir [mm];

• c_t: capacity of transfer reservoir [mm];

• k_{exc}: non-conservative water exchange flux [mm/h];

• a_{kw}: first kinematic wave routing parameter;

• b_{kw}: second kinematic wave routing parameter;

- h_i: interception reservoir state [mm];
- h_p: production reservoir state [mm];
- h_t: transfer reservoir state [mm].

Text S2. Learnable Parameterization-Regionalization operator ϕ

The LPR operator ϕ is composed of two neural networks, which are the parameterization network ϕ_1 and regionalization network ϕ_2.

The first one seeks to determine a spatio-temporal correction $\boldsymbol{f}_q$ for the internal fluxes $\boldsymbol{q}$ using neutralized rainfall and evapotranspiration $\boldsymbol{\mathcal{I}}_n = (P_n, E_n)$, along with production and transfer sates at previous time step $\boldsymbol{h} = (h_p, h_t)$:

$$\boldsymbol{f}_q(x,t) = \phi_1([P_n, E_n](x,t), [h_p, h_t](x, t-1); \boldsymbol{\rho}_1)$$

where $\boldsymbol{\rho}_1$ is the parameter vector to be optimized for ϕ_1.

The output $\boldsymbol{f}_q(x,t) = (f_{q,1}(x,t), f_{q,2}(x,t), f_{q,3}(x,t), f_{q,4}(x,t))^T$ will be applied as multiplicative factors to the GR operators internal fluxes as detailed in S3.

Here, the neural network ϕ_1 is configured as a multilayer perceptron with two hidden layers (consisting of 32 and 16 neurons, respectively), a Leaky ReLU activation function between hidden layers, and a modified Softmax function in the output layer that is bounded from 0 to 2. The modified Softmax function is used to constrain the fluxes corrector. The total number of trainable parameters of ϕ_1 is $N_{\rho_1} = 756$.

The second estimates the spatially distributed model parameters $\boldsymbol{\theta}$ from physical descriptors $\boldsymbol{\mathcal{D}} = (d_{i,i=1..7})^T$ (refer to Figure S2 for information on input descriptors):

$$\boldsymbol{\theta}(x) = \phi_2([d_1, ..., d_7](x); \boldsymbol{\rho}_2)$$

where $\boldsymbol{\rho}_2$ is the parameter vector to be optimized for ϕ_2.

The output $\boldsymbol{\theta}(x) = (c_p(x), c_t(x), k_{exc}(x), a_{kw}(x), b_{kw}(x))^T$ is composed of conceptual model parameters as detailed in S3.

Here, the neural network ϕ_2 is a multilayer perceptron with three hidden layers consisting of 96, 48 and 16 neurons, respectively. 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 boundary conditions. The total number of trainable parameters of ϕ_2 is $N_{\rho_2} = 6276$.

Text S3. Differentiable, gridded rainfall-runoff operator $\mathcal{M}_{rr}$

For a given cell $x \in \Omega$ and time step $t > 0$, $P(t)$ and $E(t)$ represent the local precipitation and potential evapotranspiration, respectively. The fluxes and states are computed as follows.

Interception (from Ficchì, Perrin, and Andréassian (2019)):

- Evapotranspiration from the interception reservoir: $E_i(t) = \min[E(t),\ P(t) + h_i(t-1)]$;
- Remaining rainfall: $P_n(t) = \max[0,\ P(t) + h_i(t-1) - c_i - E_i(t)]$;
- Remaining evapotranspiration: $E_n(t) = E(t) - E_i(t)$;
- Update of interception reservoir state: $h_i(t) = h_i(t-1) + P(t) - E_i(t) - P_n(t)$.

Production (refined from Perrin, Michel, and Andrèassian (2003) with neural network flux correction). The first order ordinary differential equation (ODE) describing the GR production store without percolation is:

$$\frac{dh_p}{dt} = \left(1 - \left(\frac{h_p}{c_p}\right)^2\right) p_n - \frac{h_p}{c_p}\left(2 - \frac{h_p}{c_p}\right) e_n$$

Assuming that instantaneous rainfall p_n and evaporation e_n are constant over a constant integration time step, one can obtain an analytical solution of infiltrating rainfall and actual evapotranspiration fluxes as follows.

- Infiltrating rainfall flux: $P_s\left(t\right) = c_p \left(1 - \left(\frac{h_p(t-1)}{c_p}\right)^2\right) \frac{\tanh\left(\frac{P_n(t)}{c_p}\right)}{1+\left(\frac{h_p(t-1)}{c_p}\right)\tanh\left(\frac{P_n(t)}{c_p}\right)}$ (analytical solution from stepwise approximation of P_n);
- Actual evapotranspiration flux: $E_s\left(t\right) = h_p\left(t-1\right)\left(2 - \frac{h_p(t-1)}{c_p}\right) \frac{\tanh\left(\frac{E_n(t)}{c_p}\right)}{1+\left(1-\frac{h_p(t-1)}{c_p}\right)\tanh\left(\frac{E_n(t)}{c_p}\right)}$ (analytical solution from stepwise approximation of E_n);
- Direct runoff flux: $P_r(t) = P_n(t) - f_{q,1} \times P_s(t)$;
- Update of production reservoir state: $h_p\left(t\right) = h_p\left(t-1\right) + f_{q,1} \times P_s\left(t\right) - f_{q,2} \times E_s\left(t\right)$.

Transfer (adapted and refined from Perrin et al. (2003) with neural network flux correction).

- Non-conservative exchange flux: $F(t) = k_{exc} \times \left(\frac{h_t(t-1)}{c_t}\right)^{7/2}$;
- Initial transfer reservoir state: $h_t(t^*) = \max\left[\epsilon,\ h_t\left(t-1\right) + 0.9 f_{q,3} \times P_r(t) + f_{q,4} \times F(t)\right]$ with $\epsilon > 0$, a fixed small constant;

The transfer reservoir is described with a first order ODE with a power law leakage source term:

$$\frac{dh_t}{dt} + c_t h_t^5 = \left(\frac{h_p}{c_p}\right)^2 p_n$$

Again, assuming that instantaneous rainfall p_n is constant over a constant integration time step, one can obtain an analytical solution of outflow flux from this transfer reservoir.

- Outflow flux from transfer reservoir: $Q_{ft}\left(t\right) = h_t\left(t^*\right) - \left(h_t\left(t^*\right)^{-4} + {c_t}^{-4}\right)^{-1/4}$;
- Update of transfer reservoir state: $h_t(t) = h_t(t^*) - Q_{ft}(t)$;
- Outflow from direct runoff: $Q_d(t) = \max\left[0,\ 0.1 f_{q,3} \times P_r(t) + f_{q,4} \times F(t)\right]$;

- Total outflow: $Q_{lat}(t) = Q_{ft}(t) + Q_d(t)$.

Kinematic wave routing (adapted from Te Chow, Maidment, and Mays (1988)):

This routing module 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 8 direction ("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 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 constants 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 one equation model is obtained:

$$\frac{\partial Q}{\partial t} + \frac{1}{a_{kw}b_{kw}}Q^{(1-b_{kw})}\frac{\partial Q}{\partial x} = \frac{1}{a_{kw}b_{kw}}Q_{lat}Q^{(1-b_{kw})}$$

This model is expressed as follows in order to be discretized with a finite difference approach (Te Chow et al., 1988):

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

Given N_u adjacent upstream cells within Ω flowing into cell x as imposed by flow direction map $\mathcal{D}_\Omega$, the upstream runoff is:

$$Q_u(x,t) = \sum_{k=1}^{N_u} Q(k,t)$$

Then, the numerical solution of the simplified mass equation by a finite difference approach, using a linearized implicit scheme (Te Chow et al., 1988), is:

$$Q(t) = \frac{\frac{\Delta t}{\Delta x} Q_u(t) + a_{kw} b_{kw} Q(t-1) \left(\frac{Q(t-1)+Q_u(t)}{2} \right)^{b_{kw}-1} + \Delta t \frac{Q_{lat}(t-1)+Q_{lat}(t)}{2}}{\frac{\Delta t}{\Delta x} + a_{kw} b_{kw} Q(t-1) \left(\frac{Q(t-1)+Q_u(t)}{2} \right)^{b_{kw}-1}}$$

Text S4. Model setup

The study zone in the Eastern Mediterranean region features 9 gauged catchments that are used for calibration while 6 others are considered as pseudo-ungauged for validation of spatial extrapolation capabilities (regionalization) (Figure S1). The SMASH model is run on a $dx = 1$ km spatial grid at $dt = 1$ h time step, i.e., at the same resolution as rainfall data. The model is forced by: (i) observed rainfall grids based on hourly ANTILOPE J+1 radar-gauge rainfall reanalysis from Météo-France (Champeaux et al., 2009); (ii) potential evapotranspiration (PET) estimated using the formula of (Oudin et al., 2005); and (iii) temperature data from SAFRAN reanalysis produced by Météo-France on a 8×8 km^2 spatial grid (Quintana-Seguí et al., 2008) downscaled to a 1×1 km^2 spatial grid. The model initial state $\boldsymbol{h}_0$ is not calibrated here and simply set such that relative saturation of production and transfer reservoir is 0.5. Discharge data at gauges over the calibration period P1 (August 2009 to July 2013) and the validation period P2 (August 2013 to July 2016) are obtained from preprocessed data sourced from SCHAPI-DGPR. Seven spatially distributed physical descriptors (Figure S2), resampled at $dx = 1$ km on the model grid, are considered as input for the regionalization algorithm, i.e., as input of the neural network ϕ_2.

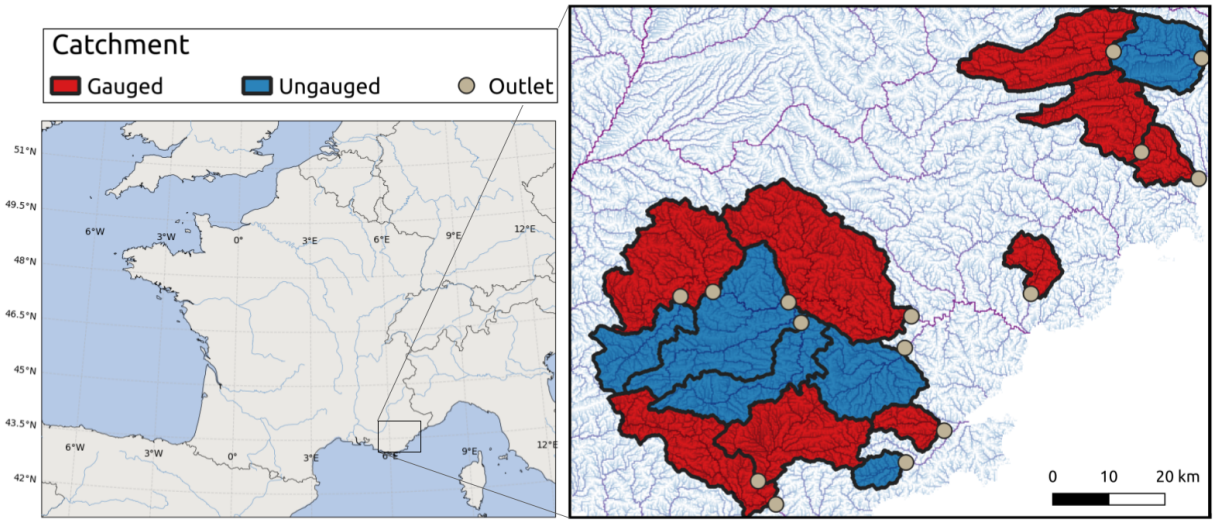

Figure S1. Study zone located in the Southeast region of France, comprising multiple gauges downstream of both nested and independent catchments. The gauged catchments (highlighted in red) are used for multi-site calibration, whereas the pseudo-ungauged catchments (highlighted in blue) serve for spatial validation. The selection of catchments is based on the availability of extensive time series with high-quality observed flow and minimal anthropogenic impacts. Nine gauged catchments are employed for calibration, with an additional six considered pseudo-ungauged for spatial validation. This area presents a challenging modeling case due to diverse hydrological properties, such as steep topography and highly heterogeneous soils and bedrock. The region is susceptible to intense rainfall, leading to nonlinear flash flood responses, and is characterized by a significant proportion of karstic zones.

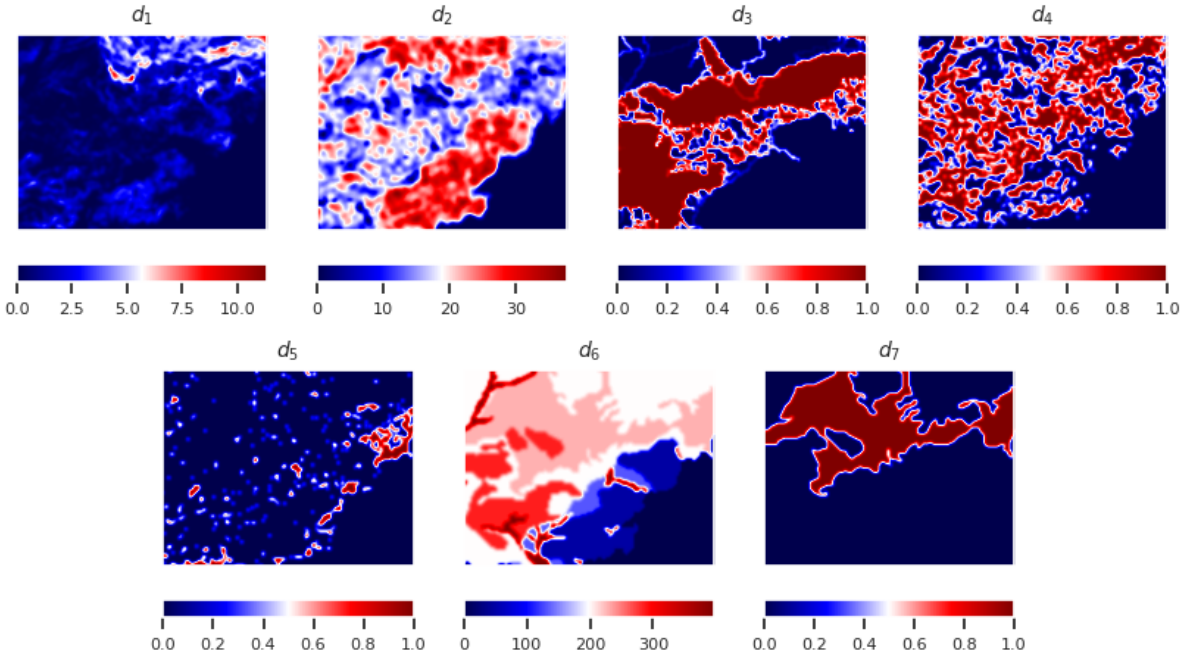

Figure S2. Maps of physical descriptors in the study area, including various types such as topography, morphology, land use, and hydrogeology. A set of seven descriptors at 0.01° in the WGS 84 projection is used as inputs for the regionalization mapping ϕ_2, consisting of: d_1 the local slope (in degrees) from the Copernicus database (version of 2016) and preprocessed by Odry (2017); d_2 the drainage density from Organde et al. (2013); d_3 the percentage of basin area in karst zone from Caruso et al. (2013); d_4 the forest cover rate and d_5 the urban cover rate both from the Corine Land Cover database (version of 2012); d_6 the potential available water reserve (in mm) from Poncelet (2016); d_7 the high storage capacity basin rate from Finke et al. (1998). Before the optimization process, all descriptors are standardized between 0 and 1 using min-max scaling.

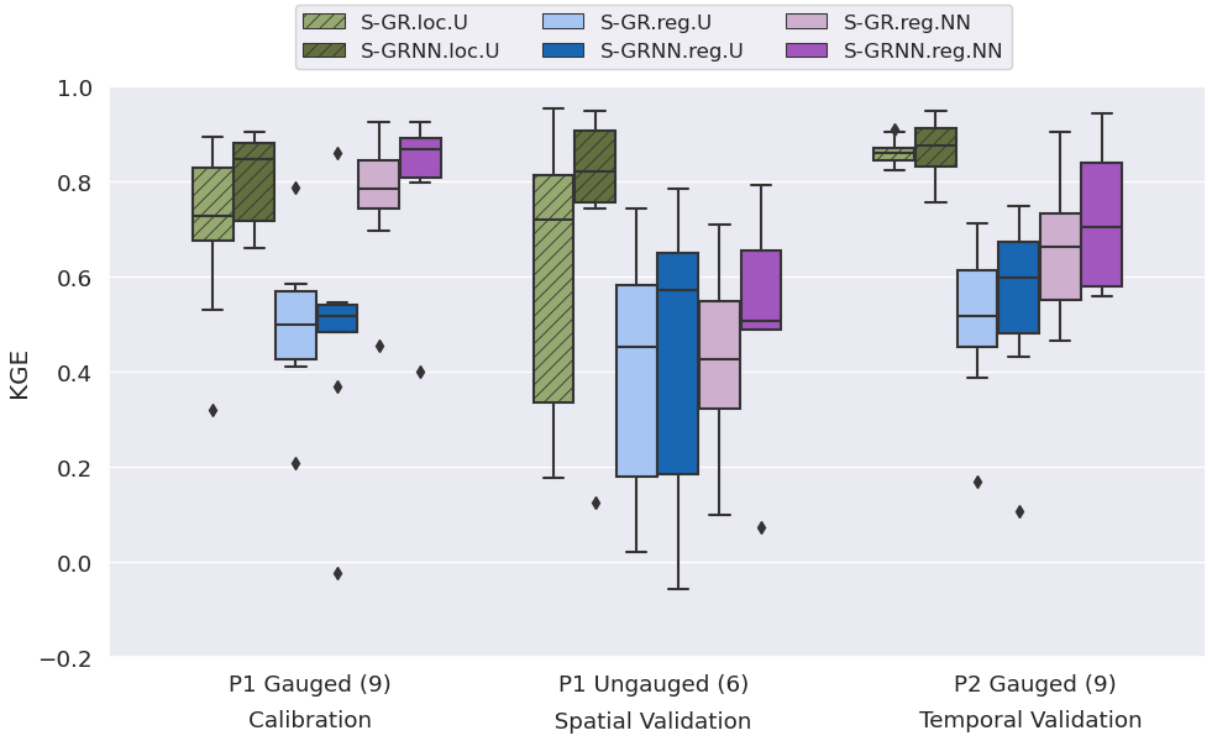

Figure S3. Boxplot of KGE scores computed over the full time series across gauged and pseudo-ungauged catchments in calibration (number of catchments in parenthesis), spatial validation, and temporal validation for four regionalization methods (reg), alongside a reference obtained from two locally calibrated methods (loc). The comparison assesses the performance in simulating discharges of different methods, hybrid models with process parameterization pipeline ϕ_1 in darker colors (S-GRNN), with regionalization pipeline ϕ_2 in different shades of pink (reg.NN).

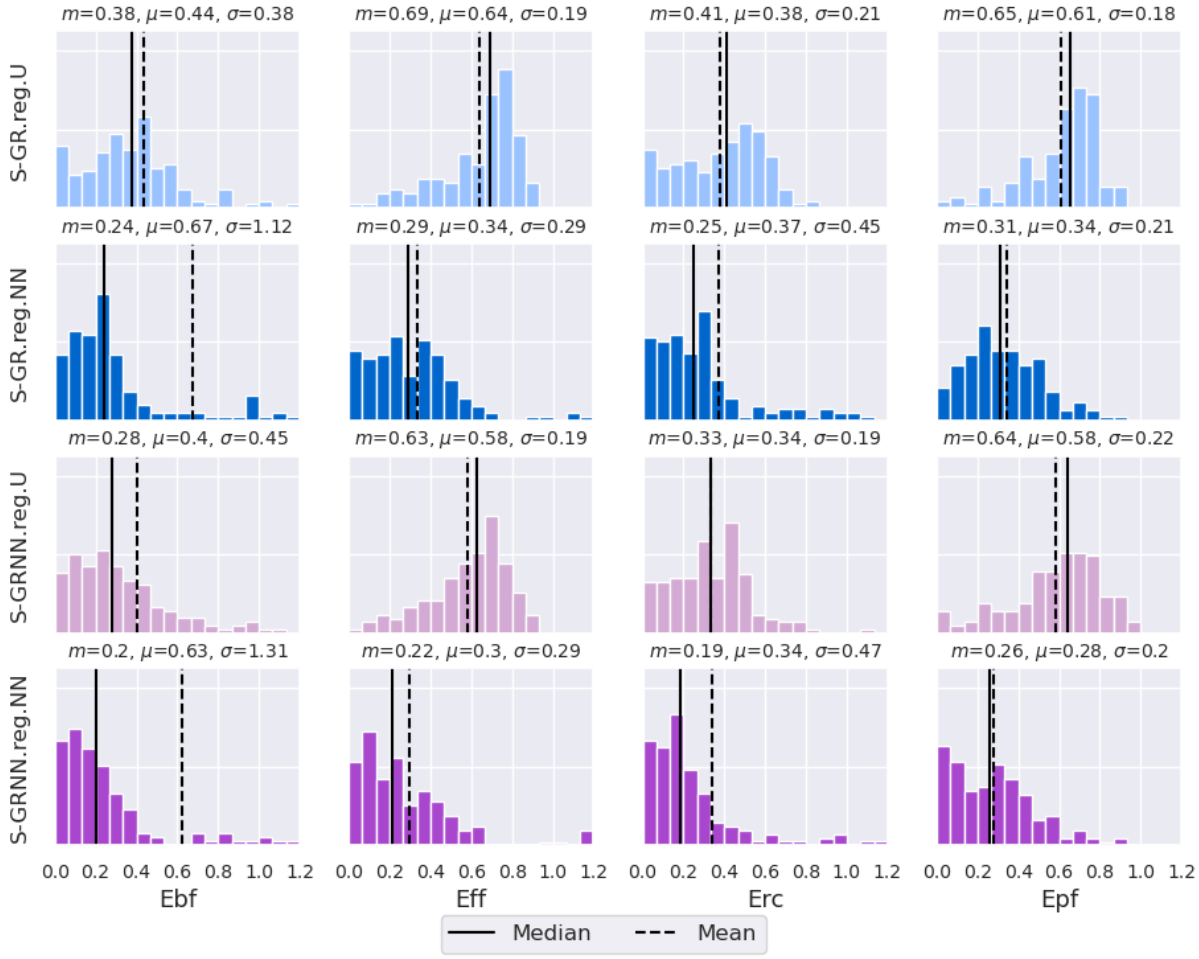

Figure S4. Distribution over 161 flood events of relative error, with an optimal value of 0, for four flood event signatures (Ebf - base flow, Eff - flood flow, Erc - runoff coefficient, Epf - peak flow). The median, mean, and standard deviation errors are respectively denoted m, μ, σ. The evaluation takes place in gauged catchments during P1 (2009-2013) for calibration.

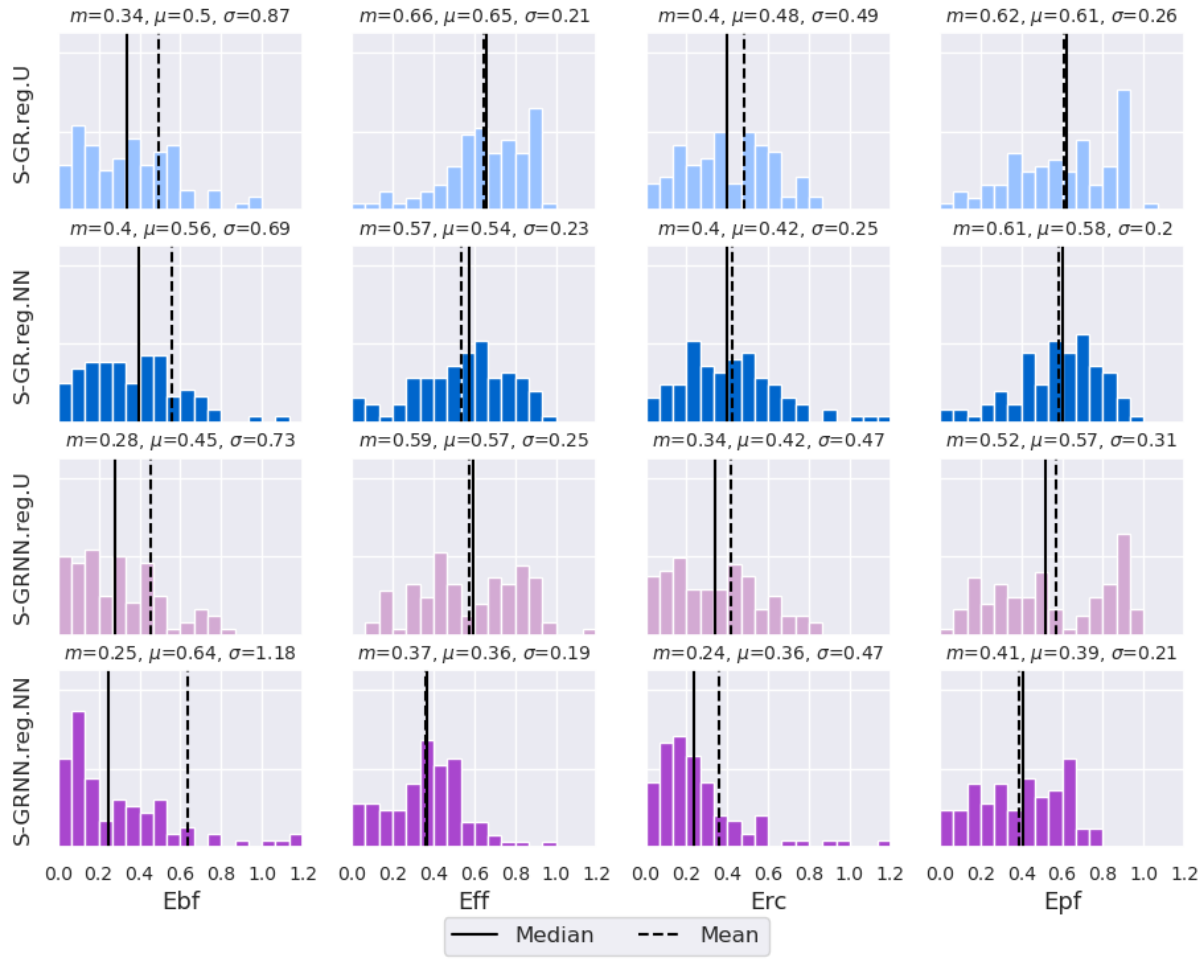

Figure S5. Distribution over 93 flood events of relative error, with an optimal value of 0, for four flood event signatures (Ebf - base flow, Eff - flood flow, Erc - runoff coefficient, Epf - peak flow). The median, mean, and standard deviation errors are respectively denoted m, μ, σ. The evaluation takes place in pseudo-ungauged catchments during P1 (2009-2013) for spatial validation.

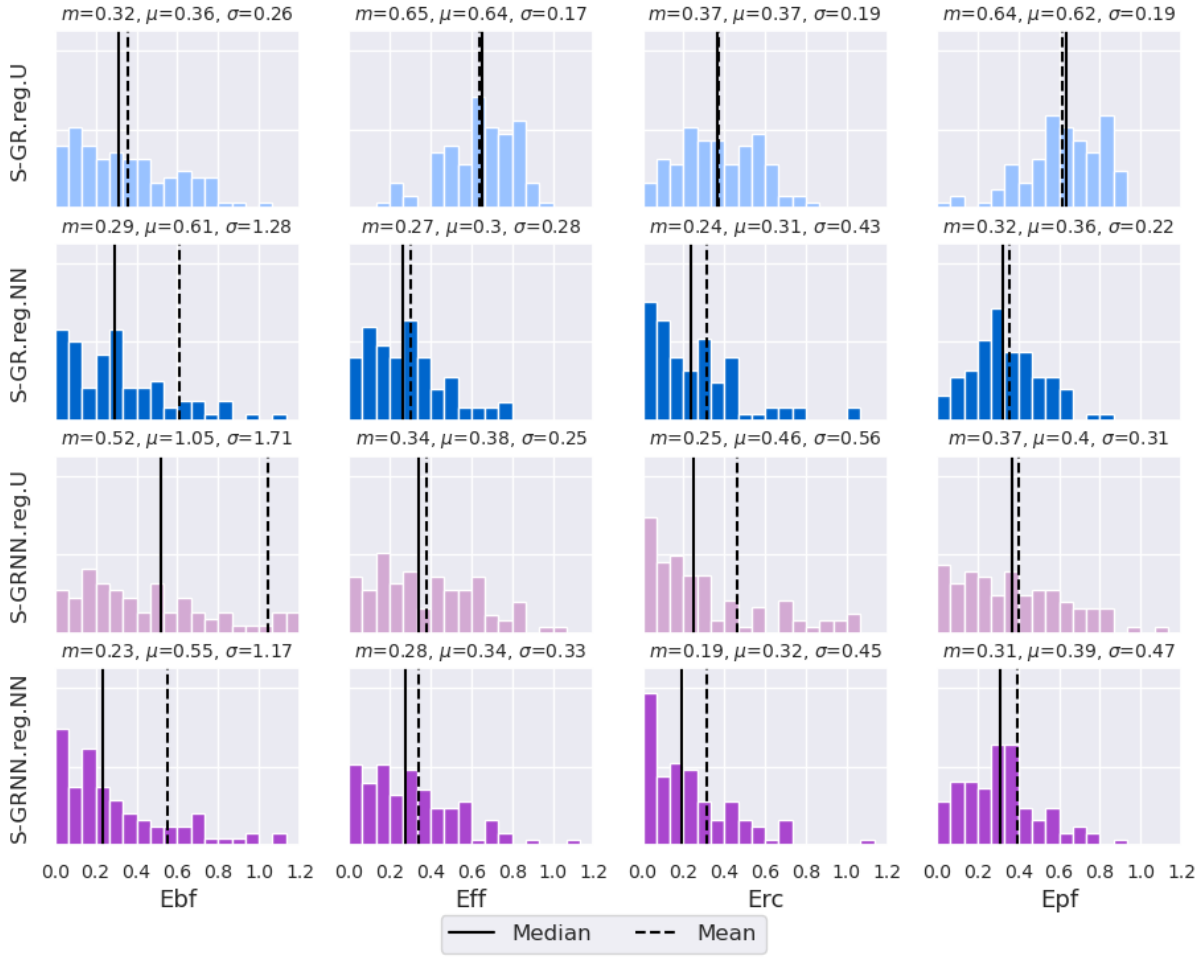

Figure S6. Distribution over 95 flood events of relative error, with an optimal value of 0, for four flood event signatures (Ebf - base flow, Eff - flood flow, Erc - runoff coefficient, Epf - peak flow). The median, mean, and standard deviation errors are respectively denoted m, μ, σ. The evaluation takes place in gauged catchments during P2 (2013-2016) for temporal validation.

References

Caruso, A., Guillot, A., & Arnaud, P. (2013). Notice sur les indices de confiance de la méthode shyreg-débit–définitions et calculs. In *Aix en provence: Irstea, convention dgpr/snrh.*

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

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

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

Odry, J. (2017). *Prédétermination des débits de crues extrêmes en sites non jaugés : régionalisation de la méthode par simulation shyreg* (Doctoral dissertation). Retrieved from http://www.theses.fr/2017AIXM0424 (Thèse de doctorat dirigée par Arnaud, Patrick Géosciences de l'environnement. Hydrologie Aix-Marseille 2017)

Organde, D., Arnaud, P., Fine, J.-A., Fouchier, C., Folton, N., & Lavabre, J. (2013). Régionalisation d'une méthode de prédétermination de crue sur l'ensemble du territoire français: la méthode shyreg. *Revue des Sciences de l'Eau*, *26*(1), 65–78.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2 towards a simple and efficient potential evapotranspiration model for rainfall–runoff mod-

elling. *Journal of hydrology*, *303*(1-4), 290–306.

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

Poncelet, C. (2016). *Du bassin au paramètre : jusqu'où peut-on régionaliser un modèle hydrologique conceptuel ?* (Doctoral dissertation). Retrieved from http://www.theses.fr/2016PA066550 (Thèse de doctorat dirigée par Andréassian, Vazken et Oudin, Ludovic Hydrologie Paris 6 2016)

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

Te Chow, V., Maidment, D. R., & Mays, L. W. (1988). *Applied hydrology*. McGraw-Hill Series in Water Resources and Environmental Engineering.