Multi-gauge Hydrological Variational Data Assimilation: Regionalization Learning with Spatial Gradients using Multilayer Perceptron and Bayesian-Guided Multivariate Regression

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

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

[2]*Institut de Mécanique des Fluides de Toulouse (IMFT), Université de Toulouse, CNRS, 31400 Toulouse, France*

**This work was supported by funding from SCHAPI-DGPR, ANR grant ANR-21-CE04-0021-01 (MUFFINS project, "MUltiscale Flood Forecasting with INnovating Solutions"), and NEPTUNE European project DG-ECO*

Context

Challenging problem: estimating parameters of spatially distributed hydrological models, requiring spatial constraints given relatively sparse discharge data, in view to model ungauged catchments.

Approach: enhance distributed and regionalized hydrological modeling (spatial extrapolation) by making the most of available information, especially physical descriptor maps. This includes adapting the complexity of forward and inverse approaches to ensure consistency between the model and data, by taking advantage of statistical learning.

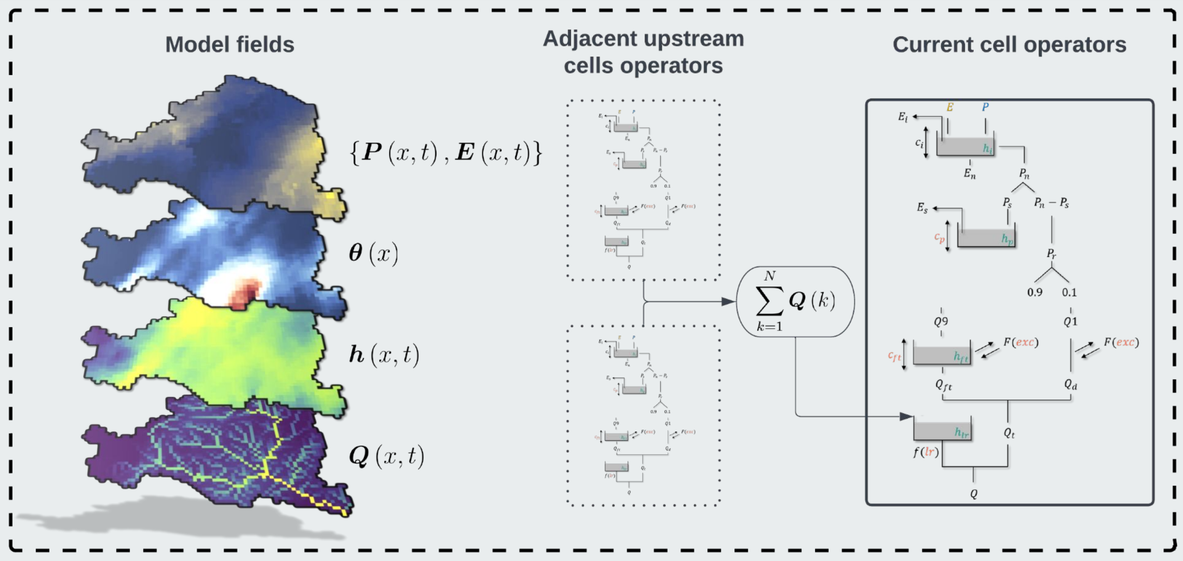
Presentation outline:

➔ New approaches for learning spatially distributed parameters of a differentiable hydrological model: descriptor to parameters operators, Bayesian like estimator for first guess in equifinality context.

➔ Study equifinality within a high-dimensional optimization framework adapted to distributed models.

➔ Application of the method in a multi-gauge calibration setup on flash flood prone area with complex hydrological responses.

➔ Perspectives on uncertainty quantification and correction, learnable hydrological operators, both in the structural modeling and the content of the data



Methodology

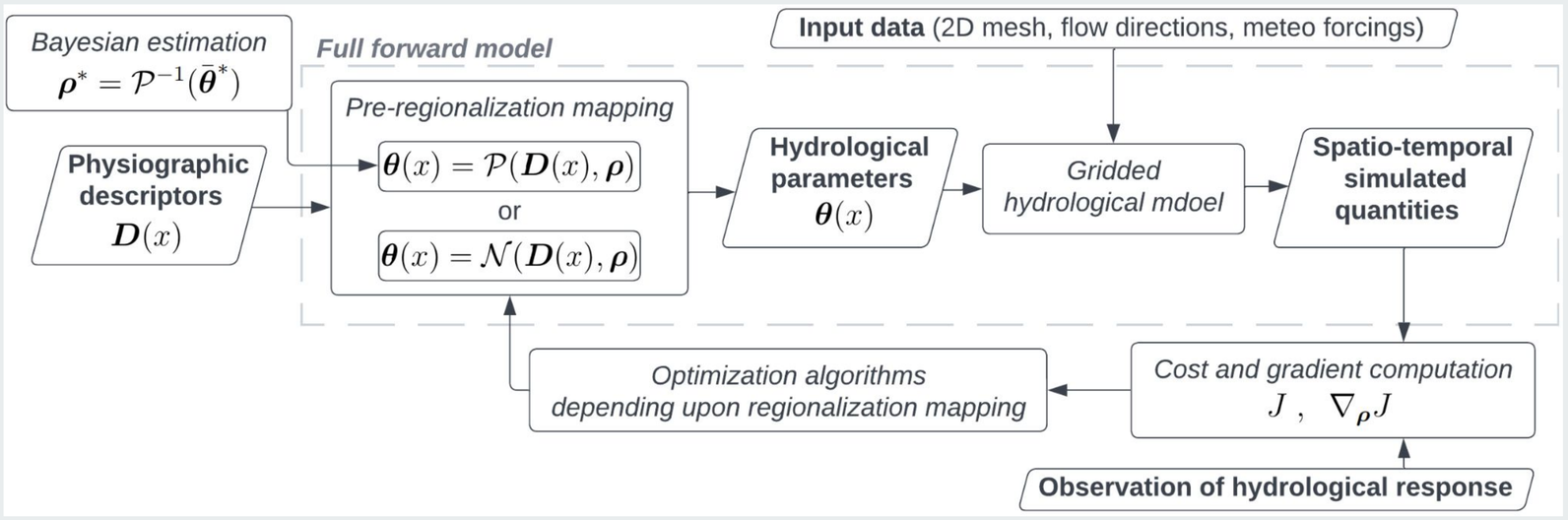
➔ Parameter regionalization within differentiable distributed hydrological model using adjoint-based gradient and high dimensional optimization algorithms

➔ Learnable regional mappings based on: (i) multilayer perceptron (ANNR) and (ii) Bayesian-guided multivariate regression (BGM2R)

➔ Combination of numerical adjoint model and neural network Jacobian, enabling high-dimensional spatialized optimization

➔ Combination of Bayesian-like estimator and VDA algorithms in the context of "resulting parametric equifinality" when dealing with spatialized controls and high-dimensional optimization problems

➔ Multi-gauge calibration cost function accounting for information from multiple observation sites



Mathematical ingredients

Forward model

The classical forward model $\mathcal{M}_{rr}$ is a dynamic operator projecting the input fields $\boldsymbol{P}(x,t)$ and $\boldsymbol{E}(x,t)$ onto the discharge field $Q(x,t)$ and states fields $\boldsymbol{h}(x,t)$ written as:

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

where $\boldsymbol{\theta}$ is the N_θ-dimensional vector of model parameters 2D fields to be estimated regionally and $\boldsymbol{h}$ is the N_S-dimensional vector of internal model states. The full forward model $\mathcal{M}$ is composed of the distributed hydrological model $\mathcal{M}_{rr}$ on top of which is applied a regionalization operator $\mathcal{F}_R$ to estimate hydrological parameters $\boldsymbol{\theta}(x)$ using physical descriptors $\boldsymbol{D}(x)$:

$$\mathcal{M} = \mathcal{M}_{rr}\left[\,.\,, \boldsymbol{\theta}(x) = \mathcal{F}_R(\boldsymbol{D}(x), \boldsymbol{\rho})\right], \forall x \in \Omega$$

Calibration cost function

$$J = \sum_{g=1}^{N_G} w_g J_g^*$$

with w_g a weighting function, J_g^* a local quadratic metric "at the station" (e.g., $1 - NSE$) involving the response of the direct model $\mathcal{M}$.

Bayesian-like estimator

A Bayesian-like estimator (following Chelil et al. 2022, Gejadze et al. 2023) is used to look at the mean of the posterior distribution of the model parameters that is more stable in context of equifinality than searching its mode, maximum a posteriori probability (MAP) search is the essence of variational data assimilation)

A prior probability distribution $f_{\bar{\theta}}$ is used to generate a sample of spatially uniform parameter sets $\boldsymbol{\theta}_i, \forall i \in 1..N$ within the hypercube defined by parameters bounds $[l_k, u_k], \forall k \in 1..N_\theta$. The likelihood function is defined as: $\mathcal{L}_i^\alpha = e^{-2^\alpha (J_i/J_{\min}-1)^2}$, where α is a parameter controlling the decay rate of this function that compares the value of $J_i = J(\boldsymbol{\theta}_i)$ to J_{min} the minimum value of J_i over the sample of N parameter sets. The posterior ensemble mean and variance are computed as follows:

$$\bar{\boldsymbol{\theta}}^{*,\alpha} = \frac{1}{K}\sum_{i=1}^{N}\left(\mathcal{L}_\alpha^i \cdot \boldsymbol{\theta}_i \odot f_{\bar{\theta}}(\boldsymbol{\theta}_i)\right)$$

$$\mathrm{Var}\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right) = \frac{1}{K}\sum_{i=1}^{N}\left(\mathcal{L}_\alpha^i \cdot (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}^{*,\alpha}) \odot (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}^{*,\alpha}) \odot f_{\bar{\theta}}(\boldsymbol{\theta}_i)\right)$$

where $K = \sum_{i=1}^{N} \mathcal{L}_\alpha^i \cdot f_{\bar{\theta}}(\boldsymbol{\theta}_i)$ and " $\odot$ " denotes the Hadamard product. The parameter α is determined using the L-curve approach, considering a parametric curve $\left\{J\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right), D^\alpha\right\}$, $\alpha = -1, ..., 10$, where $D^\alpha = \left(\mathrm{Var}\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right)\right)^{-1} \odot \left(\bar{\boldsymbol{\theta}}^{*,\alpha} - \bar{\boldsymbol{\theta}}^{0}\right) \odot \left(\bar{\boldsymbol{\theta}}^{*,\alpha} - \bar{\boldsymbol{\theta}}^{0}\right)$ is the probabilistic (Mahalanobis) distance between the estimate $\bar{\boldsymbol{\theta}}^{*,\alpha}$ and the average prior $\bar{\boldsymbol{\theta}}^{0} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\theta}_i$. The value of α is sought in a L-curve "corner" such that it minimizes both $J\left(\bar{\boldsymbol{\theta}}^{*,\alpha}\right)$ and D^α.

Regionalization mappings

Multivariate polynomial regression (following Götzinger and Bárdossy (2007), Samaniego et al. 2010):

$$\theta_k(x, \boldsymbol{D}, \rho_k) := s_k\left(\alpha_{k,0} + \sum_{d=1}^{N_D} \alpha_{k,d} D_d^{\beta_{k,d}}(x)\right), \forall k \in [1..N_\theta], \forall x \in \Omega$$

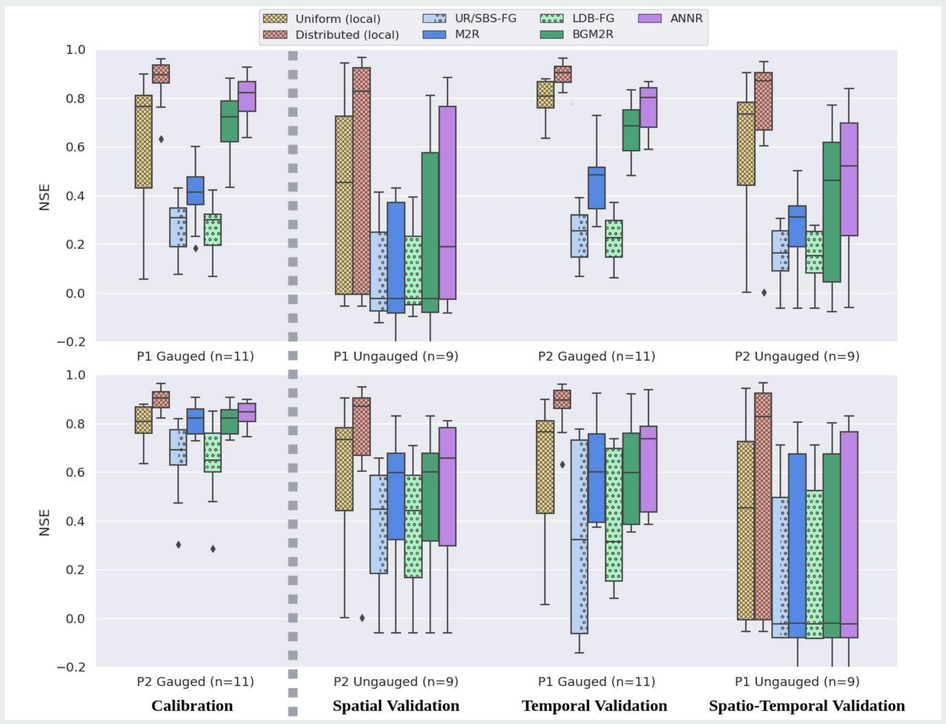
with $s_k(.)$ a Sigmoid-based transformation imposing bound constraints in the direct hydrological model. The lower and upper bounds are assumed to be spatially uniform for each parameter field θ_k of the hydrological model.

Multilayer perceptron:

$$\boldsymbol{\theta}(x, \boldsymbol{D}, \boldsymbol{\rho}) := \mathcal{N}(\boldsymbol{D}(x), \boldsymbol{W}, \boldsymbol{b}), \forall x \in \Omega$$

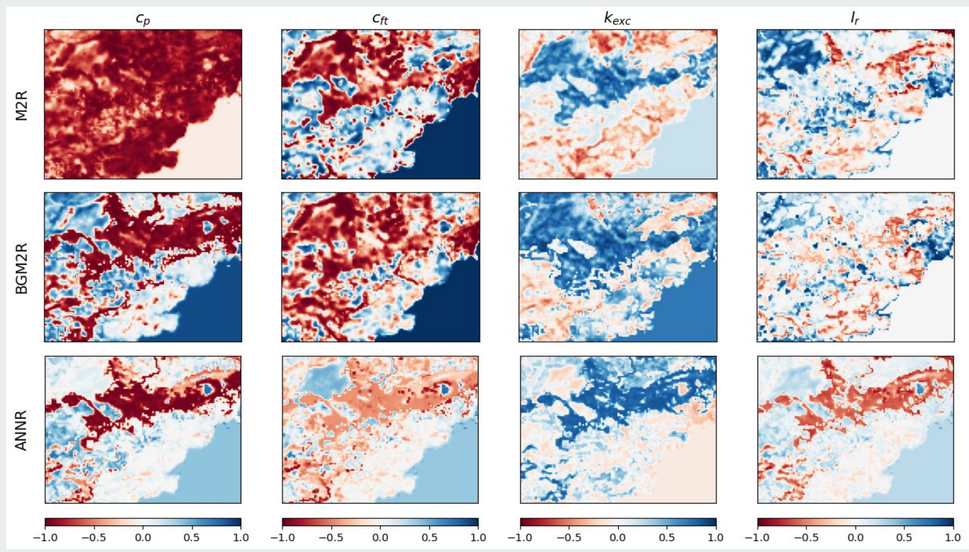
where $\boldsymbol{W}$ and $\boldsymbol{b}$ are respectively weights and biases of the neural network, whose output layer consists in a scaling transformation based on the Sigmoid function in order to impose bound constraints on each hydrological parameters.

Numerical results



Performance of all calibration and regionalization methods when calibrated on P1 (upper sub-figure) and on P2 (lower sub-figure). BGM2R uses the first guess obtained by Low-Dimensional Bayesian estimation (LDB-FG), while the first guess of M2R is obtained by the SBS algorithm (SBS-FG), which also represents the solution of the regionalization method with lumped parameters (UR). n denotes the number of study catchments.

Spatially uniform calibration (UR) leads to limited performance in calibration and poor performance in regionalization when compared to the reference local spatially distributed calibration that is overparameterized. The regionalization methods, which incorporate information from physical descriptor maps, all yield relatively satisfying performances in calibration and temporal validation at pseudo-ungauged sites (median NSE scores higher than 0.4 when calibrated on P1). The regionalization approach ANNR achieves the best results for both gauged and pseudo-ungauged catchments. In spatio-temporal validation, the most challenging case especially on this zone that is difficult to model, performance are relatively low, with regionalization methods or spatially uniform regionalization (UR/SBS-FG) and further investigations are required.



Normalized temporal stability over the periods P1 and P2 of the optimal distributed parameters for the three regionalization approaches using physical descriptors.

The ANNR leads to the most robust inference over P1 and P2, with remarkably stable average parameter values as well as spatial standard deviation over time. The priors inferred with SBS-FG or LDB-FG exhibit slight differences and also lead to a different optimum during regionalization for P1. The opposite trend is observed for P2, where data uncertainty and model adequacy might be better, resulting in similar functioning points after regionalization despite substantially different priors determined with SBS-FG or LDB-FG.

Ongoing work and perspectives

➔ Neural ODEs for introducing "learnable" model structures for hydrological flux production at the pixel level:

$$\frac{d\boldsymbol{h}}{dt} = \begin{pmatrix} \frac{dh_p}{dt} \\ \frac{dh_{ft}}{dt} \end{pmatrix} = \begin{pmatrix} \mathcal{NN}_1(I_s, h_p) \\ \mathcal{NN}_2(I_s, h_p, h_{ft}) \end{pmatrix} \text{ or } \begin{pmatrix} f_1(I_s, h_p) + \mathcal{NN}_1(I_s, h_p) \\ f_2(I_s, h_p, h_{ft}) + \mathcal{NN}_2(I_s, h_p, h_{ft}) \end{pmatrix}$$

where the deterministic source term is a system of ODEs depending on the flux production and transfer:

$$\begin{pmatrix} f_1(I_s, h_p) \\ f_2(I_s, h_p, h_{ft}) \end{pmatrix} = \begin{pmatrix} \left(1 - \left(\frac{h_p}{c_p}\right)^{\alpha_1}\right) P_n - \frac{h_p}{c_p}\left(2 - \frac{h_p}{c_p}\right) E_n \\ 0.9\left(\frac{h_p}{c_p}\right)^{\alpha_1} P_n - c_{ft} h_{ft}^{\alpha_2} + k_{exc}\left(\frac{h_{ft}}{c_{ft}}\right)^{\alpha_3} \end{pmatrix}$$

➔ Stochastic process for correcting modeling errors at multiple temporal scales:

$$\epsilon_{rr}(x,t) = Q^*(x,t) - Q_{sim}(x,t) = Q^*(x,t) - \mathcal{M}_{rr}(x,t;\mathcal{I}_{rr};\theta_{rr}) \sim \mathcal{G}(\mu, \sigma^2)$$

where: $(\mu, \sigma) := LSTM(x, t, \theta_{NN}, I_{NN})$, or
$\mu := LSTM_1(x, t, \theta_{NN_1}, I_{NN})$ et $\sigma := LSTM_2(x, t, \theta_{NN_2}, I_{NN})$