Advanced Hybrid Data Assimilation for Parameter Regionalization within a Differentiable Spatially Distributed Hydrological Model and Uncertainty Correction with Bidirectional LSTM

Ngo Nghi Truyen Huynh, Pierre-André Garambois, François Colleoni, Benjamin Renard, Hélène Roux

AGU23

WIDE. OPEN. SCIENCE.

#AGU23

Advanced Hybrid Data Assimilation for Parameter Regionalization within a Differentiable Spatially Distributed Hydrological Model and Uncertainty Correction with Bidirectional LSTM

Ngo Nghi Truyen Huynh

San Francisco, 14 December 2023

INRAE

Aix*Marseille université
Socialement engagée

Ngo Nghi Truyen Huynh

Research Engineer/PhD Student
INRAe, France

Machine Learning
Data Assimilation
Hydrology

Pierre-André Garambois

Researcher, INRAe

Thesis supervisor

Hydrology-Hydraulics
Numerical Modeling
Data Assimilation

François Colleoni

Research Engineer, INRAe

Hydrology
Numerical Modeling

Benjamin Renard

Researcher, INRAe

Co-supervisor

Stochastic Hydrology
Uncertainty Quantification

SMASH - Spatially distributed Modelling and ASsimilation for Hydrology

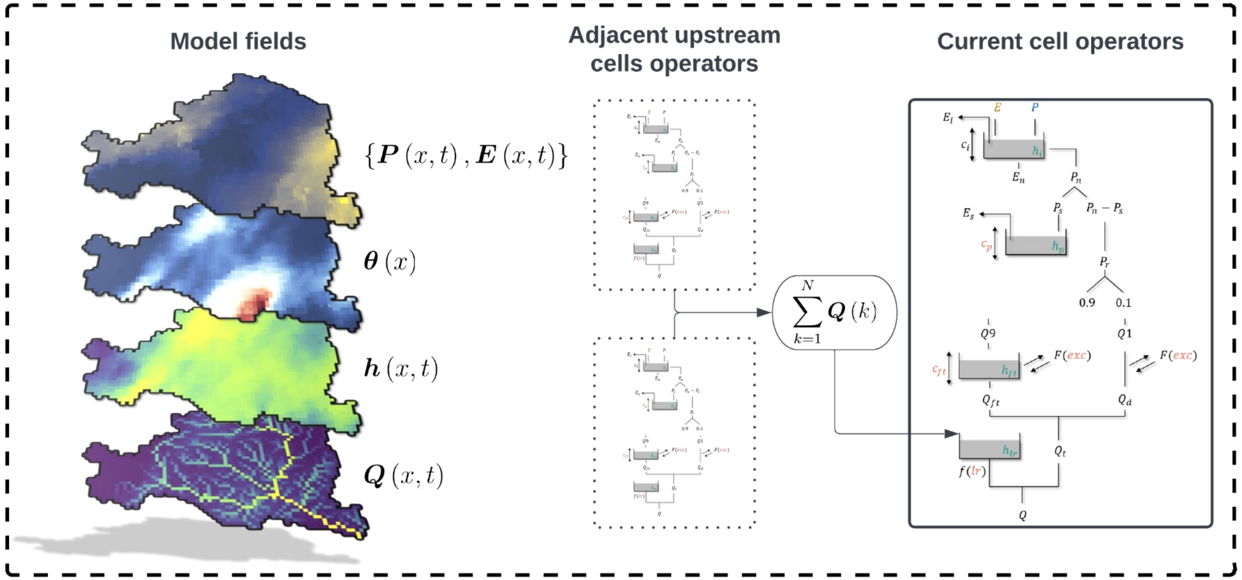
- An open source, Python library interfacing the differentiable Fortran Spatially distributed Modelling and ASsimilation for Hydrology

- Providing a variety of user-friendly routines for hydrological modeling, including data preprocessing, high-dimensional optimization tools, sensitivity and signatures analysis, all interfaced with an efficient Fortran solver

- Offering a range of advanced optimization techniques, including Variational Data Assimilation (VDA), Bayesian estimation, and Artificial Neural Network (ANN) approaches, using an adjoint model generated with an automatic differentiation engine

SMASH - Spatially distributed Modelling and ASsimilation for Hydrology

- A gridded mesh and modular design
- Different components such as: snow module, surface interception, production, transfer, percolation functions on each cell
- Different routing models with a cell-to-cell numerical scheme



SMASH GR-like conceptual hydrological model structure (Perrin et al., 2003)

Context

Estimating parameters of spatially distributed hydrological models, requiring spatial constraints given relatively sparse discharge data, in view to model ungauged catchments, and uncertainty correction:

→ New approach (HDA-PR) for learning spatially distributed parameters of a differentiable hydrological model from physical descriptors using multivariate regression and artificial neural network, with adjoint-based gradient and high dimensional optimization algorithms

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

→ Uncertainty quantification and correction, learnable hydrological operators, both in the structural modeling and the content of the data

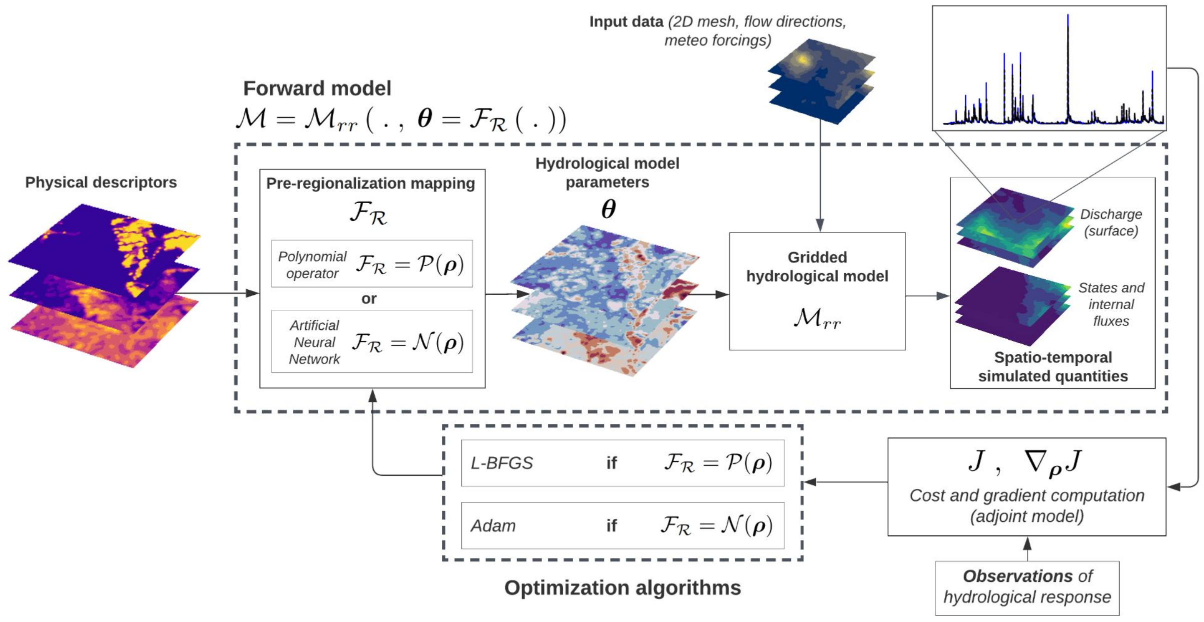
AGU23

Learnable regionalization

Previous works on regionalization within the calibration process: *Samaniego et al. (2010)*; *Beck et al. (2020)*

A new HDA-PR approach with:

- A <u>differentiable</u> spatially <u>distributed</u> hydrological model
- Introduction of learnable regionalization functions into the forward hydrological model
- Accurate spatial gradients of the cost obtained by solving the numerical adjoint model
- Model calibration with high-dimensional optimization algorithms



WIDE. OPEN. SCIENCE.

AGU23

Learnable regionalization

- Rainfall-runoff model:

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

- Full forward model with regionalization mapping:

$$\mathcal{M} = \mathcal{M}_{rr}\left(\,.\,,\ \boldsymbol{\theta} = \mathcal{F}_{\mathcal{R}}\left(\,.\,\right)\right)$$

with a regionalization function for mapping physical descriptors onto unknown conceptual hydrological parameters:

$$\boldsymbol{\theta}(x) = \mathcal{F}_{R}(\boldsymbol{D}(x),\boldsymbol{\rho}),\ \forall x \in \Omega$$

where $\boldsymbol{\rho}$ be the optimizable parameters of the regionalization mapping

- Two regionalization mappings are considered: multivariate polynomial regression and multilayer perceptron

WIDE. OPEN. **SCIENCE.**

Learnable regionalization

- Inverse problem:

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} J\left(\boldsymbol{U}\left(\boldsymbol{\rho}\right)\right)$$

$$J\left(\boldsymbol{U}\left(\boldsymbol{\rho}\right)\right) = J\big(\boldsymbol{Q}^*, \mathcal{M}_{rr}(. \,, \; \boldsymbol{\theta} = \mathcal{N}(\boldsymbol{D}, \boldsymbol{\rho}))\big)$$

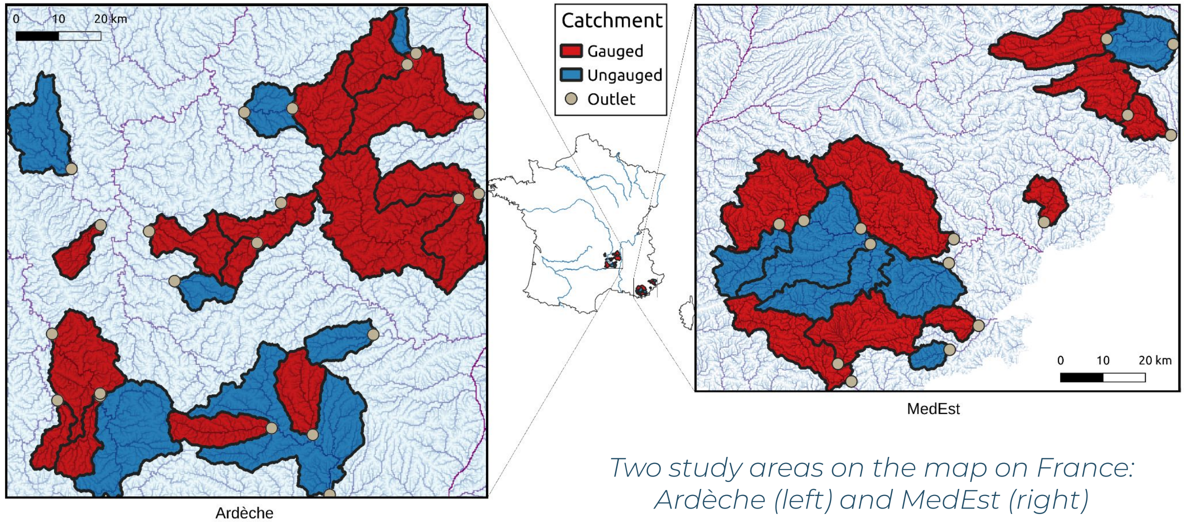
- Gradient computation:

$$\nabla_{\boldsymbol{\rho}} J = \nabla_{\boldsymbol{\theta}} J . \nabla_{\boldsymbol{\rho}} \boldsymbol{\theta}$$

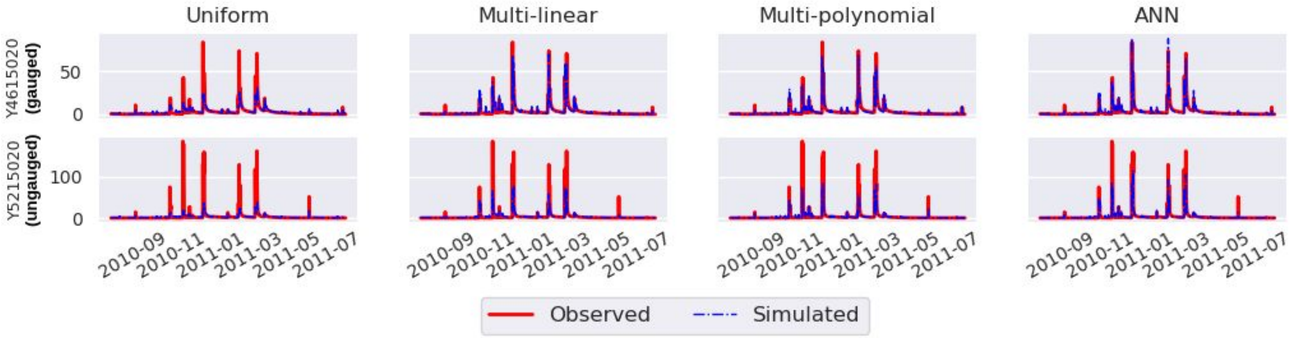
The first term can be computed via the automatic differentiation applied to the Fortran code and the second is obtained by analytical calculus applicable given the explicit architecture of the neural network

AGU23

Results



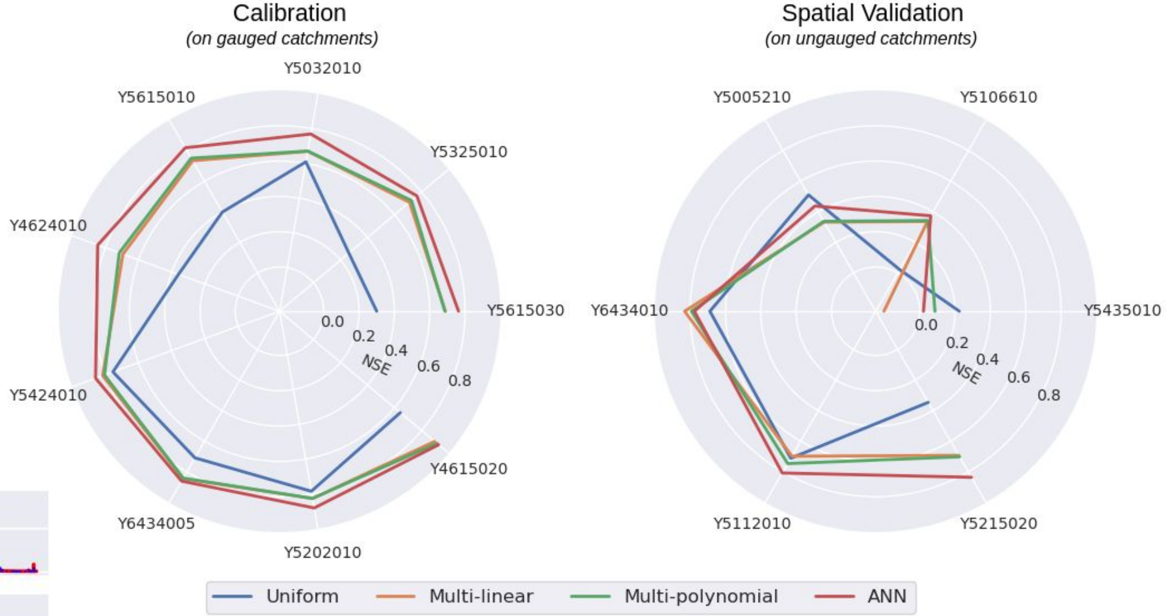
Two study areas on the map on France: Ardèche (left) and MedEst (right)



One year of observed and simulated discharges (in m³/s) at one gauged (Y4615020) and one pseudo-ungauged catchment (Y5215020)

Calibration period: 2006-2016

Validation period: 2016-2018

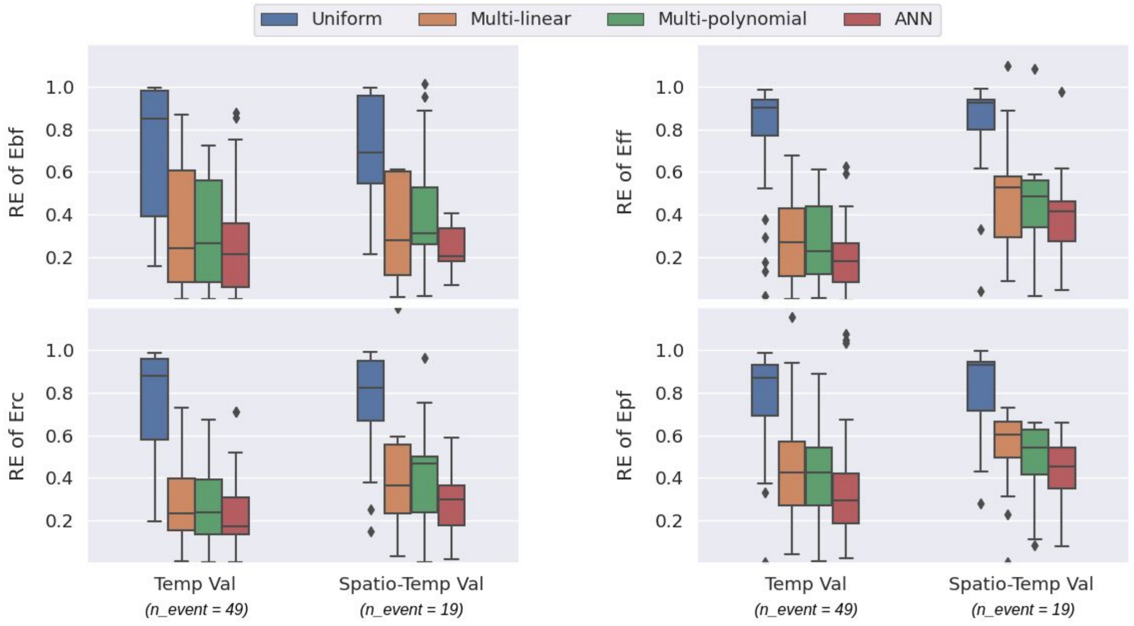


Radial plots of the NSE (optimal value = 1) in gauged catchments (left) and pseudo-ungauged catchments (right)

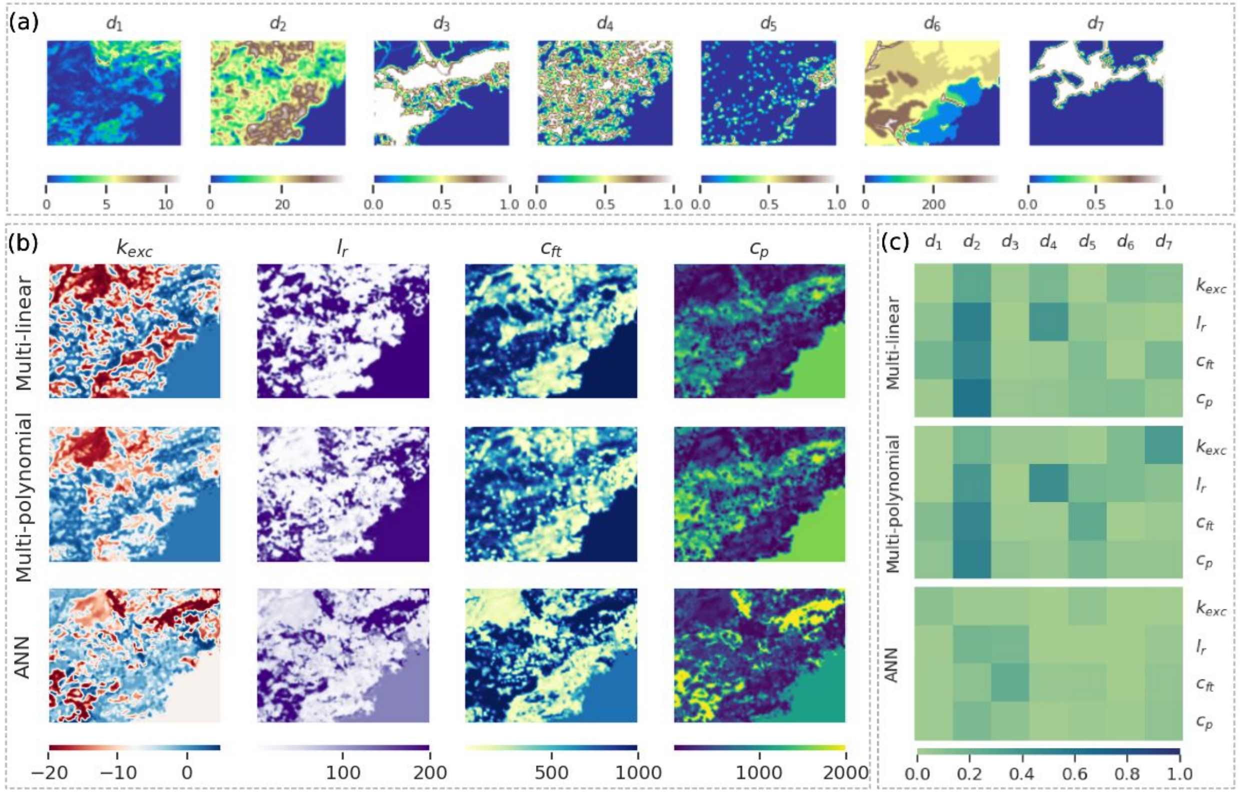
WIDE. OPEN. SCIENCE.

Results

Flood performances (*Huynh et al., 2023*)



Relative error (optimal value = 0) of 4 flood event signatures: base flow (Ebf), flood flow (Eff), runoff coefficient (Erc) and peak flow (Epf), evaluated using 49 flood events at gauged catchments for temporal validation (Temp Val) and 19 flood events at pseudo-ungauged catchments for spatio-temporal validation (Spatio-Temp Val)



Sub-figure a: Maps of input descriptors. Sub-figure b: Calibrated hydrological parameters maps for three non-uniform regionalization methods. Sub-figure c: Linear correlation between descriptor and parameter for the three regionalization methods

WIDE. OPEN. **SCIENCE.**

Stochastic process for uncertainty correction

Recent works on streamflow correction with LSTM: *Cho & Kim (2022)*; *Hashemi et al. (2023)*

A proposed stochastic process to correct simulation responses addresses:

- Observation error
- Structure modeling error

$$\epsilon(x,t) = Q^*(x,t) - \mathcal{M}_{rr}(.,x,t) \sim \mathcal{G}(\mu,\sigma^2)$$

where:

$$(\mu,\sigma) := LSTM(.,x,t)$$

The spatio-temporal error of simulated discharges is assumed to follow a Gaussian distribution with learnable parameters such as mean and standard deviation. Then, these two parameters are estimated using a bidirectional Long Short-Term Memory (LSTM) network to improve the model's ability to capture temporal dependencies and spatial patterns in uncertainty modeling

Ongoing works and perspectives

Ongoing works:

- Testing of HDA-PR and uncertainty correction with LSTM at national scales and on other continent
- Study of effective descriptor selection along with multi-gauge cost functions explicitly accounting for data uncertainties

Perspectives:

- Introduction of learnable production and transfer operators in the forward hydrological model (e.g., neural ordinary differential equations (neural ODEs) in *Hoge et al., 2022*)
- Regionalization of differentiable integrated hydrological-hydraulic networks models (*Pujol et al., 2022*)

References

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

Samaniego, L., Kumar, R., and Attinger, S. (2010), Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46, W05523

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I. J. M., & Wood, E. F. (2020). Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. Journal of Geophysical Research: Atmospheres, 125

Huynh, N. N. T., Garambois, P. A., Colleoni, F., & Javelle, P. (2023). Signatures-and-sensitivity-based multi-criteria variational calibration for distributed hydrological modeling applied to Mediterranean floods. Journal of Hydrology, 625, 129992

Cho, K., & Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. Journal of Hydrology, 605, 127297

Hashemi, R., Brigode, P., Garambois, P. A., & Javelle, P. (2022). 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(22), 5793-5816

Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., & Fenicia, F. (2022). Improving hydrologic models for predictions and process understanding using neural ODEs. Hydrology and Earth System Sciences, 26(19), 5085-5102

Pujol, L., Garambois, P. A., & Monnier, J. (2022). Multi-dimensional hydrological–hydraulic model with variational data assimilation for river networks and floodplains. Geoscientific Model Development, 15(15), 6085-6113

THANK YOU

Contact: ngo-nghi-truyen.huynh@inrae.fr or pierre-andre.Garambois@inrae.fr

Related paper:

Learning Regionalization within a Differentiable High-Resolution Hydrological Model using Accurate Spatial Cost Gradients

SMASH open source code:

https://github.com/DassHydro/smash

SMASH documentation:

https://smash.recover.inrae.fr

THANK YOU

TO OUR SPONSORS!

This work was supported by funding from SCHAPI-DGPR Météo France, ANR grant ANR-21-CE04-0021-01 (MUFFINS project, "***MU**ltiscale **F**lood **F**orecasting with **IN**novating **S**olutions*"), and NEPTUNE European project DG-ECO

VIGICRUES

METEO FRANCE

anr agence nationale de la recherche