



**HAL**  
open science

# 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

## ► To cite this version:

Ngo Nghi Truyen Huynh, Pierre-André Garambois, François Colleoni, Benjamin Renard, Hélène Roux. Advanced Hybrid Data Assimilation for Parameter Regionalization within a Differentiable Spatially Distributed Hydrological Model and Uncertainty Correction with Bidirectional LSTM. AGU Annual Meeting, 2023. hal-04389155

**HAL Id: hal-04389155**

**<https://hal.inrae.fr/hal-04389155>**

Submitted on 11 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**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*

**AGU23**

WIDE. OPEN. SCIENCE.

## Ngo Nghi Truyen Huynh

---

Research Engineer/PhD Student  
*INRAE, France*

Machine Learning  
Data Assimilation  
Hydrology



WIDE. OPEN. SCIENCE.



## 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

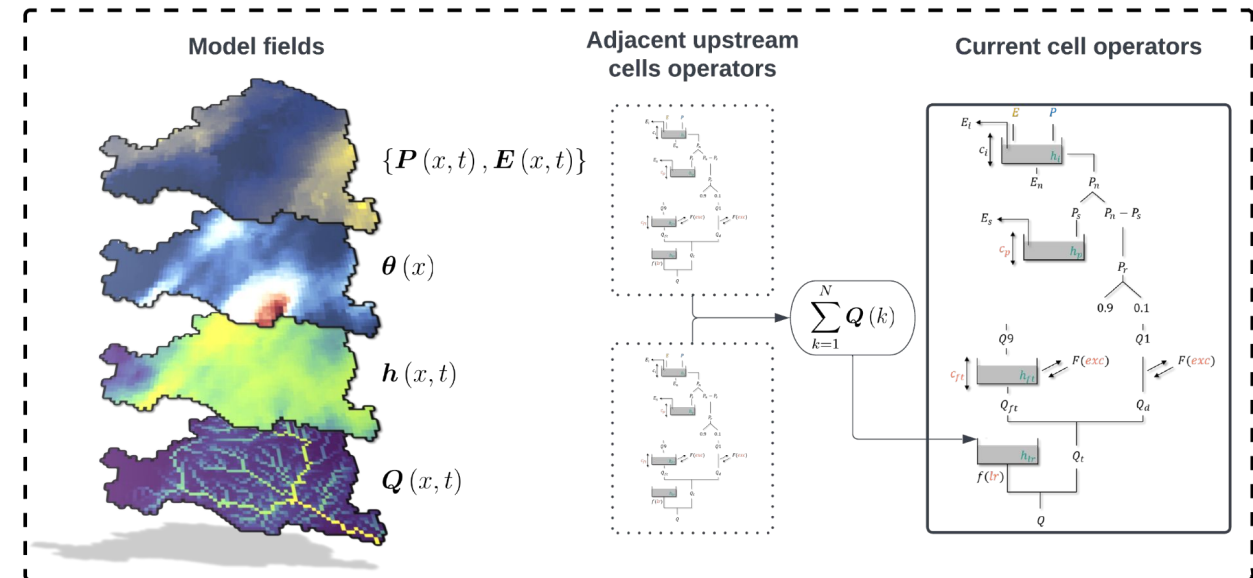
WIDE. OPEN. SCIENCE.

## 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

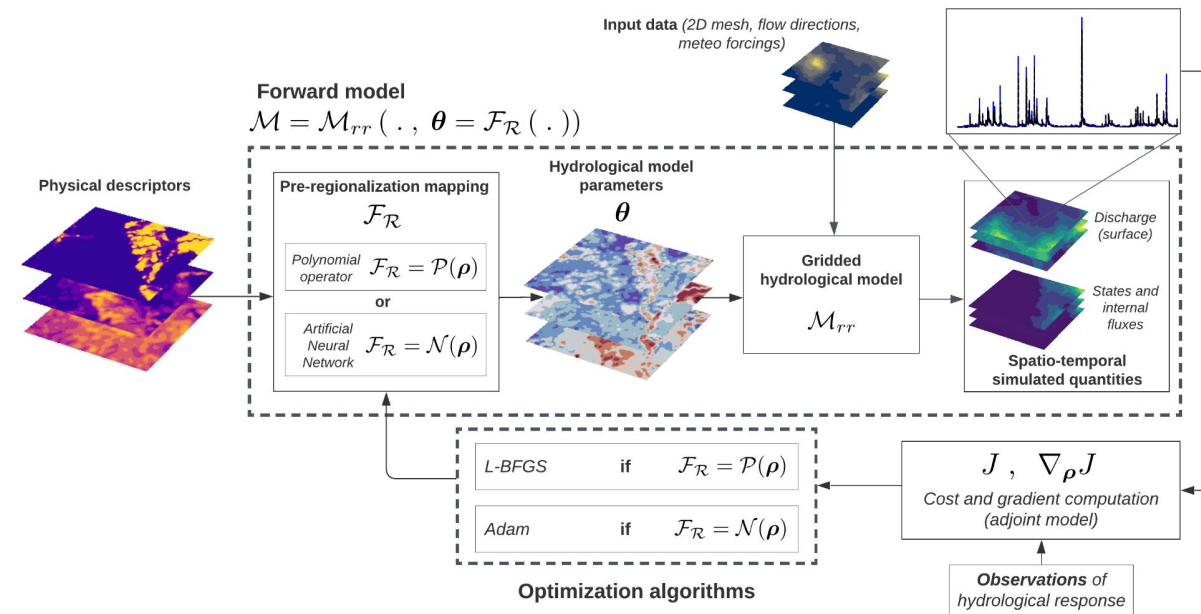


## Learnable regionalization

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

A new HDA-PR approach with:

- A differentiable spatially distributed 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



# Learnable regionalization

- Rainfall-runoff model:

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

- Full forward model with regionalization mapping:

$$\mathcal{M} = \mathcal{M}_{rr}(\cdot, \boldsymbol{\theta} = \mathcal{F}_R(\cdot))$$

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

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

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

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

# Learnable regionalization

- Inverse problem:

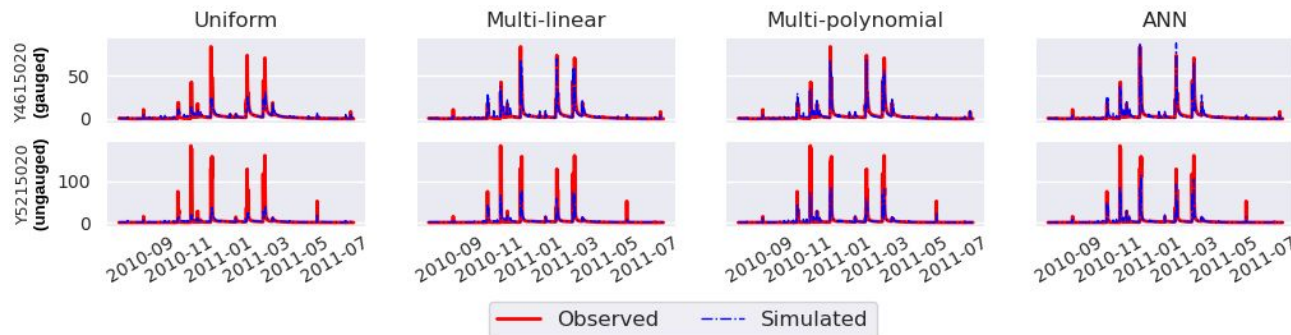
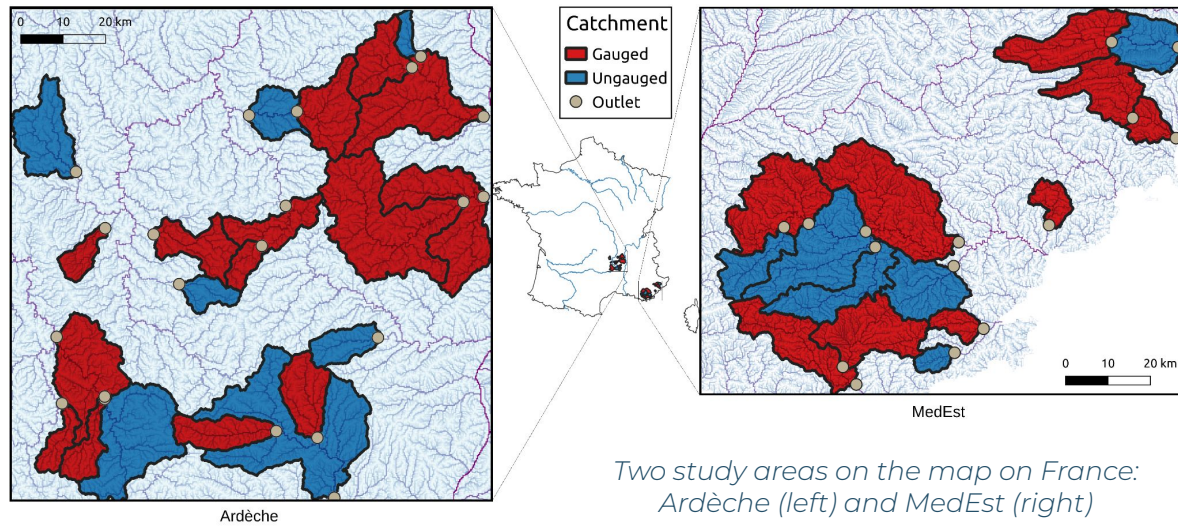
$$\hat{\rho} = \arg \min_{\rho} J(\mathbf{U}(\rho))$$
$$J(\mathbf{U}(\rho)) = J(\mathbf{Q}^*, \mathcal{M}_{rr}(\cdot, \boldsymbol{\theta} = \mathcal{N}(\mathbf{D}, \rho)))$$

- Gradient computation:

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

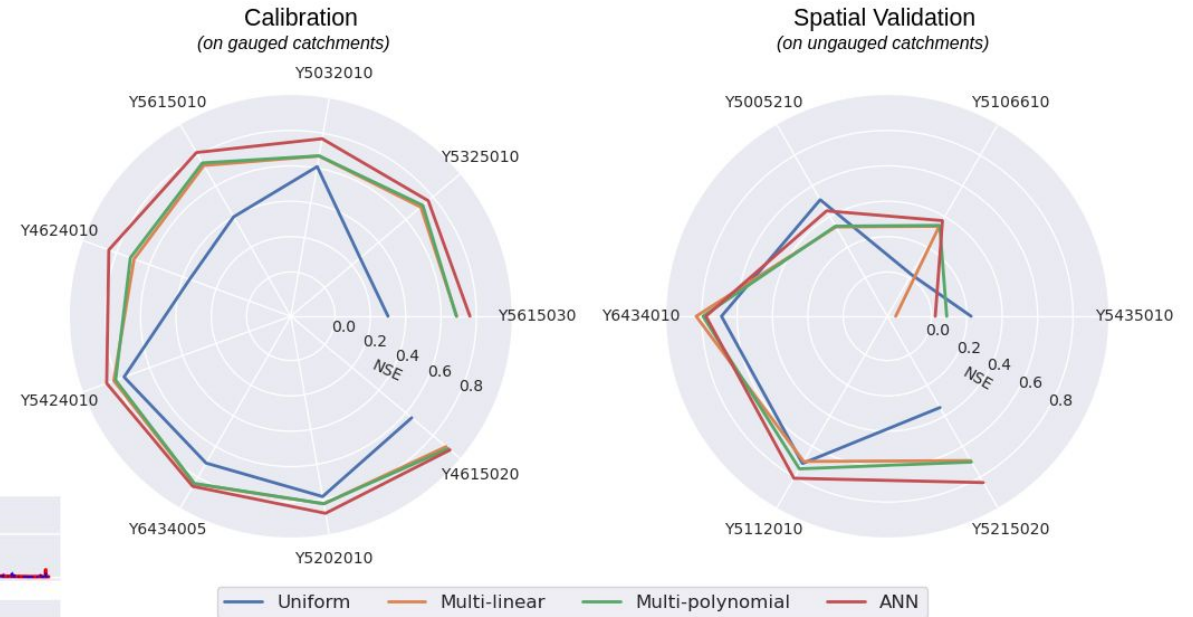
The first term can be computed via the automatic differentiation applied to the Fortran code and the second is obtained by analytical calculus applicable given the explicit architecture of the neural network

## Results



One year of observed and simulated discharges (in  $m^3/s$ ) at one gauged (Y4615020) and one pseudo-ungauged catchment (Y5215020)

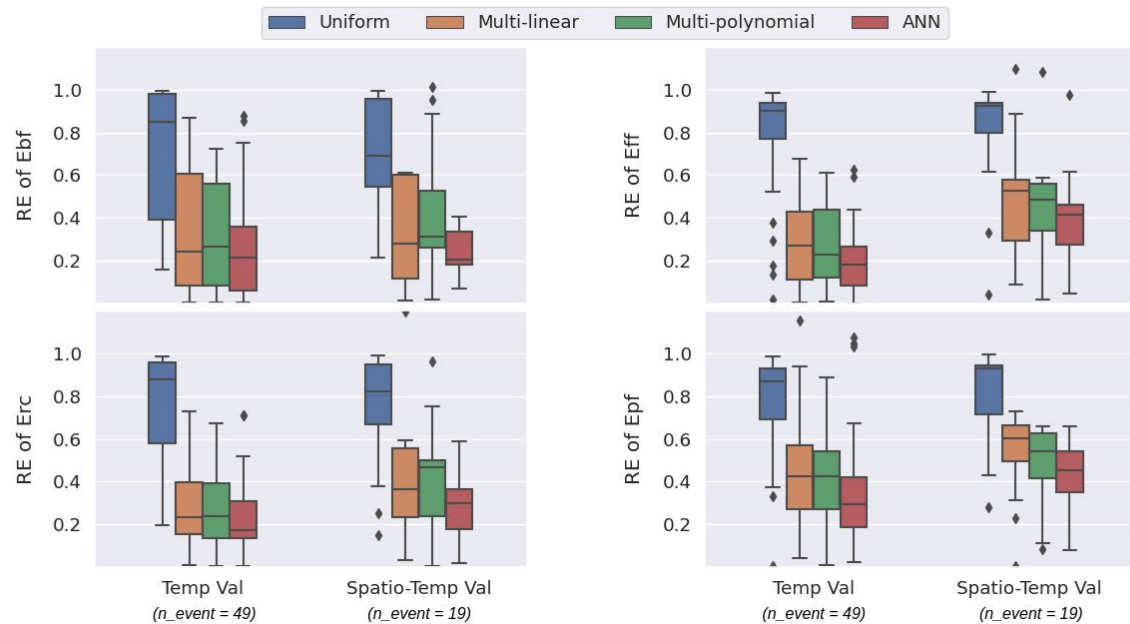
Calibration period: 2006-2016  
Validation period: 2016-2018



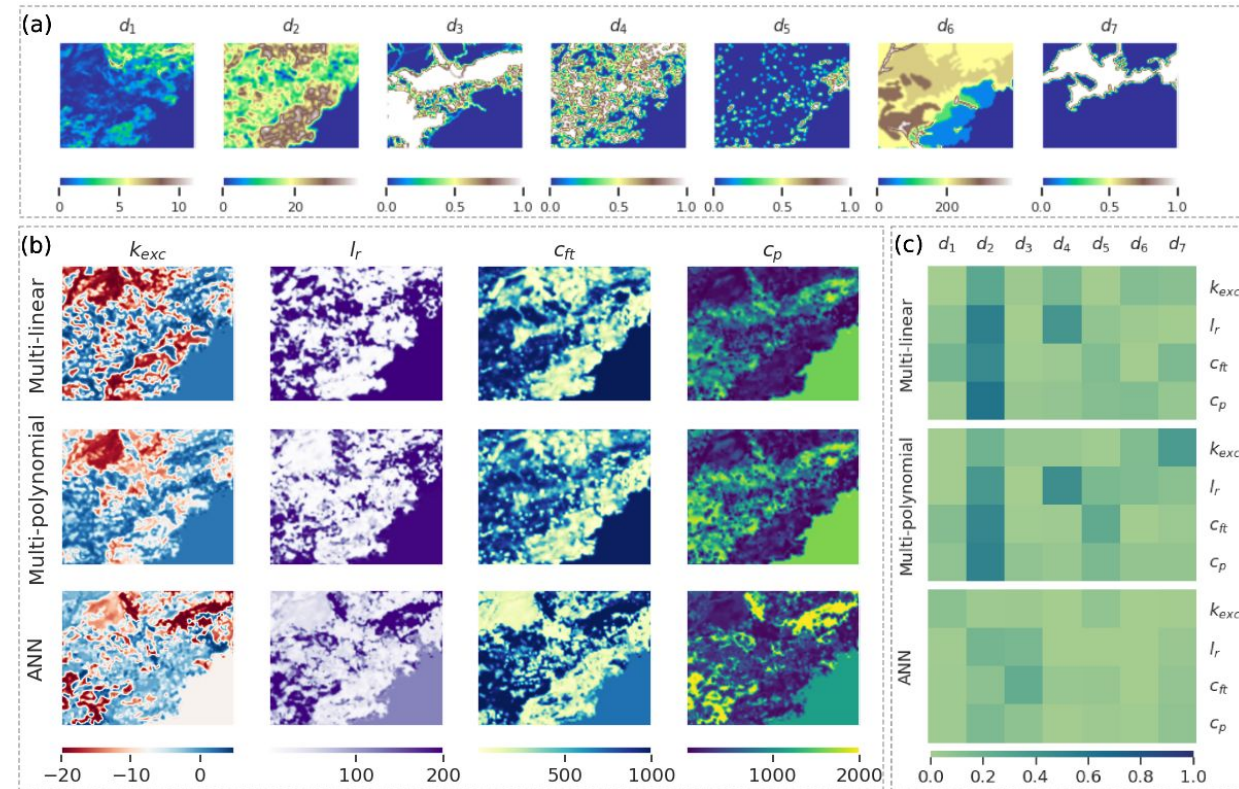
Radial plots of the NSE (optimal value = 1) in gauged catchments (left) and pseudo-ungauged catchments (right)

## 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 parameter 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

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

where:

$$(\mu, \sigma) := \text{LSTM}(\cdot, x, t)$$

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

**WIDE. OPEN. SCIENCE.**

## 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](mailto:ngo-nghi-truyen.huynh@inrae.fr) or  
[pierre-andre.Garambois@inrae.fr](mailto: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>



# AGU23

WIDE. OPEN. SCIENCE.



# 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, "*MUltiscale  
Flood Forecasting with INnovating Solutions*"), and  
NEPTUNE European project DG-ECO



# AGU23

WIDE. OPEN. SCIENCE.