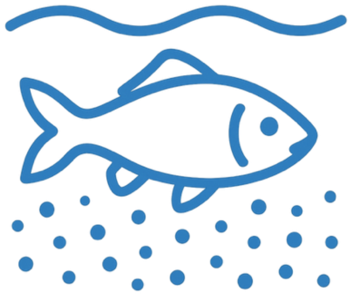
PREDICTION OF SUSPENDED SEDIMENT LOAD USING MACHINE LEARNING : RHÔNE RIVER

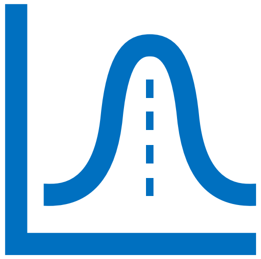
T.HAMADENE, V.NICOULAUD-GOUIN, H.LEPAGE AND M.FOULADIRAD

01

INTRODUCTION AND CONTEXT

INTRODUCTION AND CONTEXT





Suspended Sediment Load (SSL)

- Quantity of sediment transported by a stream/river
- Key variable in river management and ecosystem health

Why it matters in the Rhône river

- SSL directly impacts particulate radionuclide concentrations
- Rhône river being on of the most nuclearized rivers in the world

Modeling Approches

- Traditional methods : Empirical and physical models
- Machine and Deep Learning models

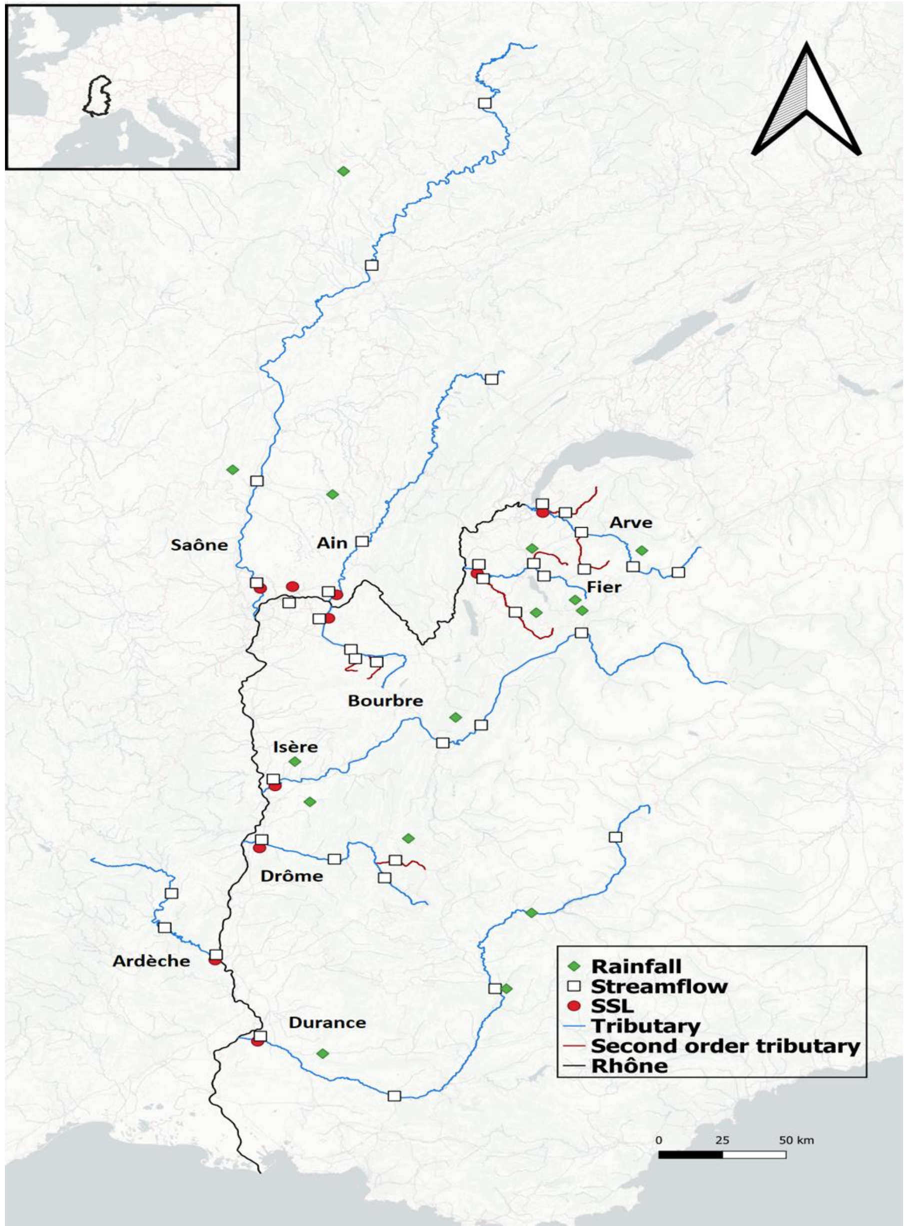
DATA AND STUDY AREA

- Rhône originates from a glacier located in the Swiss Alps at an altitude of approximately 2,208 meters and flows through Switzerland and France for about 813 kilometers before emptying into the Mediterranean Sea.

- SSL is studied in the exit of the 9 main tributaries of the Rhône river

- SSL is monitored using turbidimeters and was extracted from the OSR databank (bdoh.com)

- Rainfall and streamflow data were collected from the open-source French databank (hydroportail.com)



Rhône river and its main tributaries

02

DATA & METHODOLOGY

INPUT DATA

- Input data varies from one tributary to an other according to data avaibility

- Data is in hourly timestep
  - Models are trained on 80% of the data and evaluated on the remaining 20%

Tributary	Years	N observation	Input variables
Drôme	2019-2023	30361	4 streamflows, 1 rainfall
Durance	2014-2023	39920	4 streamflows, 3 rainfalls
Ain	2012-2015	3064	3 streamflows
Saône	2014-2023	47361	4 streamflows, 3 rainfalls
Fier	2014-2023	66516	5 streamflows, 2 rainfalls
Bourbre	2011-2013	17679	4 streamflows
Arve	2019-2023	30576	6 streamflows, 2 rainfalls
Isère	2019-2023	29761	4 streamflows, 3 rainfalls

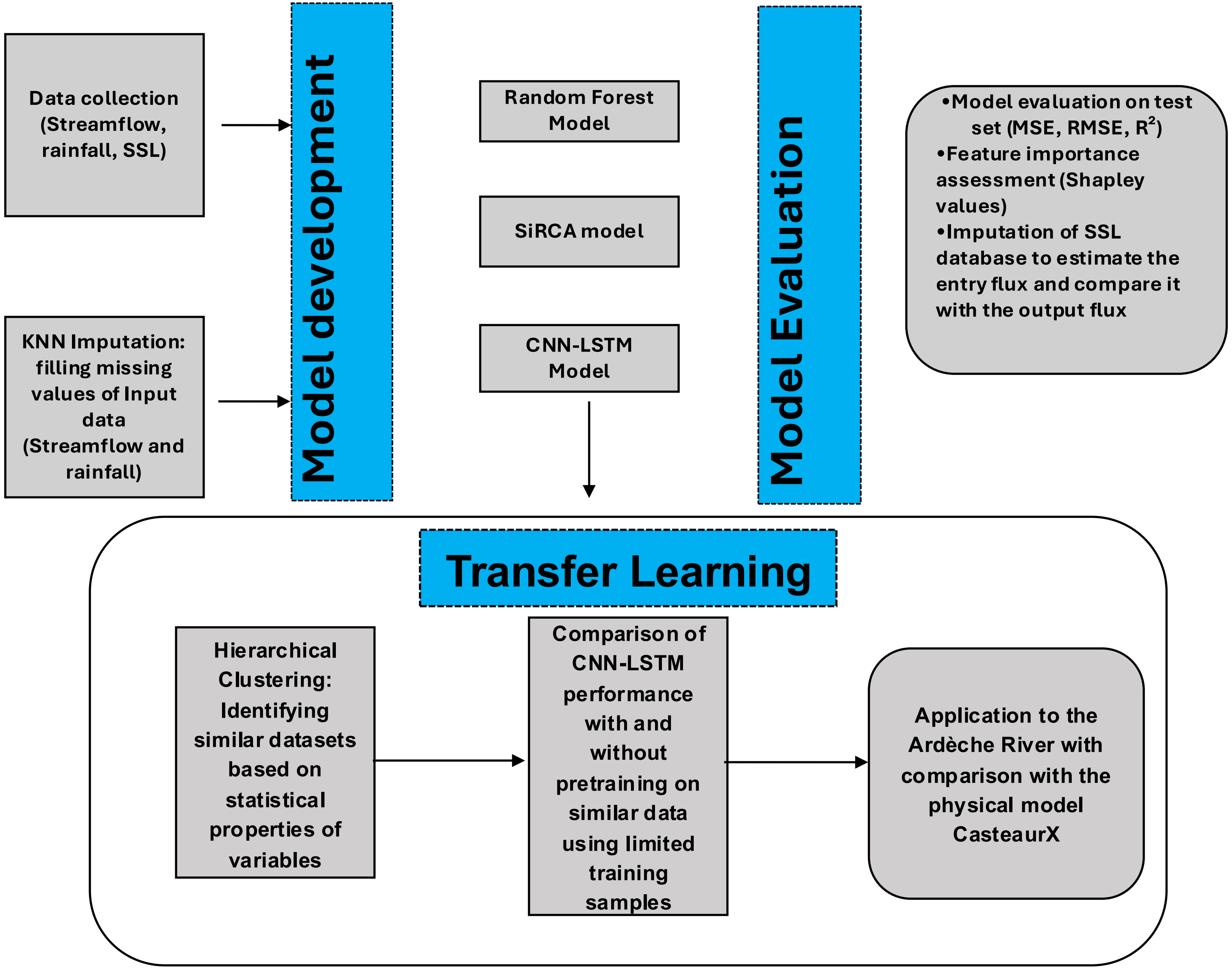
Data

METHODOLOGY

- **Streamflow and rainfall data across the river is used for every SSL variable at the exit of each tributary**

- **Evaluation of 3 models :**
  I. Empirical model : Simplified Rating curve model (SiRCA)
  II. Machine Learning : Random Forest
  III. Deep Learning : CNN-LSTM

- **For rivers where Data is scarce, a Transfer Learning approach was proposed to enhance performance of the CNN-LSTM model**



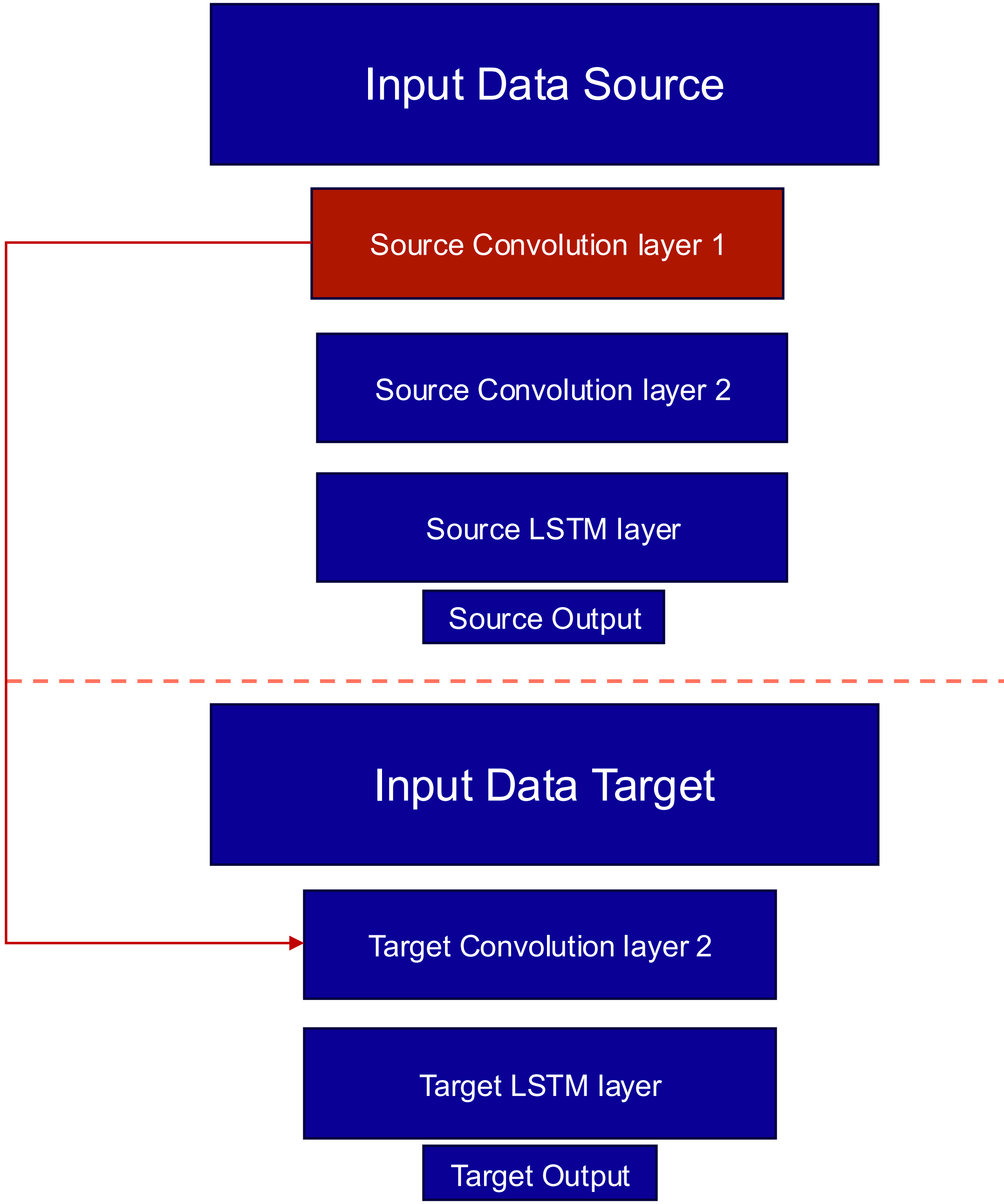
TRANSFER LEARNING

OBJECTIVES

- **Improve model performance**
  - By using the parameters *learned* from another river
- **Reduce training time**
  - By freezing one or multiple layers of the model

METHODOLOGY

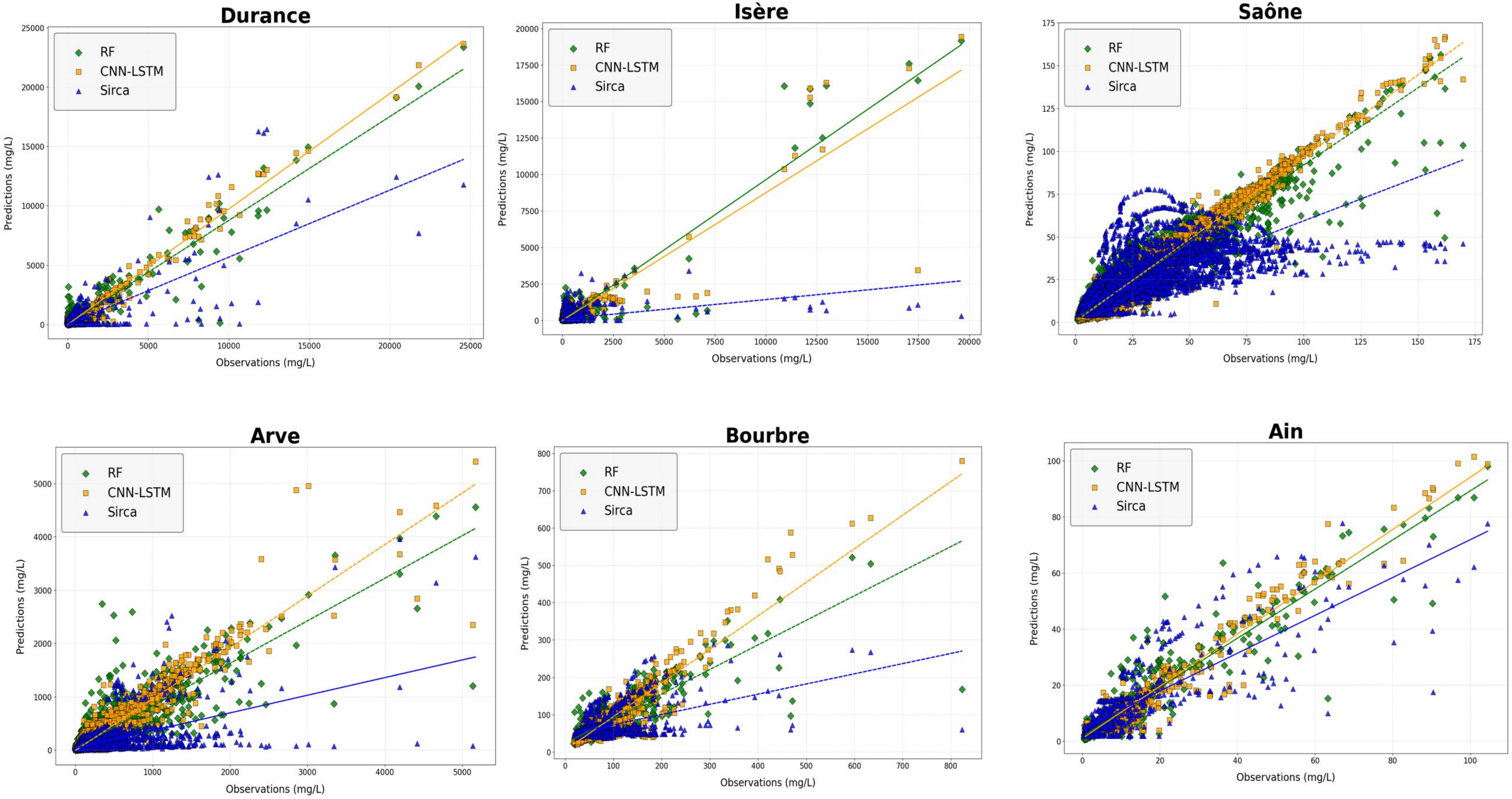
- **Identify similar rivers using hierarchical clustering**
  - The clustering is made using the statistical properties of SSL and three streamflow variables
- **Pretrain the model on a source river and finetune it on a target river**
  - The first layer of the model is frozen



03

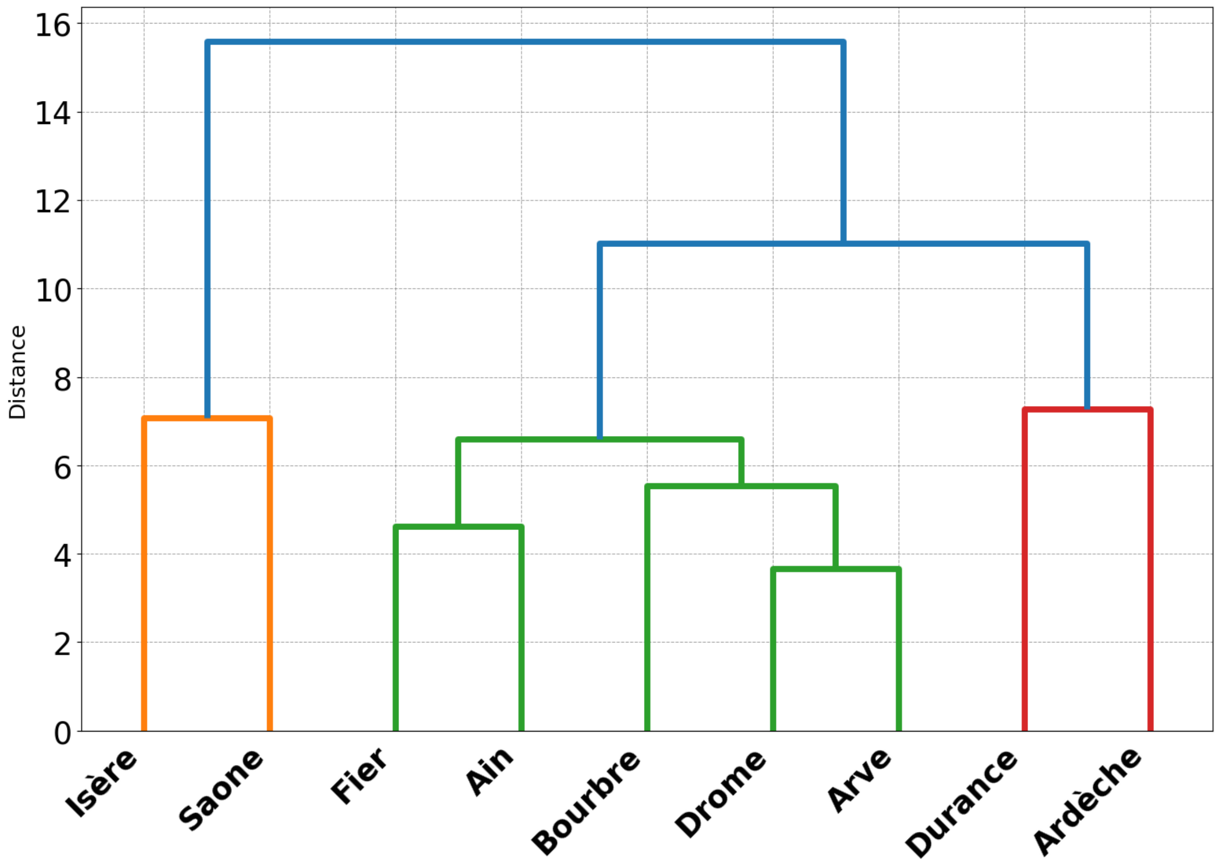
RESULTS

COMPARING THE THREE MODELS



RIVER CLUSTERING

- Consideration of 3 streamflow variables (ordered from downstream to upstream) and the SSL for classification.
  - We calculate the statistical properties: the mean, the standard deviation, the quartiles, the maximum, the skewness, and the kurtosis.
- Application of **divisive hierarchical clustering**.
  - Interpretation of the results with PCA and Kohonen networks
- Evaluation of Transfer Learning on obtained clusters

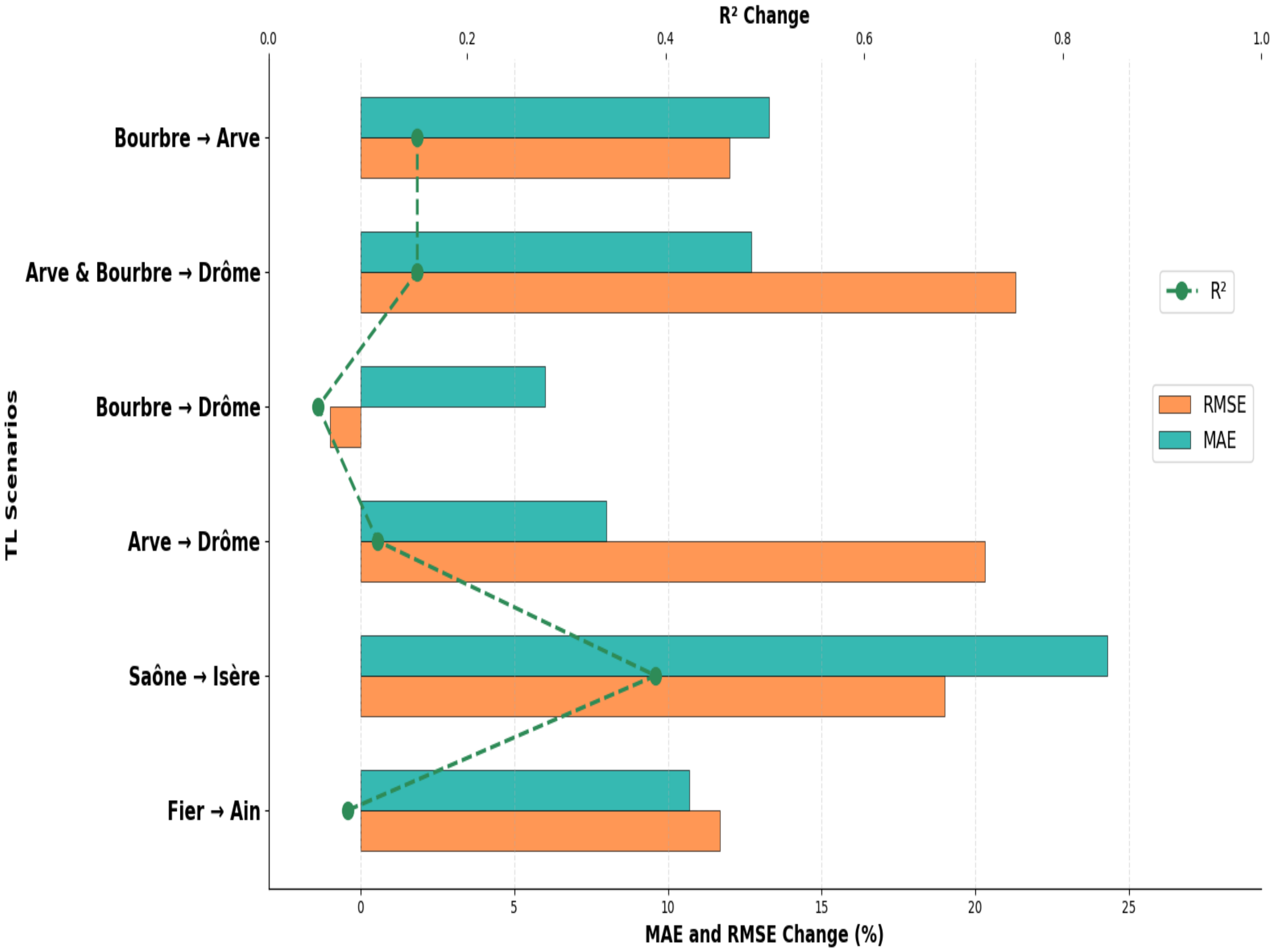


Clustering results

TL RESULTS

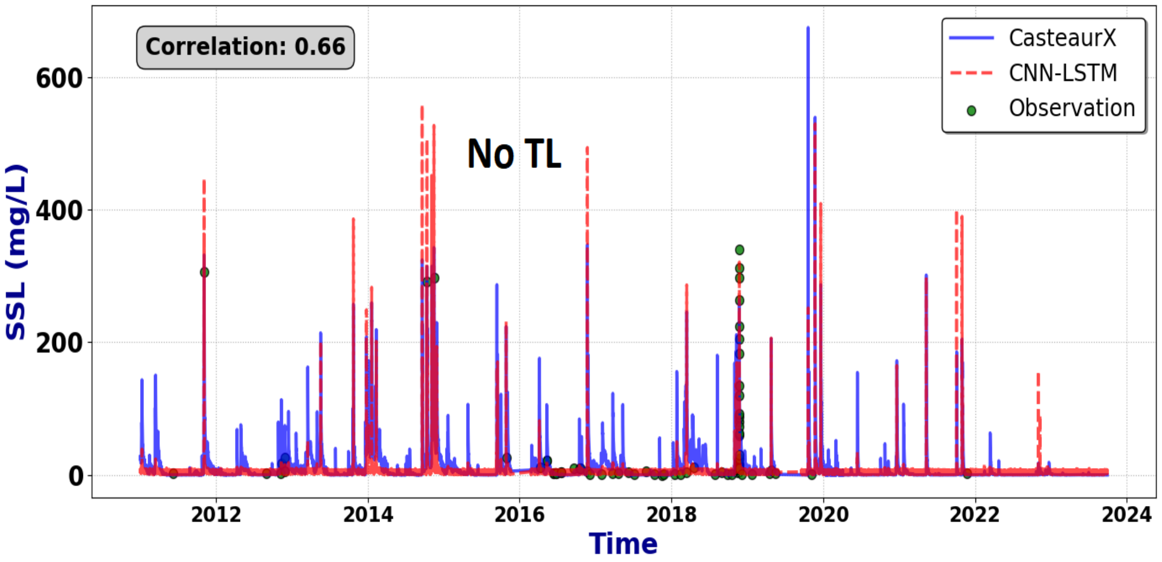
Target	Source	% Train	$\Delta(R^2)$	Δ (RMSE %)	Δ (MAE %)
Ain	Fier	5% 10% 15%	0.07 **0.11** 0.05	9% **16%** 10%	8% **14%** 10%
Isère	Saône	5% 10% 15%	0.35 **0.42** **0.42**	17% 18% **22%**	22% **27%** 24%
Drôme	Bourbre	5% 10% 15%	**0.11** 0.04 -0.01	-3% **1%** -2%	**13%** 6% -2%
	Arve	5% 10% 15%	**0.21** -0.08 0.2	**34%** -7% **34%**	**17%** -10% **17%**
	Arve & Bourbre	5% 10% 15%	**0.4** 0.06 -0.02	**25%** 22% 17%	**33%** 9% -4%
Arve	Bourbre	5% 10% 15%	0.05 **0.28** 0.12	5% **17%** 14%	7% **22%** 11%

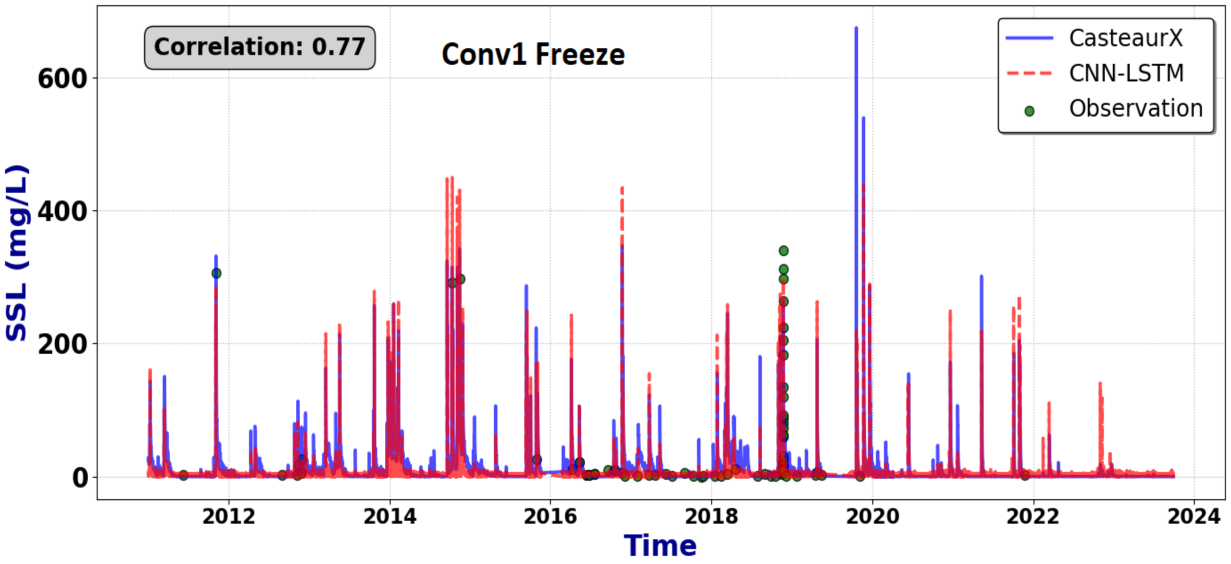
TL performance change

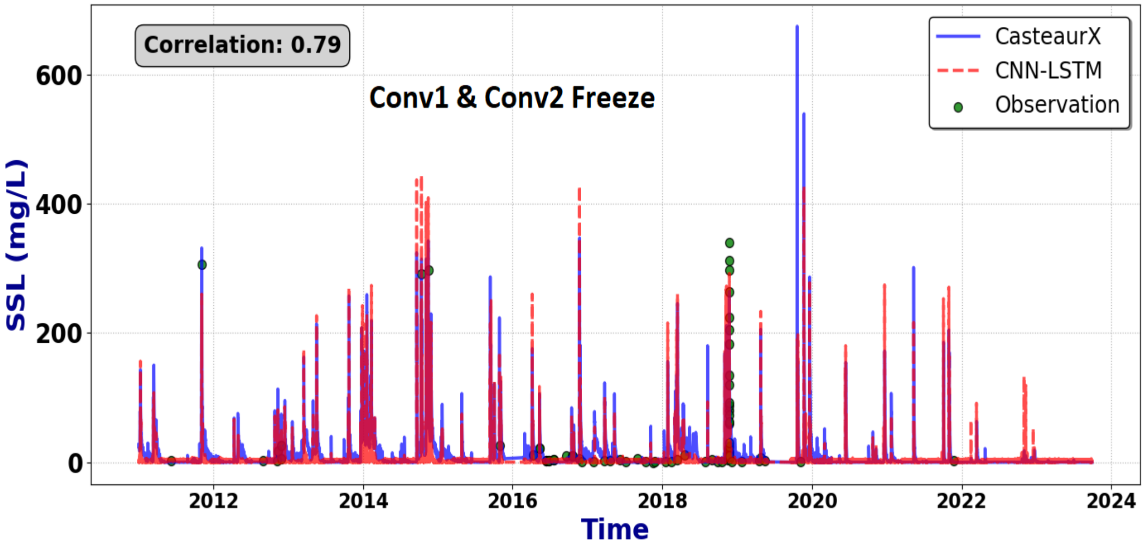


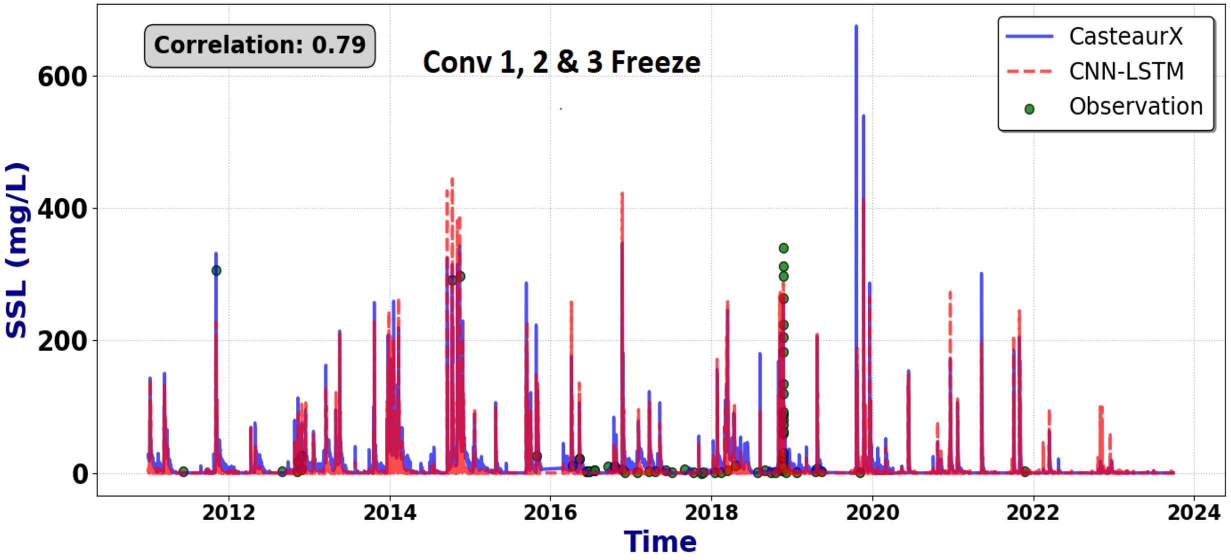
TL Mean Performance change per scenario

ARDÈCHE SIMULATION ACCORDING TO THE TL APPROACH



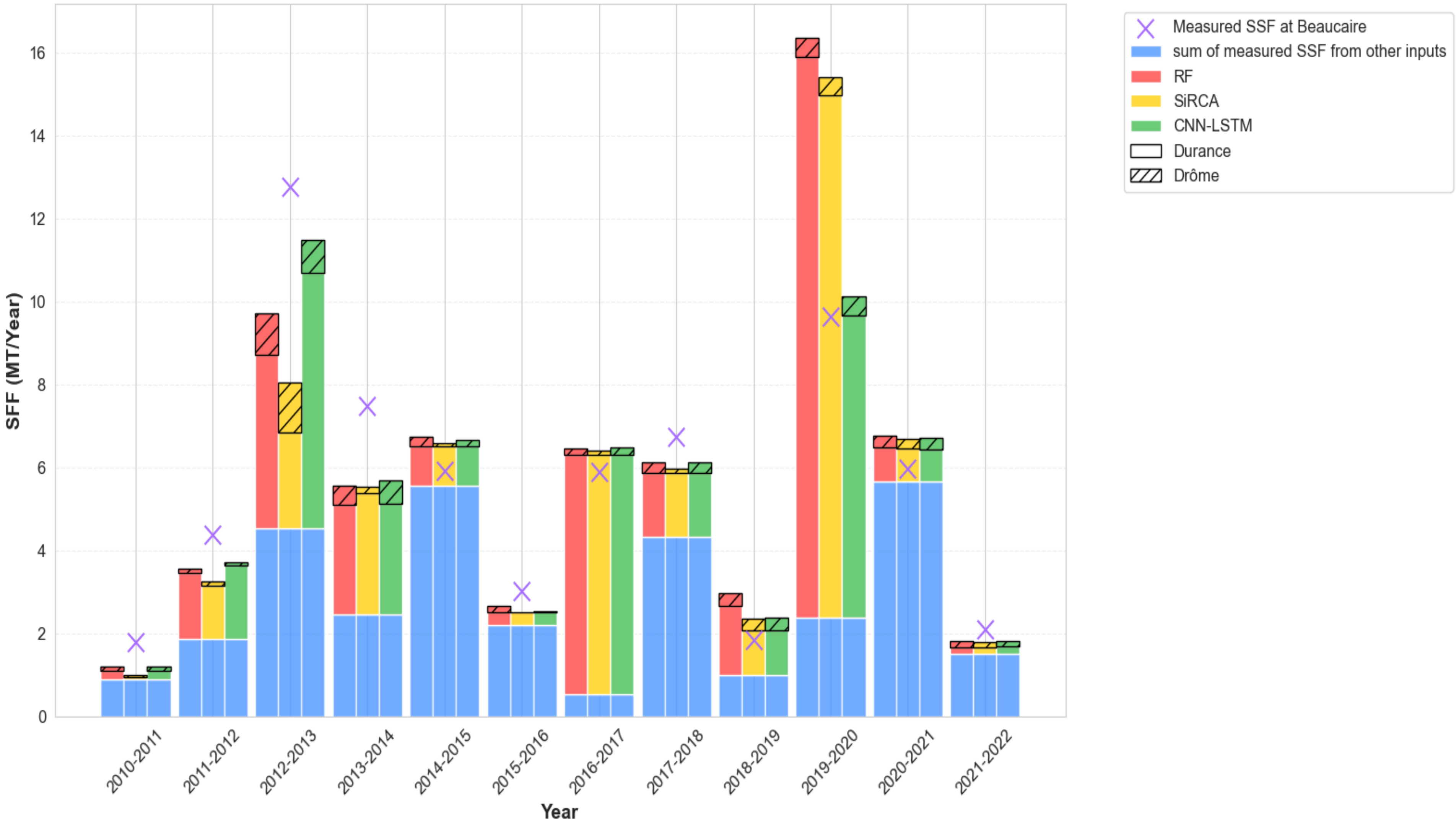






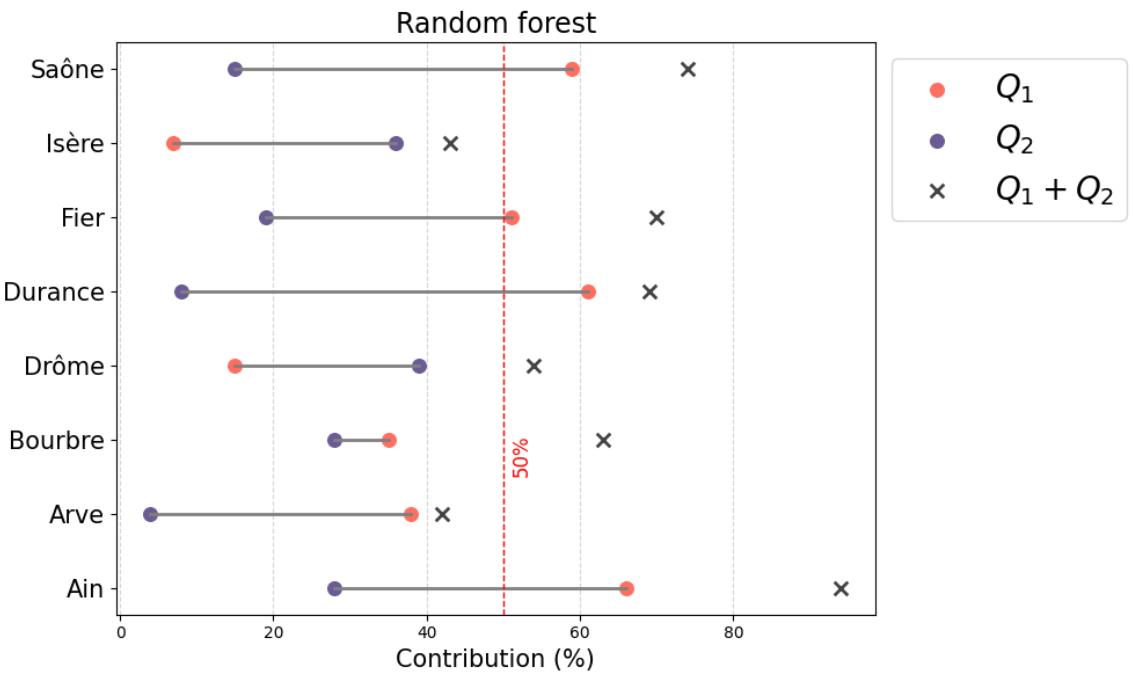
MODEL EVALUATION ON FLUX CALCULATION

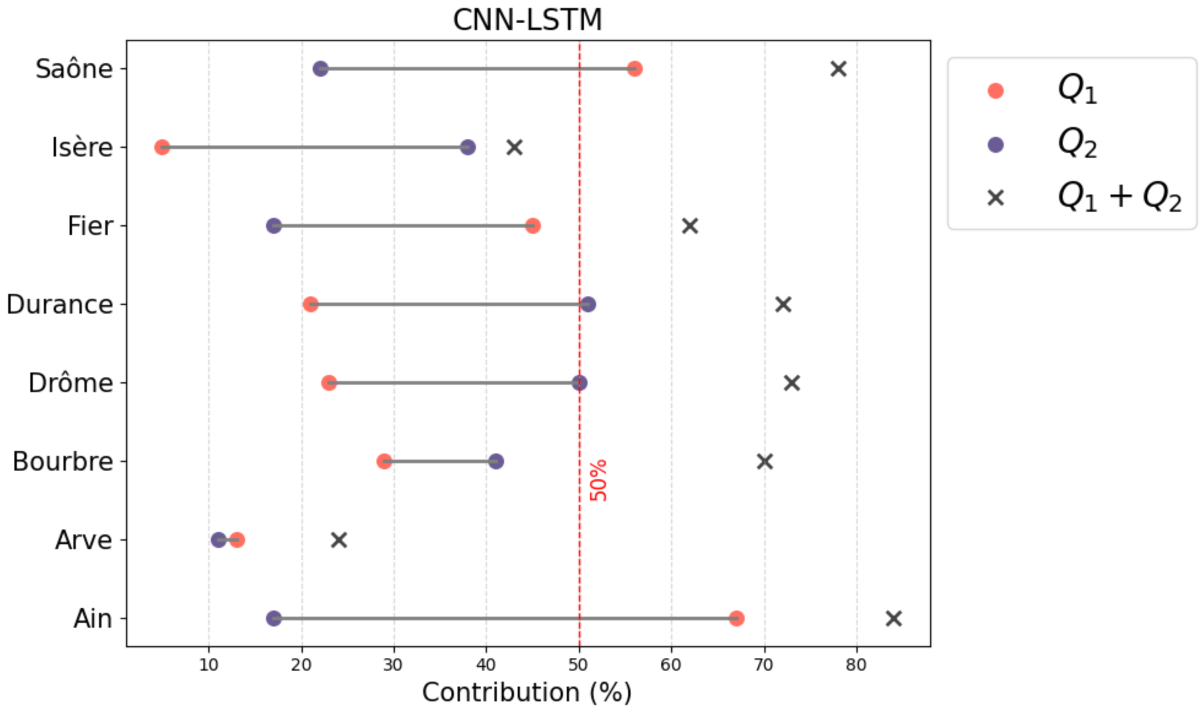
- Models are trained on the full dataset
- Every missing value is filled with the model output
- The annual flux is calculated with three models on Durance and Drôme rivers



Flux comparison : entry (coloured bars) vs exit (X mark)

FEATURE IMPORTANCE





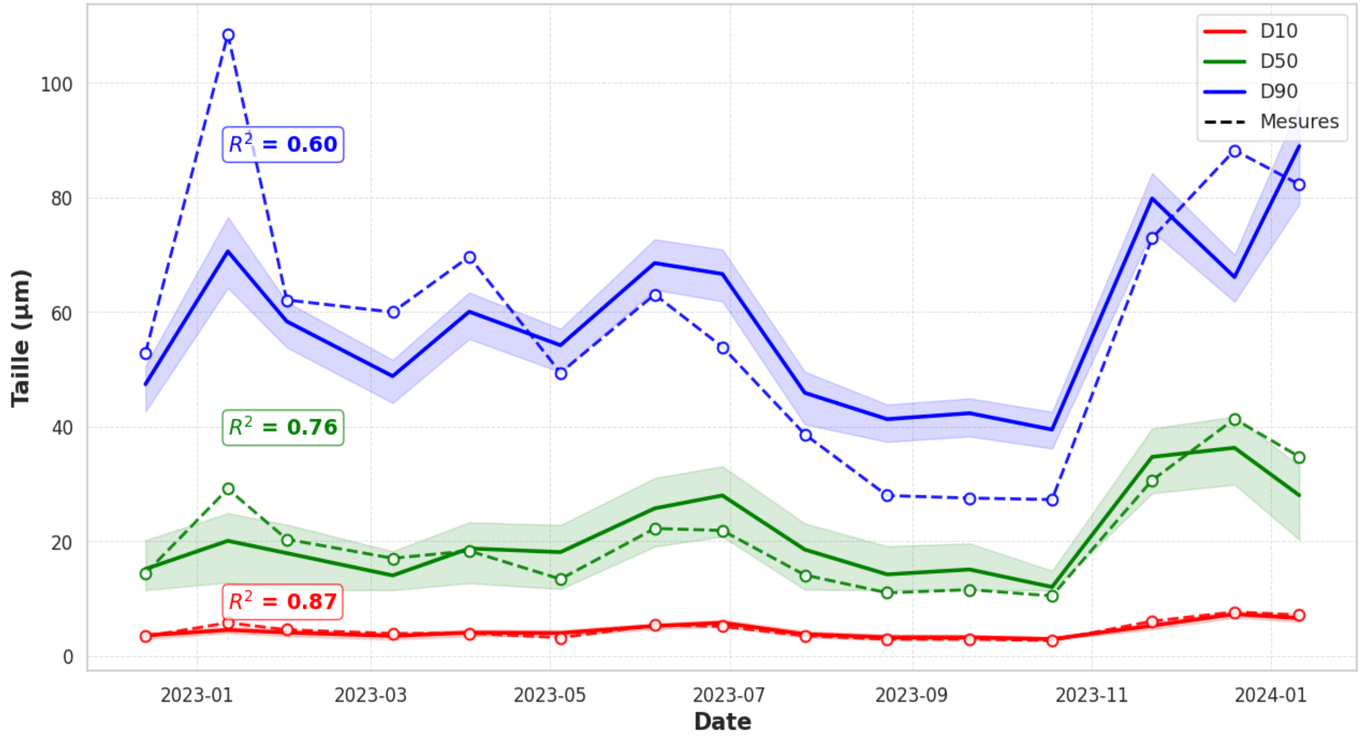
- Shapley values indicate that rainfall does not contribute much to SSL prediction (<1% in every river)
- Variable importance varies from river to river, highlighting the need to consider multiple streamflow variables
- The empirical model performs best when Q1 is the most important variable, suggesting the utility of feature importance to select the streamflow variable for rating curve equations

04

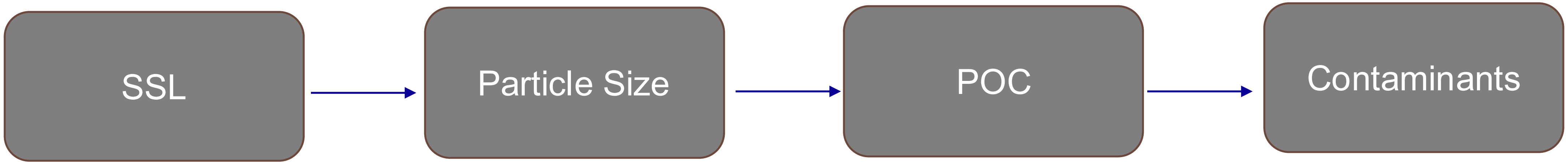
CONCLUSION & FOLLOW UP WORK

FOLLOW UP WORK

- Exploring transformer-based models as an alternative approach and compare against reference models.
- Integrating SSL data into a modelling chain of particle size, particulate organic matter, and contaminant concentrations (mainly radionuclides).
- Quantifying of the model uncertainty (Monte Carlo Dropout, Ensemble Techniques)



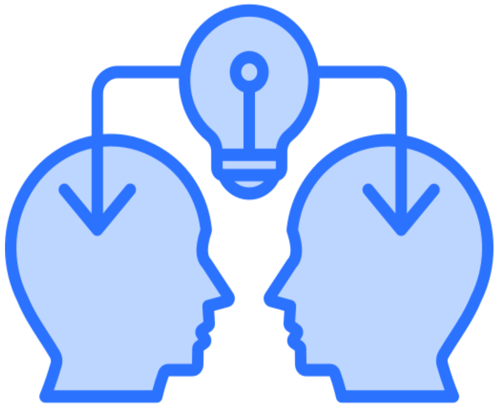
Particle size (D10, D50 and D90) prediction in test set in the Isère river



CONCLUSION & LIMITATIONS

- The CNN-LSTM model stands superior in estimating SSL and in filling the missing values to asses the flux.

- Transfer learning improves average performance when data is scarce but requires homogeneous datasets:
  - The proposed method for identifying river similarities remains relatively simple and purely data-driven.
  - In few cases, TL reduced performance, highlighting the need for more robust similarity metrics.
  - Comparing with other models (SWAT) is needed when there is no observation to validate the TL simulation.



- Shapley values revealed river specific differences, which may reflect variability in soil erosion processes across river basins

THANK YOU FOR YOUR ATTENTION