Flood Risk Forecasting using a combination of hydrological modeling and machine learning

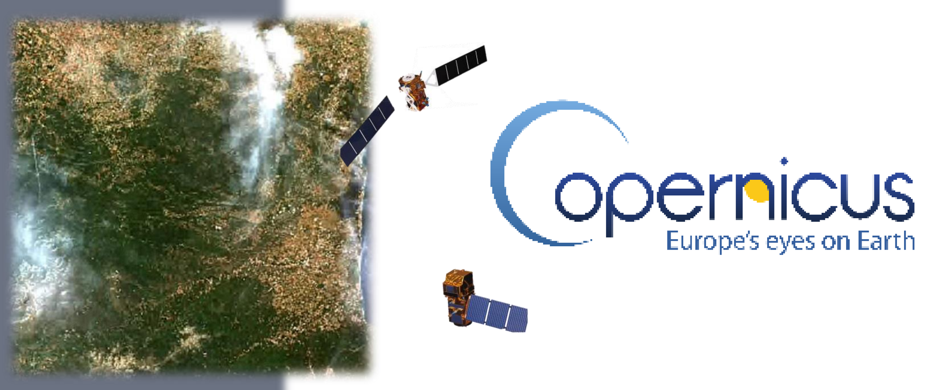
Boris Gratadoux, Imane Farouk, Laure Chaumat

THALES
Building a future we can all trust

An innovative approach combining *in-situ* & satellites imagery data and artificial intelligence for reliable forecasts of flood events

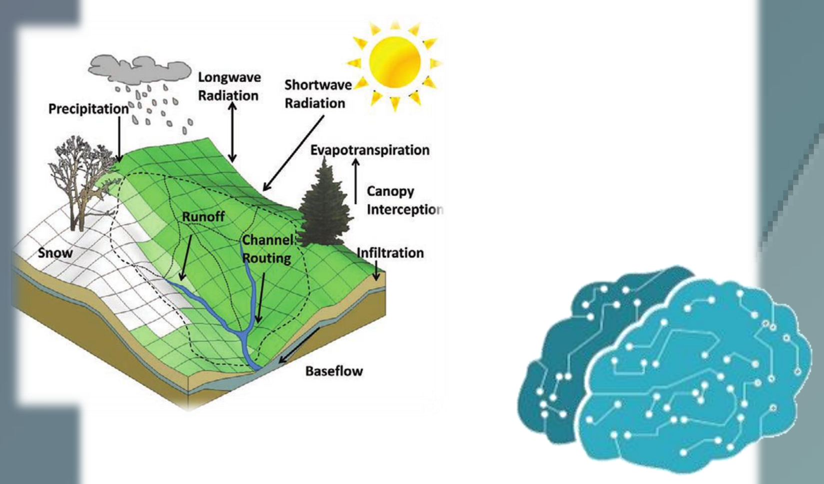
Input

- Meteorological data
- Satellite observations
- On-ground observations



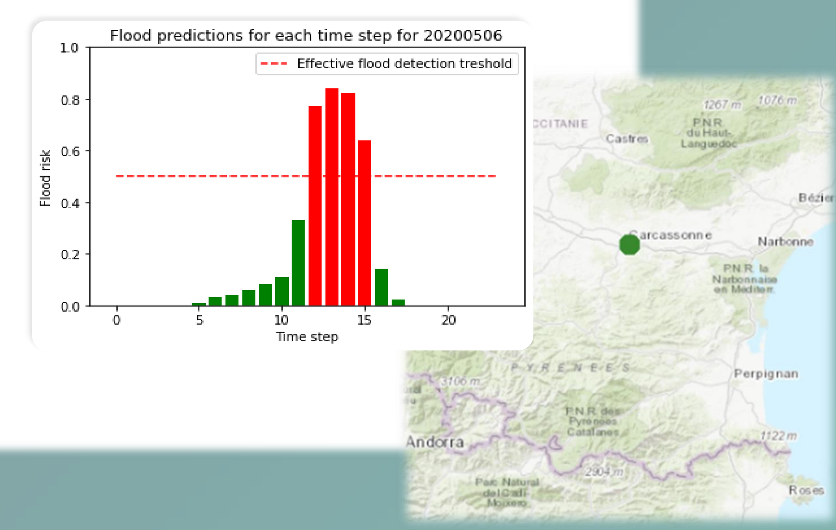
Combination

- Hydrologic modelling & Machine Learning



Output

- 5-day forecast
- Hourly time step
- 93% reliability



- <u>The hydrological model</u>

The HEC HMS model [1] used is a **conceptual semi-distributed model** which includes snowmelt. It was created by the US Army Corps of engineers and used in many studies with good results. It takes precipitation and temperature as input to provide future out-flow at the outlet of the selected basin.
To allow computation of the best estimate of the initial state of the hydrological model, particle filter data assimilation is used:

- <u>Ensemble creation method</u>

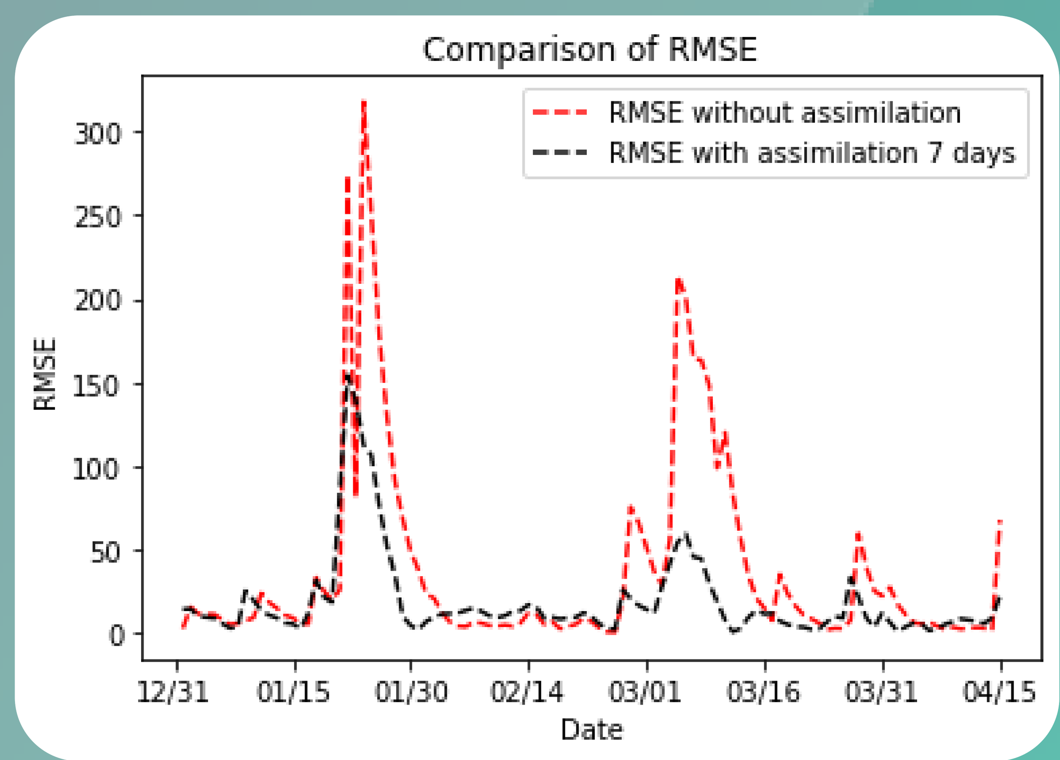
To create our ensemble, for correcting the initial state, and take into account the uncertainty of the forcing, we assume that the ideal forcing is not very far from the known one. We therefore create a set of forcing perturbed several times, this set will contain the ideal forcing. The perturbation method chosen corresponds to an **autoregressive (AR) model** :

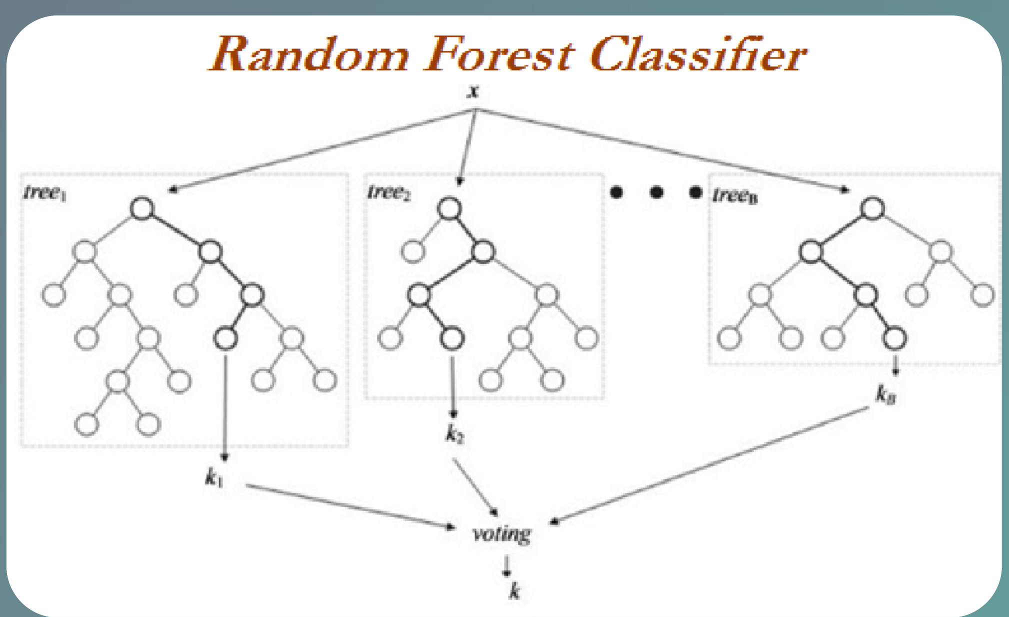
$$X(t) = \varphi X(t-1) + \varepsilon$$

With φ a parameter and ε a noise of expectation zero and variance σ^2

- <u>Data assimilation ensemble method</u>

The principle of the particle filter is to run successive simulations over short periods of time (7 days in our case), called a simulation window. Several simulations are run in parallel, each with different perturbed forcing. The resampling method used is Kitagawa resampling [2]. On the initialisation period, **RMSE decreases by 49% with the use of assimilation**.





The <u>Machine Learning algorithm</u> allows transforming flow forecast in flood information. It takes two types of input:

- The set of flows forecasted with data assimilation at the basin outlet
- The soil state observation on multi-sensor satellite products extracted mainly from **Copernicus**

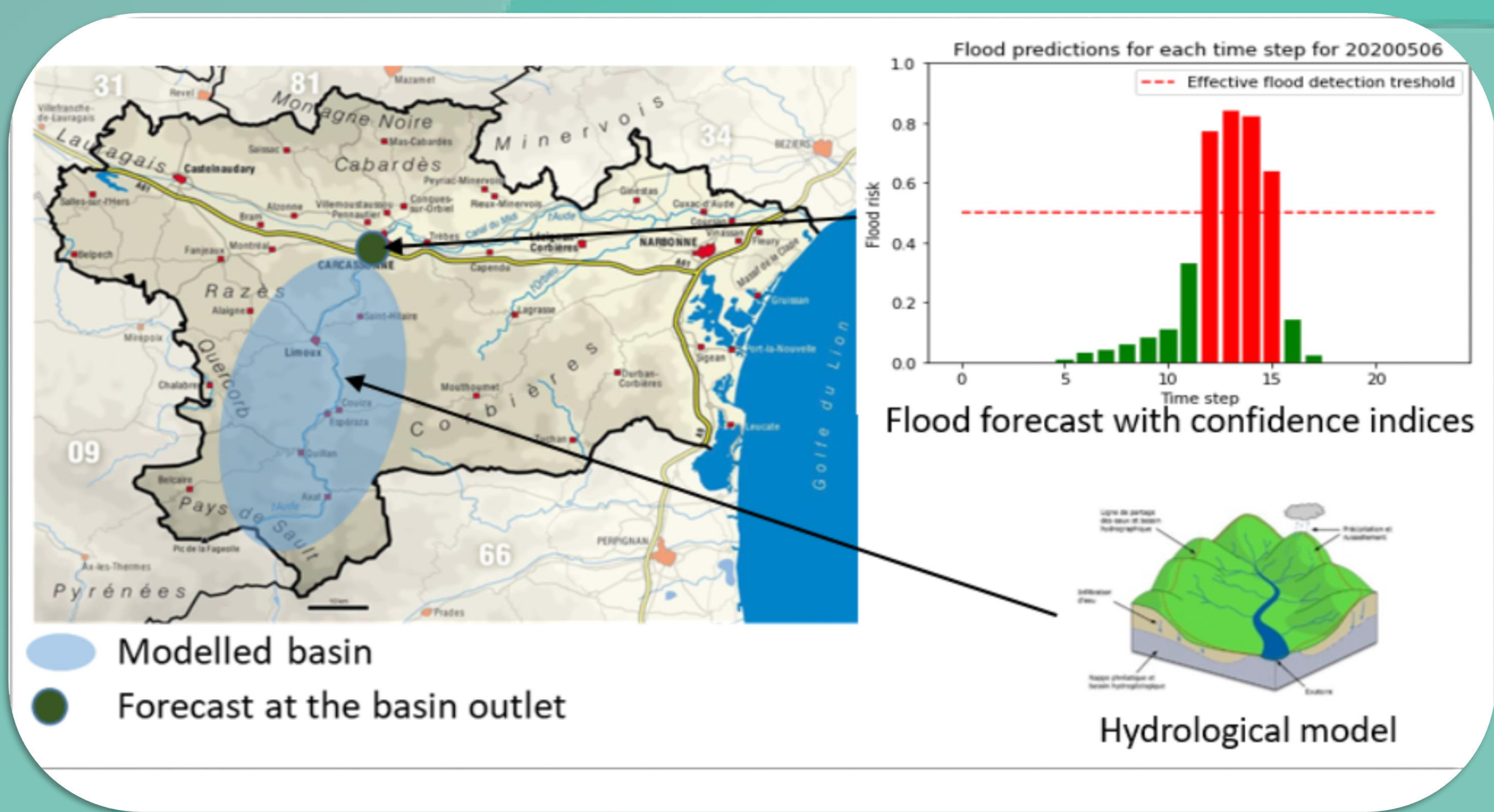
Preprocessing is needed to extract the most relevant information from the data. The use of **"lagging method"** on the flow forecast increases time information and stabilize the average closer to zero.
The satellite products are used to improve the soil state basin over the watershed:

- Moisture Stress Index (MSI) from Sentinel-2 images;
- Normalized Difference Vegetation Index (NDVI) from Sentinel-2 images;
- Normalized Difference Water Index (NDWI) from Sentinel-2 images;
- Soil Surface Moisture (SSM) from Sentinel-1 images;
- Soil Moisture (SM) from SMOS images.

The learning database of the model is composed of **past flood events and events with high flow values** but without flooding. The "truth state" of each time step is built manually.

The prototype has been tested and validated on **the French watershed of Aude**. For these test cases, we use precipitation and temperature from **ECMWF**'s ERA5 re-analysis and flow reports at the Carcassonne station made available by SCHAPI, the French national service mandated for flood forecasting in France.
Table below summarizes the score comparison obtained by cross-validation for different soil index inputs from satellite images:

	Score	Mean	Mean standard error on 6 cross-validations
Score without satellite data	AUC*	0.736	0.10
	Correct Forecast*	92.19 %	4.1 %
Score with NDWI, MSI and SSM	AUC*	0.746	0.12
	Correct Forecast*	92.77 %	4 %



Flood forecast with confidence indices

Ensemble data assimilation, where each ensemble member represents a different set of disturbed weather data, allows exploring the range of uncertainty of flood occurrence. This probability is expressed in the form of a confidence index which can be interpreted as a risk of flood.

<u>Outlook:</u> adaptation to other watersheds and addition of groundwater upwellings detection with IRT satellite observation

Definition:
*AUC= Area under the curve ROC, represents the rate of true positives over the rate of false positives, the positives are here the flood episodes
*Correct forecast is the percentage of forecasts where a flood was predicted and observed or there was no flood predicted and no flood observed and this is calculated independently on each time step

Références:
[1] M.M Patel, Prof. H.M.Gandhi, Dr. N.J. Shrimali. 2014. Literature Study on Application of HEC-HMS for Event and Continuous Based Hydrological Modeling
[2] Kitagawa G. (1996) Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models, Jounal of Computational and Graphical Statistics, Vol. 5, pp. 1-25, 1996

Farouk Imane et al., 2021: Hydrological modeling and uncertainty estimation due to atmospheric input data by means of ensemble methods. 4th Space for Hydrology Workshop in collaboration with GEOGloWS: "Inland Water Storage and Runoff: Modeling, In Situ Data and Remote Sensing", Hosted as a Virtual Event from ESA-ESRIN, Frascati (Rome), Italy, 7-11 June
Farouk Imane et al., 2022: Flood forecast combining hydrological modeling and Machine Learning (EUMeTrain). 5th H SAF User Workshop supported by EUMeTrain - Session 2: The usage of satellite products for flood and drought monitoring - EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF), held in online mode from 24th to 28th January 2022

boris.gratadoux@thalesgroup.com
laure.chaumat@thalesgroup.com

Thales Services Numériques, Labège, France