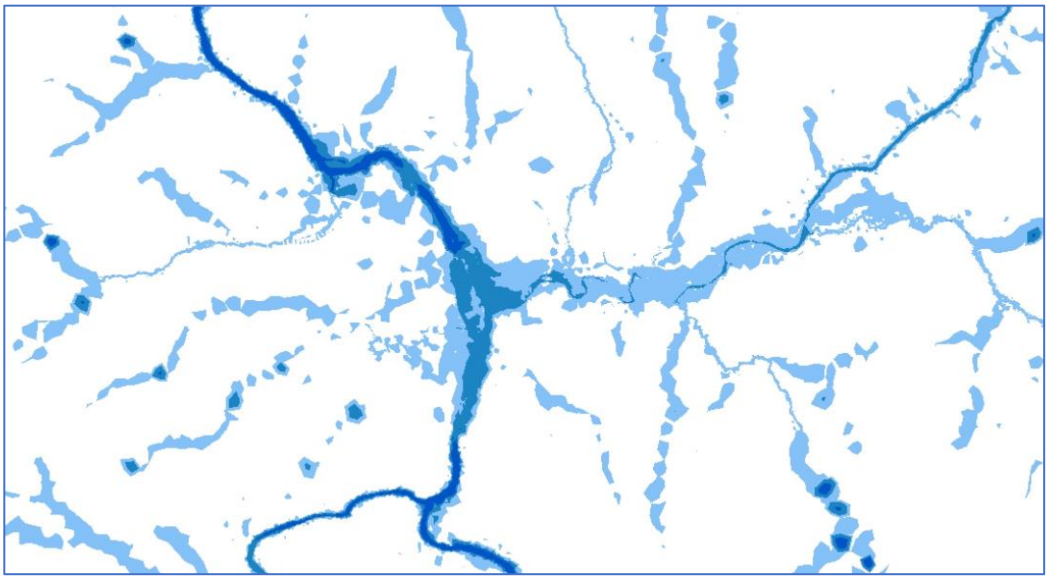
INTERREG SUDOE V-B EUROPA SUROCCIDENTAL FEDER 80%

Comparison of numerical models and application to the automatic production of flood maps in flood forecast



AVRIL 2022

Proyecto SOE3/P4/E0929 *Automatización del modelado de riesgos de inundaciones en cabeceras de cuenca a través de técnicas de inteligencia artificial y Big Data*

Authors: Philippe Sergent, Rémy Gasset, Bruno Bader, Nicolas Maquignon, Hassan Samoui (Cerema)

Thanks to the contribution of Sofiane Hadji (Sixense)

Table des matières

1. Introduction

The objective of the work is to produce flood maps automatically on the Nive basin from flow and rainfall measurements on the catchment area. The approach is based on the case-based reasoning method. We present in this report the complete methodology that is used to implement this method of case-based reasoning. It is divided into three parts which will be presented in details in the report. It is first a question of implementing an artificial neural network which makes it possible to determine Sobol indices. These Sobol indices are then used to establish the distances between scenarios of the case-based reasoning method. A hydraulic model is finally implemented in order to establish a database of referenced maps. These maps are selected according to the calculated distances and possibly averaged to provide the most reliable possible response.

2. Hydrological modeling

2.1. Nive water catchment

Nive water catchment that is located in South West of France covers more than 1000 km². Nive river is born at the foot of Mendi Zar (1,323 m), beyond the Spanish border, under the name of Harpeko erreka. Its main source is at an altitude of 360 m. With a length of approximately 80 km of linear with a general South-South-East / North-North-West orientation, the Nive results from the junction of Saint-Jean-Pied-de-Port of five main rivers :

- Nive of Arnéguy which has its source in Navarre at Puerto de Ibaneta;
- Nive of Béhérobie (or Esterençuby) which originates at an altitude of 1300m;
- Laurhibar which is born in the forest of Iraty;
- Arzuby which has its source at Bustince Iriberry;
- Nive of Aldudes which has its source upstream of Urepel.

The basin is made up of 1,300 km of permanent or temporary watercourses, including 365 km of main watercourses longer than 10 km. The hydrological regime of the Nive is of the rainy type with maximum flows observed from December to April. The average annual rainfall is 1,680 mm. Nive regime is pluvial and due to these abundant precipitations, with a very limited influence of snowmelt. Like most other Pyrenean watercourses in Adour basin, Nive is a very abundant river. Its flow was observed over a period of 42 years (1967-2008), in Cambo-les-Bains, a town located about fifteen kilometers from its confluence with Adour. The area thus observed is 870 km² or 85% of the entire catchment area of the river. The interannual hydrological flow of the river at Cambo-les-Bains is 30.2 m3/s. At low water, the minimum consecutive volume for 3 days can drop to 4.8 m3/s, in the event of a dry five-year period, which is far from being severe, and rather normal compared to the average of the Pyrenean rivers in the basin. The floods can be very important, especially since the size of the catchment area is relatively high. The maximum instantaneous two-yearly flood flow and five-yearly are respectively 390 and 530 m3/s. The Nive basin is relatively calm, because the variation of its flow depends strongly on the rain, otherwise it is stationary without oscillations around its base flow, all the more so it is appreciated for light activities such as sport fishing for salmonids and the practice of sports.

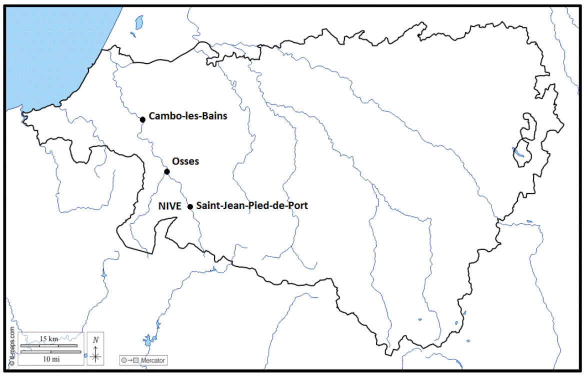


Figure 1. Adour water catchment with the main hydraulic stations on Nive River.

2.2. Artificial Neural Networks

For our problem, several types of models and architecture have been tested including hybrid models, however we will present only the results of the 5 models which gave results deemed acceptable for our problem, these models are MLP, LSTM, CNN, multi-headed CNN and Encoder-Decoder LSTM. The aforementioned models were trained and tested respecting the following distribution of data: 2014-2017 for learning and 2018-2020 for testing.

We have firstly showed that a time windows of 6 hours is sufficient for the two water discharges in Cambo and Osses. Concerning the rainfall, it was found that a time window of 3 hours in the past (observation) and 3 hours in the future (forecast) was optimal in particular for distant horizons like 6 hours. It seems moreover that multi-headed CNN has the most ability among all the models to capture distant horizons.

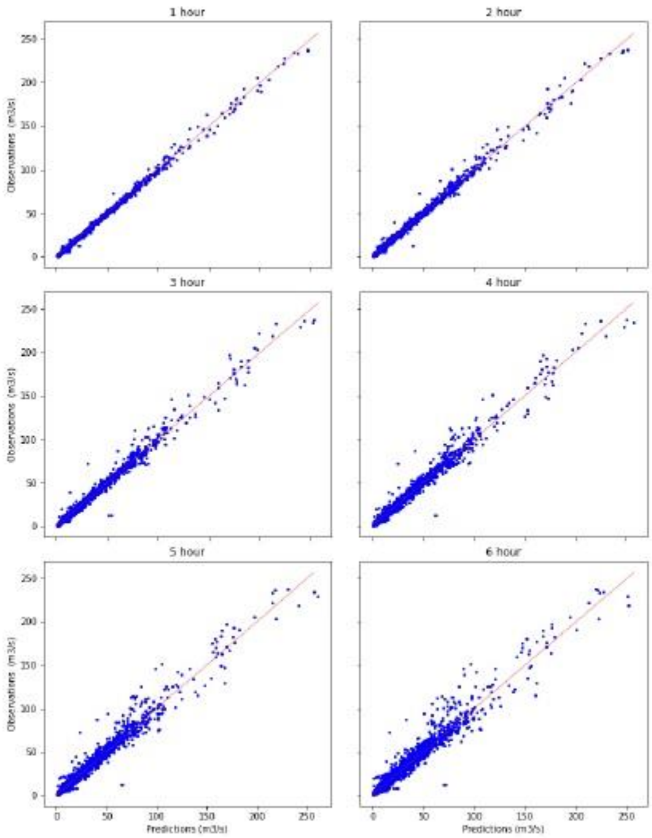


Figure 2. Analysis of results with multi-headed model in for Cambo water discharge with different horizons.

The correlation graph of figure 2 shows that there is a tightening around the first bisector for all horizons. The points that diverge the most from the first bisector are the points whose flow values are between 100 and 170 m^3/s, these values are present in the test set, particularly in 2018 but the training data would not contain not enough observations in this interval for a response of quality.

2.3. Sobol Indices

For a better understanding of the relationships between the inputs and outputs of the system and to confirm our hypotheses on the importance of variables such as the water discharge and the predicted rainfall on the different forecast horizons, a sensitivity analysis was carried out. Sensitivity analysis is the study of how uncertainty in the output of a code or system can be attributed to uncertainty in its inputs. This involves estimating sensitivity indices that quantify the influence of an input on the output. There are several methods for carrying out the sensitivity analysis, the analysis based on variance was retained, more precisely the Sobol index method. Its principle is to decompose the variance of the model output into several fractions that can be attributed to the inputs or input group. For example, given a model with two input variables and one output variable, it is possible that 70% of the output variance is caused by the variance of the first input variable, 20% by the second variable, and 10% by their interaction. This percentage is interpreted as the measure of sensitivity.

The Salib library (Python) was used for the analysis of variance. It is flexible and easy to use, it does not interact directly with the model but rather with its outputs. Salib generates inputs for the model through a random sampling function named sample and calculates sensitivity indices from model outputs relative to the generated inputs. Basically, the typical steps for performing a sensitivity analysis with Salib are: determining model inputs and their ranges of variability, generating model inputs, calculating outputs from inputs generated using the model, analyzing on the outputs obtained to calculate the different orders.

The notations are as follows:

T_0: real time

T_{-1h} : first exact hour before T_0

T_{+1h} : first exact hour before T_0

N: horizon time for forecast in hours

The three input variables are water discharge in Osses Q_{Osses}, water discharge in Cambo Q_{Cambo} and rainfall R. Since we deal with time series, we have finally 18 inputs that are sum-up in table 1 with the sobol indices for two horizon times.

The output variable is the observed water discharge in Cambo Q_{Cambo} at horizon time $Q_{Cambo}(T_{+Nh})$.

	Q_{Cambo}						Q_{Osses}						R					
Ti	-6h	-5h	-4h	-3h	-2h	-1h	-6h	-5h	-4h	-3h	-2h	-1h	-3h	-2h	-1h	+1h	+2h	+3h
N=2	1	2	3	6	24	51	0	0	5	9	7	10	0	1	2	2	0	0
N=6	1	3	3	6	6	4	5	26	22	20	32	30	3	3	4	2	2	1

Table 1. Sobol indices in percentage for two horizon times.

We notice in table 1 that Q_{Cambo} is the dominant index for the close horizon time and that Q_{Osses} is the dominant index for the distant horizon. Rainfall is never dominant since Q_{Osses} always contains indirectly information on the rainfall.

3. Case-based reasoning

3.1. Methodology

The methodology is based on distances that are built from the Sobol indices that are noted S_{iN} with I is varying from 1 to 18 and N takes value 2 or 6. The 18 times Ti are associated with sobol indices like in Table 1.
We note ΔX_i the difference between the observed variable and variable of the scenario at time Ti.
We note X_i the value of the variable of the scenario at time Ti.
For i between 1 and 6, the variable X_i is Q_{Cambo}, for i between 7 and 9, the variable X_i is Q_{Osses} and for i between 10 and 12, the variable X_i is rainfall *R.*
The distance D is calculated as follows:

$$D_N = \frac{\sum_{i=1}^{18} Si\mathrm{N}\, \Delta X_i}{\sum_{i=1}^{18} Si\mathrm{N}\, X_i} \quad (1)$$

3.2. Data base

The data base contains 6 years of data (discharges and rainfall) from 2014 to 2019. Removing periods without of low waters, the data base contains finally around 9000 hours that constitute 9000 scenarios.

3.3. Validation

The methodology is tested with the forecast of the water discharge in Cambo for three events that are described below and with increasing water discharge.

- Event n°1 lasts from 21st to 24th of January 2019 with a forecasted water discharge in Cambo of 18 m^3/s ;
- Event n°2 lasts from 21st to 24th of January 2014 with a forecasted water discharge in Cambo of 37 m^3/s ;
- Event n°3 lasts from 3rd to 7th of March 2014 with a forecasted water discharge in Cambo of 102 m^3/s.

The rise of the water discharge is presented in figures 3, 4 and 5 for the three events.
For the validation, we take the whole data base except the hour that gives a zero distance. For the first analysis we keep the 10 best scenarios that gives the smallest distances. The obtained distances for the two chosen horizon times (2 hours and 6 hours) are presented in figures 6, 7 and 8 for the three events. We find firstly that the scattering increases with the severity of the flood. That can be due to the smallest number of data that are in the database for large discharges. The accuracy is much better with the closest horizon time than for the most distant one. We also find a good correlation between the calculated distance and the accuracy. We can then expect that a distance below 0.04 can provide a good forecast.

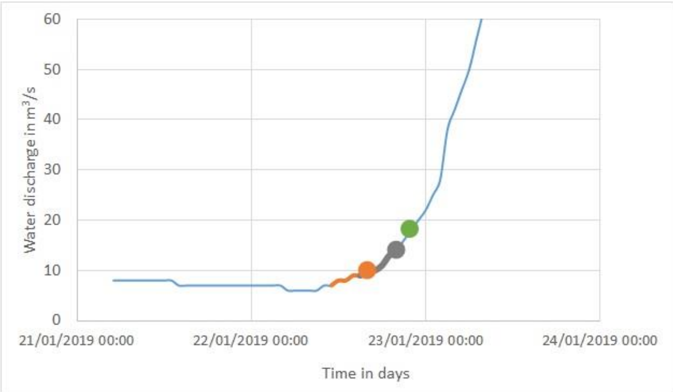


Figure 3. Water discharge evolution for event n°1.

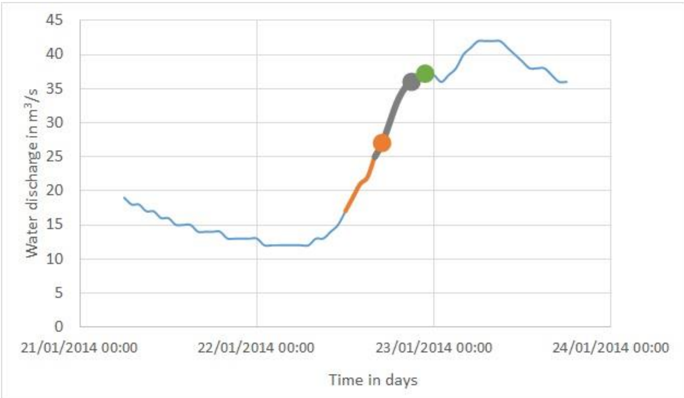


Figure 4. Water discharge evolution for event n°2.

Figure 5. Water discharge evolution for event n°3.

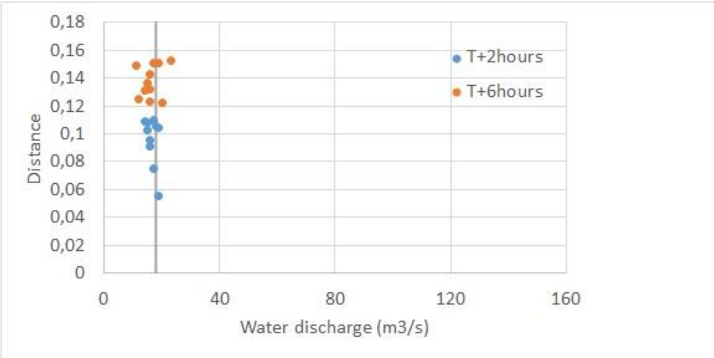


Figure 6. Distances for the 10 best scenarios for event n°1.

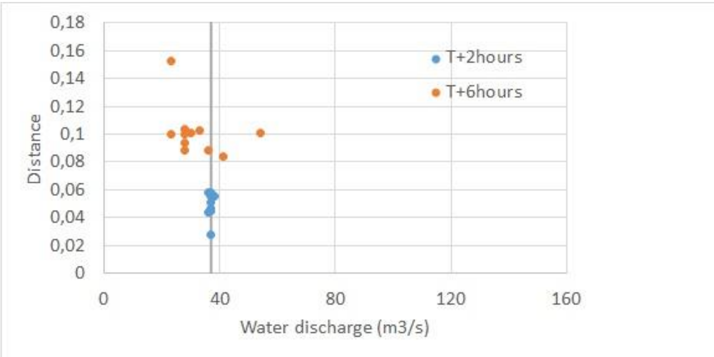


Figure 7. Distances for the 10 best scenarios for event n°2.

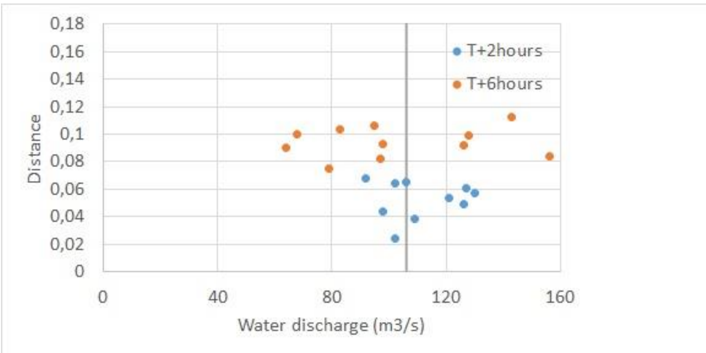


Figure 8. Distances for the 10 best scenarios for event n°3.

In figures 9, 10 and 11, the 9000 scenarios are presented in terms of clouds of distances for the three events and the two time horizons. We do not find important differences between the two time horizons. We find however a different behavior for the strongest event n°3. When the two first events present a symmetry between the inferior and superior discharges of the target discharge, the third event present a clear dissymmetry. There are more uncertainties towards the large discharges. That shows the interest of this method that can gives simultaneously a forecast with information of uncertainty.

We should recall finally that part of rainfall in the distance function is small compared to the water discharges in Cambo and Osses. That is due to the fact that information on the rainfall is already included in the values of water discharges.

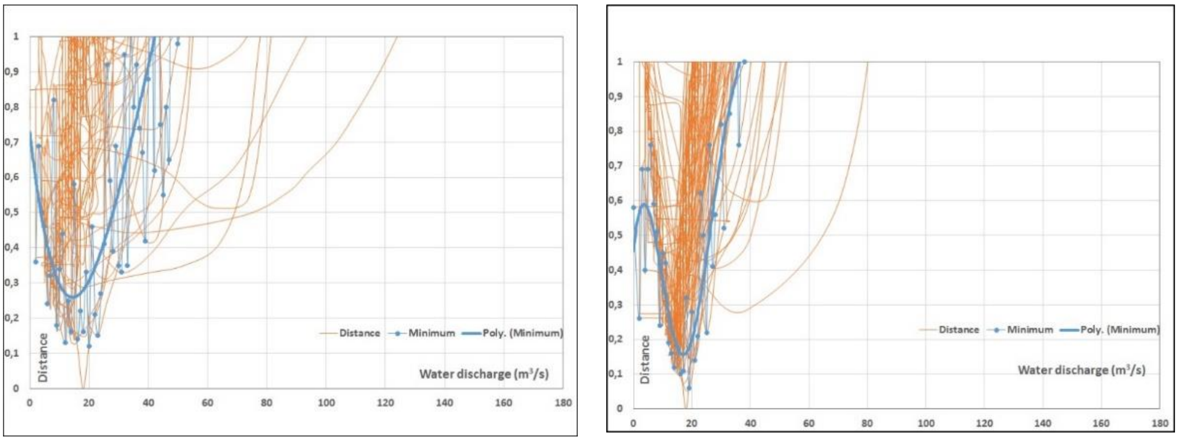


Figure 9. Q=18 m³/s (a - left) forecast T+6hours (b-right) forecast T+2hours

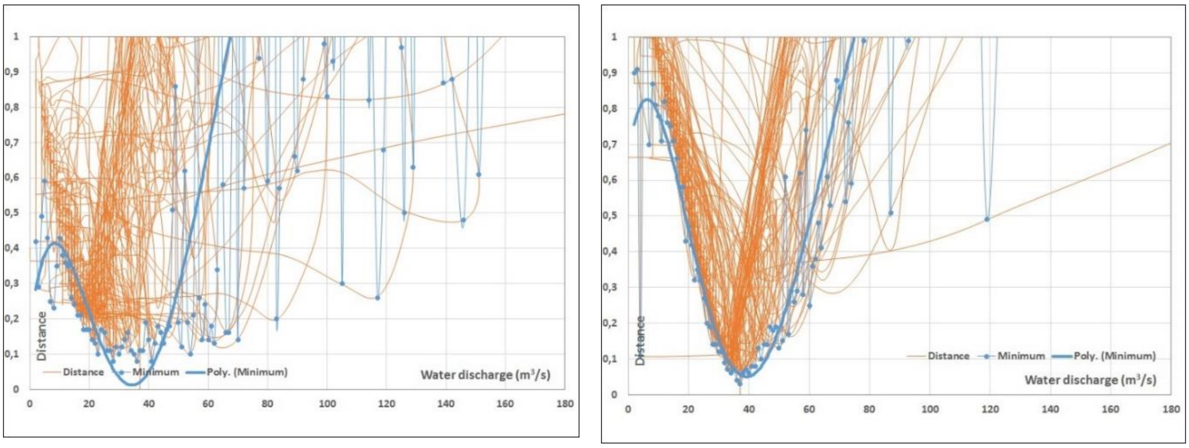


Figure 10. Q=37 m³/s (a - left) forecast T+6hours (b-right) forecast T+2hours

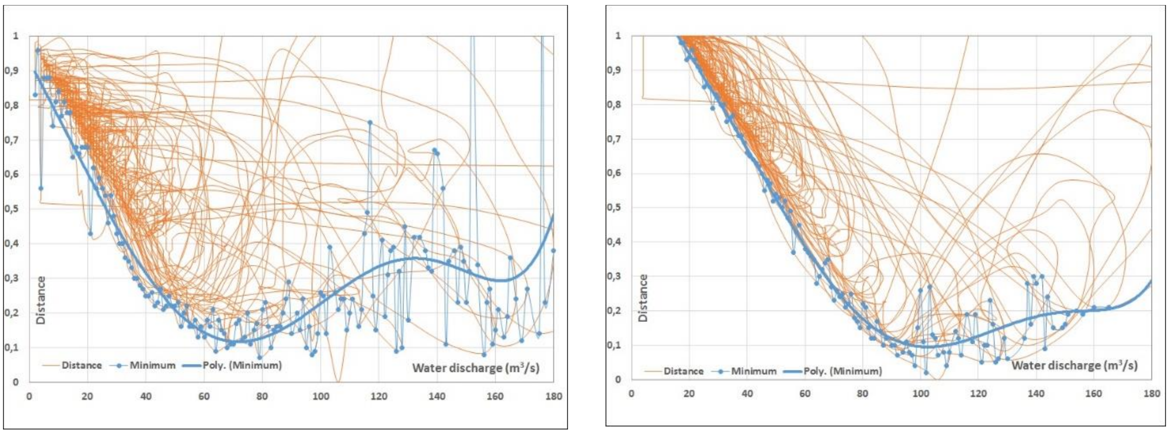


Figure 11. Q=102 m³/s (a - left) forecast T+6hours (b-right) forecast T+2hours

3.4. Consequences for the CBR method

We can conclude that the best scenario can be selected as the scenario with the smallest distance function if this distance is smaller than 0.04. If this distance is larger than 0.04 we recommend to average the water heights in the flood maps with the ten best scenarios.

4. Hydraulic modeling

4.1. Shallow water equations

Under the hypothesis of a hydrostatic pressure, shallow water equations are expressed as follows in tensor form :

$$\frac{\partial h}{\partial t} + \frac{\partial hU_j}{\partial x_j} = 0$$
$$\frac{\partial hU_i}{\partial t} + \frac{\partial hU_iU_j}{\partial x_j} = -g\frac{\partial hU_j}{\partial x_i} \qquad (2)$$

4.2. Comparison of models

Four different numerical models are tested on the Nive catchment area.

A model based on the diffusive wave was first developed and tested on different test cases. Despite its speed, the first tests showed inconclusive results. It was therefore abandoned. Three other models based on shallow water equations were then tested.

Hydraulic modeling by the Lattice Boltzman Method (LBM) was then used. The initial idea of the development was to develop a digital simulation model to draw flood maps quickly and with good precision. The LBM method was considered by its parallel side and promising results available in the scientific literature. In a first part, the numerical model was developed and pre-validations were carried out. In a second part, the case of the Nive catchment area was simulated with the bathymetry that is presented in figure 2 on a rectangular grid. We finally compared the calculation data to the station data.

In order to validate LBM modeling and compare it to another recognized numerical tool, the choice of the Nays2DFlood program, developed by a Japanese university was proposed. The Iric (International River Interface Cooperative) model offers a number of open access models, including Nays2DFlood which is a flood simulation code based on the complete Saint-Venant equation. The software has a user interface and takes as input a bathymetric model in Cartesian mesh. This model was initiated in 2007 and today has a community of users who can communicate in a forum. This code is developed on CPU and works in windows environment. It is able to use several CPU cores if the computing machine has them. For more details, see https://i-ric.org/en/solvers/nays2dflood/. The bathymetry is presented in figure 2 on a rectangular grid.

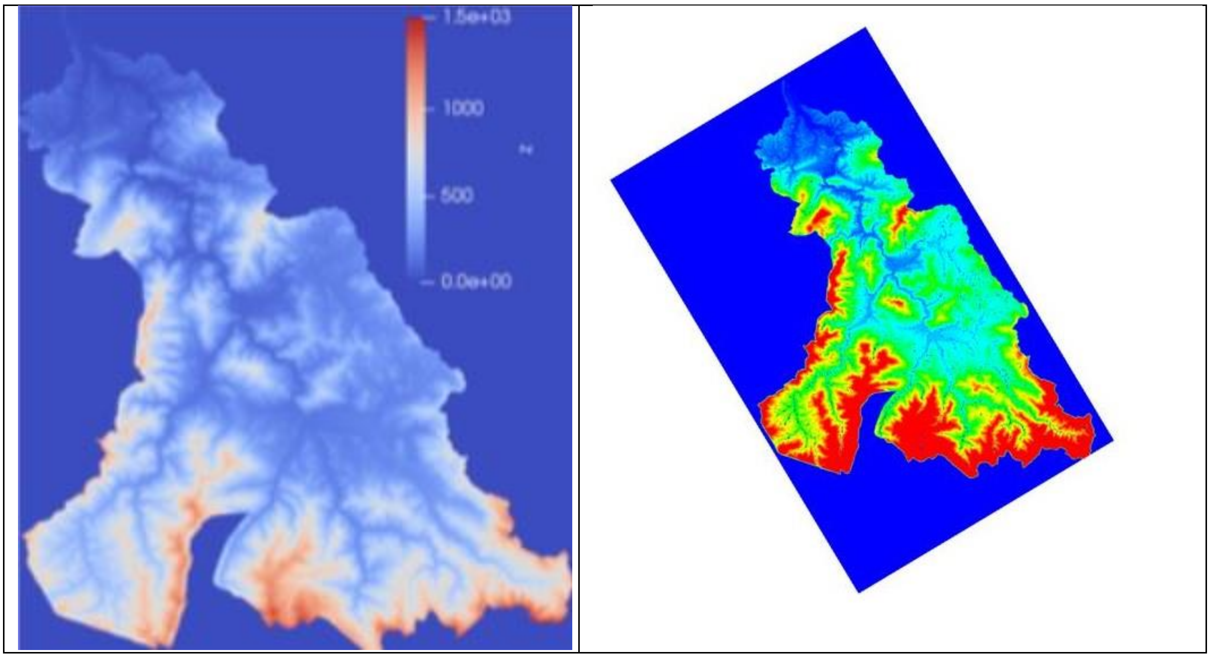


Figure 12. Bathymetry of Nive water catchment used with the LBM model and Iric model.

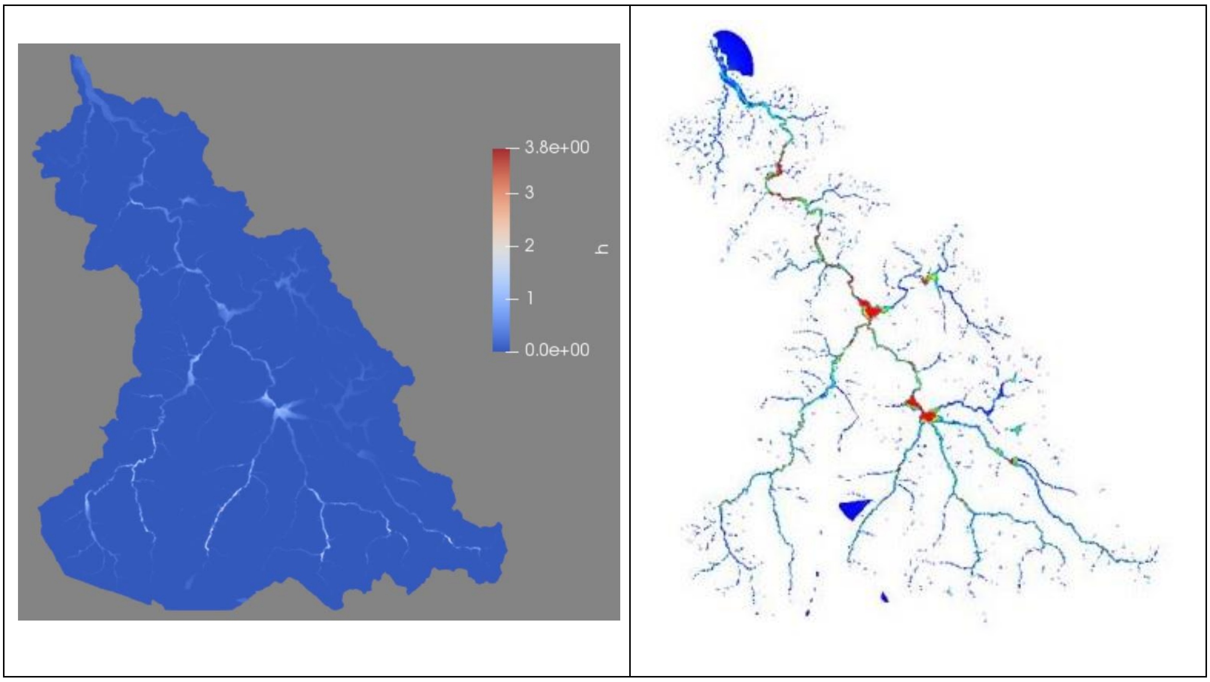


Figure 13. Example of flood map with LBM model and with Iric model.

A Telemac SCS-CN model based on a digital terrain model (DTM) provided by IGN with 5 meter resolution has finally been developed. Part of the Nive watershed is in Spain and is not covered by these data, this missing part was built from higher resolution and lower quality data and integrated into the model. The quality of the DTM is very variable because it is built from various data sources. As a result, it was necessary to make the DTM hydrologically valid by filling the small basins using the Darboux/Planchon method, which has the effect of forcing the flows and locally increasing the altitudes at the points. To function correctly and taking into account the relief of the basin, the finite element model had to be refined at the level of the minor beds. The axes of the watercourses were extracted from the DTM using a D8-type method followed by a classic accumulation calculation. The mesh was constrained on the discretization points of the minor beds every 5 meters. The boundary conditions imposed on the free exit type model did not give conclusive results and forced us to build a ditch on the periphery of the basin that could collect the volume of water falling on the area during the simulation. The mesh is made up of 1,650,000 elements. A simulation of 1000 hours with a time step of 1s was carried out in 38 hours (user time) on a machine of 48 processors. Up to 6 simulations ran at the same time. The maps are produced through images in a vector format that enables to get easily different zooms of the flood maps. An example of zooms is give in figure 13.

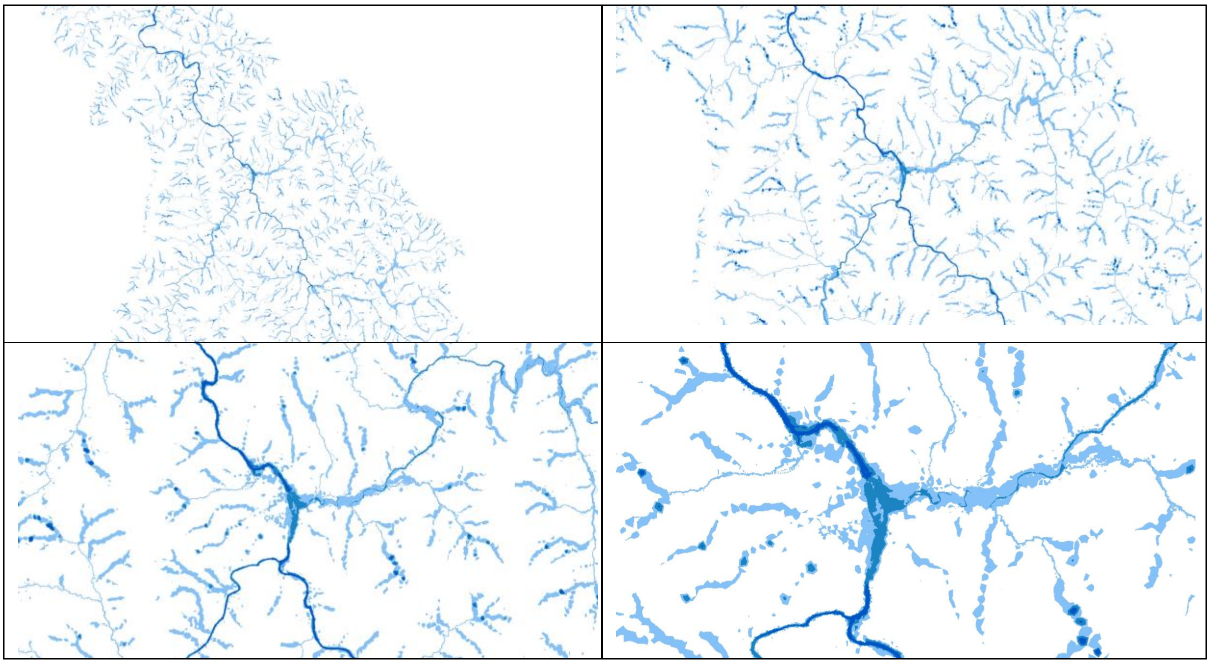


Figure 13. Flood maps with different zooms of an image in a vector format.

The last three models were compared with measurements at the Osses station. The conclusions of the study led to favoring the use of Telemac. This finite element model makes it possible to greatly refine at the level of minor beds what is essential in the valleys of mountainous areas. This model made it possible to obtain a good compromise between precision and computation time thanks to the parallelization of the code. It was therefore implemented over a period of 6 years from January 2014 to December 2019 in order to provide around 9000 flood maps.

Figure 14 shows the comparison of the Iric model and Telemac model in Osses the 4th of July 2014. Only these two model could provide coherent results. Iric model and observations fit very well. The accuracy of Telemac model was slightly less good but computation time of Iric model was prohibitive whereas Telemac model could benefit from the advantage of parallel computations.

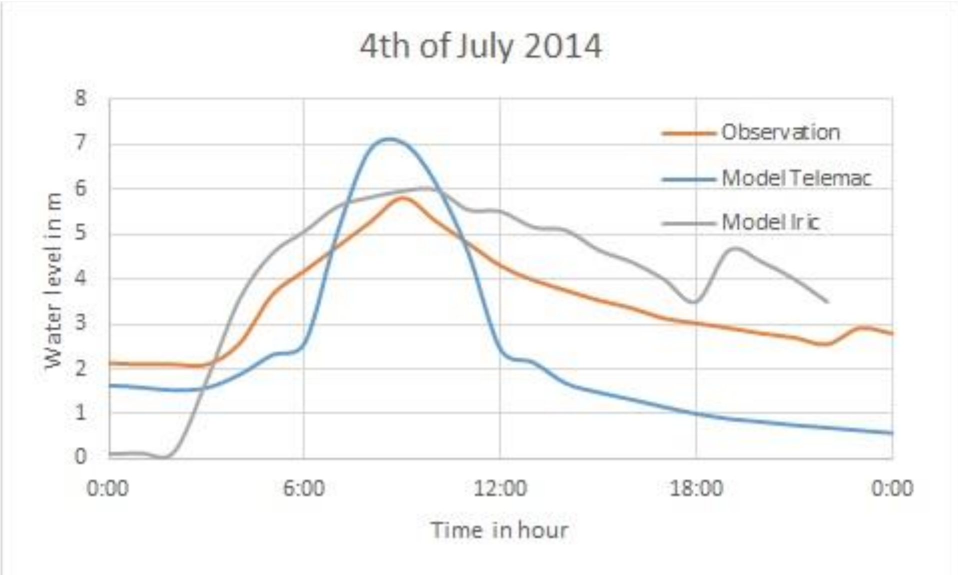


Figure 14. Water level in Osses during the flood of 4th of July 2014.

5. Conclusions

In this work, we have shown the possibility to develop a complete computation chain that enables to automatically provide flood maps for flood forecast. This work should be completed in the future by a deeper study of the numerical models and in the comparison of CBR method with other methods of artificial intelligence. An attention should finally be paid to the validation of the whole procedure.