IRSN
INSTITUT DE RADIOPROTECTION ET DE SÛRETÉ NUCLÉAIRE

Faire avancer la sûreté nucléaire

Uncertainty quantification and global sensitivity analysis with dependent inputs:
Application to the 2D hydraulic model of the Loire River

Lucie Pheulpin[1] & Vito Bacchi[2]

[1] IRSN, Fontenay-aux-Roses, France
lucie.pheulpin@irsn.fr

[2] EDF R&D, Chatou, France
vito.bacchi@edf.fr

EGU General Assembly 2020

Online | 4–8 May 2020

Session NH1.1

The NARSIS project

- Contribution to the **NARSIS** (New Approach to Reactor Safety ImprovmentS) European project initiated in 2017

- Objectives of the NARSIS project:
  - Bring contributions to the safety assessment methodologies
  - Improve the Probabilistic Safety Assessment (PSA)

- Our objectives:
  - Propose a methodology to evaluate uncertainties in 2D hydraulic models by taking into account the dependencies between inputs
  - Apply this methodology to a 2D operational model

Context of the study

- External hazards (*i.e.* flooding) assessed through numerical modelling

- Numerous uncertainties in the hydraulic models related to:
  - the chosen numerical model (Telemac-2D, HEC-RAS, *etc.*)
  - the lack of knowledge of the physical parameters describing the system
  - the model numerical parameters:
    - river geometry, roughness coefficients
    - levee physical characteristics and levee breach parameters
    - flood hydrograph, *etc.*

- Use of Uncertainty Quantification (UQ) and Global Sensitivity Analysis (GSA) to better understand these uncertainties

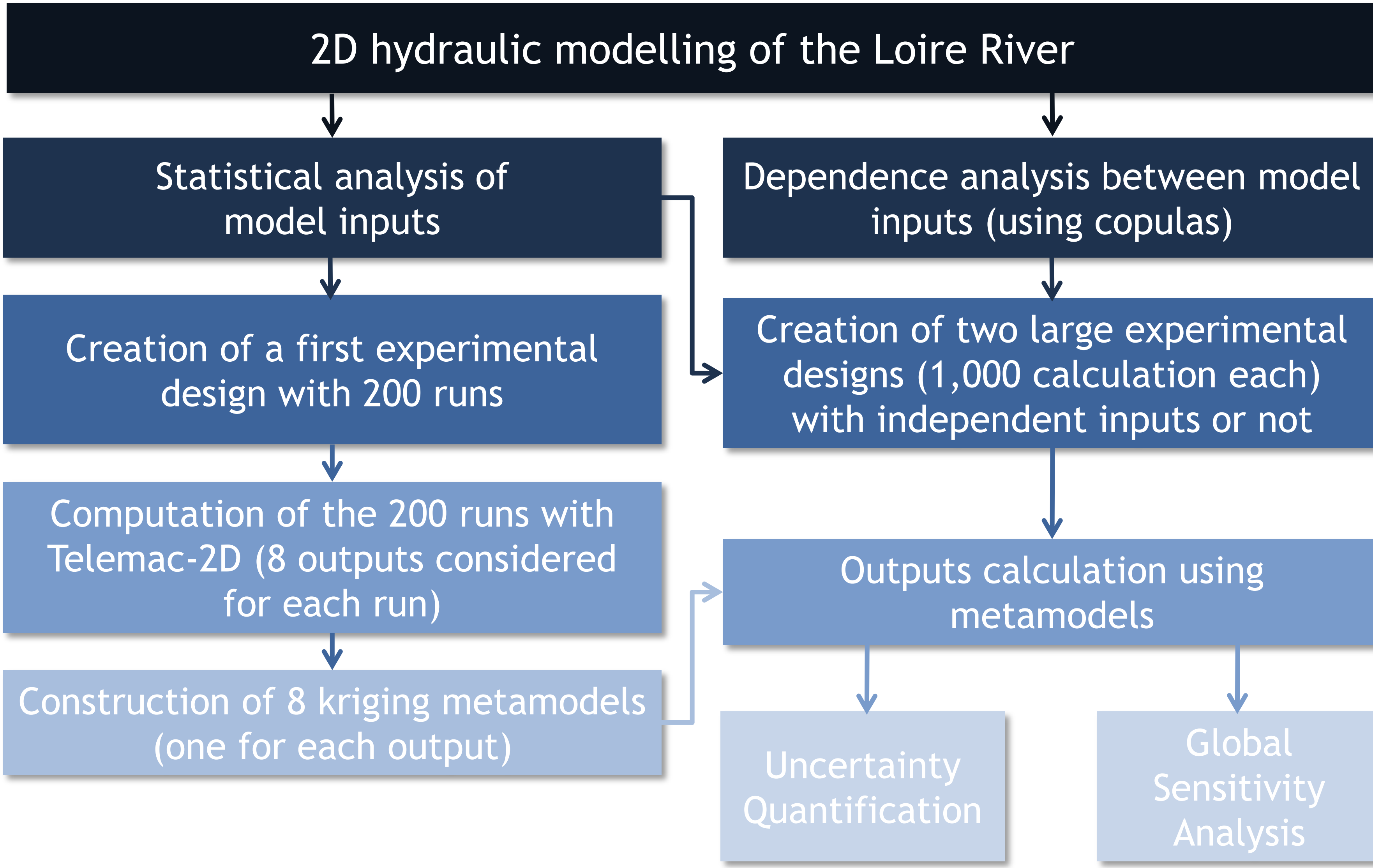
- Consideration of the dependence between model inputs (usually, inputs are considered to be independent in uncertainty quantification studies)

Case study: why the Loire River?

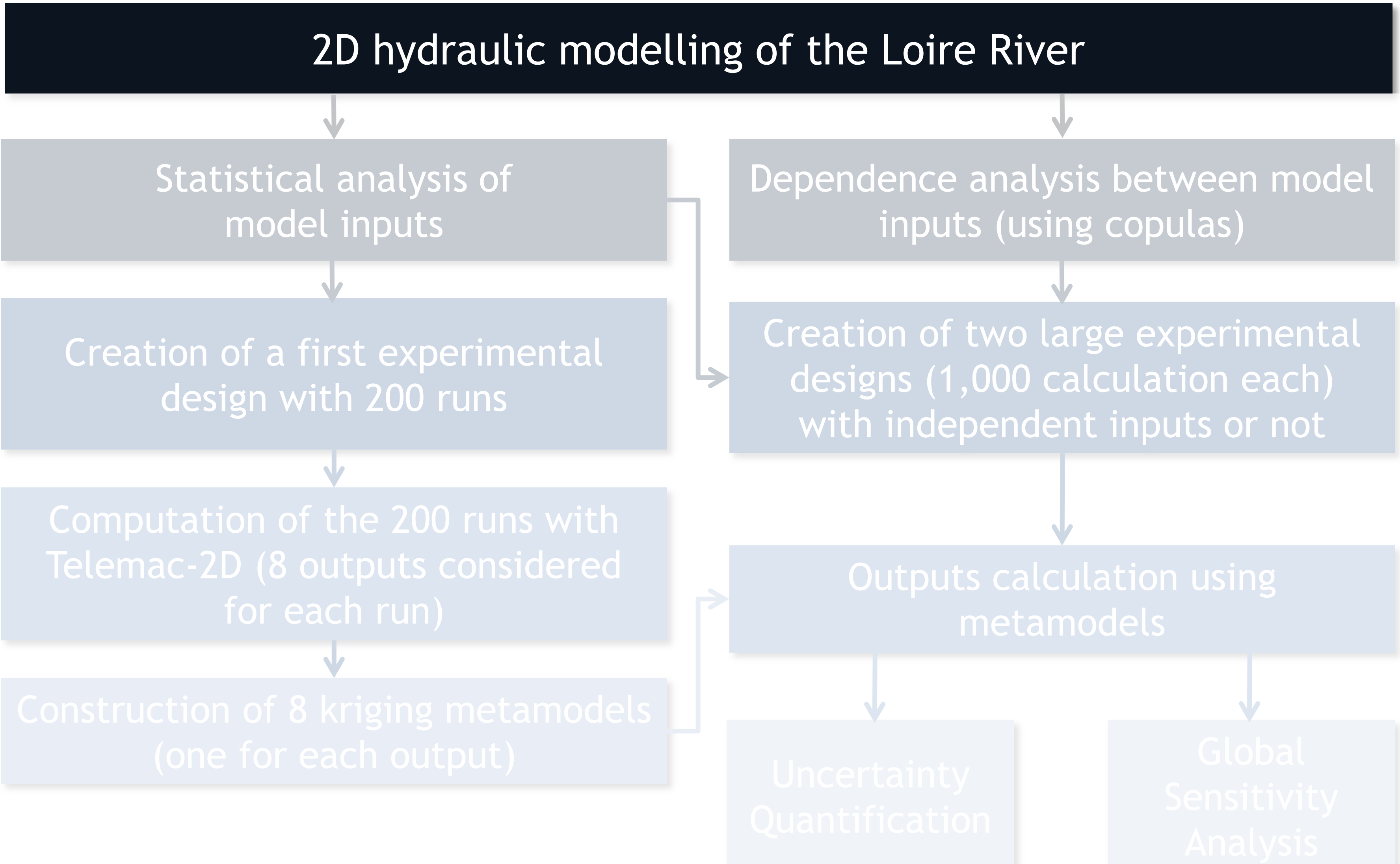
- Several historical major floods identified (1846, 1856, 1866, 1917)
- Historical sites, industrial facilities and large cities along the Loire River → **Risk of human and material damages**
- Numerous open data available

Floods in 2016 in the Centre-Val de Loire Region © France3

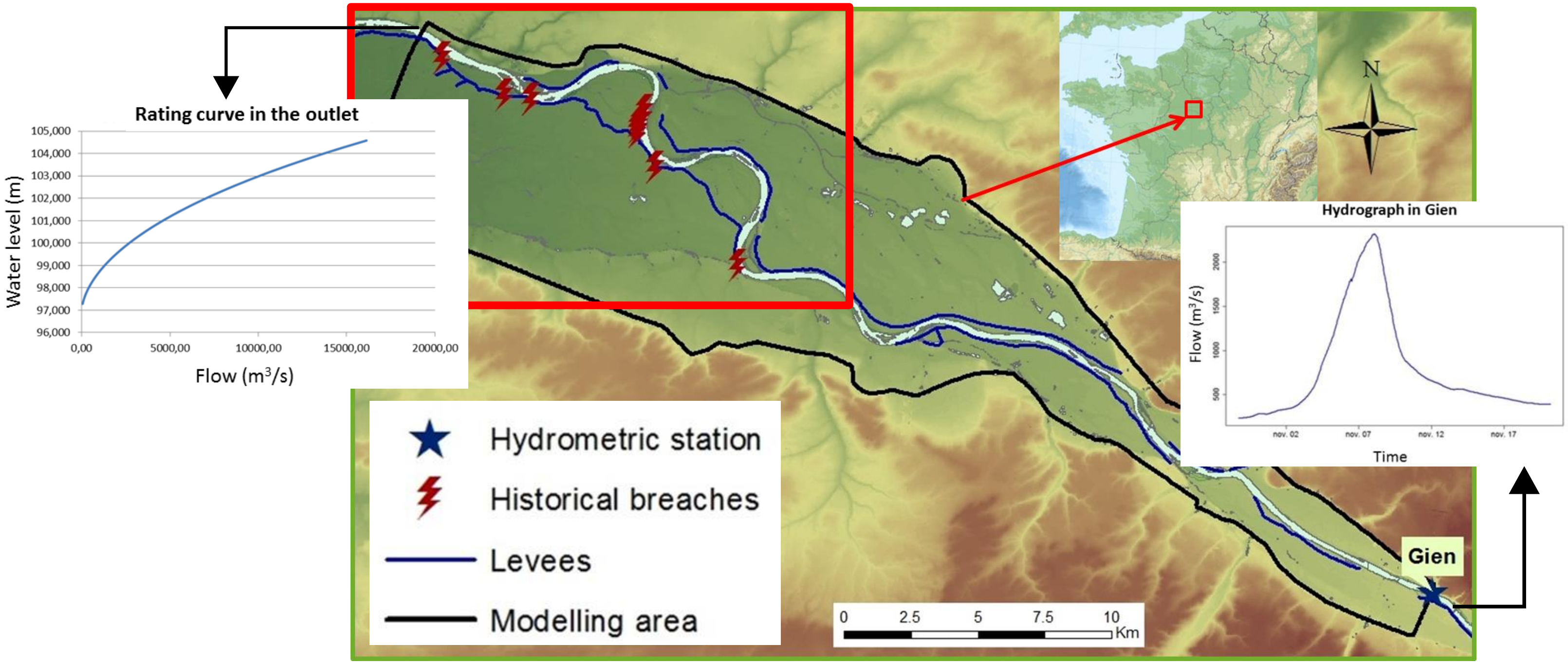
Methodology



Methodology

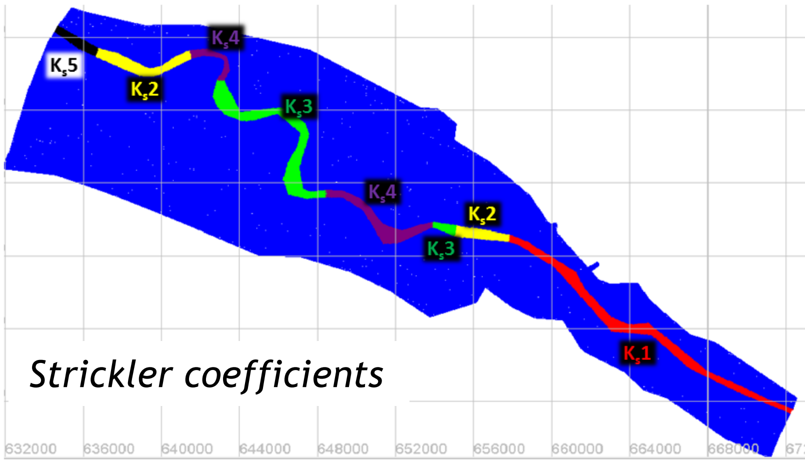


2D hydraulic modelling of the Loire River



- **50 km-long reach** modelling between Gien and Orléans
- 2D modelling with **Telemac-2D**
- **262,800 meshes**
- Computation time: **1h30** in average, depending on the flood duration
- Limit conditions: **hydrograph** in Gien and **rating curve** in the outlet
- Focus on the lower part of the model (red square)

2D hydraulic modelling of the Loire River
Uncertain inputs and outputs investigated

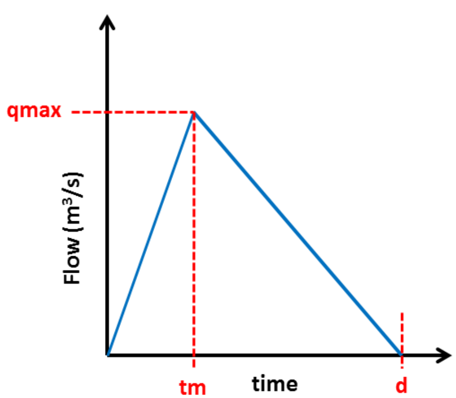


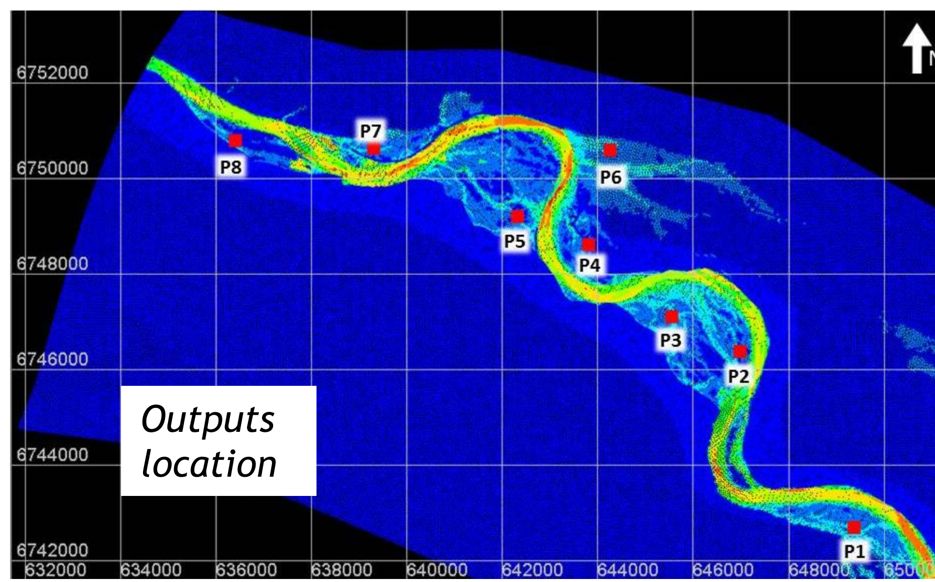
Strickler coefficients

8 Inputs:

- 5 different Strickler coefficients **(K_s1 to K_s5)**
- 3 inputs linked to the hydrograph:
  - maximum flow (**qmax**)
  - total duration of the flood (**d**)
  - rise time (**tm**)

Inputs linked to the triangular hydrograph



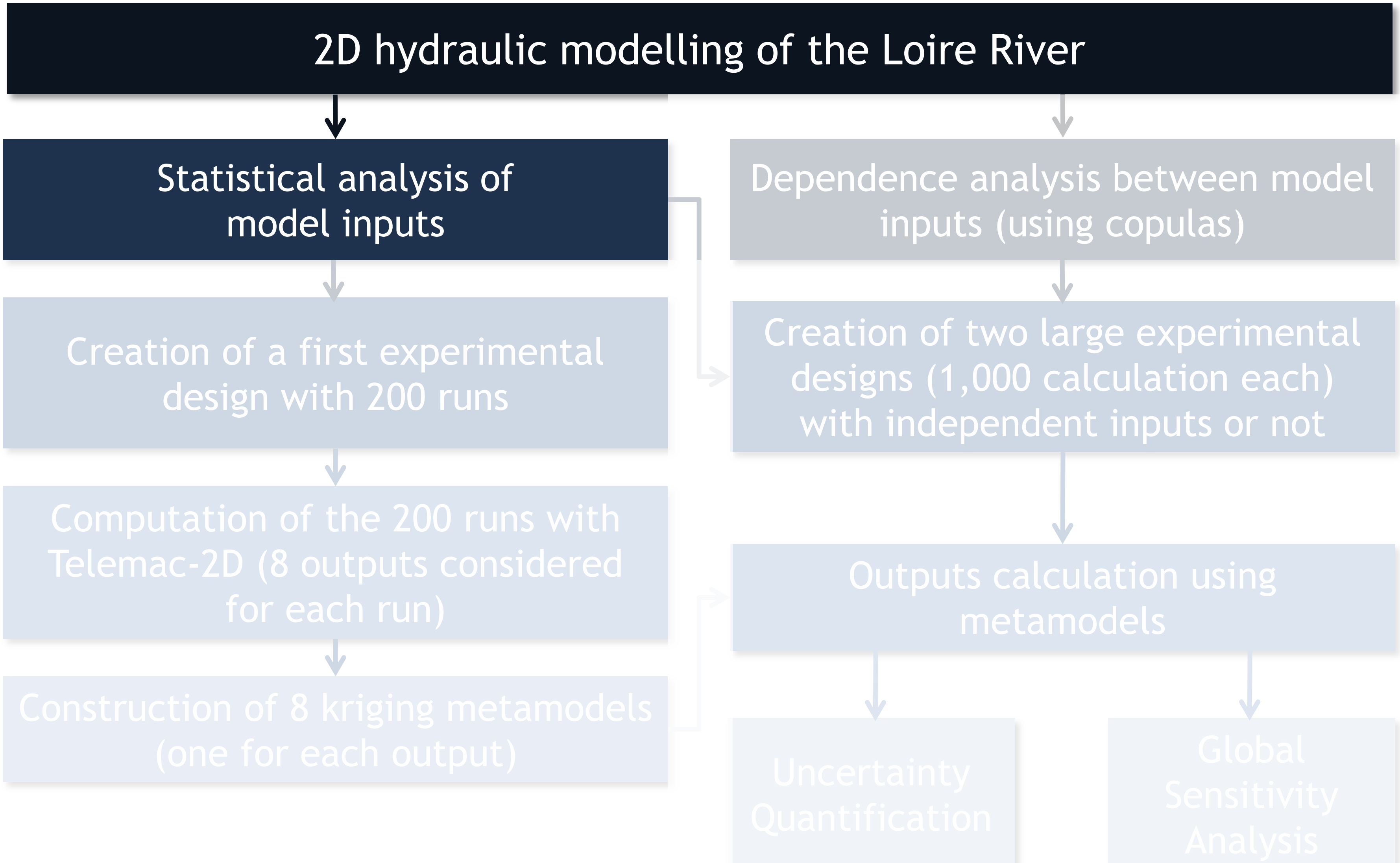


Outputs location

8 outputs (P1 to P8)

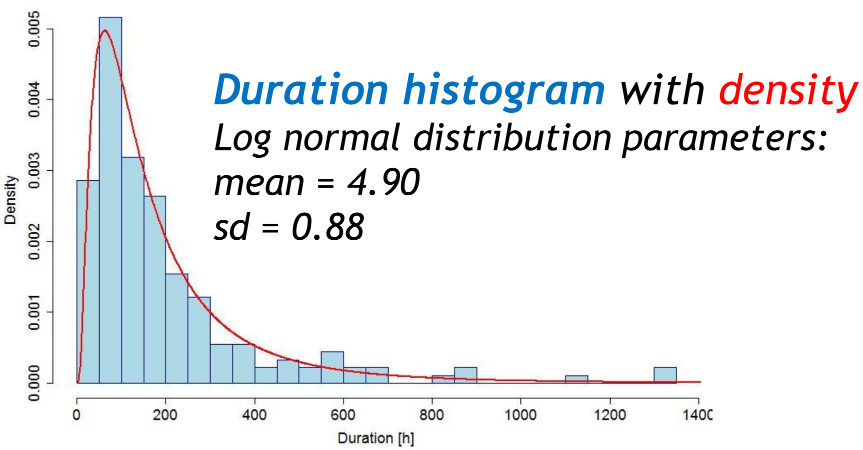
- Extraction of the maximum water level

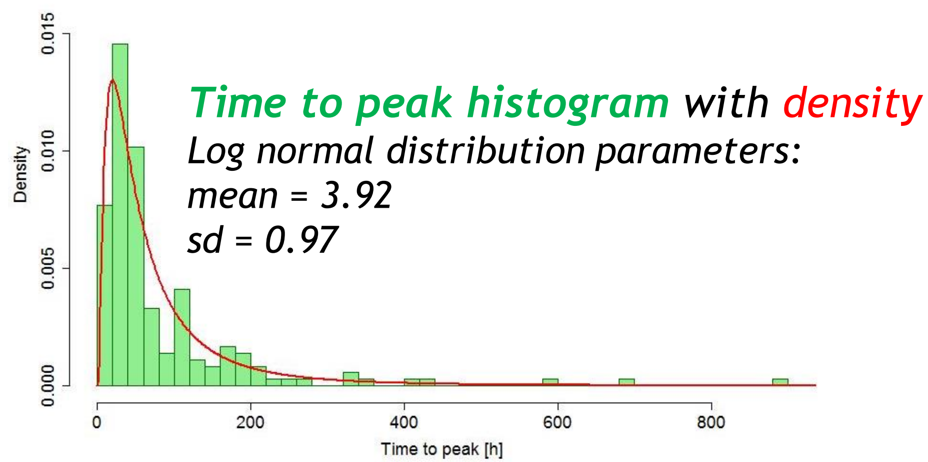
Methodology



Statistical analysis of model inputs (hydrograph parameters)

- Objective: define the probability distribution of each input
- For the duration (d) and the time to peak (tm):
  - Extraction of the major floods between 1953 and 2019 (flood considered when flow > 600 m3/s, duration > 24h and time between two floods > 24h)
  - 182 floods selected
  - For each flood, extraction of the total duration, time to peak and maximum flow
  - Research of the most accurate probability distributions for d and tm → **Log normal distributions**

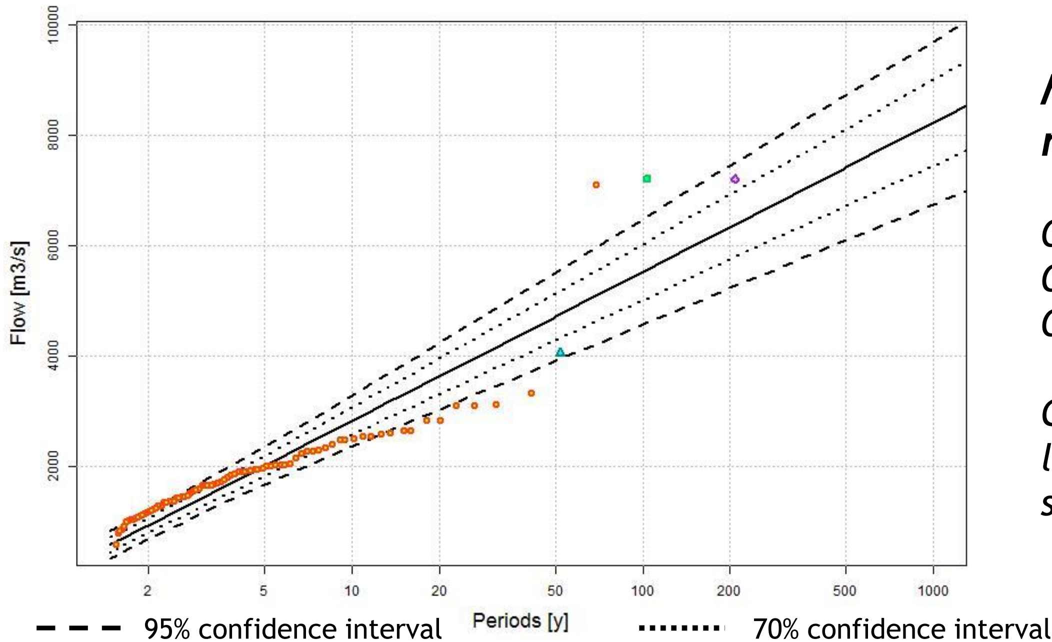




Statistical analysis of model inputs (hydrograph parameters)

For the maximum flow → **extreme value analysis**

- From the maximum annual discharges since 1936 + 4 historical floods (1846, 1856, 1866, 1917)
- Adjustment of the maximum annual discharges with a **Gumbel distribution function** (R-package *Renext*)

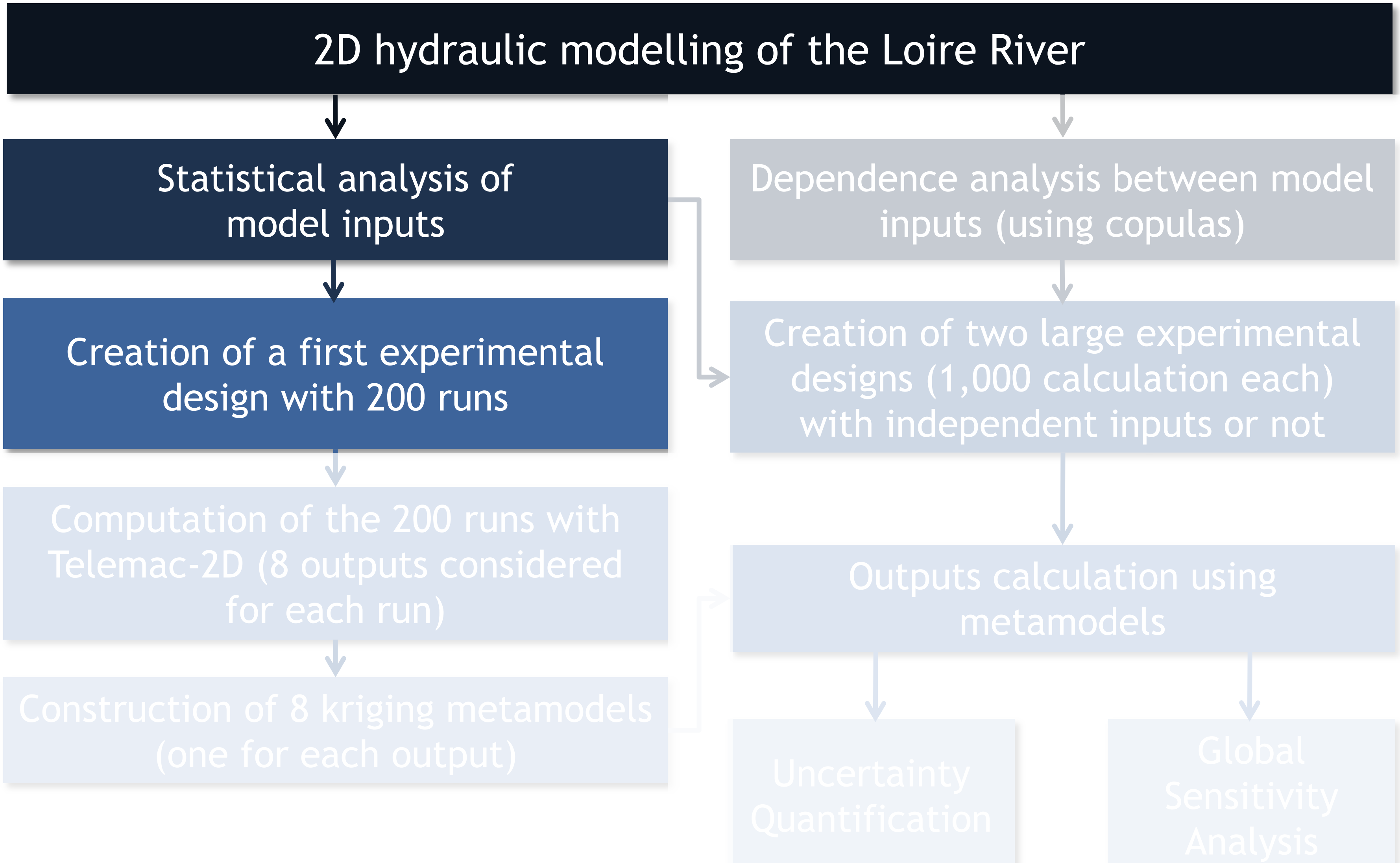


Maximum annual discharges return levels

Q10 = 2,637 m^3/s
Q100 = 5,301 m^3/s
Q1000 = 7,916 m^3/s

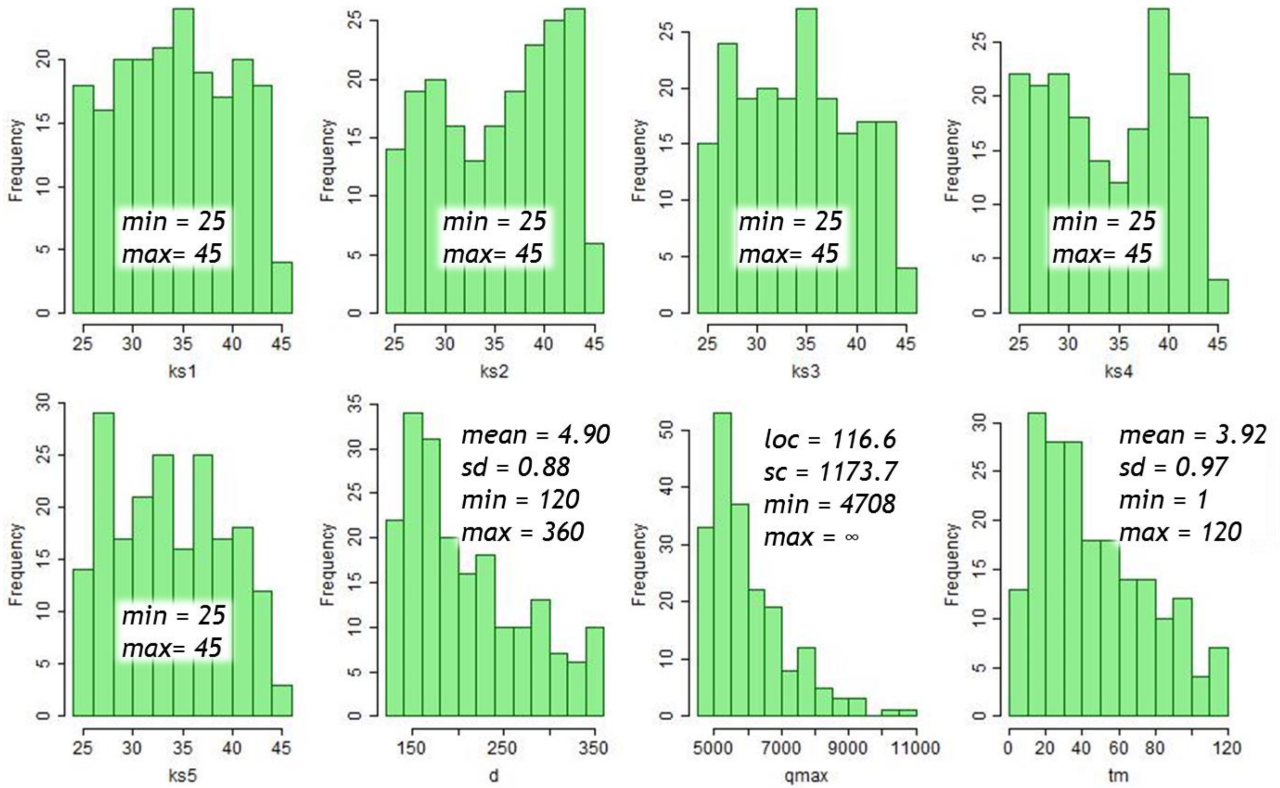
Gumbel distribution parameters:
location = 116.65
scale = 1173.74

Methodology



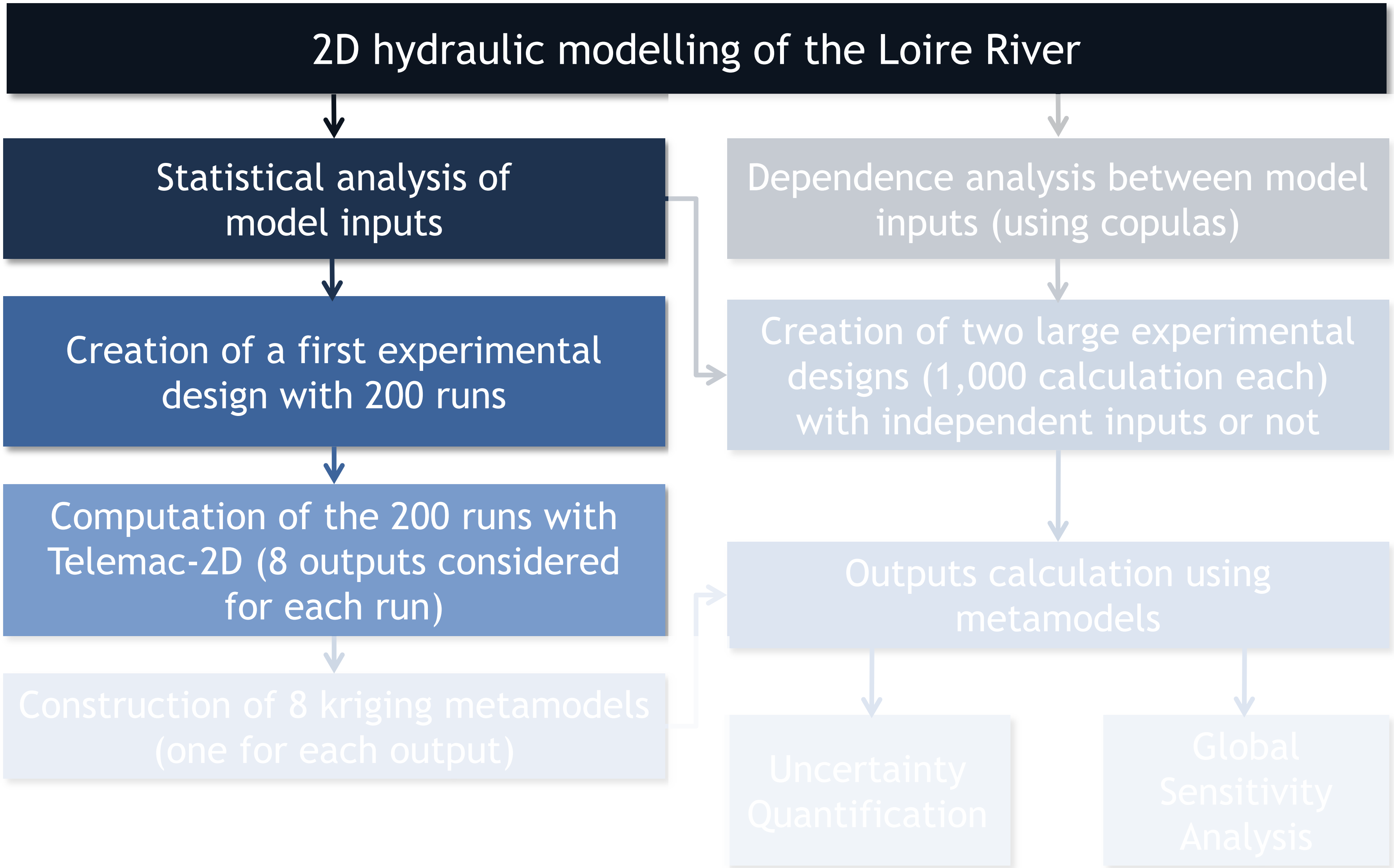
Creation of an experimental design

- Objective: create an input parameter table of **200 runs** with Telemac-2D
  - Strickler parameters (K_s1 to K_s5) sampled inside uniform distributions
  - Hydrograph parameters sampled inside the distributions previously defined (here truncated distributions are used):



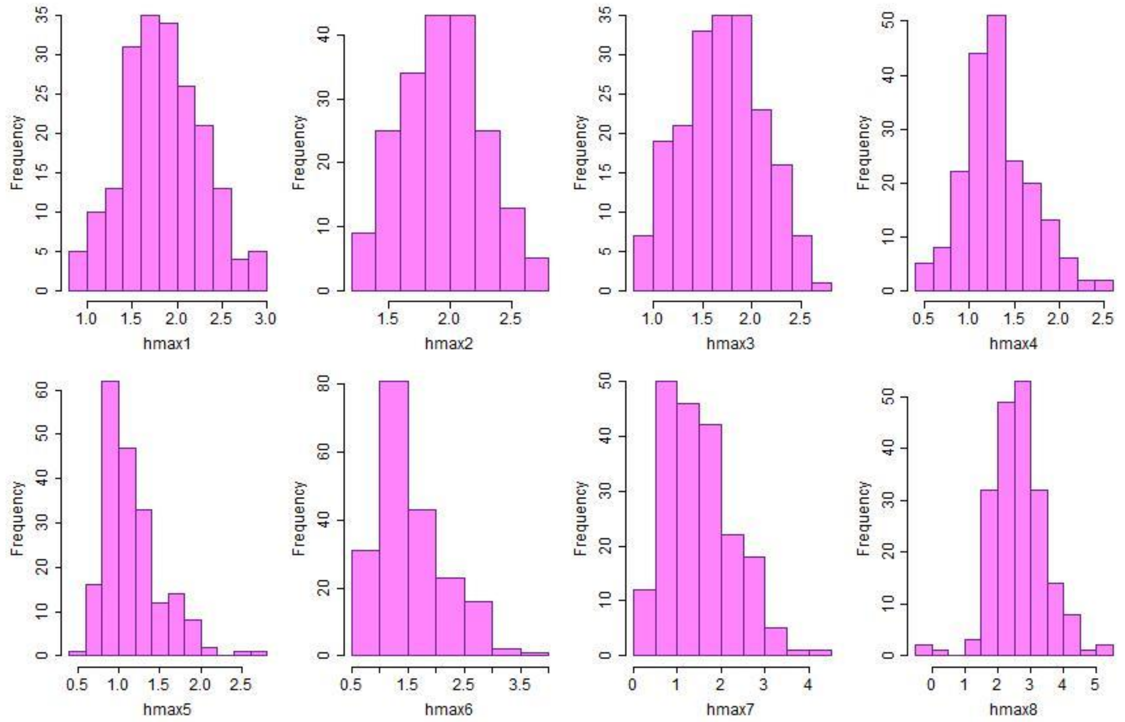
Disribution histograms of sampled inputs & value of parameters for the associated probability distributions

Methodology



200 runs with Telemac 2D

- Use of the parametric computing environment developed in IRSN: Funz (https://github.com/Funz)
- Coupling between Funz and Telemac-2D to run the 200 calculations successively
- Computation time: between 36 min and 2h30 for one run (mean = 1h20)
- In total = 260 hours (~11 days) with 38 parallel processors)



Distribution histograms of outputs

Methodology

2D hydraulic modelling of the Loire River

Statistical analysis of model inputs

Creation of a first experimental design with 200 runs

Computation of the 200 runs with Telemac-2D (8 outputs considered for each run)

Construction of 8 kriging metamodels (one for each output)

Dependence analysis between model inputs (using copulas)

Creation of two large experimental designs (1,000 calculation each) with independent inputs or not

Outputs calculation using metamodels

Uncertainty Quantification

Global Sensitivity Analysis

Construction of kriging metamodels
Generalities

- What is it?
  - Mathematical tool used to replace the original model with a function
  - Function constructed using statistical criteria (*e.g.* maximum likelihood) in order to fit the "experimental" computation of the original model

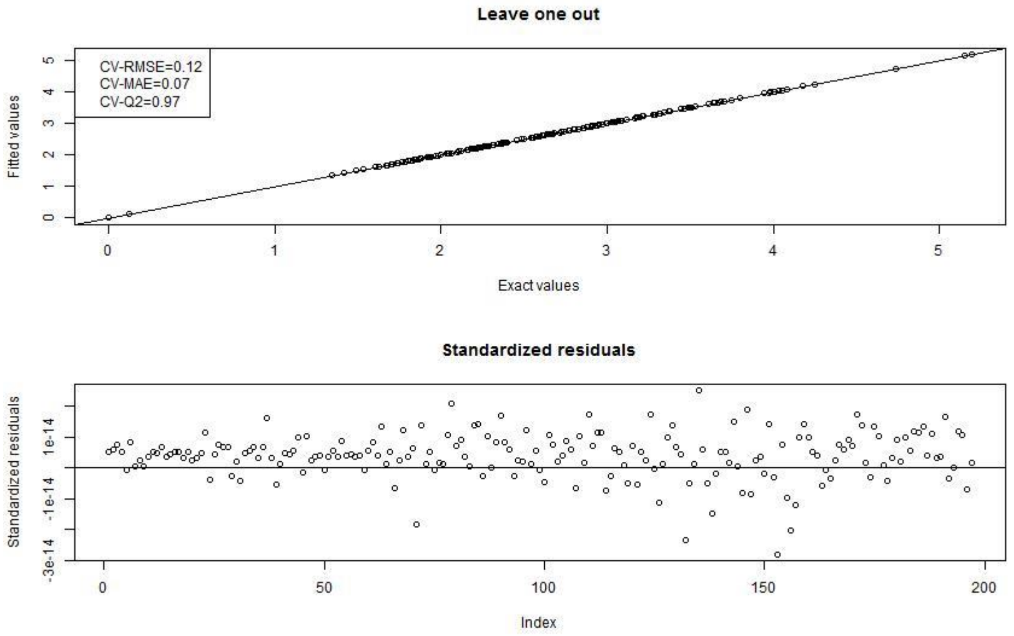
- Objective of the metamodel: reduce the computation time

- Three main steps to construct a metamodel:
  1. **Design**: creation of an "experimental" dataset used as learning basis for the metamodel
  2. **Construction**: it depends on the chosen function (*e.g.* kriging, random forest)
  3. **Validation** of the metamodel through statistical tests (*e.g.* leave-one-out & K-fold cross validation)

Construction of the kriging metamodels
Metamodels of the Loire River model

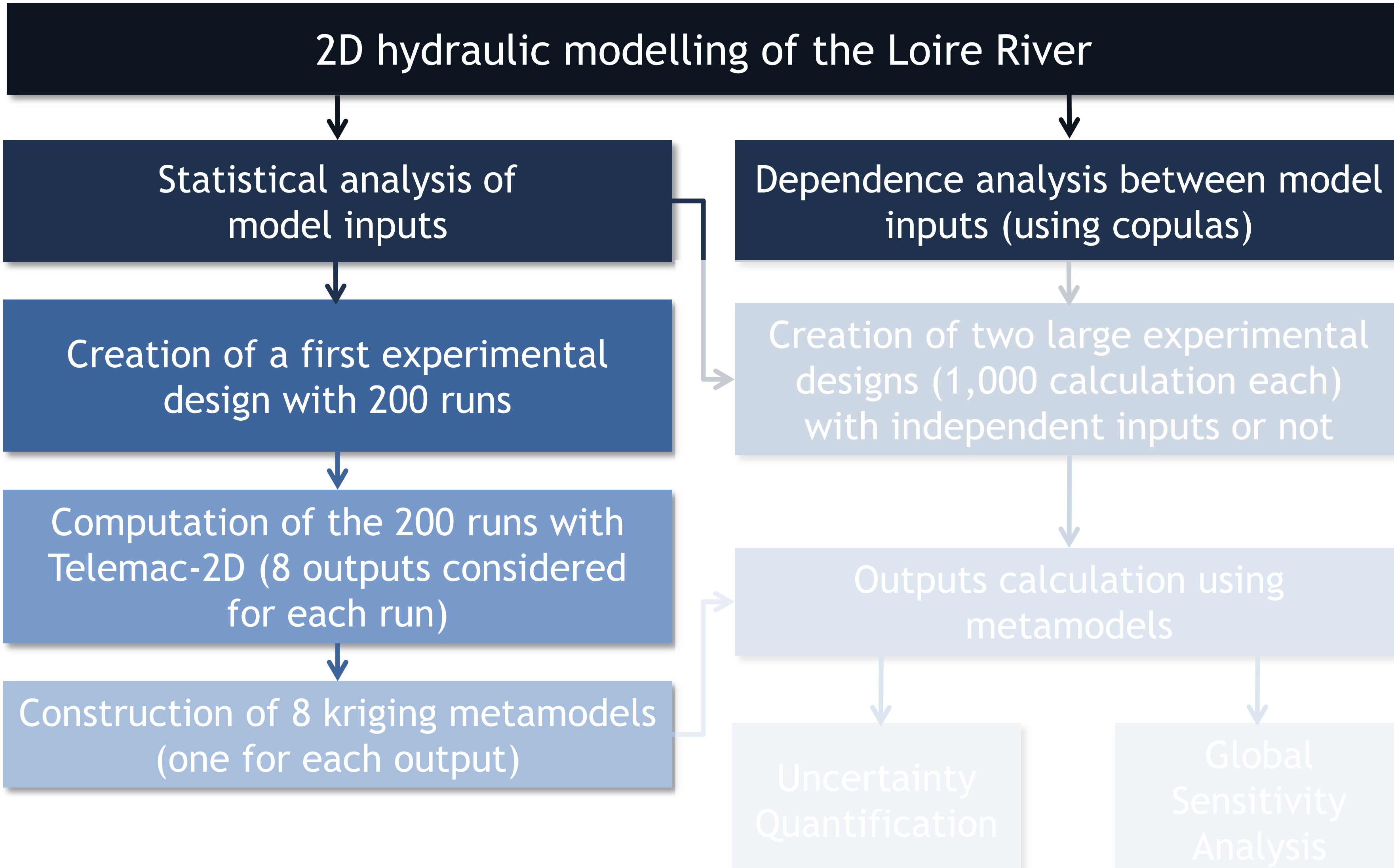
- Construction of 8 metamodels (one for each output) with the R-package *DiceEval*

- Validation: cross validation & leave one out validation
  → $R^2 > 0.97$ for the 8 metamodels



Exemple of metamodel validation for the output n°8

Methodology



Dependence analysis between inputs

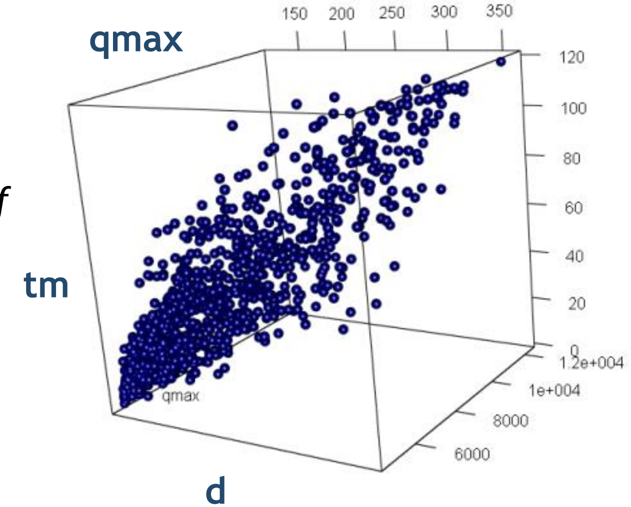
- Use of the 182 floods extracted from the flow data between 1953 to 2019

- Extraction of the 3 parameters: maximum discharge (qmax), time to peak (tm) and duration (d)

Correlation matrix between inputs (pearson coefficients)

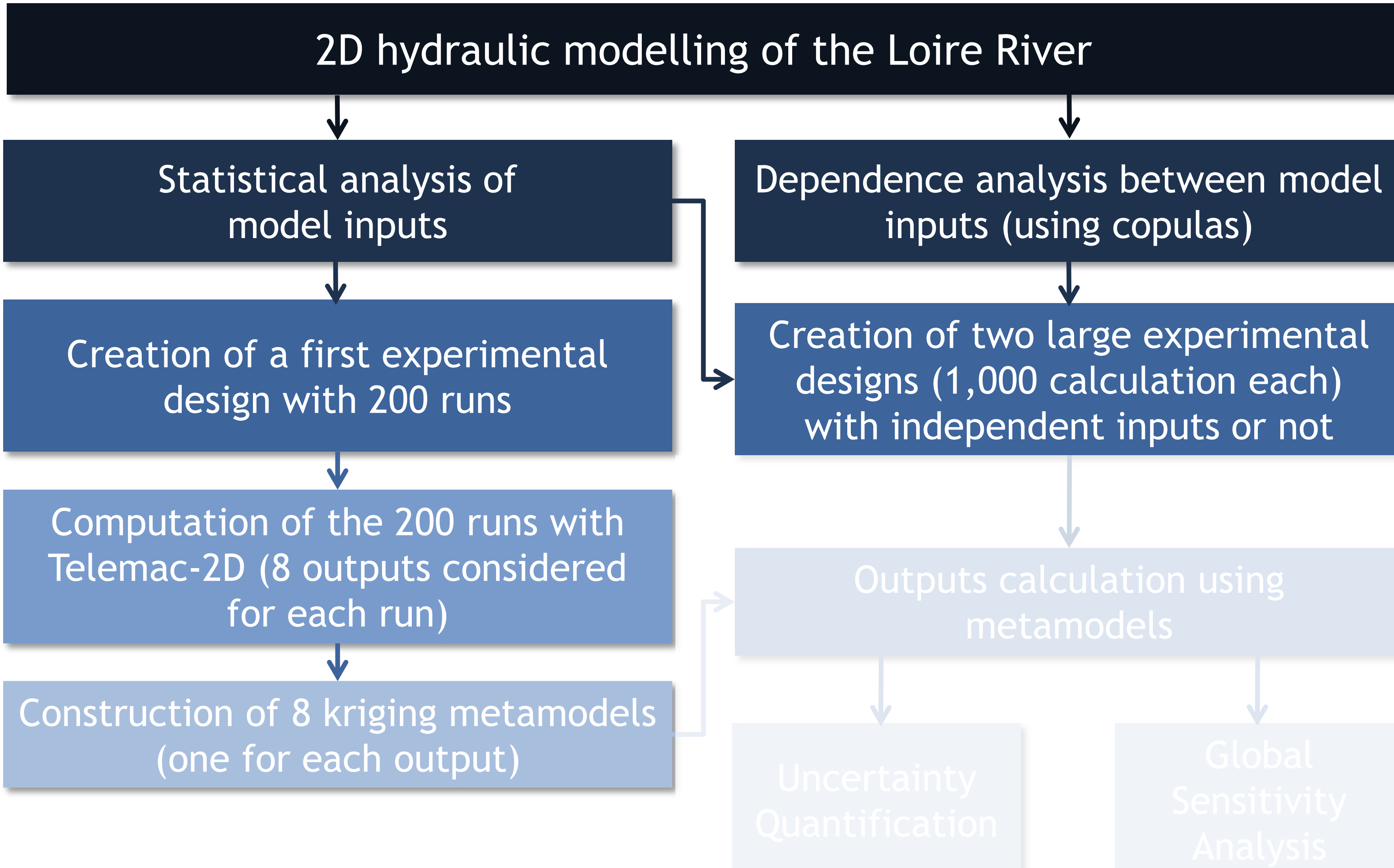
	d	qmax	Tm
d	1	0.68	0.77
qmax	-	1	0.57
tm	-	-	1

Representation of the dependence between the 3 inputs



- Research of the best copula to represent the dependence between inputs (R-package *Copula*)
  - Goodness of fit tests to select the best copula and the most adapted parameters (Cramer von Mises tests)
  - Selection of a normal copula with 3 parameters (class of meta-elliptical copula)

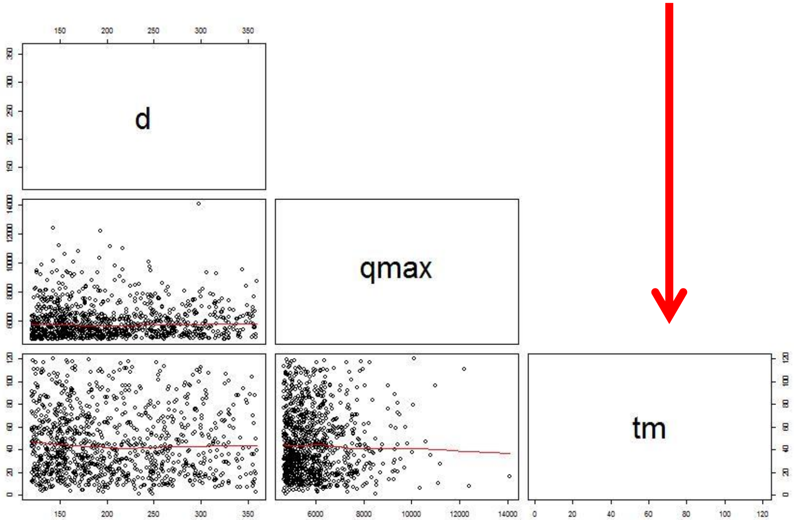
Methodology

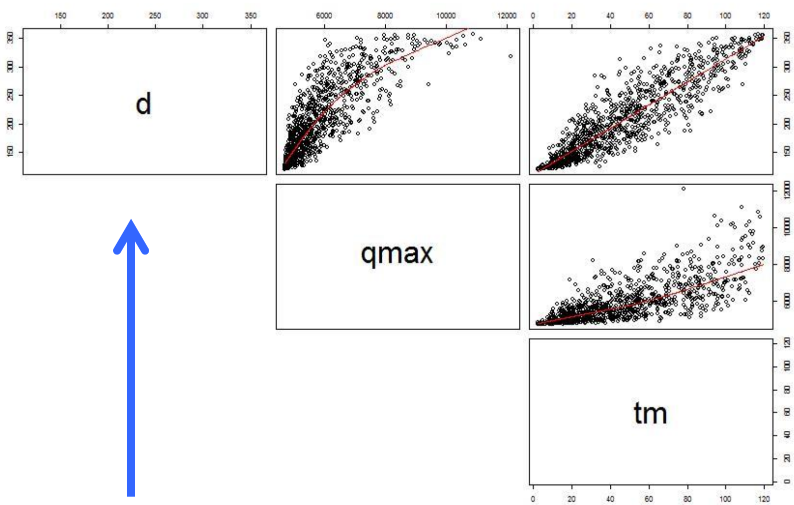


Creation of new experimental designs

First design considering independent inputs:

- For each of the 8 inputs: random sampling of 1,000 values inside their own probability distributions

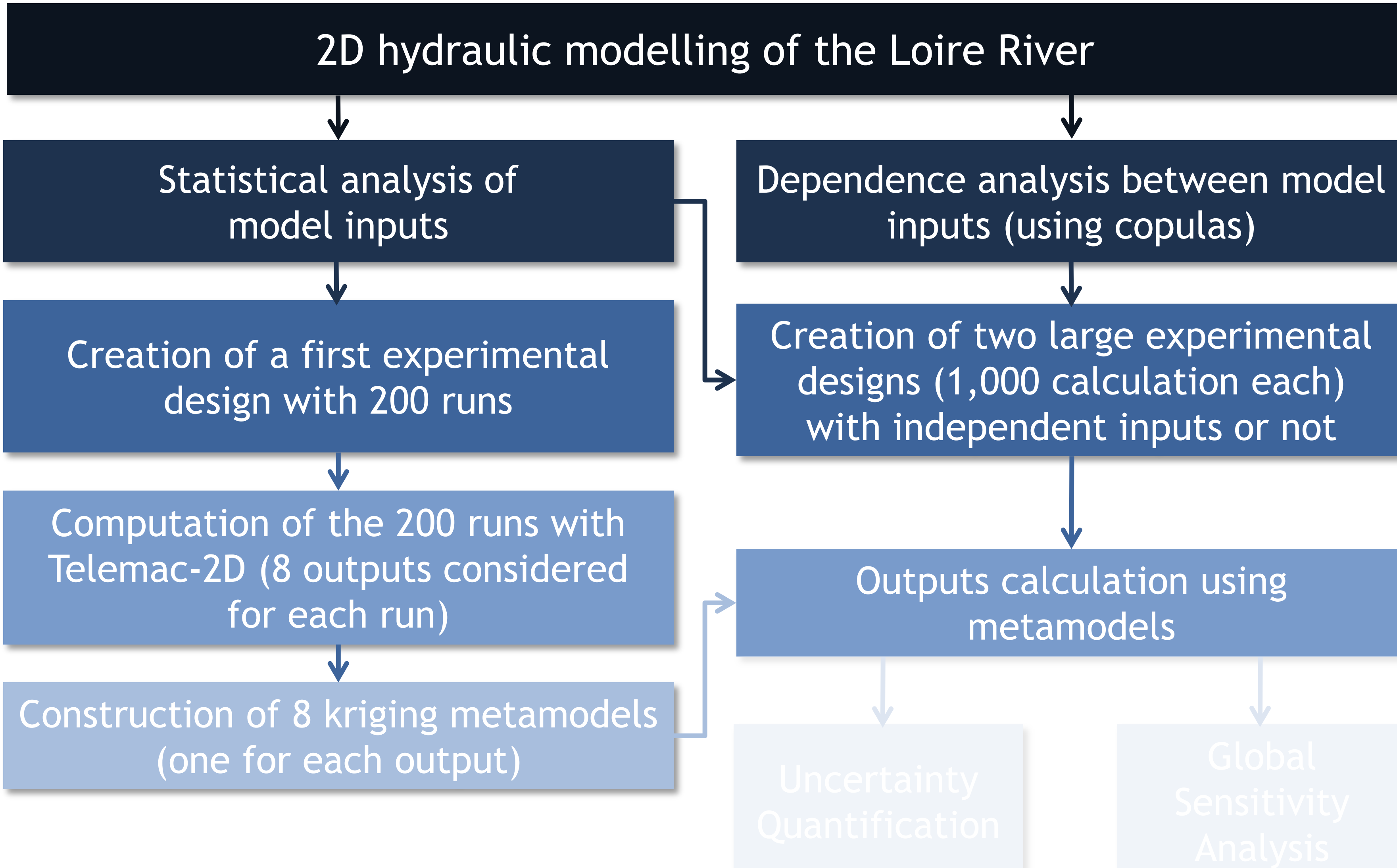




Second design considering some dependent inputs:

- For each independent input (5 Strickler coefficients, K_s1 to K_s5): random sampling of 1000 values inside their own probability distributions
- For each dependent input: random sampling of 1,000 values inside a multivariate distribution defined by the combination between the normal copula previously defined and the probability distribution of each input

Methodology

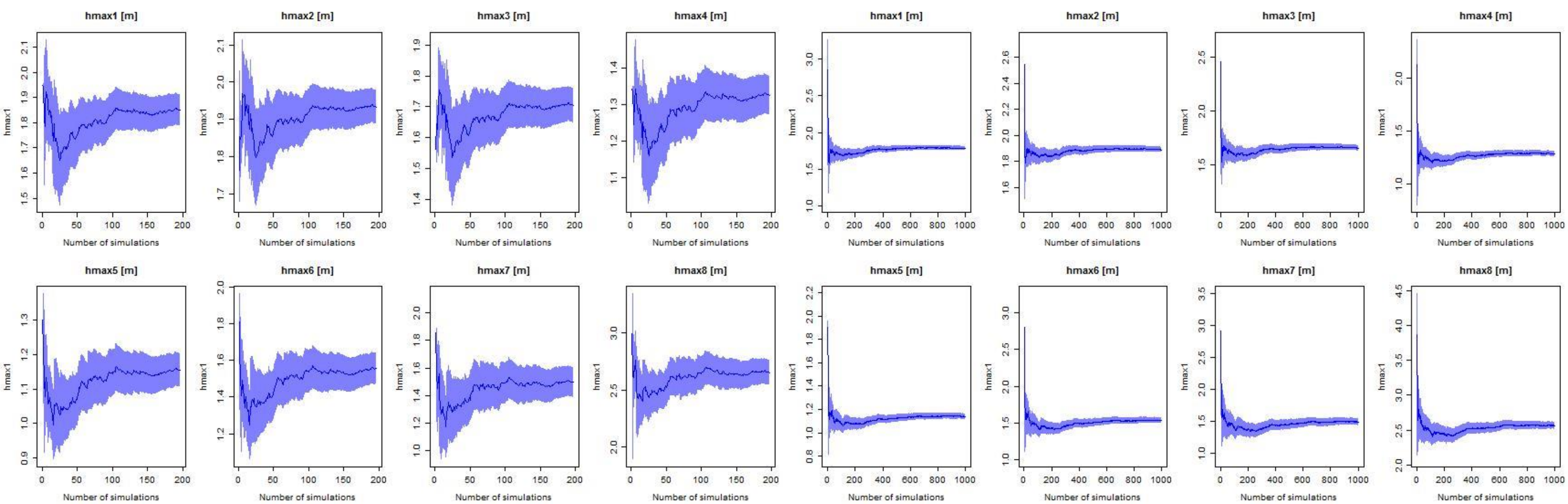


Outputs calculation using metamodels

- Calculation of the 8 outputs for the 1,000 runs of each experimental design (using the 8 kriging metamodels)

- Computation time: **less than 10 seconds !** (instead of 2,600 hours with the hydraulic model)

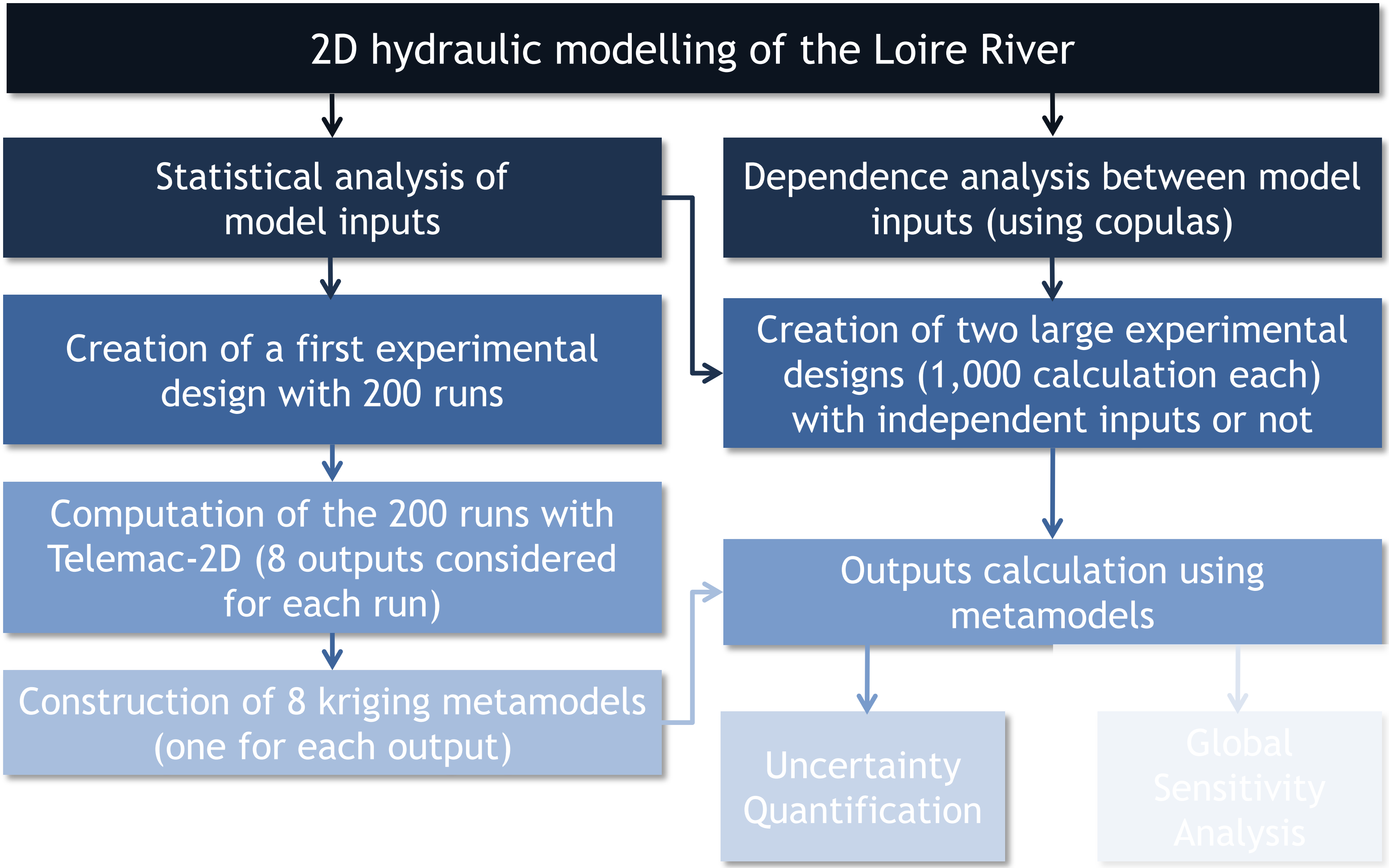
Mean convergence plots for the 8 outputs



200 runs from the Telemac-2D model

1,000 runs from the metamodels (considering independent inputs)

Methodology



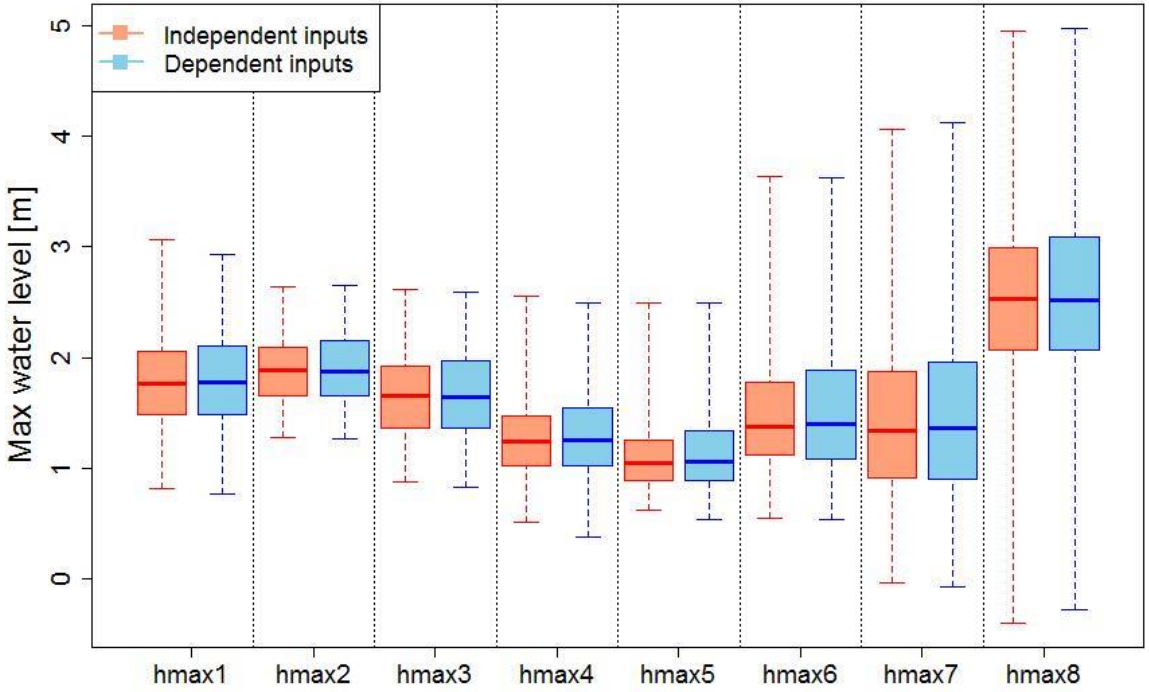
Uncertainty Quantification (1)

Generalities:

- Used to describe every possible outputs considering the input system which are not perfectly known
- Conducted with a random sampling (*e.g.* Monte-Carlo sampling) of the input parameters to obtain the distributions of the resulting outputs
- Description of the range of outputs using basic statistics (*e.g.* mean, sd), histograms, boxplots, *etc.*

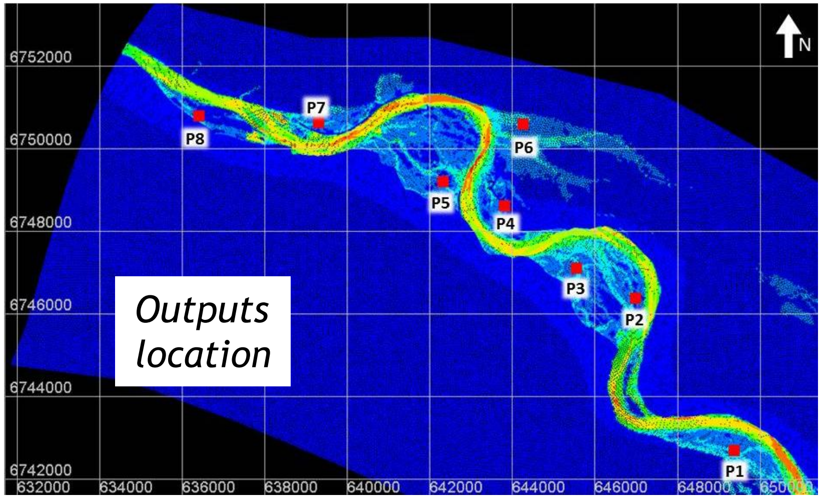
Boxplots of the outputs considering dependent or independent inputs

→ Almost no differences between the 2 cases

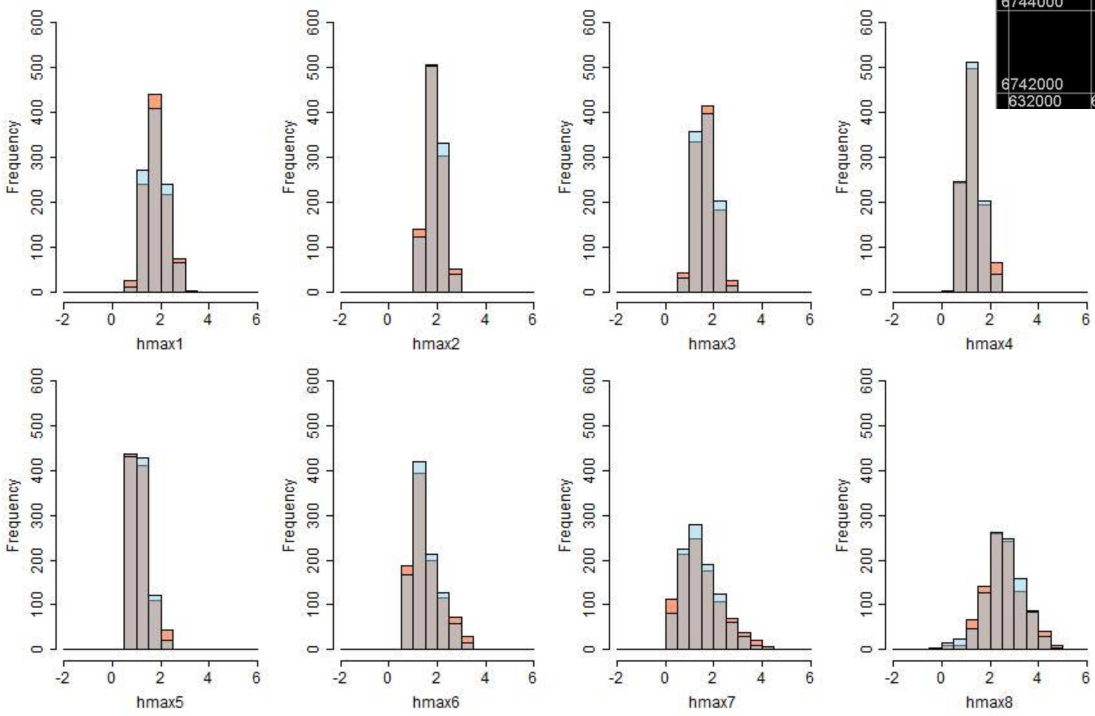


Uncertainty Quantification (2)

- Histograms of the 8 outputs considering dependent or independent inputs



Outputs location



Considering independent inputs

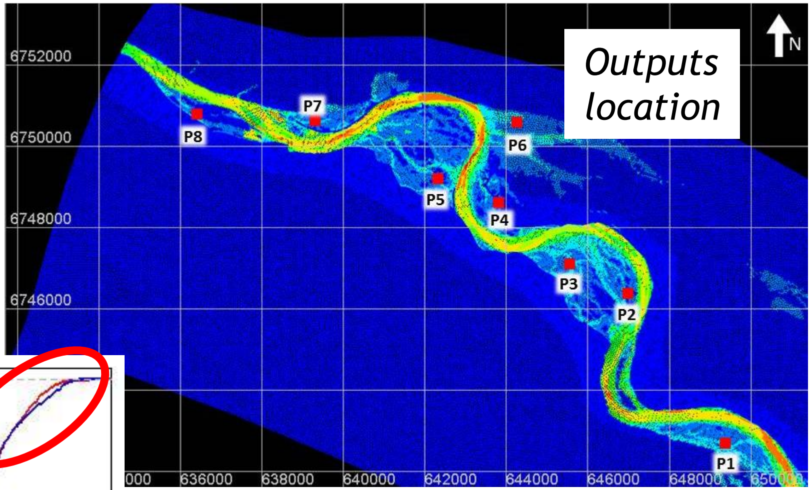
Considering dependent inputs

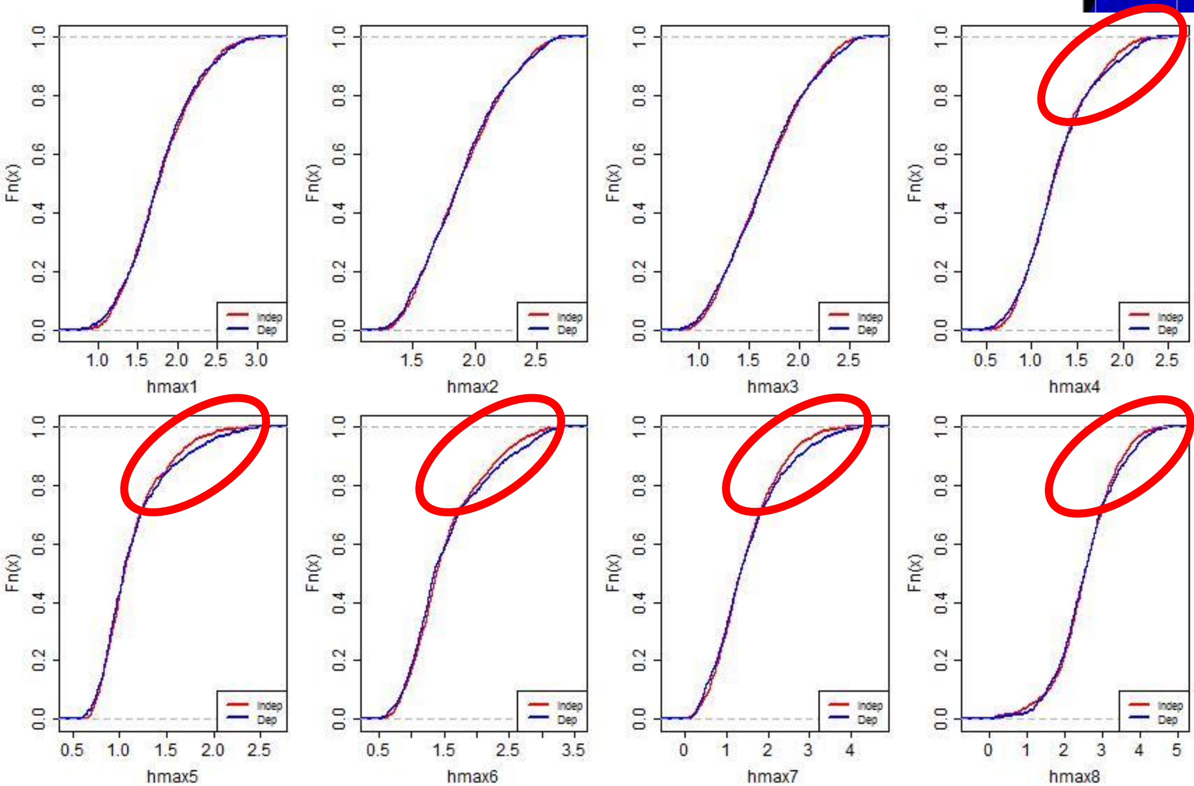
Histograms overlap

→ A few differences between both cases are observed but without any trend

Uncertainty Quantification (3)

- Empirical Cumulative Density Functions (eCDF) of the 8 outputs considering dependent or independent inputs





Considering independent inputs

Considering dependent inputs

→ Different behavior of the tail distribution for the downstream outputs

Methodology

2D hydraulic modelling of the Loire River

Statistical analysis of model inputs

Creation of a first experimental design with 200 runs

Computation of the 200 runs with Telemac-2D (8 outputs considered for each run)

Construction of 8 kriging metamodels (one for each output)

Dependence analysis between model inputs (using copulas)

Creation of two large experimental designs (1,000 calculation each) with independent inputs or not

Outputs calculation using metamodels

Uncertainty Quantification

Global Sensitivity Analysis

Global Sensitivity Analysis (1)

Generalities:

- Used to analyze the impact of the variability of inputs parameters on the variability of the outputs
- Useful to determine the most contributing variables, the non-influential ones and to rank parameters
- Use of sensitivity indices (SI) (*e.g.* Sobol' Indices) for this kind of analysis
- Indicators between 0 and 1 measuring the main effect (1st order SI) or the total effect (total order SI) of the considered input on the output

Problem:

- The SI computation is different if we consider dependent inputs or not → **traditional methods of GSA cannot be used with dependent inputs**

Use of 3 new methods to compute sensitivity indices with dependent inputs:

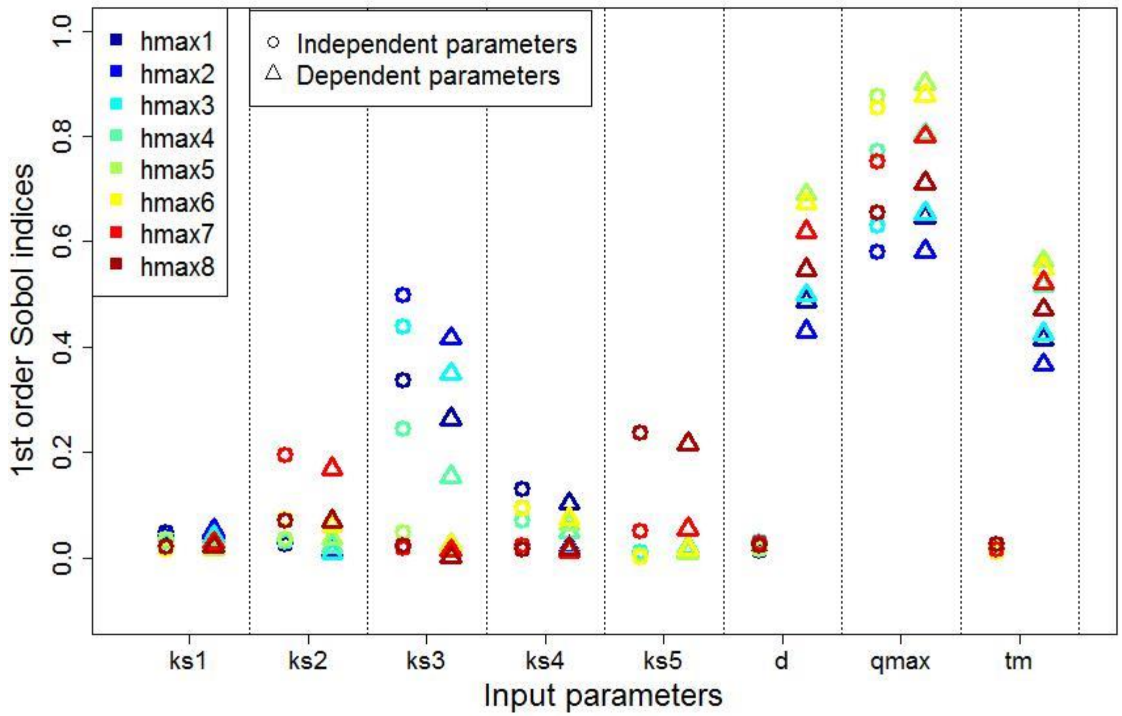
- **Li and Mahadevan, 2016**: method to directly estimate the 1st order Sobol' SI
- **McKay, 1995**: method using Latin Hypercube Sampling to compute the 1st order Sobol' SI
- **Iooss and Prieur, 2018:** method to compute Shapley effects and Sobol' SI (1st and total order) with the R-package *sensitivity*

Global Sensitivity Analysis (2)

Computation of the 1st order Sobol' SI (Li method) for all outputs:

- Depending the location of an output, the influence of each K_s coefficients differs
- For all the inputs except d and tm, the indices are almost equal considering inputs dependency or not
- For d and tm, the indices considering certain dependent inputs are much higher than considering only independent inputs → **the parameter ranking changes**

- *The Strickler coefficients (K_s1 to K_s5) are always considered to be independent*
- *d, qmax and tm are considered to be dependent in the analysis: "Δ Dependent inputs"*



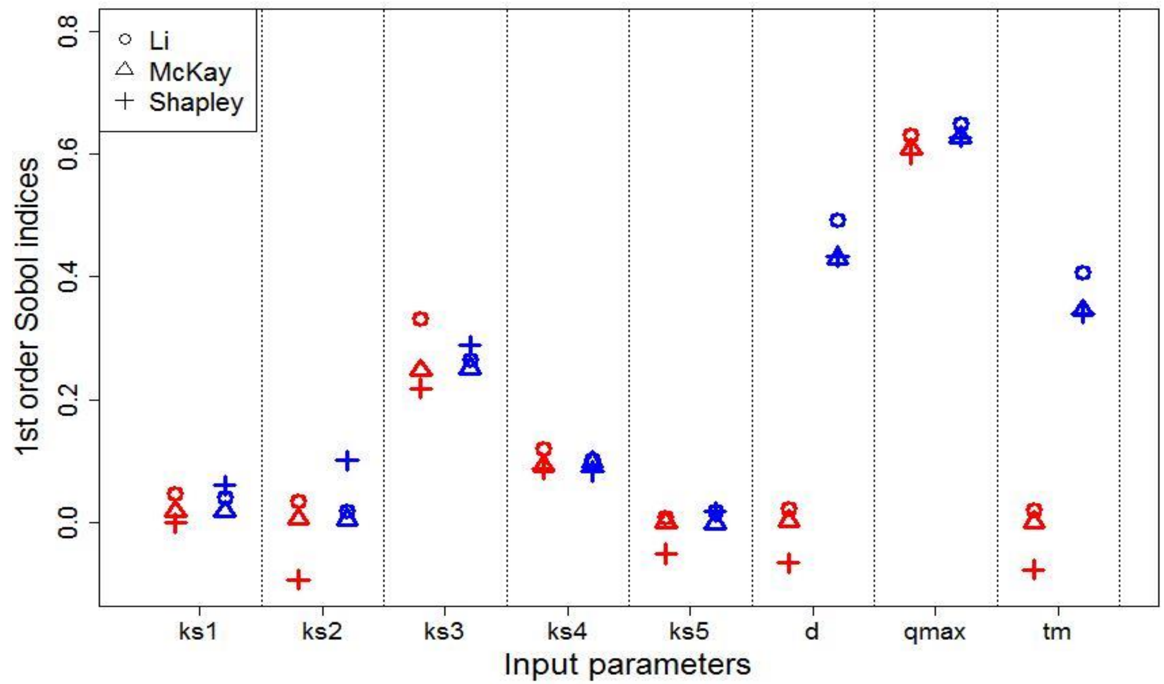
IRSN

Global Sensitivity Analysis (3)

- Comparison between the 3 methods previously cited for the output P1 (upstream)
  - Few differences between the 3 methods
  - The Li method is the fastest
  - With the Iooss and Prieur method ("Shapley") it is also possible to compute total order SI which are slightly higher than the 1st order SI

Considering independent inputs

Considering dependent inputs



Conclusion and perspectives

- Strong dependence between the hydrograph parameters (d, qmax, tm) → use of copula to model the dependencies
- Metamodel very useful for uncertainty analysis studies (almost all done during the containment with limited computation ressources)
- Limited impact of inputs dependency in uncertainty quantification in this study
- The duration and time to peak inputs have strong influence on the outputs → **The hydrograph shape should not be ignored in hydraulic studies**
- Further work: study the influence of other hydraulic parameters dependencies (*i.e.* breach levee parameters)

Thank you for your attention

Orleans archives

References

- *I. M. Sobol, « Sensitivity Estimates for Nonlinear Mathematical Models », MMCE, vol. 1, nº 4, p. 8, 1993.*
- *M. D. McKay, « Evaluating prediction uncertainty », Nuclear Regulatory Commission, 1995.*
- *C. Li et S. Mahadevan, « An efficient modularized sample-based method to estimate the first-order Sobol ′index », Reliability Engineering & System Safety, vol. 153, p. 110-121, sept. 2016.*
- *B. Iooss et C. Prieur, « Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications », p. 39, 2018.*