Uncertainty and Sensitivity Analyses with Dépendent Inputs in a 2D Hydraulic Model of the Loire River

Lucie Pheulpin, Antonin Migaud, Nathalie Bertrand

Uncertainty and Sensitivity Analyses with Dependent Inputs in a 2D Hydraulic Model of the Loire River

Lucie Pheulpin[(1)], Antonin Migaud[(1)] and Nathalie Bertrand[(1)]

[(1)]Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France
lucie.pheulpin@irsn.fr, antonin.migaud@irsn.fr, nathalie.bertrand@irsn.fr

Abstract

To assess flooding risk, hydraulic models are used but they contain many uncertainties related to the lack of knowledge of input parameters. To quantify the uncertainties and evaluate the most influential parameters of the model, Uncertainty Quantification (UQ) and Global Sensitivity Analysis (GSA) can be used. In order to implement these methods, model inputs must be independent, which is not often the case. This research aims to propose a methodology to deal with dependent inputs in UQ and GSA for hydraulic models and to reduce the computational times. To do so, a 2D hydraulic model of the Loire River built with TELEMAC-2D was used. The study methodology is carried out in the following steps: the uncertain model inputs (hydraulic and breach parameters) and the outputs of interest (the water level at given points) are set. The inputs margins and dependencies are defined by a statistical analysis using a real dataset of the Loire River. The dependency structure between inputs is represented by copulas. Since UQ and GSA require many simulations, kriging metamodels are used to increase the number of experiments in a short time period. Finally, UQ and GSA are carried out by considering the inputs dependent or not. The outputs distribution slightly differs if the inputs are considered independent or not. The number of influencing parameters increases when inputs are considered dependent. Some parameters, usually considered as not influencing, may actually be significantly impacting on the outputs. These results suggest that dependencies should not be overlooked in 2D hydraulic models.

Keywords: Uncertainty quantification, Global sensitivity analysis, Copula, Metamodel, Levees breaches

1. INTRODUCTION

In France, 17.1 million inhabitants are exposed to different consequences of flooding by river overflow. These different types of flooding can be related to each other, e.g., runoff contributes to river overflow, a marine flood can cause or aggravate a river overflow, a structure failure can cause or aggravate a river overflow or a marine flood, and conversely a river overflow of a level exceeding the protection level of the structure can cause its partial or total failure. These floods can be assessed by numerical modelling. However, the models have numerous uncertainties linked to the input parameters of the model. Indeed, the choice of hydraulic parameters such as roughness coefficients, hydrograph parameters and breach parameters (e.g. location or geometry) is a major source of uncertainty that has a strong impact on the water level at a given location. Thus, methodologies have been developed to better understand the impact of uncertain input parameters on the overtopping generated and to improve the quantification of the flood risk by considering the uncertainties. Hence, uncertainty quantification (UQ) and global sensitivity analysis (GSA) are used. Traditionally, to perform these kinds of analyses, model input parameters are supposed to be independent, which is not always true, especially in hydraulic studies. Therefore, this study aims to propose a complete methodology to analyse uncertainties in hydraulic models by taking into account the dependencies between model inputs.

To manage uncertainties in hydraulic models, Bacchi et al. (2018) highlights that the following methods based on statistical techniques can be used (Faivre et al., 2013; Saltelli et al., 2008). Where UQ aims to describe the set of possible outcomes considering the input system as not perfectly known and GSA attempts to measure the extent to which model outputs are affected by changes in model inputs and it is also used to rank parameters.

Moreover, to handle UQ with inputs that are considered independent, classical methods such as Monte Carlo are commonly used. To deal with dependent inputs, methods including copula (joint distribution defined in a d-dimensional space $[0,1]^d$ with uniform marginal distributions (Nelsen, 2007), where d is the number of parameters in the given copula) can be used. To do this, inputs are randomly sampled inside marginal distributions, taking into account the dependency structure, i.e. the copula. These mathematical objects are often used in flood frequency multivariate analyses and some authors (Balistrocchi et al., 2014; Domeneghetti et al., 2013; Vorogushyn et al., 2011) use copula approaches to model the dependence structure between hydrograph parameters. With respect to GSA, new methods have been recently developed to consider dependencies between inputs (Da Veiga et al., 2009; Jacques et al., 2006; Li and Mahadevan, 2016; McKay,

1995), as the one of Owen and Prieur (2017) and looss and Prieur, (2017) who introduced the Shapley effects.

These methods require many simulations and can be computationally time consuming especially when 2D models are used. Thus, alternative methods, such as metamodel approaches, can be used to significantly reduce the computational time (Richet and Bacchi, 2019).

In our previous studies, influence of levee breaches on flooded areas using HEC-RAS 1D (Pheulpin et al., 2019) and TELEMAC-2D (Pheulpin et al., 2020) models has been investigated. More recently, a methodology to perform a complete uncertainty analysis including UQ and GSA and taking into account the dependencies between the model inputs has been proposed here applied to a very simple case of flooding modelled using TELEMAC-2D (Pheulpin et al., 2022).

In this context, the objective of this work is to quantify the uncertainties in a 2D hydraulic model and to evaluate the effects of dependencies between some input parameters on the water level modelled of the Loire River between Given and Jargeau. This paper is divided into several parts: the second part presents the case study, the third part introduces the methodology and the fourth part shows the results of the UQ and GSA with dependent and independent inputs.

2. THE CASE STUDY: LOIRE RIVER

The selected model represents the Loire River between Gien and Jargeau over 50 km. The Loire River is lined by numerous levees on both banks and more than 20 historical breaches are known in this study area (Figure 1). The model was built by IRSN with the open-source TELEMAC-2D computational codes. The model mesh is made of approximately 180,000 nodes and takes one hour on average for one run with 38 parallel processors.

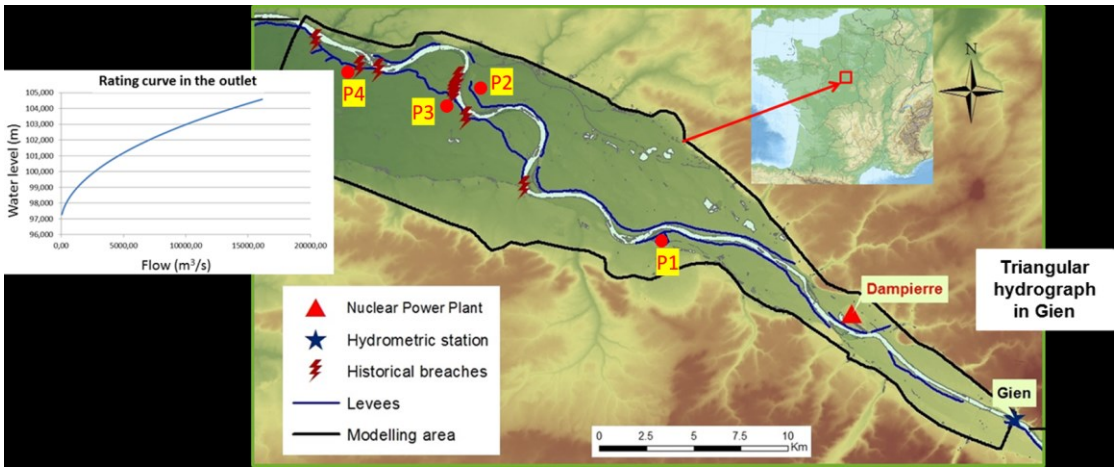

Figure 1. Modelling area of the Loire River. The 4 red points are the outputs of interest.

The boundary conditions used are an upstream hydrograph and a downstream calibration curve. The upstream hydrograph is defined by the following three parameters: time to peak (*tm*), duration (*d*) and maximum flow (*qmax*) (Figure 2). The model has been calibrated for well-known flood events; the calibration parameters are the eight roughness (Strickler) coefficients (K_s) presented in Figure 2.

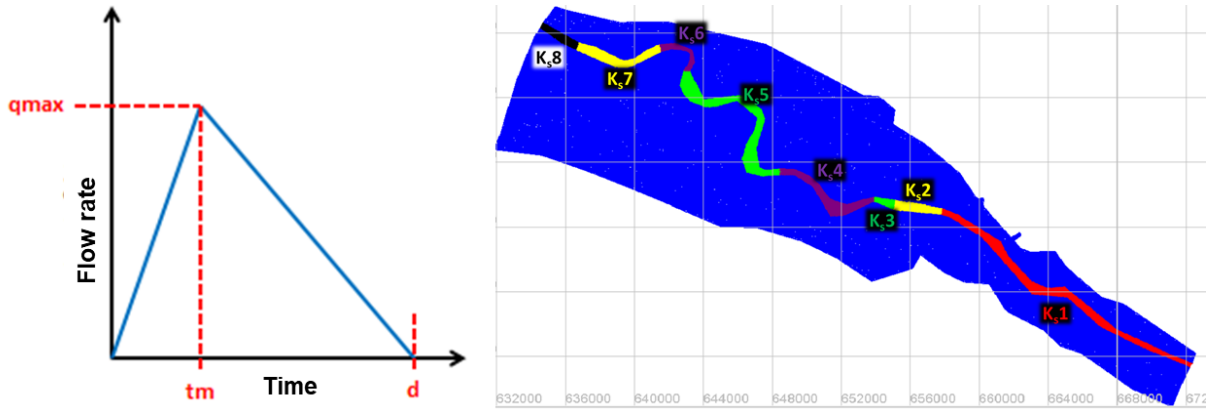

Figure 2. Designed hydrograph parameters (left) and locations of the 8 roughness coefficients used to calibrate the model (right)

To simulate breaches, three following breach parameters, introduced in Figure 3, are considered: final breach depth (*cf*), breach initiation parameter (also named control level), corresponding to the water level above the levee (*cc*) and breach opening time (*do*). A breach occurs when the water level above the levee reaches the control level (*cc*).

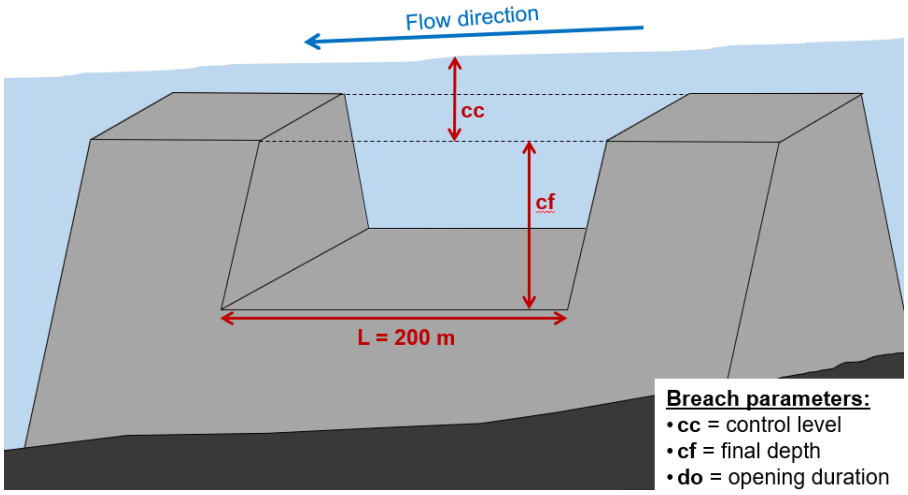

Figure 3. Breach parameters considered

For this study, we only focus on four outputs points located in the floodplain (P1, P2, P3 and P4 in Figure 1) and we are interested in the maximum water level at these points as a variable of interest .

3. METHODOLOGY

3.1. Definition of uncertain inputs and dependencies

Since the goal is to perform a large number of floods considering different combinations of randomly chosen parameters, it is therefore necessary to define the probability density function (PDF) of each input and the dependencies between some of these inputs.

Regarding hydrograph parameters, a truncated GEV distribution is used for the maximum flow rate (*qmax*), and truncated log-normal distributions are used for the duration (*d*) and time to peak (*tm*). Concerning the dependencies between the hydrograph parameters, a Vine model is used (Mazo, 2014). The PDF and the dependencies of the hydrograph parameters are defined using the Loire River dataset (daily flow rate in Gien) (see Pheulpin et al., 2022 for further details). Regarding the Strickler coefficients (K_s), as calibration parameters have been defined and the PDF cannot be defined from a real dataset, triangular distributions are used. The mode of triangular distributions corresponds to the calibration values defined for each sector presented in Figure 2. Moreover, all roughness coefficients are considered independent.

Concerning the breach parameters, the PDF and the dependencies are more complex to define because data is missing. For the opening time (*do*), a uniform distribution is retained and for breach size parameter (*cf*) and control level (*cc*), triangular distributions are used. As we have no real data to evaluate the dependencies, no copula is used to model a possible dependency.

Finally, the inputs are randomly sampled within the PDF presented in Table 1 with or without using the copula mentioned here. Minimum and maximum boundaries are arbitrarily chosen. Among others, the minimum value of *qmax* corresponds to a 50-year return period flood. In fact, as we would like to simulate breaches, the water level above the levees, therefore the maximum flow, must be high enough to cause a dyke failure. The maximum valued of *tm* and *do* must always be under *d*. *cc* max. is selected in such a way as the dyke breaches with an overflow of 20 cm max.

3.2. Simulations with TELEMAC-2D

Once the input distributions and dependencies were chosen, numerous TELEMAC-2D simulations were run. However, in order to perform our uncertainty analysis, a large dataset is required and as one simulation with TELEMAC-2D lasts one hour on average, thousands of simulations would take too much time to run. To overcome this difficulty, only 200 runs using TELEMAC-2D were made and metamodels were used to generate much larger datasets.

We created an experimental design of 200 combinations of parameters. Here, inputs are randomly sampled within uniform distributions (use of the min. and max. limits presented in Table 1) and without taking

into account the dependencies for now. To run the 200 calculations successively and to extract the maximum water level at the 4 points of interest (cf. Figure 1), we used the IRSN computational tool Funz (https://funz.github.io/) which can be coupled with the software TELEMAC-2D. The 200 runs with TELEMAC-2D lasted 194 hours (with 38 processors).

Table 1: Inputs distributions and copula. *cc* is in meters below or above the levee crest and *cf* corresponds to the proportion of the levee that breaks (0 means that there is no breach while 1 means that the levee breaks on all its height).

Inputs	Units	PDF	Distribution parameters		Copula
K_s 1 to 8	-	triangular	min max	10 40	-
d	h	truncated log-normal	mean(log) sd(log) min max	4.91 0.83 151 14,000	Vine model
tm	h	truncated log-normal	mean(log) sd(log) min max	3.83 0.72 30 150	
qmax	m^3/s	truncated GEV	location scale shape min max	1548 578 0.15 4600 ∞	
do	h	uniform	min max	0 150	-
cc	m	triangular	min max	-0.2 0.2	-
cf	-	triangular	min max	0 1	-

3.3. Metamodels building

Metamodels are currently used for UQ and GSA approaches because they allow to reduce the computation time while preserving the statistical outputs of the initial model. For our analysis, a kriging metamodel was chosen (Gratiet et al., 2015) for its good predictive capabilities already demonstrated by Marrel et al., (2008). The methodology used for the development and validation of the metamodels is fully detailed in previous studies (Roustant et al., 2012; Saltelli, 2002; Wahl, 2004).

Four metamodels were built, one for each output, following these three main steps:

1. Construction of a learning basis with 200 combinations of 14 input parameters and four outputs (cf. part 3.2);
2. Use of the learning basis to build the 4 kriging metamodels with the R package DiceEval;
3. Validation of the metamodels using *k-fold cross-validation* and *leave-one-out cross validation* methods (cf. Pheulpin et al., 2020 for more details) and check for accuracy using the R-squared.

3.4. Uncertainty Quantification (UQ)

To perform the UQ, two new experimental designs of 10,000 simulations each are built from the metamodels. Input parameters are randomly sampled from the distributions introduced in Table 1. They are considered independent for one design and dependent for the second design. Then, using the four metamodels, the outputs are calculated. Finally, the uncertainties can be quantified using histograms, boxplots, empirical cumulative distribution functions, etc.

3.5. Global Sensitivity Analysis (GSA)

The Shapley effects (looss and Prieur, 2017) were used to perform the GSA because they are suitable for dependent inputs. These indices range from 0 to 1 and the closer they are to 1, the more influential the inputs are. As the sum of the indices of all inputs is equal to 1, the indices allow to directly estimate the share of variance of the output explained by a given input. To compute the Shapley sensitivity indices, the R package sensitivity (looss et al., 2020) has been used.

4. RESULTS AND DISCUSSION

As explained above, to conduct UQ and GSA, four metamodels were built and validated. The Table 2 shows the validation criteria computed with the cross-validation method. The metamodels well represent the models as the quality criteria are good enough, so they can be used for UQ and GSA.

Table 2. Metamodels validation criteria. RMSE is the Root Mean Squared Error and MAE is the Mean Absolute Error.

Output	RMSE	MAE	R^2
P1	0.109	0.084	0.947
P2	0.046	0.036	0.981
P3	0.118	0.095	0.948
P4	0.180	0.140	0.910

4.1. Uncertainty quantification

To carry out the UQ, 10,000 simulations considering all inputs being independent and 10,000 others considering the hydrograph parameters being dependent, have been achieved. The Figure 4 illustrates the densities of the maximum water levels of the four outputs of interest (P1 to P4 illustrated in Figure 1). We do not report a significant difference between the case where all inputs are independent (in red) and the one where they are not (in blue), except for the output P4 where there is a slight difference in the center of the distribution.

Regarding the output P1, numerous simulations do not generate overflow as the median of the maximum water level is close to 0. For the other outputs, the ranges are more extended. For instance, the maximum water level varies from 0 to 4 meters in the point P4.

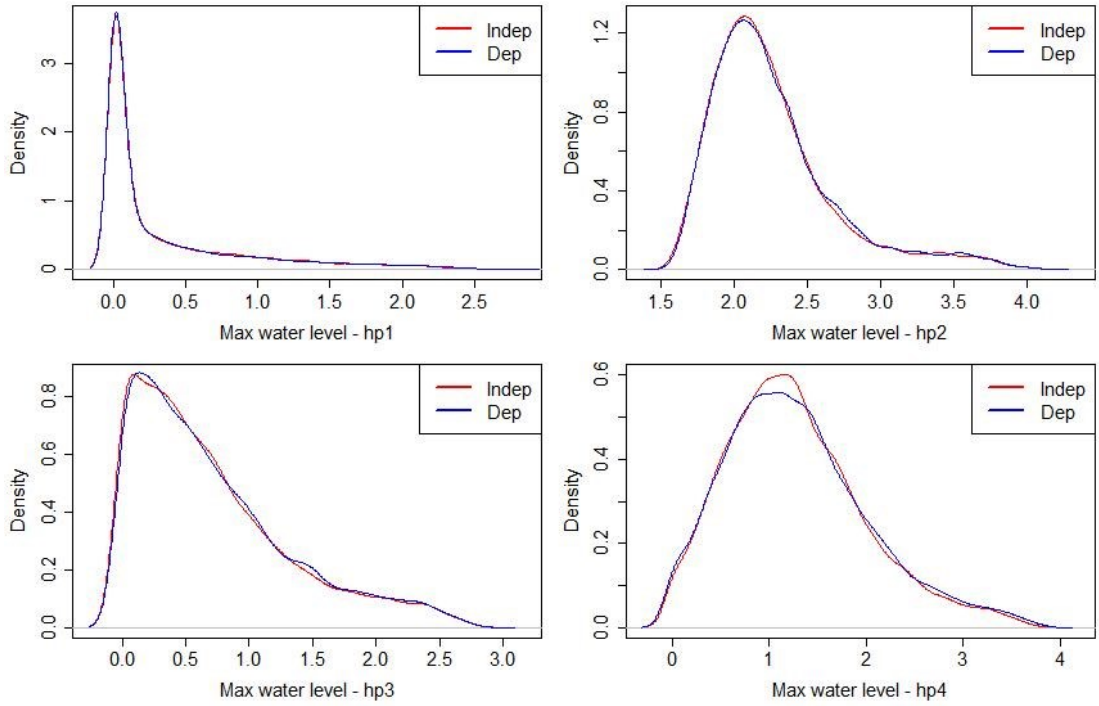

Figure 4. Maximum water level distributions at the 4 output points considering independent inputs (in red) or dependent inputs (in blue)

4.2. Global sensitivity analysis

Shapley effects, presented in Figure 5, were used to perform the GSA. They were computed using the 10,000*2 simulations (for independent and dependent inputs) coming from the metamodels. In case of independent parameters (in red), there is a strong influence of the maximum flow (*qmax*) and a lower influence of some roughness coefficients (K_s). The influent roughness coefficients (K_s6, K_s7 and K_s8) are the downstream ones, which seems usual because the downstream conditions regulate the upstream zone. If the hydrograph parameters are considered as dependent (in blue), the flood duration (*d*) and the time to peak (*tm*) also seem to have an influence on the outputs. Finally, in case of independent inputs or not, there is no

influence of breach parameters. The breaches are too small compared to the flood magnitude. The Shapley effects are similar for all outputs except for the first one (P1). In this case, the rugosity coefficient K_s8 has as much influence or more than the maximum flow *qmax*.

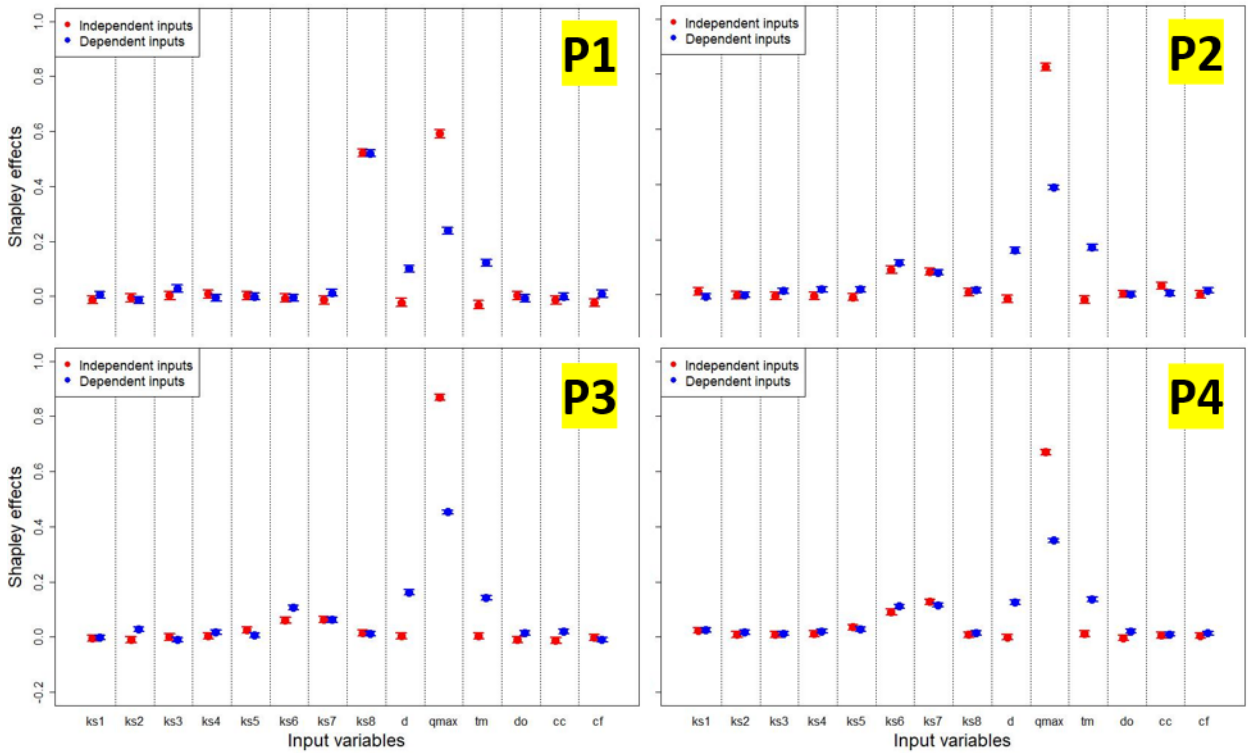

Figure 5. Shapley effects for the 14 uncertain inputs, considering independent inputs (in red) or dependent inputs (in blue)

4.3. Simplistic probabilistic hazard assessment

To make a first assessment of the probabilistic hazard, the maximum water level was computed for different return periods from 100 to 1,000 years by considering the 70% confidence interval of the maximum flow GEV distribution and the variability of other inputs. For this, for each return period, a thousand values of *qmax* are sampled within the 70% confidence interval of the GEV distribution and the other values are sampled inside their probability distributions (Table 1). The dependences are considered or not. The results for the P4 output are presented in Figure 6. It should be noted that this method is still exploratory. The results show that considering a 70 % confidence interval, the outputs can vary by an average of 30 cm in height, depending on the return period. The differences between the case with independent inputs and the one with some dependent inputs are very weak compared to the thickness of the confidence interval.

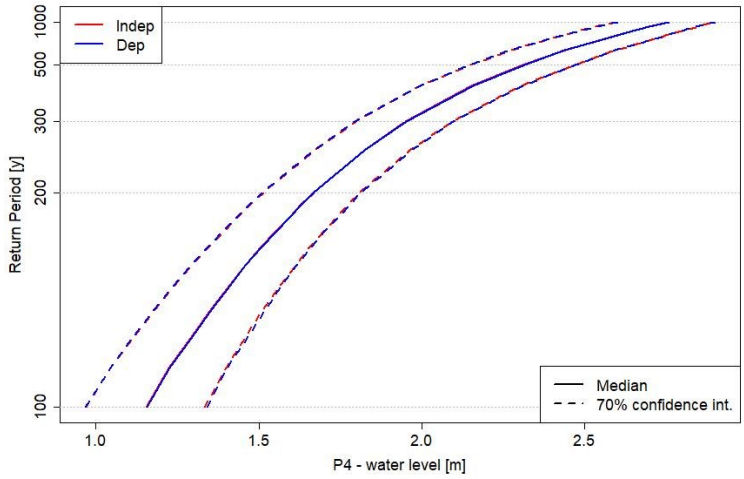

Figure 6. Maximum water levels with 70 % confidence interval, at the point P4 for different return periods.

4.4. Discussion

As the simulated flood have a return period of 50 years or more, the flow rates used in this study are particularly high, so the breach effects are insignificant or marginal. Indeed, the breach dimensions considered are too small to have a major influence on the flood magnitude. Thus, for this kind of study, where all the floodplain is massively inundated, the consideration of the realistic breaches is not meaningful.

Regarding hydraulic parameters, the GSA with dependent inputs shows that the three hydrograph parameters have a major effect. It indicates that the hydrograph shape should not be neglected. The rugosity coefficients and particularly those downstream also have a significant influence on the outputs. Usually, the rugosity coefficients are dependent on the maximum flow rate. But, as there are calibration coefficients, it is complex to model this dependency as we have no real data of rugosity coefficients.

About the UQ, it seems that the integration of the Vine model has no effect on the output densities. With or without dependencies, the outputs distributions are the same. This observation is available for this case study and for these points of outputs. For instance, in the basic case study presented in Pheulpin et al., 2022 the dependence had a significant effect on the output distributions.

The method used to build the probabilistic hazard assessment curves is very simplistic and will be improved in a near future. In fact, uncertainties on the metamodels are not considered as well as the uncertainties of other than the maximum flow rate inputs.

5. CONCLUSION AND PERSPECTIVES

The objective of this work is to quantify the uncertainties in a 2D hydraulic model and to evaluate the effects of the consideration of the dependency between some input parameters on the results. The method is applied to the 2D hydraulic model of the Loire River between Gien and Jargeau with a focus on hydraulic and breach parameters. First, a statistical analysis of the inputs was made (search for probability distributions of inputs and dependencies between inputs through copulas) and a learning basis of 200 runs was built using the coupling tool Funz/TELEMAC-2D. Then kriging metamodels were built to perform more simulations with a low computation time. Finally, UQ and GSA were treated using the metamodels and considering independent inputs or not.

For this study case, the UQ results show a few differences between the case with dependent inputs and the one without. For the concerned outputs the consideration of the dependence between hydrograph parameters is not meaningful but we have to be careful because the results could be different for other outputs.

The GSA shows more differences if the dependence is considered or not. When the hydrograph parameters are considered dependent, the three hydrograph parameters have a strong influence on the results, unlike the case of independent parameters where only the maximum flow rate have influence. Regarding the breach parameters, they have no influence on the outputs because breaches are insignificant compared to the flood magnitude.

The simplistic probabilistic hazard assessment curves are a first result providing water levels with uncertainties for given return periods. However, to increase their accuracy, these curves should integrate the uncertainties on the other parameters and on the metamodels.

In a near future, this methodology will be applied to other hydraulic models where breaches have more influence on the water levels in the flooding plain. We would like to improve some points: to take into account the shapes of the hydrographs and not only use triangular hydrographs as input data, to integrate the dike fragility curves as input data for the model and to take into account other types of breaches and not only breaches by overflow.

6. ACKNOLEDGEMENTS

This study was conducted within the NARSIS project, which has received funding from the European Union's H2020-Euratom Programme under grant agreement No. 755439. We also would like to thank Yann Richet from IRSN for building the coupling tool Funz/TELEMAC-2D.

7. REFERENCES

Bacchi, Duluc, C.-M., Bardet, L., Bertrand, N., Rebour, V., 2018. Feedback from uncertainties propagation research projects conducted in different hydraulic fields: outcomes for engineering projects and nuclear safety assessment, in: *Advances in Hydroinformatics*. Springer, pp. 221–241.

Balistrocchi, M., Ranzi, R., Bacchi, B., 2014. Multivariate statistical analysis of flood variables by copulas: two italian case studies. *3rd IAHR Europe Congress, Book of Proceedings* 12.

Da Veiga, S., Wahl, F., Gamboa, F., 2009. Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs. *Technometrics* 51, 452–463. https://doi.org/10.1198/TECH.2009.08124

Domeneghetti, A., Vorogushyn, S., Castellarin, A., Merz, B., Brath, A., 2013. Probabilistic flood hazard mapping: effects of uncertain boundary conditions. *Hydrology and Earth System Sciences* 17, 3127–3140. https://doi.org/10.5194/hess-17-3127-2013

Faivre, R., Makowski, D., Mahévas, S., looss, B., 2013. Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l'environnement. *Analyse de sensibilité et exploration de modèles* 1–352.

Gratiet, L.L., Marelli, S., Sudret, B., 2015. Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes. *Ghanem R., Higdon D., Owhadi H. (eds) Handbook of Uncertainty Quantification* pp 1-37. https://doi.org/10.1007/978-3-319-11259-6_38-1

looss, B., Prieur, C., 2017. Shapley effects for sensitivity analysis with dependent inputs: comparisons with Sobol' indices, numerical estimation and applications. *International Journal for Uncertainty Quantification* 9. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2019028372

looss, B., Veiga, S.D., Janon, A., Pujol, G., Broto, with contributions from B., Boumhaout, K., Delage, T., Amri, R.E., Fruth, J., Gilquin, L., Guillaume, J., Gratiet, L.L., Lemaitre, P., Marrel, A., Meynaoui, A., Nelson, B.L., Monari, F., Oomen, R., Rakovec, O., Ramos, B., Roustant, O., Song, E., Staum, J., Sueur, R., Touati, T., Weber, F., 2020. sensitivity: Global Sensitivity Analysis of Model Outputs, *R package*.

Jacques, J., Lavergne, C., Devictor, N., 2006. Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System Safety* 91, 1126–1134. https://doi.org/10.1016/j.ress.2005.11.047

Li, C., Mahadevan, S., 2016. An efficient modularized sample-based method to estimate the first-order Sobol' index. *Reliability Engineering & System Safety* 153, 110–121. https://doi.org/10.1016/j.ress.2016.04.012

Marrel, A., looss, B., Van Dorpe, F., Volkova, E., 2008. An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics & Data Analysis* 52, 4731–4744.

Mazo, G., 2014. Construction et estimation de copules en grande dimension (*Thèse de doctorat*). Université de Grenoble.

McKay, M.D., 1995. Evaluating prediction uncertainty. *Nuclear Regulatory Commission.*

Nelsen, R.B., 2007. An introduction to copulas. *Springer Science & Business Media.*

Owen, A.B., Prieur, C., 2017. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification* 5, 986–1002.

Pheulpin, L., Bacchi, V., Bertrand, N., 2020. Comparison Between Two Hydraulic Models (1D and 2D) of the Garonne River: Application to Uncertainty Propagations and Sensitivity Analyses of Levee Breach Parameters, in: *Advances in Hydroinformatics.* Springer, pp. 991–1007.

Pheulpin, L., Bacchi, V., Bertrand, N., 2019. Analyse de sensibilité des paramètres de rupture des digues: application au cas de la Garonne, in: *Digues Maritimes et Fluviales de Protection Contre Les Inondations 2019*, Digues 2019.

Pheulpin, L., Bertrand, N., Bacchi, V., 2022. Uncertainty quantification and global sensitivity analysis with dependent inputs parameters: Application to a basic 2D-hydraulic model. *LHB*. https://doi.org/10.1080/27678490.2021.2015265

Richet, Y., Bacchi, V., 2019. Inversion Algorithm for Civil Flood Defense Optimization: Application to Two-Dimensional Numerical Model of the Garonne River in France. *Frontiers in Environmental Science* 7, 160.

Roustant, O., Ginsbourger, D., Deville, Y., 2012. DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *Journal of Statistical Software* 51, 1–55.

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. *Computer physics communications* 145, 280–297.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global sensitivity analysis: the primer. *John Wiley & Sons.*

Vorogushyn, S., Apel, H., Merz, B., 2011. The impact of the uncertainty of dike breach development time on flood hazard. *Physics and Chemistry of the Earth,* Parts A/B/C 36, 319–323. https://doi.org/10.1016/j.pce.2011.01.005

Wahl, T.L., 2004. Uncertainty of predictions of embankment dam breach parameters. *Journal of hydraulic engineering* 130, 389–397.