



HAL
open science

Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions

Siham El Garroussi, Sophie Ricci, Matthias de Lozzo, Nicole Goutal, Didier Lucor

► To cite this version:

Siham El Garroussi, Sophie Ricci, Matthias de Lozzo, Nicole Goutal, Didier Lucor. Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions. 2020 TELEMAC-MASCARET User Conference, 2020, Antwerp (on line), Belgium. hal-03444227

HAL Id: hal-03444227

<https://hal.science/hal-03444227>

Submitted on 23 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HENRY

Hydraulic Engineering Repository

Ein Service der Bundesanstalt für Wasserbau

Conference Paper, Published Version

El Garroussi, Siham; Ricci, Sophie; De Lozzo, Matthias; Goutal, Nicole; Lucor, Didier

Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions

Zur Verfügung gestellt in Kooperation mit/Provided in Cooperation with:

TELEMAC-MASCARET Core Group

Verfügbar unter/Available at: <https://hdl.handle.net/20.500.11970/107438>

Vorgeschlagene Zitierweise/Suggested citation:

El Garroussi, Siham; Ricci, Sophie; De Lozzo, Matthias; Goutal, Nicole; Lucor, Didier (2020): Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions. In: Breugem, W. Alexander; Frederickx, Lesley; Koutrouveli, Theofano; Chu, Kai; Kulkarni, Rohit; Decrop, Boudewijn (Hg.): Online proceedings of the papers submitted to the 2020 TELEMAC-MASCARET User Conference October 2020. Antwerp: International Marine & Dredging Consultants (IMDC). S. 84-90.

Standardnutzungsbedingungen/Terms of Use:

Die Dokumente in HENRY stehen unter der Creative Commons Lizenz CC BY 4.0, sofern keine abweichenden Nutzungsbedingungen getroffen wurden. Damit ist sowohl die kommerzielle Nutzung als auch das Teilen, die Weiterbearbeitung und Speicherung erlaubt. Das Verwenden und das Bearbeiten stehen unter der Bedingung der Namensnennung. Im Einzelfall kann eine restriktivere Lizenz gelten; dann gelten abweichend von den obigen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Documents in HENRY are made available under the Creative Commons License CC BY 4.0, if no other license is applicable. Under CC BY 4.0 commercial use and sharing, remixing, transforming, and building upon the material of the work is permitted. In some cases a different, more restrictive license may apply; if applicable the terms of the restrictive license will be binding.

Verwertungsrechte: Alle Rechte vorbehalten

Assessing uncertainties in flood forecasts using a mixture of generalized polynomial chaos expansions

Siham El Garroussi¹, Sophie Ricci¹, Matthias De Lozzo², Nicole Goutal³, Didier Lucor⁴

¹CECI, CERFACS/CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1

²IRT Saint Exupéry, CS34436, 3 Rue Tarfaya, 31400 Toulouse

³EDF, LNHE, 6 quai Watier, 78400 Chatou

⁴LIMSI, Campus Universitaire bâtiment 507, Rue John Von Neumann, 91400 Orsay

siham.elgarroussi@cerfacs.fr

Abstract—Surrogate modelling based on generalized polynomial chaos expansion has emerged as a suitable alternative to standard Monte Carlo based methods that are accurate but computationally cumbersome. However, if there are non-linearities in the relationship between model inputs and model output, building a single generalized polynomial chaos expansion model leads to poor predictions. This paper investigates a Mixture-of-Experts approach based on machine-learning methods to divide the input space into subspaces that do not feature non-linearities. Then, generalized polynomial chaos expansions are built on each of these regions. This approach is applied to a reach of the Garonne River where the floodplain water height is non-linear with respect to the uncertain inputs (bottom roughness and upstream discharge), especially in locations where the topography features a strong gradient.

I. INTRODUCTION

Real-time forecasting is an important component of flood risk management, but it is subject to multiple uncertainties caused by model inputs, initial states, model structures, and model parameters [9]. Knowledge of the type and magnitude of these uncertainties is crucial to understand and interpret the model's results.

The key part of an uncertainty quantification (UQ) analysis is the propagation of the uncertainties through the simulation model [8]. Due to the high computational cost of two-dimensional hydrodynamic simulators, the direct use of methods based on Monte Carlo sampling is excluded. Surrogate models are thus used to overcome this issue [4] [16].

From the class of surrogate models, generalized polynomial chaos expansions (gPCE) has proven useful in a wide range of applications for emulating responses of computational models with random input, quantifying output uncertainty, and providing sensitivity indices with, in particular, an analytical formula for expectation, variance and Sobol' indices [3]. This surrogate model relies on a functional representation of output random variables as an expansion in terms of orthonormal basis functions and is built on an efficient space-filling sampling of the uncertain parameter space. However, the gPCE model tends to struggle when applied to problems related to unsteadiness, stochastic discontinuities, long-term integration, and large perturbation [25].

Different approaches with varying degrees of complexity have been proposed in the literature to address this kind of behavior. Examples include multi-resolution / multi-element polynomial chaos expansions [20] [25], regression trees [10] [17], multivariate adaptive regression splines [13], among others. They rely on the idea of partitioning the input parameter space into (often disjoint) subspaces, followed by the use of regression-based surrogates in each subspace with an intrusive approach. In this paper, we propose a non-intrusive Mixture-of-Experts (MoE) approach based on machine-learning tools to handle non-linearities in a gPCE-UQ workflow that stands on the “divide-and-conquer” principle. The general outlines of gPCE modeling are first introduced. This is followed by a description of the different blocks of the MoE. Finally, the Garonne River (Southwest France) hydrodynamic test case, where the floodplain water height is non-linear due to the combination of random model inputs (bottom roughness and upstream discharge) and properties of the terrain, is used to show the effectiveness of the proposed approach.

II. PROBLEM SPECIFICATION

A. Study area

The study area extends over a 50 km reach of the Garonne River (France) between Tonneins (upstream), downstream of the confluence with the River Lot, and La Réole (downstream). This part of the valley was equipped in the 19th century with infrastructure to protect the Garonne flood plain from flooding events. A system of longitudinal dykes and weirs was progressively constructed after the 1875-flood to protect the floodplains and organize submersion and flood retention areas.

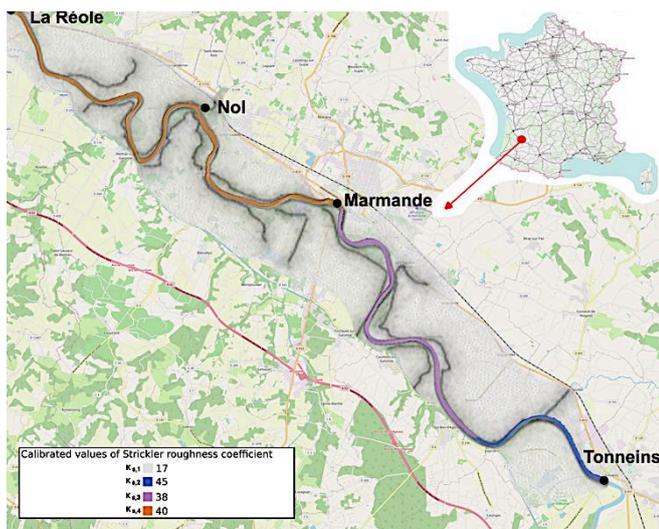


Figure 1: Zoning of the roughness coefficient over the study area.

B. Hydraulic modeling

The TELEMAC-2D (T2D) model, constituted by a triangular mesh of the study area of some 41 000 nodes with a refined mesh size near the dykes (see Fig. 1), has an upstream discharge imposed at Tonneins, and downstream, a stage-discharge rating curve corresponding to the stream gauge at La Réole. This hydraulic model has been realized by Besnard and Goutal (2008) [1]. The dynamic is solved by the 2D solver of the TELEMAC software [11] based on the resolution of the shallow water equations in the non-conservative form using the finite element method. The results of the simulation are water height and mean velocity on the vertical axis at each node in the mesh [14]. In this paper, a focus will be given to the water height in node 35067 (NoI) located on a ditch downstream of a dyke (see Fig. 1) in Sainte-Bazeille commune.

C. Model uncertainties characterization

In this study, we consider the effect of two sources of uncertainty on water height h for flood events, respectively Strickler roughness coefficient K_s which characterizes the

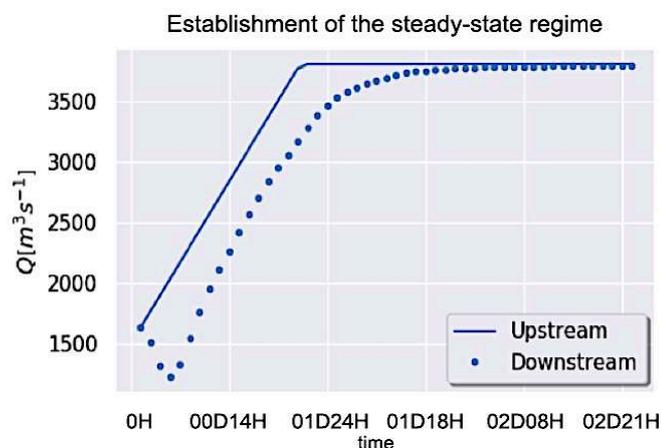


Figure 2: The uncertain input variable Q_{up} is a steady-state value reached through a ramp function.

roughness of the river bottom, and the upstream discharge Q_{up} resulting from the establishment of the steady-state regime, as shown in Fig. 2 for the node NoI. The ramp lasts one day and the constant hydraulic inflow two days at Q_{up} .

The Strickler roughness coefficient K_s is defined according to 4 areas, as shown in Fig. 1: grey for the floodplain ($K_{s,1}$), and blue, purple, and orange for the upstream ($K_{s,2}$), middle ($K_{s,3}$), and downstream ($K_{s,4}$) parts of the main channel respectively. Its distribution is assumed to be uniform, and its range is set to cover the calibration values. The upstream discharge is assumed to follow a Gaussian distribution centered around the biennial flood at Tonneins $3\,300\text{ m}^3\text{s}^{-1}$, of a standard deviation of $1\,100\text{ m}^3\text{s}^{-1}$. Moreover, to avoid too high or too low values, the probability density is truncated at $600\text{ m}^3\text{s}^{-1}$, corresponding to the annual mean discharge, and $6\,000\text{ m}^3\text{s}^{-1}$, corresponding to the vicennial flood at Tonneins. Tab. 1 summarises the distribution of uncertain model inputs.

TABLE 1: PROBABILITY DISTRIBUTION OF THE UNCERTAIN INPUT VARIABLES

Uncertain variable	Calibration values	Distribution
$K_{s,1} [\text{m}^{1/3}\text{s}^{-1}]$	17	$\mathcal{U}[5, 20]$
$K_{s,2} [\text{m}^{1/3}\text{s}^{-1}]$	45	$\mathcal{U}[40, 50]$
$K_{s,3} [\text{m}^{1/3}\text{s}^{-1}]$	38	$\mathcal{U}[33, 43]$
$K_{s,4} [\text{m}^{1/3}\text{s}^{-1}]$	40	$\mathcal{U}[35, 45]$
$Q_{up} [\text{m}^3\text{s}^{-1}]$	—	$\mathcal{N}(3\,300, 1\,100)$

D. Computing environment

CERFACS's cluster has been used to run T2D simulations. Simulating the river and the floodplain dynamics takes about 6 minutes on 15 cores, for the study case presented in Sect. II. A over three days. This cost is not practical in the context of the UQ framework requiring thousands of T2D simulations to estimate statistics. Hence the importance of replacing the numerical simulator with a surrogate model [18].

III. GENERALIZED POLYNOMIAL CHAOS EXPANSION-BASED UNCERTAINTY PROPAGATION

A. Generalized Polynomial Chaos Expansion

Let us consider a computational model $\mathcal{M}: x \in \mathcal{D}_x \subset \mathbb{R}^d \mapsto y = \mathcal{M}(x) \in \mathbb{R}$. Suppose that the uncertainty in the input parameters is modeled by a random vector X with prescribed joint probability density function (PDF) $f_X(x)$. The resulting (random) quantity of interest $Y = \mathcal{M}(X)$ is obtained by propagating the uncertainty in X through \mathcal{M} . Assuming that Y has a finite variance (which is a physically meaningful assumption when dealing with hydrodynamical systems), it belongs to the so-called Hilbert space of second-order random variables, which allows for the following spectral representation to hold [7]:

$$Y = \sum_{j=0}^{\infty} y_j Z_j. \quad (1)$$

The random variable Y is therefore cast as an infinite series, in which $\{Z_j\}_{j=0}^{\infty}$ are multivariate orthonormal polynomials in the input vector X , i.e., $Z_j = \Psi_j(X)$.

We assume that the input variables are statistically independent so that the joint PDF is the product of the d marginal distributions: $f_X(x) = \prod_{i=1}^d f_{X_i}(x_i)$, where the $f_{X_i}(x_i)$ are the marginal distributions of each variable $\{X_i, i = 1, \dots, d\}$ defined on \mathcal{D}_{X_i} . It can be proven that the set of all multivariate polynomials in the input random vector X forms a basis of the Hilbert space in which $Y = \mathcal{M}(X)$ is to be represented [7]:

$$Y = \sum_{\alpha \in \mathbb{N}^d} y_{\alpha} \Psi_{\alpha}(X), \quad (2)$$

where $\Psi_{\alpha} = \prod_{i=1}^d \phi_{\alpha_i}$ and $(\phi_{\alpha_i})_{i \geq 0}$ is an orthonormal basis for X_i .

For standard uncertain input distributions, the associated families of orthonormal polynomials are well known [5]. Given the set of these polynomial bases $(\Psi_{\alpha}(X))$, the next step is to compute the gPCE coefficients (y_{α}) . In this study, we focus on a particular non-intrusive approach based on the least-square analysis to compute the coefficients of gPCE from repeated evaluations of the existing model \mathcal{M} considered as a black-box function. By selecting an orthonormal basis with respect to the input parameter distributions, the corresponding coefficients can be given a straightforward interpretation: the first coefficient y_0 is the mean value of the model output, whereas the variance is the sum of the squares of the remaining coefficients [5]. Similarly, the Sobol' indices, commonly used for sensitivity analysis study, are obtained by summing up the squares of suitable coefficients [3].

A. Error metrics

In the present study, two standard metrics are used to measure the quality of the gPCE surrogate model: the Q_2 predictive coefficient and the Root Mean Squared Error (RMSE). The validation is carried out over an input-output validation database D_v of size N_v .

At the k^{th} mesh node, the Q_2 predictive coefficient reads:

$$Q_2 = 1 - \frac{\sum_{i=1}^{N_v} (h_k^{(i)} - \hat{h}_k^{(i)})^2}{\sum_{i=1}^{N_v} (h_k^{(i)} - \bar{h}^{(i)})^2},$$

Where $\bar{h}^{(i)} = \frac{1}{N_v} \sum_{k=1}^{N_v} h_k^{(i)}$.

B. Hydrodynamic uncertainty propagation using gPCE

[19] showed that, in a two-dimensional hydrodynamic steady regime, the gPCE model drastically reduces the number of runs needed for propagating uncertainty and could be applied to more complex studies. Moreover, [23] highlights that considering the large dimension of the water height, combining the surrogate model with a space reduction method allows good learning at a reduced computational cost.

In that respect, a reduced gPCE is used here to replace the T2D model to propagate uncertainty at a reduced cost for the transitional flow regime. Learning and validation databases of size 1 000 and 500 respectively are considered. The inputs are sampled according to an optimized LHS [26] following their

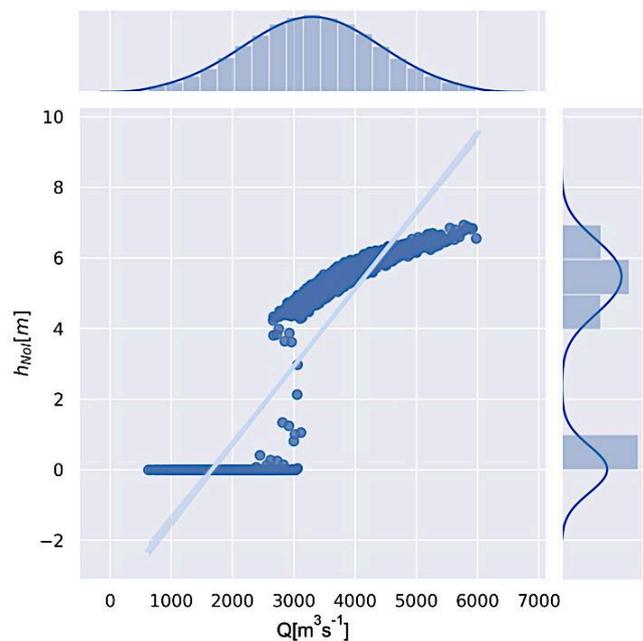


Figure 3: The PDF and the response of the water height in the node NoI to the upstream discharge.

PDF defined in Tab. 1. The gPCE surrogate is computed for each time step of the integration from time $t_{\text{initial}} = 0$ second corresponding to the time of injection of the flow discharge upstream of the study area, to time $t_{\text{final}} = 3$ days corresponding to the setting up of the steady-state regime.

While the predictive coefficient Q_2 , evaluated on the validation database considering all mesh nodes for t_{final} is equal to 1, it is significantly smaller than 1 during the transition phase before setting up the steady-state regime.

In order to understand why the gPCE surrogate model does not correctly predict the water height h simulated with the T2D during the transient phase, a time of this latter has been chosen $t_{\text{intermediate}} = 1$ day 2 hours, and only one mesh node has been selected which is the NoI node. This is the configuration where the worst RMSE value has been recorded. The gPCE degree P was varied between 0 and 24. The optimal P in this case is 4 allowing to have a Q_2 of 0.54, and an RMSE of 0.92 m, which is statistically not satisfactory.

Furthermore, as shown in Fig. 3, the PDF of h at node NoI is bimodal and the response function according to Q_{up} is non-linear. Indeed, as the node NoI is on a ditch, upstream discharge values that are lower than $3\,000\text{ m}^3\text{s}^{-1}$ lead to almost zero water depth. And upstream discharge values higher than $3\,000\text{ m}^3\text{s}^{-1}$ lead to water depth values higher than 4 m. Thus, the gPCE surrogate model is not a good choice to approach non-linear functions because of its inherent smoothness.

Advanced strategies should then be applied. In this study, a Mixture-of-Experts (MoE) approach is used in order to improve gPCE performance. This approach is based on machine-learning methods allowing to decompose the random inputs space into subspaces over which the solution varies smoothly and consequently build a global representation as a collection of smooth representations defined over subspaces.

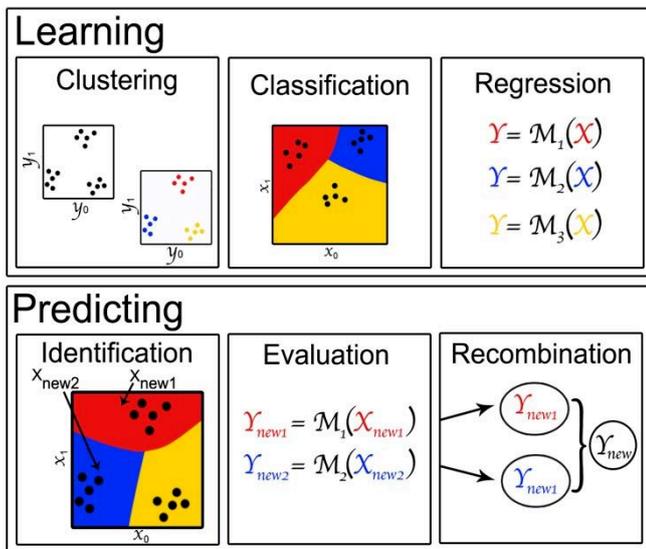


Figure 4: Illustration of the Mixture-of-Experts approach.

IV. MIXTURE OF GPCE EXPERTS FOR UQ

A. Workflow of the MoE

The proposed approach for handling non-smooth functions consists of the following steps, as illustrated in Fig. 4:

- Learning:

- 1) *Clustering*: this is the first step of the approach when the analyst attributes to each output observation $y^{(i)}, i = \{1, \dots, N\}$ a class that corresponds to an identified behavior of the system. In the ideal case, this can be done manually using expert knowledge.

In the general case, it is more convenient to rely on an automated approach where the K classes are directly learned from the data using unsupervised learning algorithms [6] as K-Means [2] and DBSCAN [15].

- 2) *Classification*: once the classes are identified, they are mapped to the input space. This latter is then split according to the labels resulting from the clustering of the output space. This can be done via the SVM algorithm [12] [24].
- 3) *Regression*: the dataset is split into the different groups identified in the previous two steps. For each group, a local gPCE $\{\hat{\mathcal{M}}_k, k = 1, \dots, K\}$ is built.

- Predicting:

Once the local models are built, it is necessary to recombine them when evaluating a new point from the input space. This is achieved in three steps:

- 1) *Identification*: this involves identifying the class to which belongs the new point. The previously built classifier can be used in that respect.
- 2) *Evaluation*: the new point is then projected in the output space using the appropriate gPCE model.

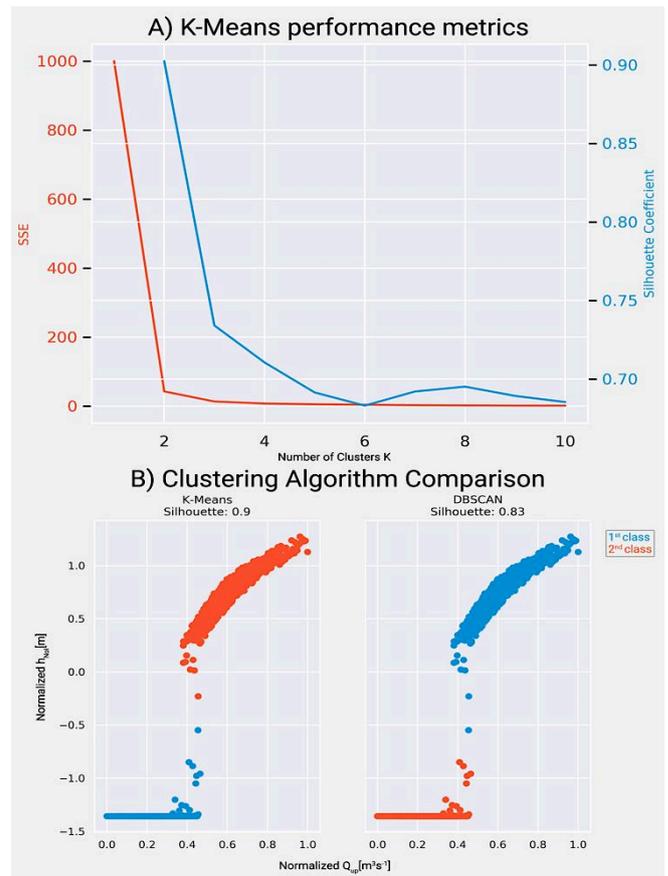


Figure 5: A) The top curve represents the evolution of the two metrics: the sum of squared errors in red, and the silhouette coefficient in blue for the K-Means clustering method. B) The result of the clustering of the water height according to the upstream discharge, left: K-Means; right: DBSCAN. In blue, the first class detected by the algorithm and in red the second class.

- 3) *Recombination*: the final approximation is obtained by combining the different predictions as follows:

$$\hat{\mathcal{M}}(x) = \sum_{k=1}^K w_k(x) \hat{\mathcal{M}}_k(x), \quad (3)$$

$w_k(x)$ are weight functions defined such that $\sum_{k=1}^K w_k(x) = 1$. There are two main types of weight functions: binary approach and weighting approach. In the present study, only the first approach was investigated.

B. Validation of the MoE on the hydraulic test case

For the K-Means clustering algorithm, data are split into K clusters. Two methods are commonly used to evaluate the appropriate number K : The elbow method [22] and the silhouette coefficient [21].

For the elbow method, the x -value of the elbow point is thought to be a reasonable trade-off between error and the number of clusters. The red curve in Fig. 5. A. shows the evolution of the sum of squared errors (SSE) of classification according to the number of clusters K , and it indicates that the elbow point in the SSE curve is located at $K = 2$.

The silhouette coefficient is a measure of cluster cohesion and separation. It quantifies how well a data point fits into its assigned cluster based on two factors:

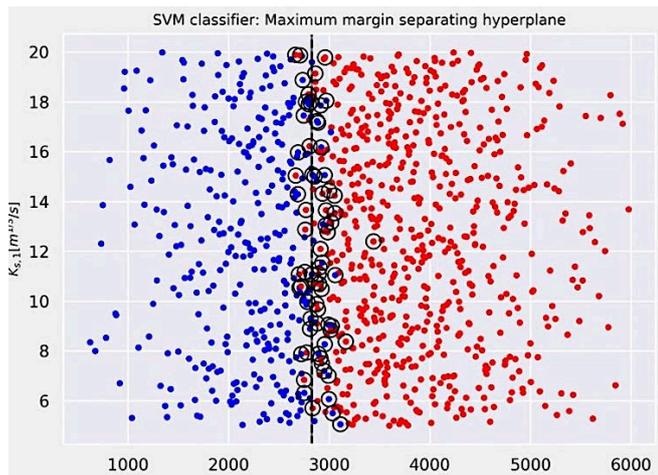


Figure 6: SVM-based classification.

- How close the data point is to other points in the cluster,
- How far away the data point is far from points in other clusters.

A high value of this coefficient indicates that samples are closer to their clusters than they are to other clusters. In Fig. 5. A., the blue curve represents the silhouette coefficient and shows that the best choice for K is 2 since it has the maximum score. Thus, the best number of classes for the K-Means based on the results of those two metrics is $K = 2$. The resulting clustering is shown in Fig. 5. B as well as the clustering of the DBSCAN algorithm. Although the two algorithms do not process in the same way, their results are similar, with more cluster refinement for K-Means (silhouette = 0.9). Therefore, the clustering algorithm retained is K-Means with $K = 2$.

Once these 2 clusters are defined according to the K-Means clustering labels, these latter are mapped to the input space as shown in Fig. 6. The input space is then divided into two classes, blue and red. The separation hyperplane is mainly dependent on the upstream discharge injected at Tonneins and corresponds to a value of $Q = 2700 \text{ m}^3/\text{s}$. This value is physically significant; it matches the bank-full discharge in the main channel. Thus, a possible interpretation of the hyperplane plotted in a black line in Fig. 6, is the stage at which the main channel and floodplain become connected.

The next step is the construction of an appropriate gPCE surrogate model in each class. To have a good gPCE model, it is necessary to have an optimal degree P, and thus to loop to get the best P. For that purpose, the polynomial degree P was varied between 1 and 9, and the optimal degree selected is the one that gives an optimal RMSE and Q_2 calculated on the validation database. The proposed approach improves the prediction, as shown in Tab. 2.

TABLE 2: COMPARISON OF THE RESULTING ERRORS

Surrogate models' performance	gPCE	MoE-gPCE
RMSE [m]	0.92	0.02
Q_2	0.54	0.98

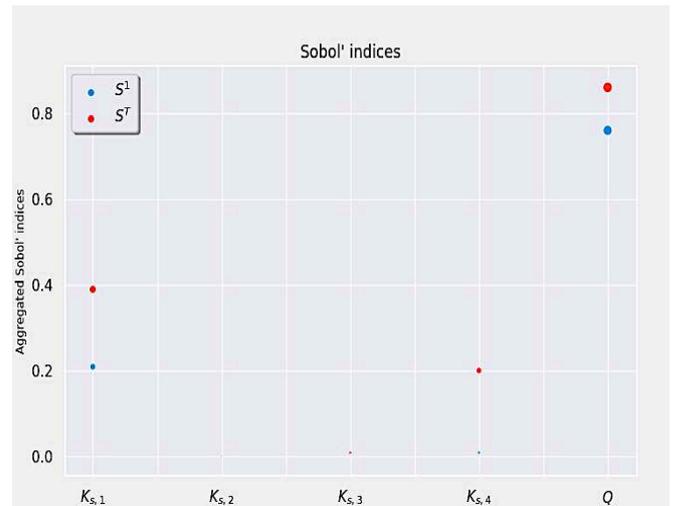


Figure 7: Sobol' indices for water height at node NoI.

Indeed, the RMSE of the MoE-gPCE model is 2 cm, contrary to the RMSE of the general model based on a single gPCE surrogate model, which is approximately equal to 1 m, and the Q_2 of the MoE-gPCE is 0.98 whereas it is 0.54 for the classic gPCE.

The computation of spatially aggregated first-order Sobol' indices for sensitivity analysis from the gPCE coefficients shows that the water height at the node NoI is mainly governed by the upstream discharge injected at Tonneins (76%) and is weakly governed by the roughness at the bottom of this node (21%) as illustrated in Fig. 7. The total order Sobol' indices highlight the interaction of the variables $K_{s,1}$ and Q with the other variables. Indeed, the total order Sobol' indices of $K_{s,1}$ is 39%, which means a contribution of the interaction of this variable with the other variables on the variability of the water height at this node.

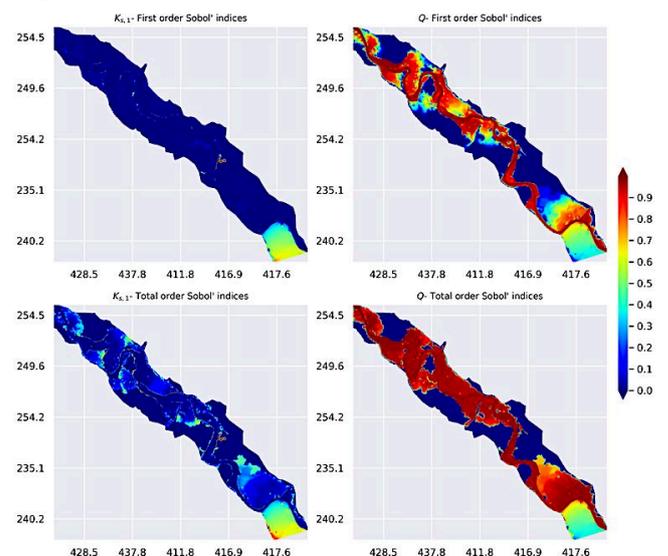


Figure 8: Spatial aggregated Sobol' indices, first at the top and total in the bottom, for the water height discretized over the mesh following the uncertain inputs: $K_{s,1}$ on the left, and Q on the right.

The MoE-gPCE proved to be robust for the NoI node where the water height is non-linear, so it was applied for all other points where the *RMSE* is greater than 0.3 m by applying a single reduced gPCE. This modeling over the whole mesh allowed to have the different statistics over the study area, in particular the mean, the upper quartile, the variance, and Sobol' indices displayed for the uncertain variables $K_{s,1}$ and Q in Fig. 8. This figure shows that, for the transitional regime studied here, the variance of the water height over the study area is mostly explained (84% on average) by the upstream discharge Q . Moreover, the water height variability in the floodplain near the liquid boundary condition at the upstream part of the numerical model is shared between the influence of the floodplain roughness coefficient $K_{s,1}$ and the upstream discharge Q , which is consistent with the numerical artefact of the T2D modeling mentioned in [1] and related to the neighbourhood of the upstream liquid boundary. The large areas next to the edges of the study area show no influence either from the upstream discharge Q either from the floodplain roughness coefficient because they are not wet at the time $t_{intermediate}$. Otherwise, the roughness coefficients of the bottom of the main channel $K_{s,2}$, $K_{s,3}$ and $K_{s,4}$ have almost no influence on the variability of the water depth over the study area.

Total order Sobol' indices highlight the multivariate effects of the different input variables. Note that the results of the sensitivity analysis are dependent on the distribution of the uncertain input variables and the degree of truncation of polynomial expansion.

V. CONCLUSION AND PERSPECTIVES

This paper presents the Mixture of generalized polynomial chaos expansion experts approach (MoE-gPCE) allowing to deal with local non-linearities and to take advantage of the benefits of generalized polynomial chaos expansion (gPCE) model. Focus is given to the particular hydrodynamic case when multiple behaviors of the water height can be observed. The proposed approach consists of first identifying such behaviors and then classifying them using a support vector machine algorithm. The resulting prediction is obtained by building local gPCE in each subspace and then recombining them using a binary scheme. When applied to the water height at a node downstream of a dyke (NoI), this approach has led to the emergence of two classes in which the water height behaves differently depending on whether the upstream discharge is less or more the bank-full discharge in the main channel. The different statistical moments as well as the Sobol' indices were computed via MoE-gPCE. The proposed approach has been extended to all points when a single gPCE gives poor predictions.

However, the accuracy of the resulting predictions relies on the accuracy of the classification step. This step can be improved using adaptive sampling to better define the boundary between the two subspaces. Furthermore, improving the performance of gPCE for all mesh nodes using the proposed approach is equivalent to construct 72 000 gPCE (3 000 nodes where a single gPCE is not accurate, 8 values of the gPCE degree P , 3 potential number of classes) which is

relatively expensive, hence the need to find a method allowing the execution of MoE-gPCE at a reduced cost in the context of high output dimension.

ACKNOWLEDGEMENT

We would like to thank EDF R&D's PRISME department for their support in using OpenTURNS. We thank all these people for the enriching discussions we were able to have on uncertainty methods for two-dimensional river hydraulics issues.

REFERENCES

- [1] A. Besnard, N. Goutal, "Comparison between 1D and 2D models for hydraulic modeling of a floodplain: case of Garonne River", *Houille blanche, Revue internationale de l'eau*, 2011, pp.42-47.
- [2] A. Likas, N. Vlassis and J. Verbeek, "The global k-means clustering algorithm", *Pattern Recognition*, 36, 2003, pp. 451-461.
- [3] B. Sudret, "Global sensitivity analysis using polynomial chaos expansion". In P. Spanos and G. Deodatis (ed), *Proc. 5th Int. Conf. on Comp. Stoch. Mech. (CSM5)*, 2006.
- [4] B. Sudret, S. Marelli and J. Wiart, "Surrogate models for uncertainty quantification: An overview", 11th European Conference on Antennas and Propagation (EUCAP), Paris, 2017, pp.793-797.
- [5] B. Sudret, "Uncertainty propagation and sensitivity analysis in mechanical models", *Contributions to structural reliability and stochastic spectral methods*, Tech. rep., Université Blaise Pascal, Clermont-Ferrand, 2007, France.
- [6] C. M. Bishop, "Pattern recognition and machine learning", Springer, 2016.
- [7] C. Soize, R. Ghanem, "Physical systems with random uncertainties: chaos representations with arbitrary probability measure", *SIAM J. Sci. Comput.*, 2004, 26(2), pp.395-410.
- [8] E. de Rocquigny, et al., "Uncertainty in industrial practice: A guide quantitative uncertainty management", January 2008.
- [9] E. M. Mockler, K. P. Chun, G. Sapriza-Azuri, M. Bruen Gonzalo and H. S. Weather, "Assessing the relative importance of parameter and forcing uncertainty and their interactions in conceptual hydrological model simulations," *Adv. in Water Resources*, 2016.
- [10] H. A. Chipman, E. I. George and R.E. McCulloch, "BART: Bayesian Additive Regression Trees", *Annals of Applied Statistics*, 2010, 4(1), pp.266-298.
- [11] J. C. Galland, N. Goutal, J. M. Hervouet, "Telemac: A new numerical model for solving shallow water equations", *Advances in Water Resources*, 1991, 14(3), pp.138-148.
- [12] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods", 1999.
- [13] J. H. Friedman, "Multivariate adaptive regression splines", *Annals of Statistics*, 1991, 19(1), pp.1-67.
- [14] J. M. Hervouet and P. Bates, "The TELEMAT Modelling System", 2000, Wiley.
- [15] K. Khan et al., "DBSCAN: Past, present and future", *ICADIWT*, 2014.
- [16] K. Sargsyan, "Surrogate models for uncertainty propagation and sensitivity analysis". In: R. Ghanem, D. Higdon, H. Owhadi (ed) *Handbook of uncertainty quantification*. Springer, Cham, 2015.
- [17] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, "Classification and regression trees", 1984, CRC press.
- [18] M. De Lozzo, "Surrogate modeling and multifidelity approach in computer experimentation", *Journal de la Société Française de Statistique*, 2015, 156(3).
- [19] N. Goutal, et al., "Uncertainty quantification for river flow simulation applied to a real test case: The Garonne valley", *SimHydro*, 2017.
- [20] O. L. Maitre, H. Najm, R. Ghanem, and O. Knio, "Multi-resolution analysis of Wiener-type uncertainty propagation schemes", *Journal of Computational Physics*, 2004, 197(2), pp.502-531.

- [21] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis", *Journal of computational and applied mathematics*, vol. 20, 1987, pp.53-65.
- [22] R. I. Thorndike, *Psychometrika*, vol. 18, 1953, pp.266-267.
- [23] S. El Garroussi, M. De Lozzo, S. Ricci, D. Lucor, N. Goutal, C. Goeury, "Uncertainty quantification in a two-dimensional river hydraulic model", UNCECOMP 2019, 3rd ECOCOMAS Thematic Conference on uncertainty quantification in computational sciences and engineering. Crete, Greece, 24-26 June 2019.
- [24] V. N. Vapnick, "The nature of statistical learning theory", Springer-Verlag, 1995.
- [25] X. Wan and G. E. Karniadakis, "Multi-element generalized polynomial chaos for arbitrary probability measures", *SIAM J. Sci. Comput.*, 2006, 28(3), pp.901-928.
- [26] G. Damblin, M. Couplet and B. Iooss, "Numerical studies of space filling designs: optimization algorithms and subprojection properties", *Journal of Simulation*, 2013, 7, pp.276-289.