Tracking the critical offshore conditions leading to inundation via active learning of full-process based models

Jeremy Rohmer, Déborah Idier, Thomas Bulteau, François Paris

Tracking the critical offshore conditions leading to inundation via active learning of full-process based models

Rohmer Jeremy[1,a], Deborah Idier[1], Thomas Bulteau[2], and François Paris[1]

[1]*BRGM, 3, av. Claude Guillemin, BP 36009, 45060 Orleans Cedex 2, France*
[2]*BRGM, Parc Technologique Europarc, 24, avenue Leonard de Vinci, 33600 Pessac, France*

Abstract. High-fidelity numerical models (full process based) facilitate accurate simulations of storm responses. For efficient implementation in probabilistic risk assessment, the simulation of a large number (>10,000s) of combinations of offshore hydrodynamic conditions (e.g. wave characteristics, offshore water level, etc.) is often necessary. To optimise this procedure, it can be of interest to concentrate the computation effort by only identifying the critical set of offshore conditions that lead to inundation on key assets for the studied territory (e.g., evacuation routes, hospitals, etc.). However, two limitations exist: 1. full-process based models have large computation time cost, typically of several hours, which often prevent from conducting several simulation scenarios; 2. the full-process based models are expected to present non-linearities (non-regularities) or shocks (discontinuities). In this study, we propose a strategy combining meta-modelling (of type Support Vector Machine) and active learning techniques to track with a limited number of long-running simulations the critical set's boundary. The developments are done on a cross-shore case, using the process-based SWASH model (computational time of 10 hours for one run). The dynamic forcing conditions are parametrized by storm surge *S* and significant wave height *Hs*. We validated the approach with respect to a reference set of 400 long-running simulations in the domain of (*S* ; *Hs*). Our tests showed that the tracking of the critical contour can be achieved with a reasonable number of long-running simulations of a few tens.

1 Introduction

Some recent storm events like Katrina in 2005 or Xynthia in 2010 (see e.g. [1]) illustrate the present-day coastal damages and injuries that can affect the coastal area, both in cyclonic and non-cyclonic environment. Katrina was one of the six most powerful hurricanes ever registered in the Atlantic, leading to 1836 deaths and damages of about 80 billion USD, whereas Xynthia was a mid-latitude storm that severely hit low-lying coasts located in the central part of the Bay of Biscay on the 27–28 February 2010, inducing 53 deaths and material damages estimated at more than one billion euros. From a statistical point of view, the wave heights generated during the Xynthia event could not be considered as extremes, but what makes this event "rare" is the combination of a high spring tide with a large storm surge (enhanced by young wind waves) reaching its maximum around the tide peak.

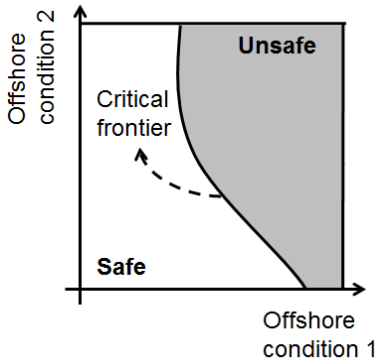

Figure 1. Schematic representation of the critical frontier in the offshore conditions domain. This frontier represents the boundary between the "safe" and "unsafe" regions (grey-coloured area), i.e. the boundary of the set of offshore forcing conditions, which lead to an inundation at the coast. Adapted from [3]

As discussed by Idier et al. [2], it can be of high interest to identify the combination of all critical set of offshore conditions that lead to inundation on key assets for the studied territory (e.g., assembly points, evacuation routes, hospitals, etc.), i.e. to track the critical frontier Γ_C, which separates the "safe" region from the "unsafe" one in the offshore conditions domain (e.g. combination of wave, tide and surge conditions): this is schematically

[a] Corresponding author: j.rohmer@brgm.fr

depicted in Fig. 1 in 2D. It should be underlined that the number of offshore conditions could be larger (>10), depending on considered physical parameters (tide, atmospheric storm surge, wave height, wave direction, wave period) and on the way time and/or spatial offshore conditions are described. Tracking such a critical frontier is the key element of an inverse methodology of coastal risk assessment.

The main idea of such an inverse risk method is the inversion of the usual risk assessment steps [2]: starting from the maximum acceptable hazard level (defined by stakeholders as the one leading to the maximum tolerable consequences) to finally obtain the return period of this threshold. Such an "inverse" approach would allow the identification of all the offshore forcing conditions (and their occurrence probability) inducing a threat for critical assets of the territory. The benefits are multiple, whether for: (1) estimating the probability of exceeding the maximum tolerable inundation height for identified critical assets, or for (2) providing critical offshore conditions for flooding in early warning systems, and for (3) raising awareness of stakeholders and eventually enhance preparedness for future flooding events by allowing them to assess risk to their territory.

However, the practicability of such an inverse risk method can be hindered, because tracking the boundary of the set of critical conditions (critical frontier) should rely on an accurate description and modelling of coastal processes: this requires the use of full-process based models for coastal flooding simulations, which might have very large computational time cost (typically of several hours per simulation). Such a computation burden often limits the analysis to a few scenarios, hence might prevent the estimation of the critical frontier. Recently, it has been shown that meta-modelling approaches can efficiently handle this difficulty (e.g., [3]). The basic idea is to replace the long-running code by an approximation constructed using only a limited number of different simulation scenarios. Rohmer & Idier [3] further extended this approach by combining the meta-model with an adaptive sampling procedure aiming at improving the local accuracy in the regions of the offshore conditions that contribute the most to the estimate of the targeted frontier.

Though the afore-described strategy proved to be very efficient (achieving a reduction by a factor 20-40 of the total number of necessary long-running simulations in the application case of [3]), it still faces a strong limitation related to the nature of the full-process based models: they are expected to present strong non-linearities (non-regularities) or shocks (discontinuities), i.e. dynamics controlled by thresholds. For instance, in case of coastal defence, the dynamics of the waterline position is characterized first by a linear behaviour (increase with increasing offshore conditions), as long as there is no overtopping, and then by a very strong increase (as soon as the offshore conditions are energetic enough to lead to wave overtopping, and then overflow). Such behaviour might make the training phase of the meta-model very tedious (see e.g., [4]).

In the present study, we propose to rely on advanced machine learning techniques to overcome the afore-described challenge related to non-regularity. We focus on the Support Vector Machines SVM (e.g., [5]). A key aspect is to optimise as much as possible the number of required simulations for estimating the critical frontier: this can be performed by relying on active learning techniques, i.e. on statistical techniques to guide and select the simulations to be run in order to predict with high accuracy the boundary of the set of critical conditions. The objective of this communication is to show the feasibility of such an approach.

In a first section, we describe the case study, which motivated the present work. In a second section, we further describe the statistical methods used to track the critical set of offshore conditions. In section 4, we apply them and discuss the results.

2 Case study

2.1 Description

The application case relies on a cross-shore configuration, as depicted in Fig. 2, and using the process-based SWASH model [6]. It should be seen as a test case for the exploration of feasibility of the proposed method to track the offshore critical conditions. It is adapted from the wave experiment by Boers [7] by adding a basin in its rightmost part: we define the critical conditions as the ones leading to flooding. The dynamic (time-varying) forcing conditions **x** are parametrised by seven factors (storm surge magnitude, significant wave height, dephasing between tide and surge, etc.) and their values are scaled with respect to the geometry of the laboratory experiment through Froude similitude and dimensional analysis (Table 1), considering a semi-diurnal tide. In the following, all results are given at the scale of the wave experiment.

Offshore Parameters	Lower bound	Upper bound
Tide magnitude A2 (m)	0.000	0.025
Storm surge magnitude S (m)	0.000	0.025
Duration of the storm surge increase TSM2 (s)	2278	9115
Duration of the storm surge decrease TST2 (s)	910	4559
Duration of the storm surge plateau TS2 (s)	910	4550
Significant wave height H_S (m)	0.000	0.375
Wave Period T_p (s)	1.415	3.400

Table 1. Parameters describing the time-varying offshore conditions.

A model run, for a given set of offshore conditions' values, takes approximately 10 hours to simulate two tide cycles. It should be noted that it includes a uniform depth channel computations to provide instantaneous water level conditions physically consistent at the offshore boundary.

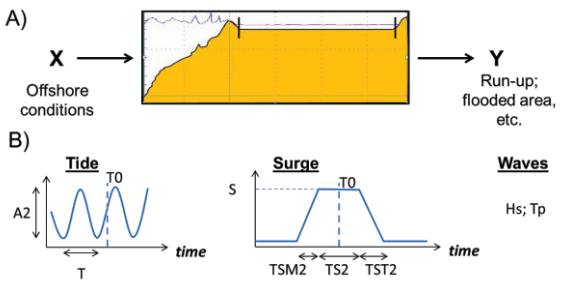

Figure 2. A) Geometry of the cross-shore model used in the study; B) Parameters used to represent the time-varying offshore conditions *X* (See table 1 for details).

2.2 Extracting the critical set using a direct approach

The purpose of the present study is to explore the feasibility of extracting the critical contour using a very limited number of simulations (a few tens). In this view, we restricted the analysis to two offshore conditions namely the significant wave height H_S and storm surge magnitude S and solved the problem using a direct approach, i.e. by relying on grid computing architecture (e.g. [8]): a total of 400 different simulations were run to estimate the run-up R by respectively varying the maximal storm surge S between 0.01 and 0.025m (dimensions with respect to Boers's experiment) and H_S between 0.000 and 0.375m while keeping other parameters constant at their median value. On this basis, the critical contour was extracted as depicted in Figure 3. This result represents our reference case to which the result using the meta-modeling approach will be compared (Section 4).

As shown in Figure 3, the relationship between the run-up R and both offshore conditions is highly non-linear: a region of the $S\times H_S$ domain does not lead to flooding, i.e. with R close to its minimum value (around 0) as highlighted by the blue-coloured region in the left bottom hand corner of Fig. 3. When both offshore conditions exceed given thresholds (which are unknown before actually running the code), overtopping occurs and the basin rapidly fills with sea water so that R steeply increases from ~0 to ~90% over a narrow region of (S; H_S) and then gently increases from ~90% to 100 % over a large region of the $S\times H_S$ domain (red coloured region in the right hand corner in Fig. 3). The objective is to estimate the black contour as depicted in Fig. 3B but with a low simulation budget.

3 Methods

Due to the highly nonlinear behaviour of the relationship between R and the offshore conditions, standard meta-modeling techniques might fail (Gaussian process, polynomial chaos expansion, etc., see e.g., [4]). A possible option is to solve the problem from a classification perspective, i.e. to identify the boundary between the offshore conditions leading to flooding (class "unsafe") and the ones which do not (class "safe"). A powerful technique is Support Vector Machine (e.g., [9]), which has the ability to explicitly define multidimensional and complex boundaries that optimally separate two classes of data. In the following, we first introduce the basic principles of this classification technique and then describe a strategy to improve its accuracy for boundary approximation.

3.1. Support Vector Machine

Consider a d-dimensional space sampled with n training points $\mathbf{x}_i$. Each point is associated with a class, either y_i=-1 (unsafe) or 1 (safe). The SVM algorithm finds the boundary that optimally separates both classes. The equation of the corresponding boundary holds as follows:

$$\Phi(x) = b + \sum_{i=1}^{n} \lambda_i y_i K(x_i, x) \quad (1)$$

where b is a constant, λ_i are Lagrange multipliers obtained from the quadratic programming optimization problem used to construct the SVM. The kernel K can have different forms such as polynomial, Gaussian radial basis function, etc. Equation (1) allows classifying any offshore conditions $\mathbf{x}$ via the sign of Φ.

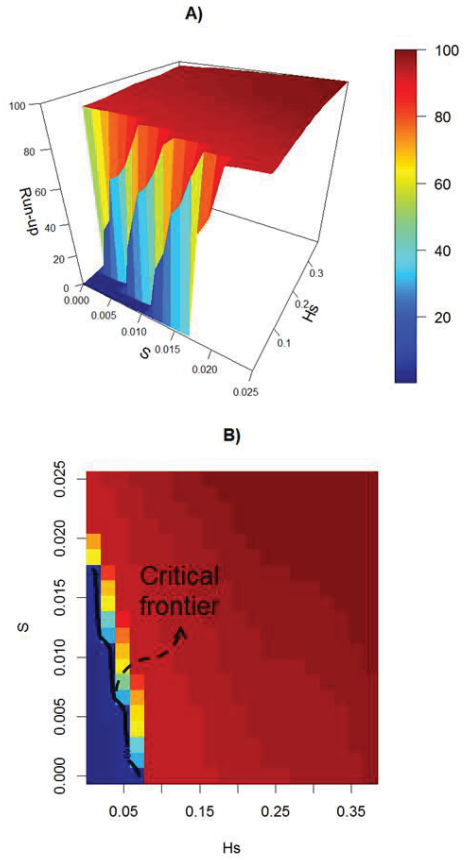

Figure 3. Run-up (in %, normalised between 0 and 1) calculated in the $S\times H_S$ domain discretised into a 40×40 grid.

3.2 Active Learning

This approach, *aka* semi-supervised machine learning, aims at sampling the training points by fulfilling two requirements:

1) the sampled training points should contribute to the accuracy improvement of the critical boundary estimate. Thus, the training point should lie where the probability of misclassification is the largest, i.e. in the vicinity of the boundary between both classes (safe and unsafe);
2) since the estimate of the boundary relies on the approximation quality of the SVM meta-model, the training points should also be selected in sparsely populated regions of the space.

For this purpose, we relied on the strategy proposed by Basudhar and Missoum [10] in the domain of structural reliability. First two samples are selected in turn by maximizing the distance to the closest training sample while lying on the SVM boundary. Then, a third sample is positioned in a region where data from one class is sparse in the vicinity of the boundary.

The whole procedure described in Table 2 is applied on the test case described in Sect. 2.2.

Step	Description
1.	Randomly sample a given set of training data $\mathbf{x}_i$ in the domain of offshore conditions (i=1 to *n*) and for each of them, estimate the Run-up by means of the long-running model;
2.	Based on these training data, construct an initial SVM meta-model;
3.	Iteratively select new training data based on the Basudhar and Missoum's algorithm
4.	Stop when the parameters (polynomial coefficients) of the SVM boundary have converged

Table 2. Main steps for approximating the critical frontier by combining active learning and SVM meta-modeling.

4 Application

In this section, we aim at assessing whether it is possible to estimate the critical frontier using a limited number of long-running simulations. The result is compared to the solution of the grid-based approach (Figure 3). We followed the different steps of Table 2:

- Step 1: An initial set of 15 training data was randomly selected using the centroidal Voronoi tessellation method (see e.g., [11]);

- Step 2: On this basis, 15 long-running simulations were conducted and the results were classified as "safe" when the normalised run-up was close to zero and unsafe otherwise. In total 14 of the simulations lead to a non-zero run-up;

- Step 3: A SVM of type linear was constructed using the 15 training data;

- Step 4: the active learning algorithm was launched and 6 new training data were iteratively selected until the SVM boundary's linear regression coefficient has converged, i.e. until the changes between the previous and the actual linear regression coefficient is no more than $1e^{-6}$.

Figure 4 illustrates the different iteration steps of the active learning strategy: the new training data are primarily selected close to the SVM boundary and, after each selection, the SVM boundary is updated and finally approximates with very satisfactory level the true critical frontier. The total number of necessary long running simulations reached only 15+6=21, namely ~5% of the simulation budget of the direct grid-based approach (of 400 simulations).

5 Concluding remarks and further work

In this study, we aimed at estimating with a low simulation budget (a few tens), the frontier, which separates the region of the offshore conditions which lead to flooding and those which do not. In order to set-up a solution to validate our method, we first solved the problem by running a large number of simulations for different offshore conditions. A strategy to identify the critical frontier with a limited number of simulation was here proposed and relied on the combination of SVM meta-modeling techniques and active learning: the comparison between both results (large number of simulations vs meta-modeling and limited number of simulations) showed that the second "SVM-active learning" strategy is very efficient and can approximate the critical frontier using only a few tens of long-running simulations.

These results are very promising: the expected results are the use of such procedure within a full flood risk assessment scheme in place of low complexity models as done for instance by Rueda et al. [12]. Yet, a necessary future step is testing the performance of the procedure by integrating a larger number of offshore conditions (for instance the ones described in Table 1).

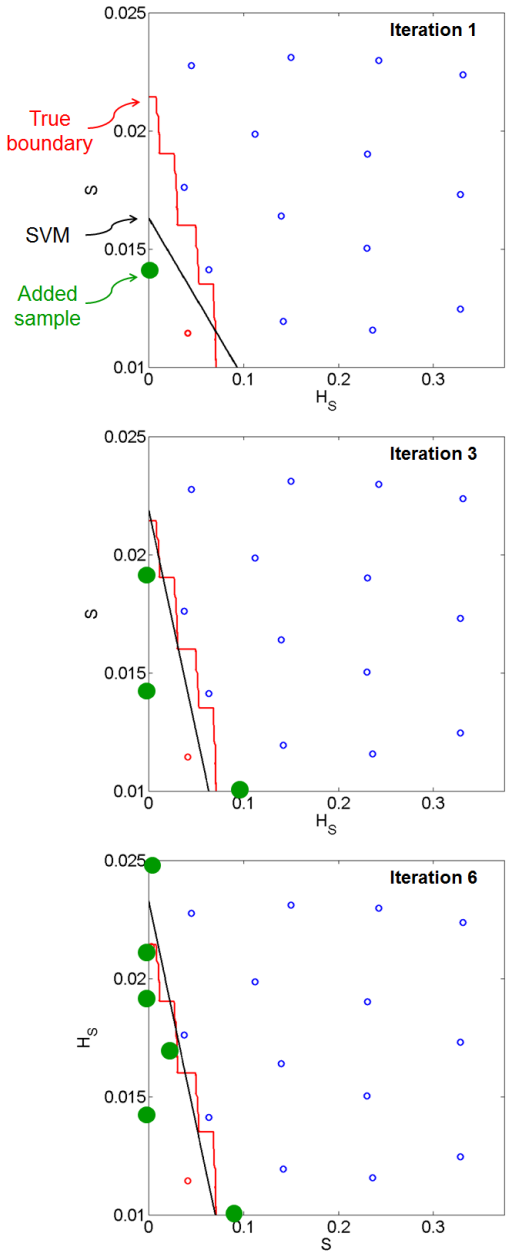

Figure 4. Iteration 1, 3 and 6 of the active learning for SVM approximation of the boundary. The dots (blue and red) are the initial training data.

6 Acknowledgements

The authors thank the BRGM-funded PROBSUB project, for contributions to the financial support of the present work. The authors are grateful to the SWASH development team (http://swash.sourceforge.net/) and more specifically to Zijlema M. (TU Delft).

7 References

1. Bertin X., Bruneau N., Breilh J. F., Fortunato A. and Karpytchev, M. (2012). Importance of wave age and resonance in storm surges: the case of Xynthia, *Ocean Model.*, **42**, 16–30.
2. Idier D., Rohmer J., Bulteau T. and Delvallée E. (2013). Development of an inverse method for coastal risk management, *Natural Hazards and Earth System Sciences*, **13(4)**, 999-1013.
3. Rohmer J. and Idier D. (2012). A meta-modelling strategy to identify the critical offshore conditions for coastal flooding, *Nat. Hazards Earth Syst. Sci.*, **12**, 2943-2955, doi:10.5194/nhess-12-2943-2012.
4. Jin R., Chen W. and Simpson T. W. (2001). Comparative studies of meta-modelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, **23(1)**, 1-13.
5. Rajasekaran S., Gayathri S. and Lee T. L. (2008). Support vector regression methodology for storm surge predictions, *Ocean Engineering*, **35(16)**, 1578-1587.
6. Zijlema M., Stelling G. and Smit P. (2011). SWASH: An operational public domain code for simulating wave fields and rapidly varied flows in coastal waters, *Coastal Engineering*, **58(10)**, 992-1012.
7. Boers M. (1996). Simulation of a surf zone with a barred beach. Report 1: Wave heights and wave breaking. *Comm. on Hydraulic and Geo. Eng.*, No. **96-05**, Delft Univ. of Technology.
8. Boulahya, F., Dubus, I. G., Dupros, F., and Lombard, P.: Footprint@work, a computing framework for large scale parametric simulations: application to pesticide risk assessment and management, Forum EGEE Enabling Grids for E-sciencE, Manchester, UK, 9–11 May 2007
9. Vapnik V. N. (1998). Statistical learning theory. Wiley, Hoboken.
10. Basudhar A. and Missoum S. (2010). An improved adaptive sampling scheme for the construction of explicit boundaries. *Structural and Multidisciplinary Optimization*, **42(4)**, 517-529.
11. Du Q., Faber V. and Gunzburger M. (1999). Centroidal Voronoi tessellations: applications and algorithms. SIAM review, **41(4)**, 637-676.
12. Rueda A., Gouldby B., Méndez F. J., Tomás A., Losada I. J., Lara J. L. and Díaz-Simal, P. (2015). The use of wave propagation and reduced complexity inundation models and meta-models for coastal flood risk assessment. *Journal of Flood Risk Management*.