Environmental Modelling & Software 104 (2018) 64–80

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Casting light on forcing and breaching scenarios that lead to marine inundation: Combining numerical simulations with a random-forest classification approach

J. Rohmer [a, *], D. Idier [a], F. Paris [b], R. Pedreros [a], J. Louisor [a]

[a] BRGM, 3 av. C. Guillemin, BP 36009, 45060 Orléans Cédex 2, France
[b] BRGM, Parc Technologique Europarc, 24, avenue Leonard de Vinci, 33600 Pessac, France

ARTICLE INFO

Article history:
Received 19 April 2017
Received in revised form
23 February 2018
Accepted 15 March 2018

Keywords:
Coastal flooding
Overflow
Random forest
Classification probability
Variable importance analysis
Partial dependence plot

ABSTRACT

Identifying the offshore forcing and breaching conditions that lead to marine inundation is of high importance for risk management. This task cannot be conducted by using a numerical hydrodynamic model due to its high computation time cost (of several minutes or even hours). In the present study, we show how the random forest (RF) classification technique can approximate the numerical model to explore these critical conditions. We focus on the Bouchôleurs site, which is located on the French Atlantic coast and exposed to overflow processes. An iterative strategy is developed for selecting the numerical simulations (a total of 200) to train the RF model. The sensitivity to the input parameters is studied using permutation-based importance measures and extended versions of the partial dependence plots. The results highlight the key interplay among the high-tide level, the surge peak and the phase difference, and the complex role of the breaching location.

1. Introduction

Recent storm events such as Katrina in 2005 and Xynthia in 2010 illustrate the present-day coastal damages and injuries that can affect coastal areas in both cyclonic and non-cyclonic environments. Katrina was one of the most powerful hurricanes ever to be registered in the Atlantic and led to more than 1,500 deaths and damages of approximately 80 billion USD (Blake, 2007), whereas Xynthia was a mid-latitude storm that severely hit low-lying coasts in the central part of the Bay of Biscay on 27–28 February 2010 and led to more than 40 deaths and one billion euros of material damages (see, e.g., Vinet et al., 2012). From a statistical point of view, the wave heights that were generated during the Xynthia event could not be considered extremes (see, e.g., Bertin et al., 2012). However, what makes this event “rare” is the combination of a spring tide with a large storm surge (enhanced by young wind waves) that reached its maximum near the tide peak.

This illustrates the importance of gaining insights into and a better understanding of the combination of offshore forcing conditions (water levels, tide peak, storm duration, wave characteristics) that lead to inundation (or not). A systematic exploration of these conditions is of primary importance for risk management based on multiple objectives, as discussed by Idier et al. (2013): 1. for crisis management purposes, since this knowledge can be used as input for constraining forecast and early warning systems (see an example of such systems in UK, Stansby et al., 2013); 2. for prevention purposes, since this knowledge can be used to support the return period calculation in determining flooding risk (see, e.g., Gouldby et al., 2014); 3. for enhancing risk culture by improving risk awareness and preparedness (regarding various possible offshore scenarios).

In addition to offshore conditions, coastal defence failures (such as failures of artificial structures such as dikes or breaching of sand dune systems) might also influence the inundation characteristics (spatial extent, inland water height, flooding time of arrival, etc.), as illustrated by the 1953 North Sea, 2005 Katrina and 2010 Xynthia flooding events. Therefore, a deep understanding of the breaching characteristics (spatial location along the coast, width, etc.) that increase the inundation likelihood is also desirable. For clarity, we use the generic term “breaching” to refer to the failure of an artificial coastal defence or the breach of natural systems such as dunes.

* Corresponding author.
E-mail address: j.rohmer@brgm.fr (J. Rohmer).

https://doi.org/10.1016/j.envsoft.2018.03.003
1364-8152/

The task of detecting and analysing such "critical" conditions in the coastal risk domain (offshore and related to coastal defence failure) shares the same objectives as the analytic approach for scenario discovery, which aims at coping with deep uncertainty (e.g., Groves and Lempert, 2007; Kwakkel et al., 2013). Scenario discovery analysis relies on exploration and analysis of numerical simulations. The objective is threefold: exhaustively sampling the space that is spanned by the input parameters, analysing the consequences via numerical simulations and identifying regions of interest in the input space. However, in the application of such a systematic search to the coastal domain, multiple difficulties are encountered, which are at the core of the present study.

First, coastal flooding assessment is typically supported by numerical simulations, which might have very large computational time cost (typically of several minutes to hours per simulation, depending on the process complexity that is taken into account by the model). Such a computational burden often restricts the analysis to a limited number of input configurations (termed as scenarios in the following), even when computing architectures are used. Meta-modelling approaches (aka surrogate modelling; see, e.g., Castelletti et al., 2012; Razavi et al., 2012) can efficiently handle the difficulty that is related to computational time cost by replacing the computationally expensive numerical code by a costless-to-evaluate statistical emulator (approximation), which is built using a limited number of computer experiments (typically a few hundreds). The meta-model-based strategy proved to be very efficient when combined with adaptive sampling strategies: a reduction by a factor 20–40 of the total number of necessary computationally expensive simulations was achieved in the case that is described by Rohmer and Idier (2012).

A second difficulty is related to the nature of the coastal processes, which are usually controlled by thresholds. As an illustration, let us consider marine flooding that is induced by overflow processes: if the water level at the coast (which results from storm surge and tide characteristics) is lower than a specified threshold, flooding cannot occur: the water height at any given location inland remains zero. Otherwise, provided that the water level slightly increases and exceeds a specified threshold, overflow-induced inundation can occur and inland locations may be flooded. In other words, the water height inland remains zero until a specific combination of conditions (surge amplitude, tide amplitude, phase difference, etc.) that lead to a flood at this location are met. In these situations, the functional (mathematical) relationship between the offshore conditions and the response is expected to be highly nonlinear and standard meta-modelling techniques might fail (Gaussian process, polynomial chaos expansion, etc.; see, e.g., Jin et al., 2001). Among the available possible options (see, e.g., Razavi et al., 2012), an alternative procedure is to use a classification strategy to identify the conditions that lead to inundation (class 1) and those that do not (class 0), as proposed, for instance, by Bourinet et al. (2011) in the field of structural safety and tested by Rohmer et al. (2016) in the field of marine flooding.

Finally, a third difficulty is related to the number of offshore conditions that depend on the considered physical parameters (tide, atmospheric storm surge, wave height, wave direction, wave period, etc.). This complicates the detection of the critical conditions (the larger the number of such conditions, the broader the input space and the more tedious the exploration). This also makes the visualisation of the potentially complex classification rules difficult. Variable importance analysis (see, e.g., Galelli et al., 2014) can overcome this issue by carefully selecting the most important parameters regarding the occurrence of the considered event.

Different statistical techniques have been proposed in the literature for dealing with these three difficulties (see, e.g., Kuhn and Johnson, 2013). In the present study, we utilize a meta-modelling strategy that is based on the random forest (denoted RF) technique (Breiman, 2001) to approximate computationally expensive numerical simulators (first difficulty) using a limited number of pre-calculated simulations (typically 100–200). Many studies have shown the high performance of RF in different domains of application (ecology, Cutler et al., 2007; water quality, Brooks et al., 2016; wildfires, Rodrigues and de la Riva, 2014, etc.). To overcome the second difficulty, the RF model is used as a classifier to predict whether a set of forcing and breaching conditions will lead to inundation (class 1) or not (class 0) at a given spatial location of the considered study site, regardless of the type of input parameters (i.e., continuous or categorical), with possible correlation among them. The third difficulty is handled by means of the variable importance measures that are provided by the RF model (Wei et al., 2015).

The objective of the present study is to address the following questions: (Q1) To what extent is the RF model a valid approximation of the true computationally expensive numerical simulator? (Q2) How can the scenarios (defined by values of the input parameters) that should be simulated to construct the RF model be efficiently selected? (Q3) How can the RF model help to discover and cast light on the critical conditions (forcing and breaching) that lead to marine inundation? For this purpose, we take advantage of recent advances that are related to the implementation of this technique in the R software (R Developement Core Team R, 2017): 1. the numerically efficient version of the RF model that was developed by Wright and Ziegler (2016) (implementation in the R package ranger); 2. the estimate of the classification probability that was proposed by Malley et al. (2012) (implementation in the R package ranger); 3. the feature selection algorithm that was developed by Kursa and Rudnicki (2010) (implementation in the R package Boruta); and 4. the visualisation tools that were proposed by Goldstein et al. (2015) for exploring the dependencies using the Individual Condition Expectation curves (implementation in the R package ICEbox).

The paper is organized as follows: First, we describe the overall strategy and the associated methods for exploring the offshore forcing and breaching conditions that lead to marine inundation (Sect. 2). Second, we describe the application case at the Bouchôleurs site, which is located along the French Atlantic coast (Sect. 3). Third, we apply the proposed strategy (Sect. 4) and address research questions Q1-Q3 by discussing the added value and the limitations of the RF-based analysis from a risk analysis perspective (Sect. 5). The list of acronyms is provided in Appendix A.

2. Methods

In this section, we first describe the overall strategy that is applied in the present study to explore and analyse the scenarios that lead to inundation (Sect. 2.1). Then, we describe the key ingredients of the strategy (Sect. 2.2–2.6).

2.1. Strategy description

An overview of the strategy is provided in Fig. 1. A preliminary (and necessary) step aims at setting up (and validating) a hydrodynamic model for simulating the flooding processes for the site of interest. On this basis, the steps are as follows:

- Step 1: A limited number n_0 of scenarios are first selected (in the domain of forcing and breaching conditions $\boldsymbol{X}$). These scenarios are used as inputs of the numerical modelling chain (described