**Natural Hazards
and Earth
System Sciences**

A meta-modelling strategy to identify the critical offshore conditions for coastal flooding

J. Rohmer and D. Idier

BRGM, 3 av. C. Guillemin, 45060 Orléans Cédex 2, France

Correspondence to: J. Rohmer (j.rohmer@brgm.fr)

Abstract. High water level at the coast may be the result of different combinations of offshore hydrodynamic conditions (e.g. wave characteristics, offshore water level, etc.). Providing the contour of the "critical" set of offshore conditions leading to high water level is of primary importance either to constrain early warning networks based on hydro-meteorological forecast or observations, or for the assessment of the coastal flood hazard return period. The challenge arises from the use of computationally intensive simulators, which prevent the application of a grid approach consisting in extracting the contour through the systematic evaluation of the simulator on a fine design grid defined in the offshore conditions domain. To overcome such a computational difficulty, we propose a strategy based on the kriging meta-modelling technique combined 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 contour. This methodology is applied to two cases using an idealized site on the Mediterranean coast (southern France): (1) a two-dimensional case to ease the visual analysis and aiming at identifying the combination of offshore water level and of significant wave height; (2) a more complex case aiming at identifying four offshore conditions (offshore water level and offshore wave characteristics: height, direction and period). By using a simulator of moderate computation time cost (a few tens of minutes), the targeted contour can be estimated using a cluster composed of a moderate number of computer units. This reference contour is then compared with the results of the meta-model-based strategy. In both cases, we show that the critical offshore conditions can be estimated with a good level of accuracy using a very limited number (of a few tens) of computationally intensive hydrodynamic simulations.

1 Introduction

Forty percent of the world population lives within 100 km of a coastline, an area that accounts for only about twenty percent of the land mass (UNEP, 2007). This urbanized area is characterized by both increasing assets and increasing coastal flooding risk induced either by extreme events and/or by sea-level rise (Nicholls et al., 2006). Some recent events like Katrina in 2005 or Xynthia in 2010 (see e.g. Bertin et al., 2012) illustrate the present-day coastal damages and injuries that can affect the coastal area. 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 EUR. From a statistical point of view, the wave heights generated during the Xynthia event could not be considered extremes, but what makes this event "rare" is the combination of a high spring tide with a large storm surge reaching its maximum around the tide peak. The processes involved in such event are identified by Bertin et al. (2012) showing a strong influence of waves in such high storm surges. This recent event provides a vivid reminder of the complexity and of the strong coupling between different hydro-meteorological processes underlying the occurrence of storm surge, and possibly of large impacts on the coastal plain.

In this context, early warning and forecasting systems are key components for a coastal flooding risk assessment and management. These are based on hydro-meteorological observations and modelling, such as the initiatives in the North Sea Region (Safecoast, 2008), and allow anticipating the co-occurrence of such high-magnitude events and setting up

accordingly preventive actions. To constrain these systems, the combinations of offshore conditions x that generate a "critical" high water level at the coast (the maximum water level at the coast and the critical one are respectively denoted ξ_m and ξ_c) are of primary importance. The "critical" characteristic of ξ_c can be judged regarding the potential associated losses and consequences.

The relationship between these critical combinations constitutes 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), as schematically depicted in Fig. 1. For instance, the combination of offshore conditions outlined by "a" in Fig. 1 leads to a water level at the coast considered "safe", whereas the combination "b" leads to an unsafe water level, as located in the set, whose boundaries correspond to Γ_c. The combination "c" is exactly on this critical frontier, and its estimation constitutes the core issue of the present study. More formally, the objective is to estimate the contour of the set $\Gamma_c = \{x$ so that $\xi_m = \xi_c\}$.

From a reliability analysis point of view, this frontier is referred to as "limit state" and determines the failure of an engineered system. It is used to estimate the probability of the system failure (see e.g. Haldar and Mahadevan (2000) for further details on system reliability analysis). Similarly as for reliability assessment, the critical frontier of offshore conditions can be used in a Monte Carlo-based approach to estimate the return period (i.e. the probability of occurrence) of the water level at the coast (see for instance the statistical approach called "response method" proposed by Garrity et al., 2006). Another modelling approach has recently been proposed by Chini and Stansby (2012) based on the statistical discretization of the offshore conditions (wave and water level), with the successive following steps: joint probability analysis of offshore conditions, discretization of extreme offshore conditions occurrence isolines, hydrodynamic modelling for each selected offshore combination and evaluation of the near-shore values (e.g. mean overtopping) in the two-dimensional space of offshore wave and water level.

From a more methodological perspective, the knowledge of Γ_c can eventually support the implementation of innovative "inverse" risk assessment procedures, which have been recently proposed either for fluvial inundation (Cunderlink and Simonovic, 2007) or for coastal flooding (Idier et al., 2010). These still under-development methodologies basically rely on the inversion of the traditional approach of risk analysis by starting from a targeted risk level judged as "acceptable" in the studied territory and by ending up with the estimate of the corresponding return period based on the knowledge of the associated hydro-meteorological conditions.

This "critical" frontier Γ_c is usually not "explicit", i.e. in many cases it does not have a simple analytical mathematical formulation and its estimate usually relies on numerical modelling. In the field of coastal flood hazard assessment,

several categories of models with different degrees of complexity depending on data requirements, resolution, physical processes and characteristics of the underlying equations are available (Fewtrell et al., 2011; Samuels et al., 2008). If we focus on the modelling of the water level at the coast (necessary step to provide final flooding hazard maps), several approaches can be used. A first class of models (class 1) is based on semi-empirical approaches, which provide total water level (tide, storm surge, wave set-up, run-up and swash) based on input parameters like offshore water level and wave conditions. An example is provided by the formula of Stockdon et al. (2006). A second model class (class 2) relies on more sophisticated attempts to model the physical processes, involving more advanced (less simplified) methods than class 1 models. As an example, spectral wave modelling (Booij et al., 1999) can be used to provide set-up at the coast, under some assumptions, but will not provide run-up information. Another example is the coupled wave action and shallow water model of McCabe and Sansby (2011), which actually allows estimating the run-up. A last model class (class 3) is based on more advanced methods including few simplifying assumptions and being more accurate but also more complicated. An example is the fully nonlinear weakly dispersive model developed by Bonneton et al. (2011), allowing to reproduce the free-surface (including, for instance, swash, run-up, set-up, etc.).

A "crude" approach for estimating Γ_c would consist in running the model on a regular design grid defined in the forcing conditions space, i.e. corresponding to an intensive parametric analysis. To illustrate, let us consider a regular design grid composed of 30 configurations of offshore water level ξ_o varying by a constant increment step between 0.25 and 1.50 m and 30 configurations of significant wave height H_s varying by a constant increment step between 0.5 and 7 m. In total, 900 simulations should be run in this case. Then, for this specific two-dimensional case, the desired critical frontier, i.e. the one for which the maximum water level at the coast ξ_m is $\xi_c = 1$ m, can be extracted in the form of a contour using, for instance, packaged software such as Matlab© function *contour*(), the accuracy of the solution being dependent on the grid increment.

Yet, the afore-described approach may rapidly become prohibitive and impractical due to the computation time cost of a single numerical simulation, especially when using the model class 3 or even model class 2 (with computation time cost varying from a few minutes to hours). Let us consider a model with moderate computation time cost of 15 min; then, the afore-described example case would require less than 9 days of computation, which is achievable using a computing cluster composed of a moderate number of computer units (CPU). If the estimate of Γ_c was carried out for four offshore conditions (e.g. significant wave height H_s, wave period T_p, direction θ and offshore water level ξ_o) using a design grid with each dimension discretized into 30 elements, the total computation time cost would reach almost
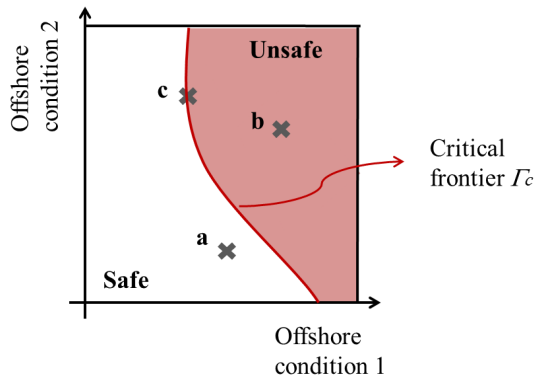
Fig. 1. Schematic representation of the critical frontier Γ_c in the offshore conditions domain. This frontier represents the boundary between the "safe" and "unsafe" regions (red-coloured area), i.e. the boundary of the set of offshore forcing conditions, which lead to a "high level" of water level at the coast.

8400 days, which is obviously not achievable and would require a computing cluster composed of at least 500 CPU to reach a computation time of about two weeks.

To tackle this challenge, two strategies can be proposed: appropriate grid computing architecture (e.g. Boulahya et al., 2007) or the use of meta-models (also named response surfaces or surrogate models; see for instance Forrester et al. (2008)). The latter approach basically consists in replacing the numerical model by a "costless-to-evaluate approximation". This has been proposed in a variety of other contexts: structural reliability problems (Bucher and Bourgund, 1990), environmental issues (Iooss et al., 2006), natural hazard assessments (Rohmer and Foerster, 2011), geological storage performance assessments (Rohmer and Bouc, 2010), etc.

The objective of the present paper is to explore the applicability of the meta-modelling techniques for identifying the critical frontier in the offshore condition domain, when using computation time consuming hydrodynamic models. To let the "crude" approach be achievable with a cluster composed of a moderate number of CPU (around 10), we propose to focus on a hydrodynamic model of class 2 characterized by a moderate computation time (of the order of 15 min). The critical frontier estimated in this manner is viewed as a reference to be compared to the one estimated using the meta-modelling approach. Both strategies are tested by applying them to an idealised site located on the Mediterranean coast in southern France considering two cases: (1) the first one is two dimensional to ease the visual analysis and aims at identifying the critical combinations $x_c = \{\xi_o ; H_s\}$; (2) the second one is more complex and aims at identifying the four-dimensional critical combinations $x_c = \{\xi_o ; H_s ; T_p ; \theta\}$.

The remainder of the paper is organized as follows. In a first section, the general strategy relying on meta-modelling is described. We focus in a second section on kriging

meta-models (Forrester et al., 2008). To improve the local accuracy of the approximation of the contour of interest, we combine the meta-modelling strategy with a sequential adaptive sampling approach relying on the recent advances of contour estimation from complex computer codes (for a recent overview, see Bect et al., 2012). In a third section, we describe the study site and the associated moderate-to-evaluate (of the order of 15 min) numerical model to compute the maximum water level at the coast. The meta-modelling strategy is then tested on both test cases (two- and four-dimensional cases).

2 Meta-modelling strategy

In this section, we describe the methodology relying on the meta-modelling technique to estimate the critical frontier from a complex and time consuming hydrodynamic code, which is referred to as "simulator" and denoted f.

2.1 Principle of meta-modelling

This technique consists in replacing f by a mathematical approximation referred to as "meta-model" (also named "response surface" or "surrogate model"). This corresponds to a function constructed using a few computer experiments (i.e. a limited number of time consuming simulations), and aims at reproducing the behaviour of the "true" model in the domain of model input parameters (here the offshore conditions) and at predicting the model responses with a negligible computation time. In this manner, any approach relying on intensive multiple simulations (such as global sensitivity analysis, in-depth uncertainty analysis or intensive parametric analysis for the case described in the present article) is made achievable at a "reasonable" computation time cost.

2.2 Methodology

The main steps of the methodology are summarized in Table 1. A meta-model is constructed and validated with an initial budget of training data (Steps 1 to 3). It is then updated (Step 4) using candidate points selected to improve the local accuracy in the regions that contribute the most to the estimate of Γ_c. Using this costless-to-evaluate function, Γ_c can be obtained from a fine grid on the offshore conditions space (Step 5) using for instance packaged software such as Matlab© function *contourc*(). Steps 1 to 4 are detailed in the next paragraphs.

2.2.1 Step 1: setting the training data

The approximation is constructed by running f given a limited number n_0 of different model input parameters (i.e. of offshore conditions vectors x), named training samples. Hence, the objective is to create a mapping (named training data) between the offshore conditions x and the quantity

Table 1. Sequential strategy for extracting the critical frontier.

Step	Description
1	Evaluate ξ_m considering n_0 offshore conditions randomly selected using a LHS technique;
2	Based on the initial training data, construct the meta-model;
3	Assess the quality of the approximation through cross-validation procedures;
4	While the total affordable simulation budget is not exhausted: 4.1: Select the next best candidate configuration based on an appropriate criterion; 4.2: Evaluate the corresponding ξ_m; 4.3: Update the training data and re-construct the meta-model;
5	Extract the Γ_c using a fine design grid in the offshore conditions space.

of interest, namely the maximum water level at the coast ξ_m. A trade-off should be found between maximizing the exploration of the offshore conditions domain and minimizing the number of simulations, i.e. a trade-off between the accuracy of the approximation and the computation time cost. To fulfill such requirements, we propose to randomly select the training samples by means of the Latin hypercube sampling (LHS) method (McKay et al., 1979) in combination with the "maxi-min" space-filling design criterion (Koehler and Owen, 1996).

2.2.2 Step 2: construction of the meta-model

Using the training data, the approximation can be carried out relying on several types of meta-models, either using simple polynomial regression techniques, non-parametric regression techniques (Storlie et al., 2009), kriging modelling (Forrester et al., 2008), artificial neural networks (Papadrakakis and Lagaros, 2002), or polynomial chaos expansions (Ghanem and Spanos, 1991), etc. The choice of the meta-model type is guided by the a priori non-linear functional form of the simulator, as well as the number of input parameters. In the present study, we focus on kriging meta-modelling. This non-parametric technique presents several attractive features. It is flexible to any kind of functional form of the simulator. In particular, it introduces less restrictive assumptions on the functional form of the simulator than a polynomial model would imply; it is an exact interpolator, which is an important feature when the simulator is deterministic; it provides a variance estimate of the prediction, the latter being very useful to guide the selection of future training samples according to the target of the optimization problem (see Sect. 3). Yet, it should be noted that these advantages come at the expense of

a parameter learning stage, which may pose additional computational difficulties (see e.g. Langewisch and Apostolakis, 2010).

2.2.3 Step 3: validation of the meta-model

Usually the differences between the approximated and the true model output (or residuals) are used to estimate the quality of the approximation. Yet, for interpolating meta-models (e.g. splines, kriging, etc.), the residuals at the sampled points are null and an alternative is to estimate the expected level of fit (i.e. quality of prediction) to a data set that is independent of the original training data that were used to construct the meta-model, i.e. to "yet-unseen" data. A first approach would consist in using a test sample of new data. Though the most efficient, this might be often impracticable as additional simulator runs are costly to collect. A possible option to overcome such a situation, coined by Tukey (1954) as "uncomfortable science", is to rely on the cross-validation procedures such as the "leave-one-out" cross validation technique (e.g. Hastie et al., 2009). Further details are provided in Appendix A.

2.2.4 Step 4: adaptive sampling

Once constructed and validated, the meta-model can replace the simulator to extract the desired contour through an intensive parametric analysis on a regular fine design grid in the offshore conditions space. For instance, the choice of the initial number of training data is advocated by Ranjan et al. (2008) to reach at least 25 to 35 % of the total budget, but no exact rule exists.

Yet, this approach may not be efficient in several cases: when the allowable budget of simulations n_0 is low; when the number of offshore conditions is high (i.e. requiring a large number of training data to explore the offshore condition domain); and when the relationship between x and ξ_m is highly non-linear for instance.

Based on the basic idea that not all offshore conditions configurations will contribute to the estimate of Γ_c, sampling effort should then be concentrated in the regions of the offshore conditions, which provide the maximum information regarding Γ_c, i.e. both to improve the local accuracy of the approximation of Γ_c and to explore the design space in a global manner, where desired contour may also exist (see Sect. 3.2). Hence, to efficiently use the available resources, we propose to complete steps 1 to 3 by a sequential sampling approach to adaptively select the best training data candidates fulfilling the afore-mentioned requirements.

The definition of a selection criterion to solve this "contour estimate" problem has only been recently addressed (see Bect et al., 2012 and references therein). This type of optimization problem differs from the traditional problem of searching the global optimum (minimum or maximum) of a black-box function, which has given rise to numerous

research works over the last two decades (see Forrester and Keane, 2009). One major difficulty is the nature of the solution of the "contour" optimization problem, which corresponds to a set of solutions, whereas it corresponds to a unique solution (or at least of finite number) for the traditional optimization problem. The comparison of the existing criteria is beyond the scope of this article, and we focus on the sampling criterion developed by Ranjan et al. (2008) (see Sect. 3).

To summarize, the meta-model is initially constructed and validated based on a first training set. The meta-model is then iteratively updated using the candidate selected based on the sampling criterion, and the procedure is continued until the allowable number of simulations is exhausted.

3 Adaptive sampling using kriging meta-model for contour estimate

In this section, we focus first on kriging meta-modelling (Forrester et al., 2008), for which selection criteria adapted to contour extraction from computer codes have been recently defined to sequentially guide the sampling effort (Bect et al., 2012). Next, we more specifically focus on the selection criterion developed by Ranjan et al. (2008).

3.1 Introduction to kriging meta-model

We introduce here the basic concepts of kriging meta-modelling, which can be viewed as an extension to computer experiments of the kriging method used for spatial data interpolation and originally developed by Krige (1951) for mining applications. For a more complete introduction to kriging meta-modelling and full derivation of equations, the interested reader can refer to Sacks et al. (1989); Jones et al. (1998); Jones (2001); Forrester et al. (2008).

The kriging model considers the deterministic (i.e. not random) response of the simulator $\xi_m = f(x)$ as a realization of a Gaussian stochastic process F so that $f(x)=F(x,\omega)$, where ω belongs to the underlying probability space Ω. In the following, we use the notation $F(x)$ for the process and $F(x,\omega)$ for one realization. The process F results from the summation of two terms:

- $f_0(x)$, the deterministic mean function, which is usually modelled by a constant or a linear model and represents the trend of f;

- $Z(x)$, the Gaussian-centred stationary stochastic process (with zero mean and covariance described below), which describes the deviation (i.e. departure) of the model from its underlying trend f_0.

The stochastic process Z is characterized by the covariance matrix $\mathbf{C}$, which depends on the variance σ_Z^2 and on the correlation function R, which governs the degree of correlation through the use of the vector of length-scale parameters ω between any input vectors. The covariance between $Z(u)$ and $Z(v)$ is then expressed as $C(u;v) = \sigma_Z^2 \cdot R(u;v)$, where $u = (u_1; u_2; \ldots; u_d)$ and $v = (v_1; v_2; \ldots; v_d)$ are two input vectors of dimension d.

A variety of correlation (and covariance) functions has been proposed in the literature (see e.g. Stein, 1999). The commonly used model is the Gaussian correlation function defined as follows:

$$R(u; v) = \exp \left(-\sum_{i=1}^{d} \|\|u_i - v_i\|\|^2/\omega_i\right) \tag{1}$$

where the vector ω determines the rate at which the correlation decreases as one moves in the i-th direction (with i from 1 to d). Intuitively, if $u=v$, then the correlation is 1, whereas if the distance between both vectors tends to the infinity, then the correlation tends to 0. In this article, we chose to focus on the Matérn covariance family. The Matérn correlation function presents more flexibility than Gaussian correlation function, as it allows regulating the degree of differentiability of the underlying random process in addition to the length-scale parameters. This correlation model is described in further details in the Appendix B.

Let us define $\mathbf{X}_D$ the design matrix composed of the vectors of offshore conditions x selected in step 1 of the methodology (i.e. the training samples) to be simulated, so that $\mathbf{X}_D=(x^{(1)}; x^{(2)}; \ldots; x^{(n0)})$ and ξ_D the vector of simulated maximum water levels at the coast associated with each selected training samples, so that $\xi_D=(\xi_m^{(1)} = f(x^{(1)}); \xi_m^{(2)} = f(x^{(2)}); \ldots; \xi_m^{(n0)} = f(x^{(n0)}))$.

Under the afore-described assumptions, the distribution of the maximum water level for a new input vector of offshore conditions x^* follows a Gaussian distribution conditional on the design matrix $\mathbf{X}_D$ and of the corresponding simulated water levels ξ_D with expected value $\tilde{\xi}_m$ for the new configuration x^* and variance s^2 respectively defined by Eqs. (2) and (3):

$$\tilde{\xi}_m (x^*)$$
$$= E(F(x^*)|\mathbf{X}_D; \xi_D) = \hat{\tilde{\xi}}_m + r(x^*)^T \cdot \mathbf{R}_D^{-1} \cdot (\xi_D - \mathbf{I} \cdot \hat{\tilde{\xi}}_m) \tag{2}$$

with the constant $\hat{\tilde{\xi}}_m = (\mathbf{I}^T \cdot \mathbf{R}_D^{-1} \cdot \mathbf{I})^{-1} \cdot (\mathbf{I}^T \cdot \mathbf{R}_D^{-1} \cdot \xi_D)$

$$s^2(x^*) = \mathrm{Var}(F(x^*)|\mathbf{X}_D; \xi_D) = \hat{\sigma}^2 \cdot \left(1 - r(x^*)^T \cdot \mathbf{R}_D^{-1}\right.$$
$$\left.\cdot r(x^*) + \frac{(1 - \mathbf{I}^T \cdot \mathbf{R}_D^{-1} \cdot r(x^*))^2}{\mathbf{I}^T \cdot \mathbf{R}_D^{-1} \cdot \mathbf{I}}\right) \tag{3}$$

with the constant $\hat{\sigma}^2 = (\xi_D - \mathbf{I} \cdot \hat{\tilde{\xi}}_m)^T \cdot \mathbf{R}_D^{-1} \cdot (\xi_D - \mathbf{I} \cdot \hat{\tilde{\xi}}_m)^T / n_0$ where $r(x^*)$ is the correlation vector between the test candidate x^* and the training samples; $\mathbf{R}_D$ is the correlation matrix of the training samples $\mathbf{X}_D$ and $\mathbf{I}$ is the unit matrix of size n_0.
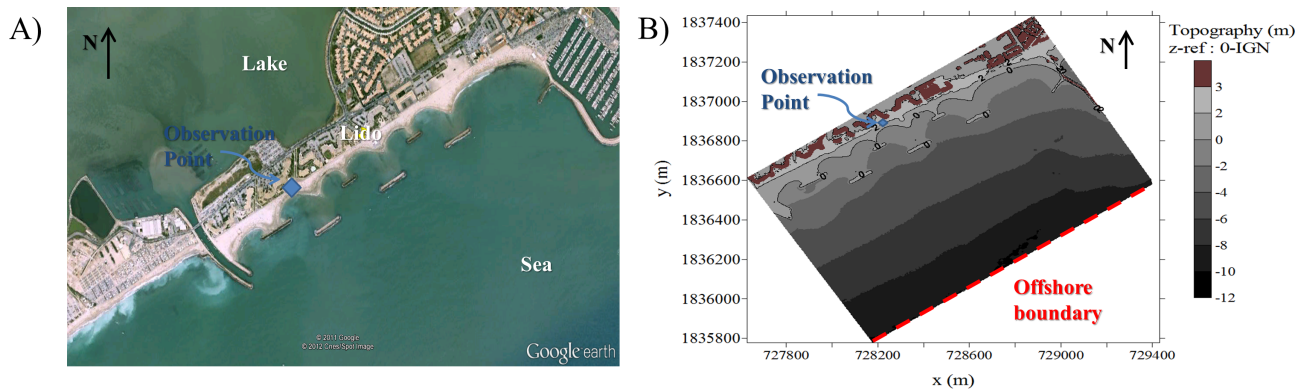
Fig. 2. (A) Location of the study site and observation point where the calculations are carried out; **(B)** SWAN 2-D domain model of the study site.
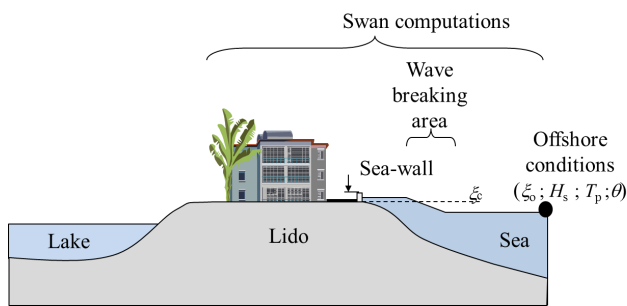
Fig. 3. Schematic cross shore representation of the processes accounted for in the SWAN models for the idealized study site.

The conditional mean in Eq. (2) is used as a predictor, i.e. the "best estimate". The conditional variance in Eq. (3) can be used to estimate the mean square error of the predictor and to deduce a confidence interval associated with the prediction. The regions of the offshore conditions space where few data are available are underlined with higher variance, so that Eq. (3) also provides a local indicator of the prediction accuracy useful to guide sampling effort as described in Sect. 3.2.

The above Eqs. (2) and (3) are categorized as "ordinary kriging" and are the most common version of kriging used in engineering (Forrester et al., 2008). A more general form of kriging equations exists, known as "universal kriging", and allows computing the deterministic mean function f_0 as a polynomial regression (generally of low-order) with unknown coefficients. For sake of clarity, these equations are presented in Appendix C.

In practice, the parameters of the kriging model corresponding to the constant (or the coefficients of regression) used to model the mean function f_0, the variance σ_Z^2 and the length scale parameters ω are determined through maximum likelihood estimation procedure (details can be found in Santner et al., 2003). In this article, we use the package

named "DiceKriging" developed by Roustant et al. (2012) of the R software (R Development Core Team, 2011). In particular, this package implements the efficient likelihood maximization algorithm of Park and Baek (2001) to overcome numerical instabilities, multimodality and dimensionality issues associated with kriging model parameters estimation.

3.2 Definition of the selection criterion

The sampling strategy described in Sect. 4 requires the definition of an appropriate selection criterion. Following the same spirit as the developments done by Jones et al. (1998) for the optimization (maximizing or minimizing) of expensive black-box computer codes, Ranjan et al. (2008) defined a selection criterion adapted to contour estimation based on the following improvement function I for any offshore conditions vector $\boldsymbol{x}$:

$$I(\boldsymbol{x}) = \varepsilon^2(\boldsymbol{x}) - \min\left((F(\boldsymbol{x}) - \xi_c)^2; \varepsilon^2(\boldsymbol{x})\right) \tag{4}$$

where ε is proportional (for some constant α) to the mean square error using Eq. (3) and defines a neighbourhood (the first term in the "min" function being the Euclidean distance) around the targeted contour Γ_c associated with ξ_c.

Jones et al. (1998) and Jones (2001) demonstrated that merely maximizing I would be inefficient and advocated to select the best candidate by maximizing the expected improvement function $E(I)$ over the offshore conditions domain using the fact that the simulator output $f(\boldsymbol{x})$ follows a Gaussian distribution with moments $\tilde{\xi}_m$ and s^2 respectively defined by Eqs. (2) and (3). $E(I)$ is the summation of three terms A_1, A_2 and A_3 as expressed by Eq. (5a, b, c) for any offshore conditions vector $\boldsymbol{x}$:

$$A_1(\boldsymbol{x}) =$$
$$\left(\varepsilon(\boldsymbol{x})^2 - (\xi_c - \tilde{\xi}_m(\boldsymbol{x}))^2\right) \cdot \left(\Phi(u^+) - \Phi(u^-)\right) \quad \text{(5a)}$$

$$A_2(\boldsymbol{x}) =$$
$$-2 \cdot s(\boldsymbol{x}) \cdot \left(\xi_c - \tilde{\xi}_m(\boldsymbol{x})\right) \cdot \left(\phi(u^+) - \phi(u^-)\right) \quad \text{(5b)}$$

$$A_3(\boldsymbol{x}) = -s(\boldsymbol{x})^2 \cdot \int_{u^-}^{u^+} w^2 \cdot \phi(w) dw \quad \text{(5c)}$$

where $u^\pm = (\xi_c - \tilde{\xi}_m(\boldsymbol{x}) \pm \varepsilon(\boldsymbol{x}))/s(\boldsymbol{x})$; ϕ is the Gaussian probability density distribution and Φ is its cumulative form.

Ranjan et al. (2008) provides an intuitive interpretation of each of the three terms:

1. The first term A_1 dominates the expression when $f(\boldsymbol{x})$ is close to ξ_c, hence corresponding to the points lying close to the regions of the offshore conditions specified by $\xi_c \pm \varepsilon(\boldsymbol{x})$, i.e. in the vicinity of Γ_c;

2. A_2 dominates when $f(\boldsymbol{x})$ is further away from ξ_c so that the sampling is guided outside the regions fulfilling $\xi_c \pm \varepsilon(\boldsymbol{x})$, where the uncertainty in the prediction is high (i.e. where the prediction variance is high);

3. the last term A_3 guides the sampling only within the regions fulfilling $\xi_c \pm \varepsilon(\boldsymbol{x})$, where the uncertainty in the prediction is high.

The three terms taken together allows to guide sampling in the vicinity of Γ_c where little information is available (A_1 and A_3) and to explore outside the vicinity of Γ_c (term A_2) in regions scarcely sampled where the contour may also exist. Thus, $E(I)$ provides a selection criterion to balance local (in the vicinity of Γ_c) and global exploration of the offshore conditions domain.

The interested reader can refer to the recent works of Bect et al. (2012) for the definition of other selection criteria formulated following either similar principles using an improvement function, or a modified version of the traditional integrated mean square error, or the Bayesian formalism.

In practice, the best candidates for simulation are selected in Step 4.1 of the methodology (see Table 1) by maximizing Eq. (4). In the present article, the package "KrigInv" developed by Picheny and Ginsbourger (2012) of the R software (R Development Core Team, 2011) is used.

## 4 Application to the identification of critical offshore conditions

In this section, we present the application of the meta-modelling strategy on an idealised site (description in Sect. 4.1 and associated model in Sect. 4.2) considering two cases (respectively Sects. 4.3 and 4.4): (1) the first one is two dimensional to ease the visual analysis and aims at identifying the critical combinations $\boldsymbol{x}_c = \{\xi_o; H_s\}$; (2) the second one is more complex and aims at identifying the four-dimensional critical combinations $\boldsymbol{x}_c = \{\xi_o; H_s; T_p; \theta\}$.

4.1 Study site

The application on the idealized site is considered for demonstration purposes only. The study site located on the Mediterranean coast is characterised by a lido, which is of primary importance both at an environmental and economic level for the region (see Fig. 2). This lido is protected from the sea, characterised by significant touristic activities. At the seafront, there are pedestrian areas, which have already been flooded, at least during two storms: 6–8 November 1982 and the 4 December 2003. This pedestrian area is limited by a sea-wall having a height of about 50 cm above the pedestrian street as schematically represented in Fig. 3. In this context, the critical maximum water level at the coast, including storm surges, tide, wave set-up and run-up, should not exceed the foot of the sea-wall, which is located at 2.15 m above the vertical reference (IGN). In this manner, the sea water is prevented from invading the area through the sea-wall openings, and specifically at the observation point (Fig. 2), which is just in front of connexion roads with the lido.

4.2 Model set-up and parameters

First, we should keep in mind that the water level at the coast results from the offshore water level ξ_o (tide and storm surge) and the run-up resulting from two dynamical processes, which are the wave set-up η and the swash S. Here, we set a modelling strategy to account for these three contributions ($\xi_o; \eta; S$). We focus the analysis on a numerical model of class 2 (see introduction) based on the following assumptions. First, the sea level induced by storm surges and tide is assumed uniform in the local area. Second, the wave set-up is computed using the code SWAN (Booij et al., 1999). It solves the wave action equation and provides wave spectra and wave set-up. At the offshore boundary (see Fig. 2b), the offshore conditions $\boldsymbol{x} = \{\xi_o; H_s; T_p; \theta\}$ are imposed. The bathymetric and topographic data used for the modelling have been obtained through lidar measurements (data provided by SMNLR), as well as bathymetric surveys (data provided by SHOM). The grid size is 2 m and the 2-D domain represented in Fig. 2b. Finally, to compute the swash-induced elevation S, we use the formulation of Stockdon et al. (2006). Based on the assumption of reflective beach, the analysis of
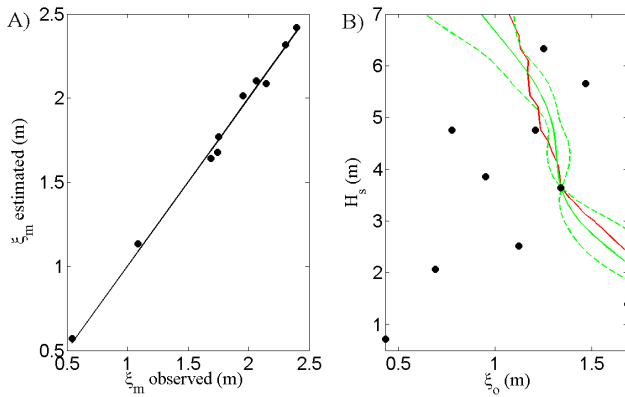
Fig. 4. (A) leave-one-out cross-validation for the construction of the initial kriging model; black dots are the water level estimations. **(B)** Best estimate of Γ_c in the $\{\xi_o ; H_s\}$ domain associated with $\xi_c = 2.15$ m (green solid line) using a kriging model constructed with 10 training data (black dots) selected randomly from a LHS technique. The associated pair of contours at confidence level 95 % (defined as the kriging mean (Eq. 2) $\pm$ twice the square root of the variance (Eq. 3)) is represented by dashed green lines. The reference contour (estimated using computation grid architecture) is represented by a red solid line.

the equations proposed in Stockdon et al. (2006) shows that we can, as a first approximation, assume that S is equal to the set-up η. Thus the maximum sea-level ξ_m at the beach can be assumed to approximately equal $\xi_o + 2\eta$. All processes and parameters are schematically depicted in Fig. 3.

Noteworthy for modelling simplicity purposes, we assume that the lido is not connected to the sea, and thus we neglect the influence of this lido on the submersion phenomena.

4.3 Two-dimensional test case

In this section, we tackle the problem of estimating the critical frontier Γ_c considering the combination of two offshore conditions $\boldsymbol{x}_c = (\xi_o ; H_s)$ leading to the critical water level at the coast $\xi_c = 2.15$ m (at the observation point). We assume that the offshore wave direction and period are constant and fixed at their most frequent values of respectively 135° N and 7 s, which have been observed in the studied area region.

Due to the moderate computation time cost of 15 min (in average depending on the input parameter values), the identification of the critical frontier can be solved, in a first approach, using a cluster of 10 CPU (2 GHz Pentium 4PC) considering 900 input parameters configurations selected within the domain [0.25; 1.50 m]×[0.5; 7 m] (each dimension being discretized in 30 elements). The total computation time reached less than one day to finally extract Γ_c (red solid line, Fig. 4b).

In a second approach, we simulate ξ_m considering 10 different offshore conditions randomly selected within the domain [0.25; 1.50 m]×[0.5; 7 m] based on a LHS
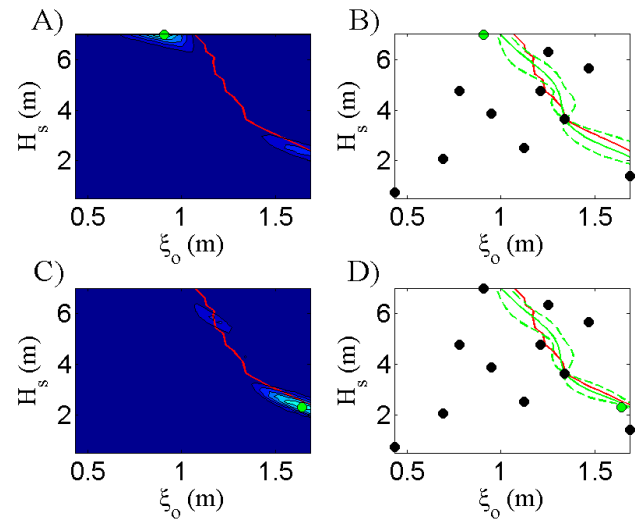
Fig. 5. (A) Expected improvement function values (scaled between 0 and 1) for the first iteration. The best candidate to be selected for simulation is depicted by a green dot. The reference contour is represented by a red solid line; **(B)** the critical frontier Γ_c estimated using a kriging model constructed with the initial 10 training data and the selected training data. The associated pair of contours at confidence level 95 % (defined as the kriging mean (Eq. 2) $\pm$ twice the square root of the variance (Eq. 3)) is represented by dashed green lines. **(C)** Expected improvement function values for the second sampling iteration; **(D)** the critical frontier Γ_c estimated using a kriging model constructed with the initial 10 training data and the selected training data for the two first sampling iterations.

technique (see Sect. 2, step 1). On this basis, a kriging model is constructed using the Matérn covariance function ($\nu = 5/2$; see Appendix B) and a linear trend (see Sect. 3). This meta-model is then validated, and the leave-one-out cross validation procedure provides a coefficient of determination $R_{CV}^2 = 99.7$ % (see Appendix, Eq. A1), which demonstrates the high level of prediction quality of the kriging model. Figure 4a depicts the relationship between the 10 simulated water level ξ_m (named "observed") and the corresponding estimations using a kriging model constructed following the procedure described in Sect. 2.3. The closer the dots are to the first bisector, the better the quality of prediction.

The best estimate of Γ_c, as provided by Eq. (2), is depicted by the green solid line in Fig. 4b. This allows the comparison and the reference contour. We can note that the reference contour lies within the region bounded by the pair of contours (green dashed lines, Fig. 4b) computed at confidence level 95 %. These correspond to the kriging mean (Eq. 2) $\pm$ twice the square root of the variance (Eq. 3). However, there is a poor agreement between the reference contour and the best estimate with several discrepancies. Besides, the width of the pair of contours computed at confidence level of 95 % appears quite large.

On this basis, the sequential sampling strategy described in Sect. 3 is performed so that 10 additional input parameter configurations are selected in the vicinity of the targeted contour defined assuming a constant $\alpha = 0.5$ (see Eq. 4). Figure 5 shows the two first iterations of the sequential sampling strategy. Figure 5a provides the mapping of the expected improvement function values (scaled between 0 and 1) in the offshore conditions space: the best candidate for simulation corresponds to the maximum of this function. Figure 5b provides the updated estimate of Γ_c (green solid line), i.e. the one estimated using a kriging model constructed on training data enriched with the selected training point. Interestingly, the selection of a training point implies the decrease of the width of the pair of contours at level of confidence 95 % (see iteration no. 2 in Fig. 5d).

To test the efficiency of using the adaptive sampling strategy, we estimate Γ_c using a "crude" meta-modelling approach consisting in selecting training samples only relying on LHS to explore the offshore conditions space. Figure 6a provides the estimation of Γ_c using a kriging model with 20 training data selected from LHS, whereas Fig. 6b provides Γ_c extracted using a kriging model with 10 training samples selected from LHS completed by 10 training data selected using the adaptive sampling strategy. Noteworthy, the adaptively selected training samples are preferably located in the vicinity of the contour, indicating that both terms A_1 and A_3 in Eqs. (5a) and (5c) (linked with the local search; see Sect. 3.2) dominate the selection criterion.

Clearly, the best estimate using the adaptive sampling strategy deviates very little from the reference. Interestingly, the total computation cost reached 5 h (20 simulations × 15 min) using a single CPU to be compared to about 1 day using the computing grid architecture of 10 CPU. The training data of the kriging model constructed with no adaptive sampling should be completed at least with 20 additional training data, i.e. based on a total number of 40 training data, to reach an agreement between the best estimate and the reference contour as good as the one reached with adaptive sampling strategy.

4.4 Four-dimensional test case

In this section, we tackle the problem of estimating the critical frontier Γ_c considering the four offshore conditions $x_c = (\xi_o; H_s; T_p; \theta)$ associated with $\xi_c = 2.15$ m at the observation point. This problem is more complex than the one tackled in Sect. 4.3 as the input space to be explored is four dimensional, hence far larger. Furthermore, contrary to Sect. 4.3, estimating the reference contour in four dimensions using a grid of moderate number of CPU is hardly achievable, as outlined in the introduction.

To demonstrate the efficiency of the meta-modelling strategy, we propose to extract a sub-set of the four-dimensional Γ_c in the $\{\xi_o; H_s\}$ space considering a fixed wave period $T_p = 7$ s and wave direction $\theta = 135°$ N. In this manner, the
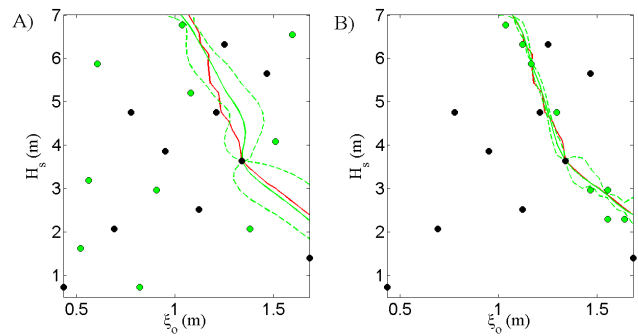
Fig. 6. Best estimate of the two-dimensional Γ_c (green solid line) in the $\{\xi_o; H_s\}$ domain. The corresponding pair of contours at confidence level of 95 % (defined as the kriging mean (Eq. 2) ± twice the square root of the variance (Eq. 3)) is represented by green dashed lines; The red solid line is the reference contour. **(A)** The training data are composed of an initial budget of 10 samples (black dots) and 10 additional samples (green dots) both selected using LHS technique. **(B)** The training data are composed of an initial budget of 10 samples (black dots, as for the case with no adaptive sampling) and 10 additional samples (green dots) adaptively selected.
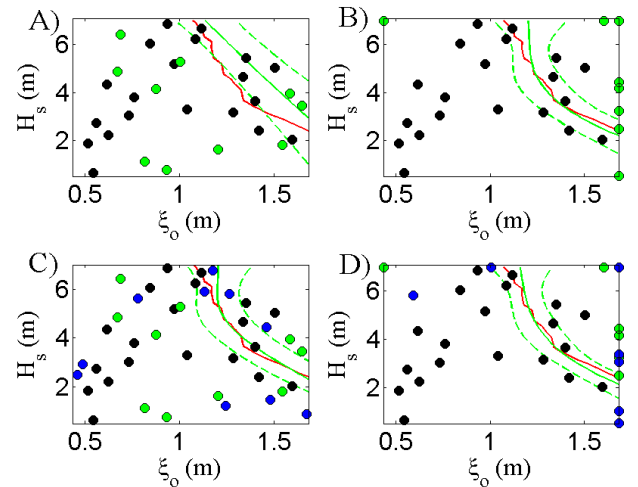
Fig. 7. Best estimate of sub-set of the four-dimensional Γ_c (green solid line) in the $\{\xi_o; H_s\}$ domain defined for fixed $T_p = 7$ s and $\theta = 135°$ N using a four-dimensional kriging model. The corresponding pair of contours at confidence level of 95% (defined as the kriging mean (Eq. 2) ± twice the square root of the variance (Eq. 3)) is represented by green dashed lines; the red solid line is the reference contour. **(A)** Sampling strategy with no adaptive selection. The training data are composed of an initial budget of 20 samples (black dots) and 10 additional samples (green dots); **(B)** sampling strategy with adaptive selection. The number of samples in **(B)** is identical to **(A)**; **(C)** sampling strategy with no adaptive selection. The training data correspond to the one of **(A)** completed by 10 additional adaptively selected training data (blue dots); **(D)** sampling strategy with adaptive selection. The training data correspond to the one of **(B)** completed by 10 additional adaptively selected training data (blue dots).
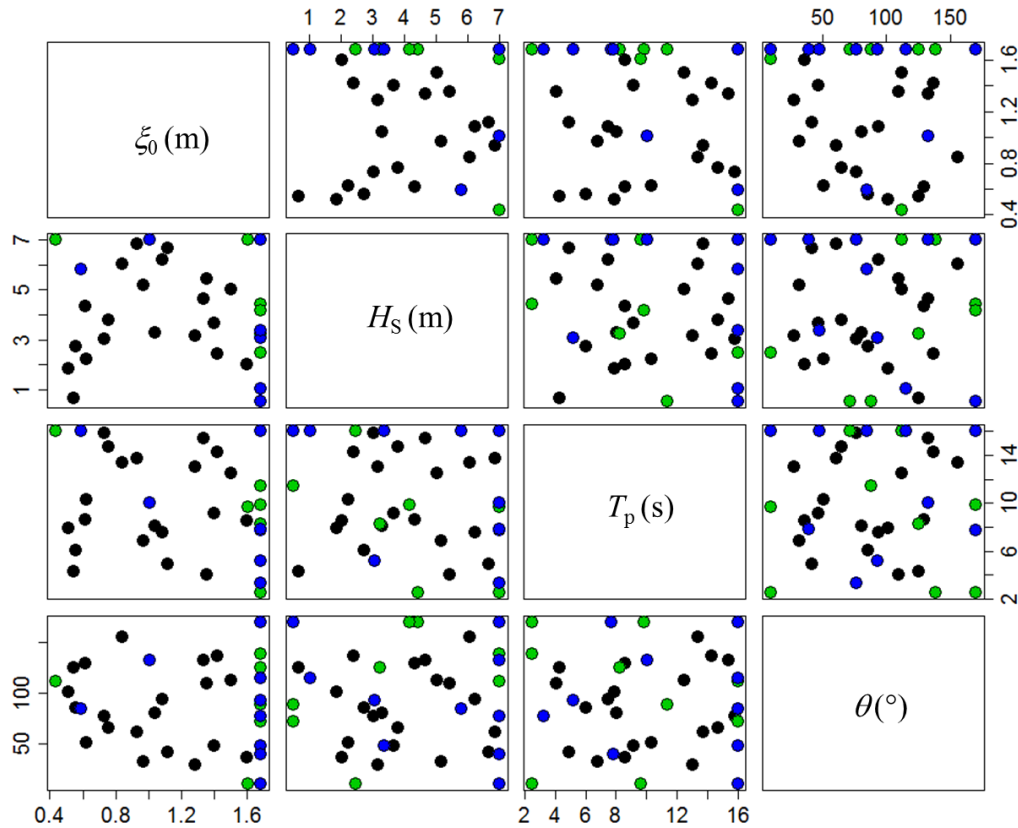
Fig. 8. Matrix of scatter plots showing the training data in the four-dimensional domain of offshore conditions: initial (black dots), 10 first adaptively selected training data (green dots), and 10 additional ones (blue dots).

"reference" contour, as calculated using the computing grid architecture, can be re-used for comparison.

An initial four-dimensional kriging model is constructed using 20 training data randomly selected within the following offshore conditions domain [0.25 ; 1.5 m]×[0.5 ; 7 m]×[2.5 ; 16 s]×[40° N ; 200° N] and using the LHS technique. The cross-validation procedure provides R^2_{CV}=91.6 % (see Appendix, Eq. A1), which demonstrates a satisfactory level of prediction quality of the kriging model.

Similarly to Sect. 4.3, two meta-modelling strategies are conducted: the first one considers no adaptive sampling and merely selects training samples using LHS. Figure 7a depicts the comparison between the sub-set of the four-dimensional Γ_c estimated in the $\{\xi_o; H_s\}$ space considering a fixed wave period $T_p = 7$ s and wave direction $\theta = 135°$ N (green solid line) and the reference contour (red solid line). Two training sample sizes are used: 10 training samples in addition to the 20 initial training ones (Fig. 7a) and 20 training samples in addition to the 20 initial training ones (Fig. 7c). For the first sample size, the visual inspection of Fig. 7a shows a rather poor agreement, the reference contour being even outside the region bounded by the pair of contours estimated at confidence level of 95 %. For the second training sample size, the agreement of the estimate with the reference contour

is equivalent to the one between the estimate using a kriging model constructed with 20 initial training samples and 10 additional adaptively selected ones (Fig. 7b). The accuracy of the approximation using the adaptive strategy is improved by increasing the number of samples (Fig. 7d).

Figure 8 provides the matrix of scatter plots showing the distribution of the selected training data in the four-dimensional domain of offshore conditions. This indicates that the second term A_2 (see Eq. 5b), linked with the global exploration of the design space, dominates the selection. The exploration of the domain preferably operates by fixing the first and second coordinates in the vicinity of the bounds ($H_s = 7$ m and $\xi_o = 1.683$ m) and by exploring the two remaining dimensions (T_p and θ).

A second validation test is performed as follows. The four-dimensional kriging model (constructed with 20 initial training samples and 20 additional adaptively selected ones, Fig. 7b, bottom) provides the best estimate of the critical frontier Γ_c. A set of new test data has then been generated by randomly selecting 60 different combinations in the four-dimensional domain of onshore conditions, which satisfy $\tilde{\xi}_m(x) = \xi_c = 2.15$ m, i.e. located on the best estimate of Γ_c. For each selected combination, a hydrodynamic simulation is effectively run. The numerically computed water levels at

the coast are then compared to $\xi_c = 2.15$ m. The root-mean-square error reaches 0.0107 m, hence demonstrating the "satisfactory" prediction quality of the kriging meta-model. Interestingly, this good performance is achieved by running the numerical model only 40 times, hence corresponding to a total computation cost of about 12 h using a single CPU. This should be compared to the computation time cost of about two weeks that would have been achieved if we had used a computation grid architecture composed of at least 500 CPU.

Finally, though the results are satisfactory using the idealized site on the Mediterranean coast, it should be underlined that for any real-case application, carefully addressing the robustness of the sequential sampling procedure with respect to the procedure of estimation of the covariance parameters of the kriging meta-model remains the key ingredient of the proposed meta-modelling strategy.

5 Concluding remarks and further works

In the present article, we described a methodology to overcome the computation challenge of extracting the critical frontier, corresponding to the set of offshore conditions leading to high water level at the coast, from complex computational, time-consuming, hydrodynamic simulations using a fine design grid. Such information is of primary importance either to constrain early warning systems based on hydro-meteorological forecast or observations, or to contribute to the return period calculation of the hazard level at the coast. Besides, it can eventually support the implementation of innovative "inverse" risk assessment methodology as recently proposed for fluvial inundation (Cunderlink and Simonovic, 2007) or for coastal flooding (Idier et al., 2010).

The proposed strategy relies on the kriging meta-modelling combined with an adaptive sampling approach using the expected improvement function developed by Ranjan et al. (2008) for contour extraction from complex computer codes. The strategy was applied on a simplified study site at the Mediterranean coast in southern France. The efficiency of the meta-modelling strategy was demonstrated by comparing the results: (1) from a strategy relying on computing grid architecture composed of a moderate number of CPU; (2) from a "crude" meta-modelling approach with no adaptive sampling. For both test cases (two and four dimensional), we showed that the critical offshore conditions can be estimated with a good level of accuracy compared to the reference contour using a very limited number (of a few tens) of computationally intensive hydrodynamic simulations.

Yet, the computation provided in this study was performed for a single observation point at the coast, whereas risk management of coastal flooding requires the spatial extent of the inundation. In future research, it would be worthwhile to address this spatial issue relying for instance on meta-models with spatially distributed outputs and taking advantage of the recent advances in the meta-modelling community (see e.g. Marrel et al., 2011).

Appendix A

Leave-one-out cross validation procedure

The "leave-one out" cross validation technique (see e.g. Hastie et al., 2009) involves using a single observation from the initial training data as the validation data, and the remaining samples as the new training data (i.e. composed of the original samples minus the one selected). A meta-model is then constructed with the new training data to predict the validation data. This procedure is repeated so that each observation in the initial training data is used once as validation data. The coefficient of determination R^2_{CV} for the procedure can be computed as follows:

$$R^2_{CV} = 1 - \frac{\sum\limits_{i=1}^{n_0} (\tilde{\xi}_{mi} - \xi_{mi})^2}{\sum\limits_{i=1}^{n_0} (\xi_{mi} - \bar{\xi}_m)^2} \tag{A1}$$

where ξ_{mi} corresponds to the observations in the validation set (for i from 1 to n_0), $\bar{\xi}_m$ to the corresponding mean and $\tilde{\xi}_{mi}$ to estimated values using the meta-model.

The coefficient R^2_{CV} provides a metric of the quality of prediction so that a coefficient R^2_{CV} close to 1 indicates that the meta-model is successful in matching the validation data.

Appendix B

Matérn correlation model

Let us define the vectors $\boldsymbol{u}=(u_1; u_2; \ldots; u_d)$ and $\boldsymbol{v}=(v_1; v_2; \ldots; v_d)$ of dimension d so that $\boldsymbol{h}=\boldsymbol{u}-\boldsymbol{v}=(h_1; h_2; \ldots; h_d)$. The correlation function between $\boldsymbol{u}$ and $\boldsymbol{v}$ can be defined as $R(\boldsymbol{u}; \boldsymbol{v}) = \prod\limits_{i=1}^{d} g(h_i; \omega_i)$ where ω is the vector of length-scales.

Following the Matérn model $\nu=3/2$, the function g holds as follows (with $i=1; 2; \ldots; d$):

$$g(h_i; \omega_i) = (1 + \sqrt{3}|h_i|/\omega_i) \cdot \exp(-\sqrt{3}|h_i|/\omega_i). \tag{B1}$$

Under this model, the process is once differentiable.

Following the Matérn model $\nu=5/2$, the function g holds as follows (with $i=1; 2; \ldots; d$):

$$g(h_i; \omega_i) = (1 + \sqrt{5}|h_i|/\omega_i + 5h_i^2/(3\omega_i^2)) \tag{B2}$$
$$\cdot \exp(-\sqrt{5}|h_i|/\omega_i).$$

Under this model, the process is twice differentiable.

Appendix C

Universal kriging equations

We introduce here the basic concepts of kriging meta-modelling. For a more complete introduction to kriging meta-modelling and full derivation of equations, the interested reader can refer to Sacks et al. (1989); Jones et al. (1998) and Jones (2001); Forrester et al. (2008).

Let us consider the deterministic response of the simulator $\xi_m = f(x)$ as a realization of a Gaussian stochastic process F so that $f(x)=F(x,\omega)$, where ω belongs to the underlying probability space Ω. In the following, we use the notation $F(x)$ for the process and $F(x,\omega)$ for one realization. The process F results from the summation of two terms:

- $f_0(x)$, the deterministic mean function, which takes the general form $f_0\ (x) = \sum_{i=1}^{p} \beta_i \cdot h_i\ (x)$, where $h(x)=(h_1(x)\ ;\ h_2(x);\ \ldots;\ h_p(x))$ is a vector of p linearly independent known functions (named basis functions), and $\boldsymbol{\beta}$ is a vector of unknown real coefficients;

- $Z(x)$, the Gaussian-centred stationary stochastic process characterized by a zero mean and the covariance matrix $\mathbf{C}$, which depends on the variance σ_Z^2 and on the correlation function R, which governs the degree of correlation through the use of the vector of length-scale parameters $\boldsymbol{\omega}$ between any input vectors.

The covariance between $Z(u)$ and $Z(v)$ is then expressed as $C(u;v) = \sigma_Z^2 \cdot R(u;v)$, where $u=(u_1;u_2;\ldots;u_d)$ and $v=(v_1; v_2;\ldots;v_d)$ are two input vectors of dimension d.

Let us define $\mathbf{X}_D$ the design matrix composed of the vectors of offshore conditions x selected in step 1 of the methodology (i.e. the training samples) to be simulated so that $\mathbf{X}_D=(x^{(1)}\ ;\ x^{(2)};\ \ldots\ ;\ x^{(n0)})$ and $\boldsymbol{\xi}_D$ the vector of simulated maximum water levels at the coast associated with each selected training samples so that $\boldsymbol{\xi}_D=(\xi_m^{(1)} = f(x^{(1)})\ ; \xi_m^{(2)} = f(x^{(2)})\ ;\ldots; \xi_m^{(n0)} = f(x^{(n0)}))$.

Under the afore-described assumptions, the distribution of the maximum water level for a new input vector of offshore conditions x^* follows a Gaussian distribution conditional on the design matrix $\mathbf{X}_D$ and of the corresponding simulated water levels ξ_D with expected value $\tilde{\xi}_m$ for the new configuration x^* and variance s^2 respectively defined by the universal kriging Eqs. (C1) and (C2):

$$\tilde{\xi}_m\ (x^*) = h(x^*)^T \cdot \hat{\boldsymbol{\beta}} + c(x^*)^T \cdot \mathbf{C}_D^{-1} \cdot (\boldsymbol{\xi}_D - \mathbf{H}_D \cdot \hat{\boldsymbol{\beta}}) \quad \text{(C1)}$$

with $\hat{\boldsymbol{\beta}} = (\mathbf{H}_D^T \cdot \mathbf{C}_D^{-1} \cdot \mathbf{H}_D)^{-1} \mathbf{H}_D^T \cdot \mathbf{C}_D^{-1} \cdot \boldsymbol{\xi}_D$ the vector of generalised least square estimates of $\boldsymbol{\beta}$.

$$s^2(x^*) = \sigma_Z^2 - c(x^*)^T \cdot \mathbf{C}_D^{-1} \cdot c(x^*) + (h(x^*)^T - c(x^*)^T \cdot \mathbf{C}_D^{-1} \cdot \mathbf{H}_D)$$
$$\times (\mathbf{H}_D^T \cdot \mathbf{C}_D^{-1} \cdot \mathbf{H}_D)^{-1} \cdot (h(x^*)^T - c(x^*) \cdot \mathbf{C}_D^{-1} \cdot \mathbf{H}_D)^T \quad \text{(C2)}$$

where $\mathbf{H}_D=[h(x^{(1)})\ ;\ h(x^{(2)})\ ;\ \ldots\ ;\ h(x^{(n0)})]^T$ is the design matrix of n_0 rows and p columns, and $c(x^*)$ is the covariance vector between the test candidate x^* and the training samples.

Note that the above equations can be easily written using the correlation matrix $\mathbf{R}_D$ using its relationship between $\mathbf{C}_D$ and σ_Z^2.

The specific case, where the basis functions reduce to a unique constant function, corresponds to the "ordinary kriging" as described in Eqs. (2) and (3).

References

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E.: Sequential design of computer experiments for the estimation of a probability of failure, Stat. Comput., 22, 773–793, doi:10.1007/s11222-011-9241-4, 2012.

Bertin, X., Bruneau, N., Breilh, J. F., Fortunato, A., and Karpytchev, M.: Importance of wave age and resonance in storm surges : the case of Xynthia, Ocean Model., 42, 16–30, 2012.

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.

Bonneton, P., Barthelemy, E., Chazel, F., Cienfuegos, R., Lannes, D., Marche, F., Tissier, M.: Recent advances in Serre-Green Naghdi modelling for wave transformation, breaking and runup processes, Eur. J. Mech. B.-Fluid, 30, 589–597, 2011.

Booij, N., Ris, R. C., and Holthuijsen, L. H.: A third-generation wave model for coastal regions, Part I: Model description and validation, J. Geophys. Res., 104, C4, 7649–7666, 1999.

Bucher, C. G. and Bourgund, U.: Fast and efficient response surface approach for structural reliability problems, Struct. Saf. 7, 57–66, 1990.

Chini, N. and Stansby, P.: Extreme values of coastal wave overtopping accounting for sea level rise and climate change, Coast. Eng., 65, 27–37, doi:10.1016/j.coastaleng.2012.02.009, 2012

Cunderlink, J. M. and Simonovic, S. P.: Inverse flood risk modelling under changing climatic conditions, Hydrol. Process., 21, 563–577, 2007.

Fewtrell, T. J., Duncan, A., Sampson, C. C., Neal, J. C., and Bates, P. D.: Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data, Phys. Chem. Earth, 36, 281–291, 2011.

Forrester, A. I. J., Sobester, A., and Keane, A. J.: Engineering Design via Surrogate Modelling: A Practical Guide, John Wiley and Sons Ltd., Chichester, UK, 2008.

Forrester, A. I. J. and Keane, A. J.: Recent advances in surrogate-based optimization, Progr. Aerospace Sci., 45, 50–79, 2009.

Garrity, N. J., Battalio, R., Hawkes, P. J., and Roupe, D.: Evaluation of the event and response approaches to estimate the 100-year coastal flood for Pacific coast sheltered waters. Proceeding of 30th International Conference of Coast. Eng., ASCE, 1651–1663, 2006.

Ghanem, R. G. and Spanos, P. D.: Stochastic finite elements-a spectral approach, Springer Verlag, New York, 214 pp., 1991.

Haldar, A. and Mahadevan, S.: Probability, reliability, and statistical methods in engineering design, Wiley, New York, 320 pp., 2000.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edn., Springer-Verlag, New York, 746 pp., 2009.

Idier, D., Rohmer, J., Turpin, V., and Magnan, A.: A risk based approach for coastal risk assesment : development of an inverse methodology, Proc. of Journées Impacts du Changement Climatique sur les Risques Côtiers, Orléans, France, 15–16 November 2010.

Iooss, B., Van Dorpe, F., and Devictor, N.: Response surfaces and sensitivity analyses for an environmental model of dose calculations, Reliability Eng. Syst. Saf., 91, 1241–1251, 2006.

Jones, D. R., Schonlau, M., and Welch, W. J.: Efficient Global Optimization of Expensive Black-Box Functions, J. Global Optim., 13, 455–492, 1998.

Jones, D. R.: A Taxonomy of Global Optimization Methods Based on Response Surfaces, J. Global Optim., 21, 345–383, 2001.

Koehler, J. R. and Owen, A. B.: Computer experiment, in: Handbook of Statistics, edited by: Ghosh, S. and Rao, C. R., Elsevier Science, NewYork, USA, 13, 261–308, 1996.

Krige, D.: A statistical approach to some basic mine valuation problems on the Witwatersrand, J. of the Chem., Metal. Mining Soc. South Africa, 52, 119–139, 1951.

Langewisch, D. R. and Apostolakis, G. E.: A comparison of polynomial response surfaces and Gaussian processes as metamodels for uncertainty analysis with long-running computer codes, Proceedings of 10th International Probabilistic Safety Assessment & Management Conference, Seattle, WA, USA, 7–11 June 2010.

Marrel, A., Iooss, B., Jullien, M., Laurent, B., and Volkova, E.: Global sensitivity analysis for models with spatially dependent outputs, Environmetrics, 22, 383–397, 2011.

McCabe, M., Stansby, P. K., and Apsley, D. D.: Coupled wave action and shallow-water modelling for random wave runup on a slope, J. Hydr. Res. (special issue Coastal and Maritime Hydraulics), 49, 515–522, doi:10.1080/00221686.2011.566253, 2011.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979.

Nicholls, R. J., Hanson, S. E., Lowe, J. A., Vaughan, D. A., Lenton, T., Ganopolski, A., Tol, R. S. J., and Vafeidis, A. T.: Improving methodologies to assess the benefits of policies to address sea-level rise, Report to OECD, Paris, 25 June 2006.

Papadrakakis, M. and Lagaros, N. D.: Reliability-based structural optimization using neural networks and Monte Carlo simulation, Comput. Methods in Appl. Mech. Eng., 191, 32, 3491–3507, 2002.

Park, J. and Baek, J.: Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential covariogram, Comput. Geosci., 27, 1–7, 2001.

Picheny, V. and Ginsbourger, D.: R cran module KrigInv: Kriging-based Inversion for Deterministic and Noisy Computer Experiments, http://cran.r-project.org/web/packages/KrigInv/index.html, 2012.

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org/, 2011.

Ranjan, P., Bingham, D., and Michailidis, G.: Sequential Experiment Design for Contour Estimation from Complex Computer Codes, Technometrics, 50, 527–541, 2008.

Rohmer, J. and Bouc, O.: A response surface methodology to address uncertainties in cap rock failure assessment for CO_2 geological storage in deep aquifers, Int. J. Greenhouse Gas Control, 4, 198–208, 2010.

Rohmer, J. and Foerster, E.: Global sensitivity analysis of large scale landslide numerical models based on the Gaussian Process meta-modelling, Comput. Geosci., 37, 917–927, 2011.

Roustant, O., Ginsbourger, D., and Deville, Y.: R cran module DiceKriging: Kriging methods for computer experiments, available at: http://cran.r-project.org/web/packages/DiceKriging/index.html, 2012.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P.: Design and analysis of computer experiments, Stat. Sci., 4, 409–435, 1989.

Safecoast: Coastal Flood Risk and trends for the Future in the North Sea Region, synthesis report, The Hague, 136 pp., 2008.

Samuels, P., Huntington, S., Allsop, W., and Harrop, J.: Flood risk management: research and practice, CRC Press, Oxford, UK, 332 pp., 2008.

Santner, T., Williams, B., and Notz, W.: The Design and Analysis of Computer Experiments, Springer, New York, USA, 283 pp., 2003.

Stein, M. L.: Interpolation of Spatial Data. Some Theory for Kriging, Springer, New York USA, 247 pp., 1999.

Stockdon, H. F., Holman R. A., Howd P. A., and Sallenger, A. H.: Empirical parameterization of setup, swash, and runup, Coast. Eng., 53, 573–588, 2006.

Storlie, C. B., Swiler, L. P., Helton, J. C., and Sallaberry, C. J.: Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models, Reliability Engineering and System Safety, 94, 1735–1763, 2009.

Tukey, J. W.: Unsolved Problems of Experimental Statistics, J. Am. Stat. Assoc., 49, 706–731, 1954.

UNEP/GRID-Arendal: Coastal population and altered land cover in coastal zones (100 km of coastline), UNEP/GRID-Arendal Maps and Graphics Library, http://maps.grida.no/go/graphic/coastal-population-and-altered-land-cover-in-coastal-zones-100-km-of-coastline, 2007.