

Implementation/optimization of moving least squares response surfaces for approximation of hurricane/storm surge and wave responses

Alexandros A. Taflanidis · Gaofeng Jia · Andrew B. Kennedy · Jane M. Smith

Received: 1 November 2012 / Accepted: 11 December 2012 / Published online: 27 December 2012
© Springer Science+Business Media Dordrecht 2012

Abstract One of the important recent advances in the field of hurricane/storm modelling has been the development of high-fidelity numerical simulation models for reliable and accurate prediction of wave and surge responses. The computational cost associated with these models has simultaneously created an incentive for researchers to investigate surrogate modelling (i.e. metamodeling) and interpolation/regression methodologies to efficiently approximate hurricane/storm responses exploiting existing databases of high-fidelity simulations. Moving least squares (MLS) response surfaces were recently proposed as such an approximation methodology, providing the ability to efficiently describe different responses of interest (such as surge and wave heights) in a large coastal region that may involve thousands of points for which the hurricane impact needs to be estimated. This paper discusses further implementation details and focuses on optimization characteristics of this surrogate modelling approach. The approximation of different response characteristics is considered, and special attention is given to predicting the storm surge for inland locations, for which the possibility of the location remaining dry needs to be additionally addressed. The optimal selection of the basis functions for the response surface and of the parameters of the MLS character of the approximation is discussed in detail, and the impact of the number of high-fidelity simulations informing the surrogate model is also investigated. Different normalizations of the response as well as choices for the objective function for the optimization problem are considered, and their impact on the accuracy of the resultant (under these choices) surrogate model is examined. Details for implementation of the methodology for efficient coastal risk assessment are reviewed, and the influence in the analysis of the model prediction error introduced through the surrogate modelling is discussed. A case study is provided, utilizing a recently developed database of high-fidelity simulations for the Hawaiian Islands.

A. A. Taflanidis (✉) · G. Jia · A. B. Kennedy
Department of Civil and Environmental Engineering and Earth Sciences,
University of Notre Dame, 156 Fitzpatrick Hall, Notre Dame, IN 46556, USA
e-mail: a.taflanidis@nd.edu

J. M. Smith
US Army Corps of Engineers, Research and Development Center, 3909 Halls Ferry Road,
Vicksburg, MS 39180, USA

Keywords Hurricane wave and surge · Coastal hazard · Storm surge · Tropical cyclones · Response surface approximations · Surrogate modelling optimization · Risk assessment

1 Introduction

Hurricane/storm risk assessment has received increased attention in the past decade, partly in response to the destructive 2004, 2005 and 2008 hurricane seasons (Dietrich et al. 2010; Kennedy et al. 2011a, b). A significant recent advance in this field has been the development of high-fidelity numerical simulation models for reliable and accurate prediction of surge responses for specific hurricane events or regions (Resio and Westerink 2008). These models permit a detailed representation of the hydrodynamic processes, albeit at a greatly increased computational effort, since they typically require thousands of computational hours for each simulation. This high computational cost has created the incentive to adopt surrogate modelling (also frequently referenced as metamodeling) and interpolation/regression methodologies that use information from existing high-fidelity simulations, that is, a database, to efficiently approximate hurricane/storm responses for different scenarios that do not belong in this database. This has been further motivated by the fact that various such databases are constantly created and updated for regional flooding and coastal hazard studies (Resio et al. 2007; Niedoroda et al. 2010; Kennedy et al. 2012).

Though different potential implementations of the aforementioned approximation approaches exist, one of the most important ones (Song et al. 2012) is within the context of the joint probability method (JPM) hurricane risk assessment framework (Ho and Myers 1975; Myers 1975), a framework that has become increasingly popular in recent years (Resio et al. 2009; Toro et al. 2010a, b; Condon and Sheng 2012). JPM relies on a simplified description of hurricane scenarios through a small number of model parameters. Description of the uncertainty in these parameters, through appropriate probability models, leads to a probabilistic characterization of the coastal risk. This risk is ultimately expressed as a probabilistic integral over the uncertain parameter space, and its estimation requires numerical evaluation of the hurricane responses for a large number of scenarios resulting from the adopted probabilistic description for the model parameters (Toro et al. 2007; Resio et al. 2009; Toro et al. 2010a). Adoption of the aforementioned high-fidelity models significantly increases the associated computational cost, making surrogate modelling approximations a critical tool (but note not the only available approach) for efficient implementation of JPM.

Motivated by these facts, low-cost, dimensional surge response functions were developed by Irish et al. (2009), but they only addressed the variation with respect to hurricane storm size, intensity, and track and were restricted to hurricane surge only and limited to specific locations of interest on the Texas coast. Udoh and Irish (2011) presented preliminary discussions for extending these surge response functions to address additional hurricane model parameters, the forward speed and heading, whereas Song et al. (2012) recently investigated the influence of regional changes in bathymetry on the surge response functions. Das et al. (2010) developed a methodology for selecting the most appropriate storm within some given database, without extending their approach to interpolation within the database. In Taflanidis et al. (2012), the use of response surface approximations (RSAs) was proposed as a surrogate model. Though this approach cannot provide the physical insight that the aforementioned dimensional surge response functions can, it has some important advantages: (1) responses for a given hurricane scenario may be calculated

rapidly for hundreds of thousands of locations in the coastal region of interest, and for any modelled quantity representing hurricane impact (e.g. surge, wave height, runup), (2) variation for any parameter used to describe the hurricane characteristics may be considered, and (3) new information (when available) can be directly and immediately added in the database of high-fidelity simulations to improve accuracy. In the aforementioned study by Taflanidis et al. (2012), though, the direct use of RSAs was considered, adopting existing general guidelines for their implementation. There was no effort to investigate the optimization of their characteristics (such as selection of basis functions or normalization techniques of the response) for the specific application of interest, to discuss validation procedures or to address potential differences in the implementation for different hurricane/storm outputs.

This paper seeks to address this latter knowledge gap by discussing the systematic implementation and more importantly the optimization of moving least squares (MLS) RSAs for hurricane wave and surge predictions. We offer significant advancements in that (1) the optimal selection of the basis functions for the response surface and of the parameters of the MLS character of the approximation is examined, (2) the impact of the number of high-fidelity simulations informing the surrogate modelling is investigated, (3) different normalizations of the response are proposed to improve accuracy, (4) special attention is given to predicting the surge in inland locations for which the possibility of the location remaining dry needs to be additionally addressed, and (5) the model prediction error introduced through the surrogate modelling is explicitly optimized, and its potential impact in the analysis is investigated. As a case study, an application to the Island of Oahu in Hawaii is discussed, using the high-fidelity simulation database described in Kennedy et al. (2012).

In the following section, the general hurricane modelling approach is presented. In Sect. 3, RSAs are first reviewed, with enough information to support the understanding/insight for the optimization problem introduced later, and then, details for implementation for the problem of interest are discussed, extending to both the prediction of the surge in inland locations as well as to statistical approaches for quantifying the accuracy of RSAs for problems involving a large number of response quantities. Section 4 introduces the optimization framework, including both the formulation of the objective function and numerical tools for solving the associated optimization problem, and finally, Sect. 5 discusses the case study considered and investigates different trends in the RSA implementation.

2 Modelling and approximation assumptions

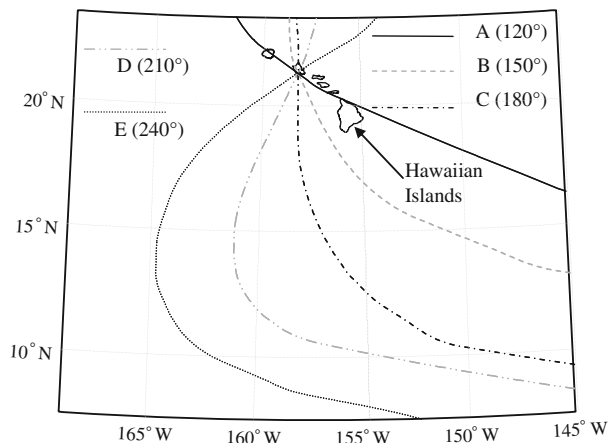
The foundation for the approximation of hurricane/storm responses through surrogate/interpolation models (Resio et al. 2009) is the parameterization of each hurricane/storm event by a small number of model parameters, corresponding to its characteristics at landfall such as (1) landfall location x_o , (2) track heading θ , (3) central pressure c_p , (4) forward speed v_f , (5) radius of maximum winds R_m , (6) tide level e_r , and (7) Holland B parameter B . Typically, a constant tide level is assumed, and the Holland B term is related (using regional approximations) to the rest of the hurricane parameters (Resio et al. 2009), leading to the following definition of the model parameter vector, \mathbf{x} , describing each hurricane scenario (for both the high-fidelity and surrogate models)

$$\mathbf{x} = [x_o \quad \theta \quad c_p \quad v_f \quad R_m]^T \quad (1)$$

where $[\]^T$ denotes a matrix transpose. Note that the discussions in this paper can be directly extended to include the tide level and Holland term (or any other appropriate hurricane characteristic) as model parameters. The implicit assumption is, though, that \mathbf{x} is low dimensional (includes <8–10 parameters). This is necessary for facilitating an efficient description through a surrogate model, without requiring an overly large database to adequately describe the variability with respect to all the model parameters. In this context, the temporal and spatial variability of the hurricane track and characteristics prior to landfall are typically addressed (Resio et al. 2009) by appropriate selection of the track-history prior to landfall (e.g. giving the track a typical curvature over time for the appropriate landfall heading), so that expected variations based on historical data are adequately described. Figure 1 shows the different tracks from the study by Kennedy et al. (2012) that will be later used in the illustrative example.

For characterizing, now, the hurricane/storm response in the coastal region examined multiple variables may be of interest. Examples include (1) the storm surge (ζ) (still water level, defined as the average sea level over a several minute period), (2) the significant wave height (H_s) (possibly along with the corresponding peak period T_p), (3) the wave runup level) defined as the sea level including runup of wind waves on the shore, or (4) the time that normally dry locations are inundated. Response variables (1)–(3) may refer to maximum responses over the entire hurricane history, or to responses at specific time instances prior to landfall. In this setting, let $\mathbf{z}(\mathbf{x})$ denote the vector of response quantities of interest (wave, surge, and so forth) throughout all locations of interest. This vector, used ultimately to describe the hurricane/storm impact, will be referenced herein as the response vector. The i th component of vector $\mathbf{z}(\mathbf{x})$ is denoted by $z_i(\mathbf{x})$ and pertains to a specific response variable [any of the (1)–(4)] for a specific coastal location. The augmentation of these responses ultimately provides the n_z -dimensional vector $\mathbf{z}(\mathbf{x})$. One significant advantage of the RSA advocated here is that no constraint is imposed for n_z (i.e. multiple response outputs can be simultaneously addressed). Also, no special characteristics are exploited for $\mathbf{z}(\mathbf{x})$ thus providing versatility to the developed methodologies. Such characteristics could have been, for example, assumptions for specific variation patterns with respect to the components of \mathbf{x} (as is utilized in dimensional response function approximations).

Fig. 1 Basic storm tracks (A–E) considered in the study (Kennedy et al. 2012). In *parenthesis*, the angle of final approach (track heading) θ (clockwise from South) is indicated



For a specific hurricane/storm scenario, described by the model parameter vector \mathbf{x} , $\mathbf{z}(\mathbf{x})$ can be accurately estimated by numerical simulation, once an appropriate high-fidelity model is established (Resio and Westerink 2008; Westerink et al. 2008). Surrogate modelling methodologies can provide an approximation to $\mathbf{z}(\mathbf{x})$ using an available database of such high-fidelity simulations. If $\hat{\mathbf{z}}(\mathbf{x})$ denote this approximation to $\mathbf{z}(\mathbf{x})$, then the relationship between $z_i(\mathbf{x})$ and $\hat{z}_i(\mathbf{x})$ is ultimately

$$z_i(\mathbf{x}) = \hat{z}_i(\mathbf{x}) + \varepsilon_i \tag{2}$$

where ε_i is the prediction error between the surrogate model and the high-fidelity model, assumed to be zero mean (since the contrary indicates a bias) Gaussian random variable with standard deviation σ_{ε_i} . This choice of probability distribution incorporates the largest amount of uncertainty (Taflanidis and Beck 2010), in terms of Information Entropy, under the constraints that only the mean and variance are known. The standard deviation σ_{ε_i} , assumed herein to be independent of \mathbf{x} (note that different assumptions could be made, that would though increase the complexity of estimating this standard deviation), can be finally approximated by comparison of the high-fidelity and surrogate models over a set of hurricane scenarios chosen to serve as validation points. This then completely defines the prediction error ε_i and the relationship between $z_i(\mathbf{x})$ and $\hat{z}_i(\mathbf{x})$ in a probabilistic sense. Note that additional modelling errors can be incorporated in the relationship (2), for example the errors introduced by the approximate description of hurricane by a small dimensional vector \mathbf{x} (Resio et al. 2009; Taflanidis et al. 2012) or the errors associated with the assumptions made about tides.

As discussed in the introduction, an important application of these concepts for modelling and approximation of hurricane/storm impacts is within the context of probabilistic hurricane risk assessment. Appendix 1 reviews briefly this implementation, focusing on the implications of using surrogate modelling methodologies and addressing the prediction error ε_i .

3 Response surfaces for hurricane/storm response approximation

3.1 Problem formulation

Let $\mathbf{x} = [x_1 \cdots x_{n_x}] \in \mathfrak{R}^{n_x}$ be the n_x dimensional vector of free variables defining the hurricane characteristics [with $n_x = 5$ for definition (1)], $\mathbf{z} = [z_1 \cdots z_{n_z}] \in \mathfrak{R}^{n_z}$ the augmented vector of response quantities of interest, and assume that a database is available consisting of N_h evaluations (frequently referenced as observations) of the response vector $\{\mathbf{z}^h; h = 1, \dots, N_h\}$ for different hurricane scenarios $\{\mathbf{x}^h; h = 1, \dots, N_h\}$, obtained through an appropriate high-fidelity simulation model. An important characteristic for this database is the need to adequately describe the variability with respect to all hurricane parameters (i.e. all components of vector \mathbf{x}). When developing such databases, this is established by considering for each hurricane characteristic x_i a sufficient number of parameter values and then combining these values for all components of \mathbf{x} to create a canonical grid, excluding combinations that are deemed improbable based on regional historical data. Further insight for this task may be found in (Resio et al. 2009).

Once the database is established, response surface methodologies can use its information to provide an approximation to \mathbf{z} . For this purpose, the database is partitioned into three different subsets (not necessarily exclusive), a larger subset (this is the main subset)

of N_S simulations $\{\mathbf{x}^I; I = 1, \dots, N_S\}$ used to establish the surrogate model (frequently called support points), and two other small subsets, one with N_c simulations used for optimizing the characteristics (tuning the parameters) of the response surface (frequently called control points) $\{\mathbf{x}^c; c = 1, \dots, N_c\}$, and one with N_E simulations $\{\mathbf{x}^p; p = 1, \dots, N_E\}$ used to evaluate the accuracy of the surrogate model (frequently called validation points). Note that when finally implementing the RSA (i.e. not at the development stage), the control and validation points should be added to the support points (expand the available database for informing RSA).

Instead of directly providing an estimate for \mathbf{z} , the response surface can be established for a normalized or transformed response variable vector $\mathbf{y} = [y_1 \cdots y_{n_z}] \in \mathfrak{R}^{n_z}$, where the transformation is intended to improve accuracy and avoid numerical problems, for example when \mathbf{z} is strictly a positive quantity but it is difficult to establish a surrogate model that can satisfy this requirement. Two common transformations we will investigate are

$$y_i = \frac{z_i - \mu_i}{\sigma_i}$$

$$y_i = \log\left(\frac{z_i}{\sigma_i}\right); \quad z_i > 0 \tag{3}$$

where μ_i and σ_i are the mean and standard deviation, respectively, for z_i over the observation set $\{z_i^h; h = 1, \dots, N_h\}$

$$\mu_i = \frac{1}{N_h} \sum_{h=1}^{N_h} z_i^h$$

$$\sigma_i = \sqrt{\frac{1}{N_h} \sum_{h=1}^{N_h} (z_i^h - \mu_i)^2} = \sqrt{\frac{1}{N_h} \sum_{h=1}^{N_h} (z_i^h)^2 - \mu_i^2} \tag{4}$$

Note that the first transformation in (3) leads to a zero mean (over the observation set) variable y_i with unit standard deviation, whereas the second can be implemented only when z_i is strictly positive. These two transformations will be referenced herein as linear and log, respectively. They are both invertible, thus provide a one-to-one correspondence between each y_i and z_i , and once the approximation for y_i has been established, the corresponding z_i can be directly calculated.

3.2 Review of moving least squares response surface approximation for a scalar output

Response surfaces (Myers and Montgomery 2002) express any $y_i(\mathbf{x})$, through $j = 1, \dots, N_B$ preselected basis functions $b_j(\mathbf{x}): \mathfrak{R}^{n_x} \rightarrow \mathfrak{R}$ through introduction of coefficients $a_{ij}\{\mathbf{x}\}$, that can be constant or depend on the location \mathbf{x} for which the interpolation is established. The notation $\{\cdot\}$ is introduced to denote the latter potential dependence, and the RSA is ultimately expressed as

$$\hat{y}_i(\mathbf{x}) = \sum_{j=1}^{N_B} b_j(\mathbf{x})a_{ij}\{\mathbf{x}\} = \mathbf{b}(\mathbf{x})^T \mathbf{a}_i\{\mathbf{x}\} \tag{5}$$

where $\mathbf{b}(\mathbf{x})$ is the vector of basis functions and $\mathbf{a}_i\{\mathbf{x}\}$ is the vector containing the coefficients for the basis functions. A common choice for basis functions is a full second order polynomial

$$\hat{y}_i(\mathbf{x}) = a_{i0}\{\mathbf{x}\} + \sum_{j=1}^{n_x} a_{ij}\{\mathbf{x}\}x_j + \sum_{j=1}^{n_x} \sum_{k \geq j}^{n_x} a_{ijk}\{\mathbf{x}\}x_jx_k; \quad N_B = \frac{n_x(n_x + 3) + 2}{2} \quad (6)$$

leading to the following definitions for vectors \mathbf{b} and \mathbf{a}_i .

$$\begin{aligned} \mathbf{b}(\mathbf{x}) &= [1 \quad x_1 \quad \cdots \quad x_{n_x} \quad x_1^2 \quad x_1x_2 \quad \cdots \quad x_{n_x}^2] \\ \mathbf{a}_i\{\mathbf{x}\} &= [a_{i0}\{\mathbf{x}\} \quad a_{i1}\{\mathbf{x}\} \quad \cdots \quad a_{in_x}\{\mathbf{x}\} \quad a_{i11}\{\mathbf{x}\} \quad a_{i12}\{\mathbf{x}\} \quad \cdots \quad a_{in_xn_x}\{\mathbf{x}\}] \end{aligned} \quad (7)$$

The coefficients $\mathbf{a}_i\{\mathbf{x}\}$ are calculated by selecting a set of $N_S > N_B$ support points, $\{\mathbf{x}^l; l = 1, \dots, N_S\}$ (subset of the available high-fidelity simulations) and by minimizing a weighted mean squared error over these points between $y_i(\mathbf{x})$ and its approximation, leading to the following choice (Taflanidis et al. 2012).

$$\mathbf{a}_i\{\mathbf{x}\} = \mathbf{M}^{-1}\{\mathbf{x}\}\mathbf{L}\{\mathbf{x}\}\mathbf{F}_i \quad (8)$$

where the following quantities were defined

$$\begin{aligned} \mathbf{M} &= \mathbf{B}^T\mathbf{W}\{\mathbf{x}\}\mathbf{B} \quad \text{and} \quad \mathbf{L}\{\mathbf{x}\} = \mathbf{B}^T\mathbf{W}\{\mathbf{x}\} \\ \mathbf{B} &= [\mathbf{b}(\mathbf{x}^1) \cdots \mathbf{b}(\mathbf{x}^{N_S})]^T; \quad \mathbf{F}_i = [y_i(\mathbf{x}^1) \cdots y_i(\mathbf{x}^{N_S})]^T \\ \mathbf{W}\{\mathbf{x}\} &= \text{diag}[w(d(\mathbf{x}; \mathbf{x}^1)) \quad w(d(\mathbf{x}; \mathbf{x}^{N_S}))] \end{aligned} \quad (9)$$

and $\text{diag}[\cdot]$ stands for a diagonal matrix. $\mathbf{W}\{\cdot\}$ ultimately corresponds to the diagonal matrix of weights $w(d(\mathbf{x}; \mathbf{x}^l))$ for each support point \mathbf{x}^l , and these weights are dependent on the distance d between that point and the point \mathbf{x} for which the approximation is established

$$d(\mathbf{x}; \mathbf{x}^l) = \sqrt{\sum_{k=1}^{n_x} (\mathbf{x}_k - \mathbf{x}_k^l)^2 v_k^2} = \sqrt{[\mathbf{v}(\mathbf{x} - \mathbf{x}^l)]^T \mathbf{v}(\mathbf{x} - \mathbf{x}^l)}; \quad \mathbf{v} = \text{diag}(v_1 \cdots v_{n_x}) \quad (10)$$

with v_k representing the relative weight for each component of x_k (Taflanidis 2012).

The introduction of the distance-dependent weights establishes for each configuration \mathbf{x} a weighted local averaging of the information obtained by the support points that are closer to it. Without these weights, the coefficient vector, \mathbf{a}_i , would be constant over the whole domain for \mathbf{x} , which means that a global approximation would be established (global least squares). The accuracy of global approximations depends significantly on the selection of the basis functions, which should be chosen to resemble $y_i(\mathbf{x})$ as closely as possible, something that is not always straightforward. The MLS circumvents such problems by establishing the aforementioned weighted local approximation for $\mathbf{a}_i\{\mathbf{x}\}$ around each point in the interpolation domain (Breitkopf et al. 2005). Still, the efficiency of the MLS interpolation depends on the weighting function chosen, which should prioritize support points that are close to the approximation point and should vanish after an influence radius D_d . This radius should be selected so that a sufficient number of neighbouring supporting points are included to avoid singularity in the solution for $\mathbf{a}_i\{\mathbf{x}\}$, or equivalently so that \mathbf{M} in (9) is invertible (this means that at least N_B support points need to be included within the domain defined by D_d , though typically a much higher value is suggested). The exponential type of weighting function will be used in this study

$$w\{d\} = \left[e^{-\left(\frac{d}{c}\right)^{2k}} - e^{-\left(\frac{d}{c}\right)^{2k}} \right] / \left[1 - e^{-\left(\frac{d}{c}\right)^{2k}} \right] \quad \text{if } d < D_d \tag{11}$$

$$= 0 \text{ else}$$

where c and k are free parameters to be selected (optimised) for improved accuracy. The relative weights v_k ultimately define the moving character of the approximation within the different directions in the \mathbf{x} space and should be chosen to (1) establish a normalization for the different components of \mathbf{x} , but more importantly (2) provide higher importance for components that have larger influence on the values of $z_i(\mathbf{x})$ (Taflanidis 2012).

Finally combining Eqs. (5) and (8) yields the MLS RSA expression

$$\hat{y}_i(\mathbf{x}) = \mathbf{b}^T(\mathbf{x})\mathbf{M}^{-1}\{\mathbf{x}\}\mathbf{L}\{\mathbf{x}\}\mathbf{F}_i \tag{12}$$

which involves simple matrix manipulations and requires for each configuration \mathbf{x} evaluation only of the basis vector $\mathbf{b}(\mathbf{x})$ at that point as well as of matrices $\mathbf{M}^{-1}\{\mathbf{x}\}$ and $\mathbf{L}\{\mathbf{x}\}$ through Eq. (9). Once y_i is obtained, then, z_i may be directly calculated by the inverse of the initial respective transformation in Eq. (3). Note that the information from the database for z_i (and ultimately y_i) influences only vector \mathbf{F}_i ; thus, updated information (from new high-fidelity simulations) can be directly incorporated into it. It is important to also stress that this response surface approach uses no implicit assumptions for the characteristics of z_i , as such can be used for any response variable of interest.

Figure 2 presents an illustrative example for a 2-D problem; it includes a plot of the exact function $z_i(\mathbf{x})$ [part (a)], the location of the support points and the value of $z_i(\mathbf{x})$ at these points [part (b)] and finally, the global and MLS approximations using quadratic basis functions [parts (c), (d)], respectively. In this case, a quadratic approximation provides a poor fit to $z_i(\mathbf{x})$, and so the global response surface cannot efficiently describe $z_i(\mathbf{x})$. On the other hand, the MLS approach provides an accurate response surface approximation to $z_i(\mathbf{x})$.

3.3 Evaluation of fit and relationship to prediction error

The fit of the RSA may be judged by evaluating some statistical measure of accuracy for the difference between \hat{z}_i and z_i for some database of validation points $\{\mathbf{x}^p; p = 1, \dots, N_E\}$ (subset of the available high-fidelity simulations). Two such measures commonly used are the coefficient of determination R_i^2 and the mean percent error ME_i . The first is given by

$$R_i^2 = 1 - \frac{SSE}{SST}; \quad SSE = \sum_{p=1}^{N_E} (z_i(\mathbf{x}^p) - \hat{z}_i(\mathbf{x}^p))^2, \tag{13}$$

$$SST = \sum_{p=1}^{N_E} (z_i(\mathbf{x}^p) - \bar{z}_i)^2, \quad \bar{z}_i = \frac{\sum_{p=1}^{N_E} z_i(\mathbf{x}^p)}{N_E}$$

and denotes the proportion of variation in the data that is explained by the response surface model. A larger value of R_i^2 (i.e. a value closer to 1) indicates a good fit. The mean percent error ME_i is

$$ME_i = \frac{\sum_{p=1}^{N_E} |z_i(\mathbf{x}^p) - \hat{z}_i(\mathbf{x}^p)|}{\sum_{p=1}^{N_E} |z_i(\mathbf{x}^p)|} \tag{14}$$

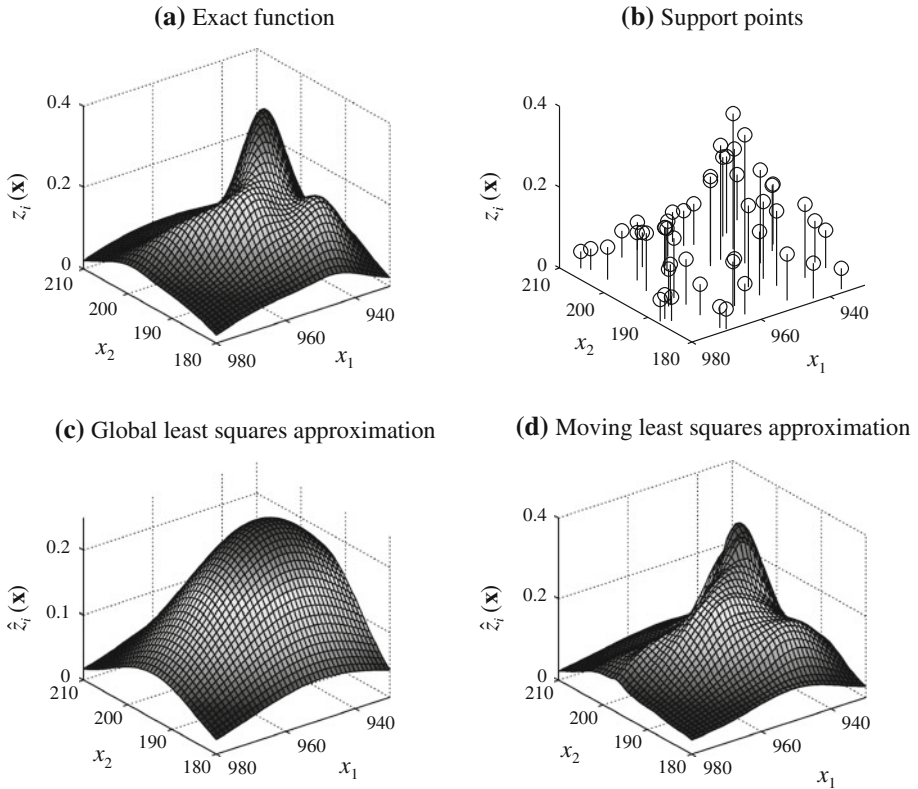


Fig. 2 Illustrative example for response surface approximation implementation

Smaller values (i.e. close to zero) for ME_i correspond to better fit in this case.

Finally, the standard deviation of the prediction error ε_i can be estimated based on these validation points as (Grimmett and Stirzaker 2001)

$$\sigma_{\varepsilon_i} = \sqrt{\frac{1}{N_E} \sum_{p=1}^{N_E} (z_i(\mathbf{x}^p) - \hat{z}_i(\mathbf{x}^p))^2} \tag{15}$$

This completely defines the probability model for the prediction error ε_i since it has been already assumed to follow a zero mean Gaussian distribution. This standard deviation provides a third measure for evaluating the fit of the response surface approximation, one that is directly related to the statistical relationship between z_i and \hat{z}_i as expressed by Eq. (2). Smaller values for σ_{ε_i} indicate better fit. For meaningful evaluation, this fit σ_{ε_i} needs to be normalized with respect to some proper threshold (note that contrary to the R_i^2 and ME_i , this third fit-measure σ_{ε_i} is not unitless, and thus, it needs to be normalised for proper comparisons) with the standard deviation for z_i over the observation database, σ_i , providing an appropriate candidate. This leads then to the normalized prediction error standard deviation

$$NS = \frac{\sigma_{\hat{e}_i}}{\sigma_i} = \sqrt{\frac{\frac{1}{N_E} \sum_{p=1}^{N_E} (z_i(\mathbf{x}^p) - \hat{z}_i(\mathbf{x}^p))^2}{\frac{1}{N_h} \sum_{h=1}^{N_h} \left(z_i(\mathbf{x}^h) - \frac{1}{N_h} \sum_{h=1}^{N_h} (z_i(\mathbf{x}^h)) \right)^2}} \tag{16}$$

3.4 Response surface approximations for entire response vector \mathbf{z}

The approximation for the entire vector \mathbf{z} , consisting of the response quantities of interest in our coastal region (as discussed in Sect. 2), is finally established by approximating each y_i through Eq. (12) and is ultimately expressed in a simple matrix form as

$$\hat{\mathbf{y}} = \mathbf{b}^T(\mathbf{x})\mathbf{M}^{-1}\{\mathbf{x}\}\mathbf{L}\{\mathbf{x}\}\mathbf{F} \quad \text{where } \mathbf{F} = [\mathbf{F}_1 \quad \mathbf{F}_2 \quad \dots \quad \mathbf{F}_{n_z}] \tag{17}$$

The initial output vector \mathbf{z} is then calculated by the inverse of the respective transformation in Eq. (3). Thus, the MLS response surface approximation simultaneously provides the *entire* response vector of interest (that is a separate surrogate model does not need to be implemented independently for each response variable). Note that the expressions for $\mathbf{b}(\mathbf{x})$ and $\mathbf{M}\{\mathbf{x}\}$ are identical to the case for scalar y_i ; thus, the computational burden for approximating the entire \mathbf{y} is similar to the one for approximating simply one of its components y_i .

The overall fit of the response surface approximation can be then evaluated by averaging the coefficient of determination and mean error (ME) over all different response variables. This then provides the average coefficient of determination, AR^2 , and average mean error, AME, given, respectively, by

$$AR^2 = \frac{1}{n_z} \sum_{i=1}^{n_z} R_i^2 \quad AME = \frac{1}{n_z} \sum_{i=1}^{n_z} ME_i \tag{18}$$

Similarly, with respect to the prediction error statistics, we have the average standard deviation, AS, or the average normalized standard deviation ANS

$$AS = \frac{1}{n_z} \sum_{i=1}^{n_z} \sigma_{\hat{e}_i} \quad ANS = \frac{1}{n_z} \sum_{i=1}^{n_z} \frac{\sigma_{\hat{e}_i}}{\sigma_i} \tag{19}$$

3.5 Implementation for surge prediction for inland locations

Though the response surface approximation described in Sect. 3.2 is a generalized one, that is, it does not depend on the characteristics of the response quantity z_i , its implementation for inland locations needs to additionally address the challenge that they do not *always* get inundated (in other words dry locations might remain dry for some storms). Though some scenarios in the database include full information for the storm surge ζ_i (when location i gets inundated), some provide *only* the information that the location remained dry.

One solution to this problem is to develop a surrogate model (Borges 1998) for the binary response quantity describing the condition of the location, that is, either wet or dry. This would require, of course, a separate surrogate model formulation for those components of the response vector (surge in inland locations). If we additionally need to know the storm surge, ζ_i , some alternative approach needs to be established, one that allows each scenario in our database to provide full, exact or approximate information for ζ_i . To facilitate this, the following approach is proposed. The storm surge ζ_i is described with respect to the mean sea level as reference point. When a location remains dry, the

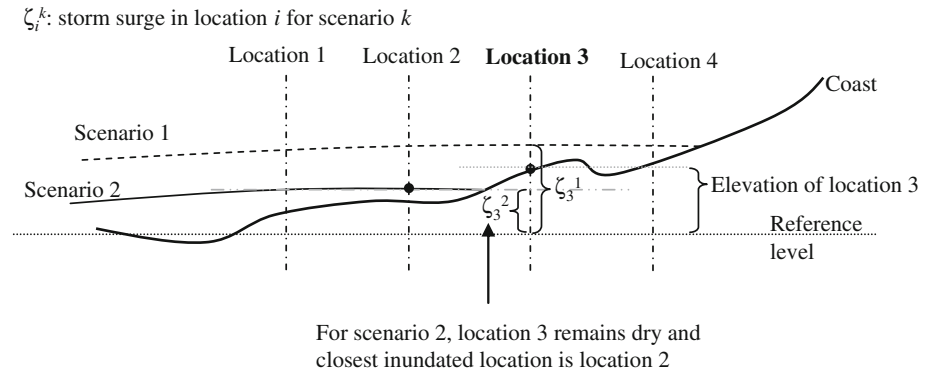


Fig. 3 Illustration of selection of storm surge for an example with a 1-D transect. Corrections for the surge in location 3 are shown

incomplete information for the exact storm surge for it is resolved by selecting (approximating) as its storm surge the one corresponding to the nearest location (nearest node in our high-fidelity numerical model) that was inundated. Figure 3 illustrates this approximation for an example with a 1-D transect (note that this approach is ultimately applied in the context of the full high-fidelity model grid, not simply with respect to 1-D transects). Once the database is adjusted for ζ_i (for the scenarios for which the location remained dry), the response surface approximation for ζ_i follows directly the steps discussed in Sect. 3.2. Comparison of ζ_i to the elevation of the location (defined with respect to the mean sea level as a reference point) provides finally the answer as to whether the location was inundated or not, whereas the storm surge elevation for the inundated locations is calculated by subtracting these two quantities. The latter is defined here as the height of inundation with respect to the ground. Thus, this approach allows us to gather simultaneous information about both the inundation (binary answer; yes or no) as well as the storm surge elevation. More importantly, it falls within the generalized response surface model discussed in the Sect. 3.2, as such it can be simultaneously implemented along with the other components of response vector \mathbf{z} , as presented in Sect. 3.4. Of course, it does involve the approximation illustrated in Fig. 3 for enhancing the database with complete information for ζ_i for all hurricane scenarios.

4 Optimal selection for response surface characteristics

For a specific set of support points $\{\mathbf{x}^I; I = 1, \dots, N_S\}$, the effectiveness, that is, fit, of the response surface approximation given by Eq. (17) depends on

- (a) Selection of type of basis functions $\mathbf{b}(\mathbf{x})$.
- (b) Selection of the parameters of the weighting function $w\{d(\mathbf{x}; \mathbf{x}_j)\}$, for example c, k and D_d for the weighting function given by Eq. (11).
- (c) Selection of the weighting vector \mathbf{v} for the distance norm in Eq. (10).

Let \mathbf{s} correspond to the vector representing these choices, then an optimal selection for \mathbf{s} may be obtained by optimizing some expression that measures the fit of the response surface approximation. The choice for \mathbf{s} should be the same for the entire response vector \mathbf{z} as it is computationally impractical to implement Eq. (12) for each of the response

quantities z_j separately, rather than for the entire vector as in Eq. (17). Thus, the selected measure should quantify the fit over all response quantities of interest. Any of the expressions provided in the previous section in Eqs. (18) or (19) can be used for this purpose, but the prediction error statistics are a more appropriate choice since they directly characterize the difference between z_i and \hat{z}_i . These expressions, though, should be evaluated for a set of control points (from the available high-fidelity simulations) $\{\mathbf{x}^c; c = 1, \dots, N_c\}$, that are different from the validation points. This is important so that the effectiveness of the response surface approximation is evaluated using a different part of the database from the one used to optimize its characteristics.

Using the relationships in Eq. (19), two different objective functions are thus defined

$$f(\mathbf{s}) = \frac{1}{n_z N_c} \sum_{i=1}^{n_z} \sum_{c=1}^{N_c} (z_i(\mathbf{x}^c) - \hat{z}_i(\mathbf{x}^c))^2 \tag{20}$$

$$f_n(\mathbf{s}) = \frac{1}{n_z N_c} \sum_{i=1}^{n_z} \frac{1}{\sigma_i^2} \sum_{c=1}^{N_c} (z_i(\mathbf{x}^c) - \hat{z}_i(\mathbf{x}^c))^2 \tag{21}$$

where \mathbf{s} ultimately influences the approximation $\hat{z}_i(\mathbf{x})$, a dependence that is not explicitly noted to simplify notation. The first objective function [Eq. (20)] corresponds to summation of square errors (estimated over the control points) for all response variables, whereas the second [Eq. (21)] additionally incorporates weights for each of the response variables, chosen as the variance over the initial observation points. Thus, the second choice incorporates a normalization for each of the response variables towards the overall objective function. Similarly, weights can be incorporated for each of the response quantities, to prioritize accuracy for the responses in a specific region. This can be established by combining the term $(1/\sigma_i^2)$ with additional weights q_i for each z_i . Response components for which it is considered critical to have better accuracy (for example, surge in locations of importance such as emergency evacuation routes) should be given higher weights q_i to increase their contribution towards the overall objective function value.

The selection for the optimal \mathbf{s} , denoted \mathbf{s}^* , corresponds then to a nonlinear minimization problem

$$\mathbf{s}^* = \arg \min_{\mathbf{s}} f(\mathbf{s}) \quad [\text{or } f_n(\mathbf{s})] \tag{22}$$

which may be solved through any appropriate numerical optimization algorithm. Note that the relationship between $f(\mathbf{s})$ and \mathbf{s} is not necessarily convex neither smooth. An easy explanation for the latter is the following. Selection of \mathbf{s} impacts the distance d and thus the exact support points included in the domain D_d for each \mathbf{x} (note that since the weights for the remaining support points are zero they ultimately do not inform the surrogate model for that \mathbf{x}). Thus, changes in \mathbf{s} can lead to abrupt changes in the predictions for some z_i (when the support points informing the surrogate model change) and ultimately abrupt changes to $f(\mathbf{s})$ (depending on the contribution of those z_i 's to the total objective function). Thus, algorithms appropriate for global, non-smooth optimization problems—such as direct-search methods or genetic algorithms— should be chosen. In this study, the powerful numerical optimization toolbox TOMLAB is used (Holmstrom et al. 2009), which offers a wide range of state-of-the-art algorithms appropriate for nonlinear, global optimization applications.

It should be pointed out that this optimization approach, relying on a single set of control points, creates some dependency of the optimization results on the exact control

points chosen, especially if these are not appropriately selected to represent the entire range of the observation points. Thus, it may promote a surrogate model that provides a targeted fit for scenarios with specific characteristics (the ones corresponding to these control points). Two alternative approaches exist for circumventing this problem (Loweth et al. 2010). In the first one, called leave-one-out cross-validation method, each observation point is removed from the data set to serve as a control point; then, the response surface is built using the remaining points, and the error between the approximated and actual response is calculated for the removed point. This is repeated for each of the observation points, and the total value of the error is calculated by summing all these individual contributions. The second approach, called k -fold method, is similar, but instead of systematically leaving out each observation point in turn, a random subset of k points is used as control points whilst the remaining points are used to build the surrogate model. The error is calculated between the approximate and actual values at each of the k control points, and this process is repeated multiple times with the total error calculated by summing over all repeated k -folds. Both these approaches, though, increase significantly the cost for evaluating the objective function (and performing the optimization) as they require formulation of the surrogate model multiple times, for different parts of the initial database. In this study, since we are primarily interested in exploring different trends (i.e. examine multiple cases), the single set of control points approach is adopted, as it reduces the computational burden of the optimization effort.

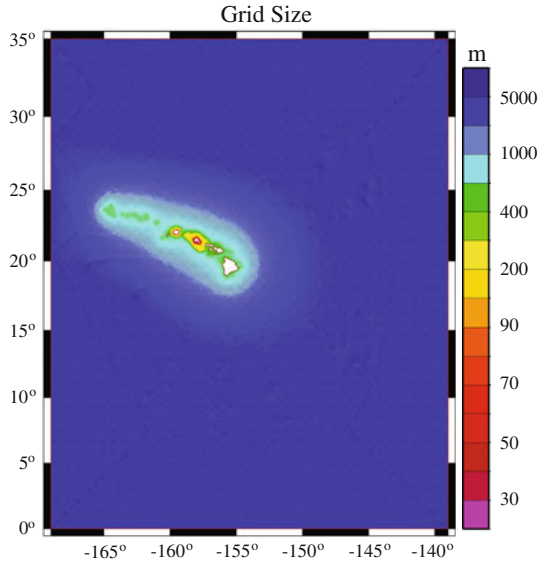
5 Case study

The proposed framework for response surface implementation for hurricane (and storm) wave/surge approximation is demonstrated here for an application to the Hawaiian Islands, focusing on the region around Oahu and using the database described in the study by Kennedy, et al. (2012). Both maximum (over the hurricane duration) significant wave heights H_s and surge levels ζ will be used as response quantities of interest, and different cases will be considered for the response surface implementation to investigate (a) the impact of the transformations in Eq. (3), (b) the influence of the different objective functions (20) and (21), (c) the differences in the approximation for different response quantities, and (d) the impact of the number of support points informing the surrogate model. A brief review of the characteristics for the high-fidelity simulations is given first, and then, the discussion focuses on the surrogate modelling optimization. Finally, illustrative results are shown for implementation of the developed surrogate model for risk assessment, emphasizing the computational benefits.

5.1 High-fidelity model details

The model used in the development of the database of high-fidelity data to predict the surge and wave response was a combination of the ADCIRC and SWAN numerical models (Kennedy et al. 2012). The computational domain for the high-fidelity model encompasses a large portion of the northern Pacific Ocean from 0 to 35° North and from 139 to 169° West. The unstructured Hakou-2010-r2 grid (Kennedy et al. 2012) resolves the deep ocean with 10-km elements, incorporates all the main Hawaiian Islands (Hawai'i, Maui, Kaho'olawe, Lana'i, Moloka'i, Oahu, Kauai, and Ni'ihau), represents Oahu with significant detail up to 30 m resolution and extends inland to the 4-m-elevation contours. Figure 4 shows the grid for the high-fidelity model, which incorporates 1,590,637 vertices

Fig. 4 Grid size for high-fidelity computational model



and 3,155,738 triangular elements. The coarsest resolution at the domain edge is up to 10 km, and finest resolution of 30 m is found in complex coastal areas and overland. For the numerical simulation, SWAN applies 10-min time steps, while ADCIRC applies 1-s time steps. A SWAN + ADCIRC simulation runs in 16 wall clock min/day of simulation on 1,024 cores on Diamond, a 2.8 GHz dual quad core-based cluster with a 20 Gb/s InfiniBand network (<http://www.ercd.hpc.mil/>). On the average, each simulation required over 1500 CPU hours to complete. The Hakou-2010-r2 model was validated by simulating tides and by hindcasting hurricane-Iniki (1992) and comparing to measured water levels as well as wave data. More details on the model itself and the validation may be found in Kennedy et al. (2012).

5.2 High-fidelity simulation database and partitioning

The available database consists of hurricane scenarios corresponding to five basic storm tracks, depicted in Fig. 1, representing different angles of final approach (i.e. track headings) θ with histories selected based on information from regional historical storms. Three different values for the central pressure c_p were used, 940, 955 and 970 mbar, three different forward velocities v_f were considered, 7.5, 15 and 22 knots, and three radius of maximum winds, R_m , 30, 45 and 60 km, were utilized in the development of the database. Landfall x_o was defined as the longitudinal point where each hurricane crosses 21.3° North, and in the database used for this study, corresponding to hurricanes with important impact on Oahu, 12 different landfall locations are considered corresponding to landfalls of 157.60, 157.85, 158.10, 158.30, 158.60, 158.90, 159.20, 159.50, 159.80, 160.00, 160.40, 160.70° West. A total of 603 storms are available in the database, which is partitioned into three different sets. One set consisting of 23 storms (chosen randomly) that serve as the control points $\{\mathbf{x}^c; c = 1, \dots, N_c\}$ for the optimization of the response surface characteristics, another set of 20 storms (also chosen randomly) that serve as the validation points $\{\mathbf{x}^p; p = 1, \dots, N_E\}$ for evaluating the accuracy of the approximation, and the remaining set of 560 storms that correspond to the support points $\{\mathbf{x}^I; I = 1, \dots, N_S\}$ used to develop

the surrogate model. An additional subset is considered for the support points, by excluding some landfall locations, to obtain a subset with a reduced number of support points. This subset consists of only 370 storms, corresponding to landfall locations of 157.60, 158.10, 158.30, 158.90, 159.20, 159.80, 160.10, 160.40° West (landfall locations evenly chosen among the initial database). The observation points that were not considered as support points are added in this case to the validation set, leading to $N_S = 370$ and $N_E = 210$. The control points remain the same for this subset as for the one with $N_S = 560$.

5.3 General characteristics of the response surfaces

Two different types of hurricane response characteristic will be examined. The first one is the significant wave height in the regions extending from 156 to 160° West and 20 to 23° North on a canonical grid consisting of a total $n_z = 9,053$ points (excluding the inland areas of the Hawaiian Islands). The second one is the surge in $n_z = 77,175$ locations around the coast of Oahu. These locations correspond to inland nodes (up to the 4 m elevation contour) in our high-fidelity numerical model that were inundated in at least one storm of our database. The response surface approximation for each of these outputs will be separately examined, to provide insight about the differences between them.

Polynomials up to second degree (quadratic) are considered as basis functions $\mathbf{b}(\mathbf{x})$ [with full quadratic basis functions corresponding to case illustrated in Eq. (7)] with the exact degree for the polynomial for each hurricane parameter corresponding to a variable to be optimized (this ultimately corresponds to a constraint for the degree of the polynomial with respect to each x_i). This leads to $N_B = 21$ basis functions based on Eq. (6) if second order polynomials are considered for all components of \mathbf{x} . Parameters for which the optimal basis is identified to be linear will be denoted as ‘linear’ (the remaining will be quadratic). For the weight function, the one given by Eq. (11) is adopted with D_d adaptively selected so that it includes 60 support points with c, k corresponding to additional variables to be optimized. This choice falls within the common guidelines for selection of D_d , to include a range of points 2–4 times N_B [this relates to having \mathbf{M} in Eq. (9) be invertible while giving priority only to support points in close proximity] (Choi et al. 2001). Note that slightly higher values for this number are suggested in the literature when the MLS RSA is intended to be used for moderate extrapolations (i.e. to predict the impact for scenarios with model parameters that fall outside the domain of the support points), but this topic falls outside our scope here (we are primarily interested in using RSA for interpolation). Finally, the weight v_k used for the distance norm of Eq. (10) is selected for each parameter as $v_k = r_k/\sigma_k$, where σ_k is the standard deviation for x_k in the database $\{\mathbf{x}^h; h = 1, \dots, N_R\}$ and is used to normalize the distance norm, whereas r_k is a weight variable, to be optimized, used to provide an importance factor to the model parameters. Larger values for r_k indicate stronger influence along that direction in the model parameter space for evaluating the distance norm. Note that the approximation in Eq. (8) is not impacted by scaling of $w\{d\}$; thus, only the relative ratio between the v_k 's, and equivalently r_k 's, (and not their exact values) influence the established response surface. This means that only four out of the five r_k 's need to be optimized, with the exact selection of the r_k to have a certain value or that specific value having no influence on the established surrogate model. In our case, r_1 is always chosen as 1.

Thus, the vector \mathbf{s} of variables that can be optimized for the response surface approximation includes (1) the degree of the basis function for each hurricane parameter, (2) the variables c, k for the weighting function, and (3) the weight (importance) variable r_k 's for determining v_k . The optimization is actually performed in two stages. In the first stage, a

specific value is set for variables (1), and the optimal values of variables (2) and (3) are obtained through the optimization described in Eq. (20). In the second stage, different choices for (1) are considered, and the overall optimal solution \mathbf{s}^* is finally identified. Apart from the full optimization (for entire vector \mathbf{s}) sub-optimization problems are also investigated. The first one considers full quadratic basis functions (so no optimization for the polynomial degree) and allows for optimization of only c , k and r_k 's, and the second additionally assumes $r_k = 1$ and considers only optimization for c and k .

Finally, with respect to transformation for response and objective function selection, three different cases are examined here. The first two correspond to the linear transformation in Eq. (3) and consider Eq. (20) [first case] and (21) [second case] as the objective function. They are denoted as linear and linear_n, respectively. The third one corresponds to the log transformation in Eq. (3) and considers Eq. (21) as the objective function. It is denoted as log_n.

For all cases, only the results for the fit of the response surface approximation are presented here with respect to all the four quantities discussed in Sect. 3.4. The optimal \mathbf{s}^* is then presented in Appendix 2 (Table 7) which also includes (Table 6) a summary of all cases considered along with their characteristics, that is, the number of support points, transformation for response and objective function selections, and optimization problem solved. These tables include all the different optimal models from the different optimization sub-problems discussed above. To simplify the presentation, each case considered is given an identification (ID) number; then, the optimal response surface characteristics are referenced in the Appendix 2 with respect to that ID number.

5.4 Significant wave height

For the significant wave height, Fig. 5 presents the contours for the mean and standard deviation over the database. Table 1 presents the results for the optimal response surface for the three different optimization cases (linear, linear_n and log_n) and for two different sets of support points. Table 2 presents the results for the different optimization sub-problems for case linear_n. It also includes the instance that no optimization is performed, and typical

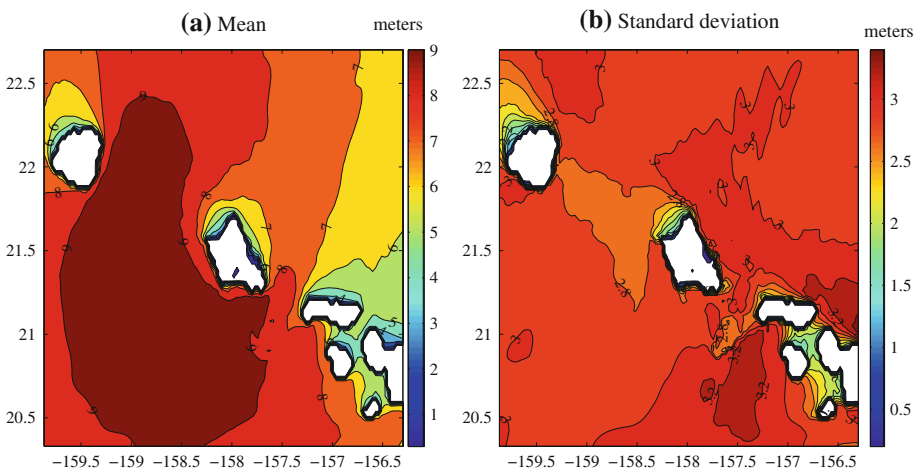


Fig. 5 Statistics for the wave height over the database. **a** Mean, **b** Standard deviation

Table 1 Fit for the optimal response surface for wave height for different number of support points and optimization cases (transformation of response variables or objective function selections)

Number of support points	Optimization case	ID number	AR ²	AME (%)	AS (m)	ANS
560	Linear	1	0.987	2.38	0.297	0.102
	Linear _n	2	0.987	2.37	0.297	0.102
	Log _n	3	0.988	2.34	0.284	0.098
370	Linear	4	0.921	6.51	0.810	0.284
	Linear _n	5	0.926	6.37	0.786	0.276
	Log _n	6	0.883	9.24	0.999	0.349

Table 2 Fit for different optimization sub-problems for wave height for different number of support points for linear_n optimization case

Number of support points	Optimization sub-problem	ID number	AR ²	AME (%)	AS (m)	ANS	
560	Optimal-full quadratic	2	0.987	2.37	0.297	0.102	
	Full quadratic basis	c, k, r_i 's	2	0.987	2.37	0.297	0.102
		c, k	7	0.955	5.14	0.552	0.190
		None	8	0.948	5.32	0.594	0.204
370	Optimal— c_p, v_f, R_m linear	5	0.926	6.37	0.786	0.276	
	Full quadratic basis	c, k, r_i 's	9	0.771	10.57	1.376	0.481
		c, k	10	0.419	10.83	2.992	1.041
		None	11	0.867	8.27	1.009	0.352

values (Choi et al. 2001) are used for the response surface characteristics (full quadratic function, r_k 's = 1, $c = 0.4, k = 1$). Finally, Fig. 6 presents the contours for the ME for the optimal linear_n surrogate model for (a) 560 [case ID = 2] (b) 370 support points [case ID = 5].

It is evident from Table 1 that when the entire database is used (560 support points), the developed surrogate models provide a very good approximation, with average MEs less than 2.5 % and coefficient of determination over 98 %. It should be pointed out that the latter demonstrates that the developed approximate model describes the variability within the data very well, a feature that is very important for risk assessment applications for which the accuracy over an ensemble of scenarios is more crucial than the accuracy on individual basis (as multiple scenarios contribute ultimately to the total risk). When the number of support points is reduced (370), the accuracy of the response surfaces significantly reduces, especially for the log_n transformation. This shows that the density of landfall locations in the initial database was important for adequately describing the variability of the response in this case study. The linear_n transformation is proven to provide an overall better fit, which is the reason it is chosen for presenting the results in Table 2 for the sub-optimization problems (note though that based on the studies we performed for the other transformations, the trends are fairly similar). Comparing the linear and linear_n cases shows that the normalization of the different response quantities in the objective function definition provides some accuracy/robustness to the optimization results; this is expected as it ultimately prevents the optimization to be dominated by response quantities (wave

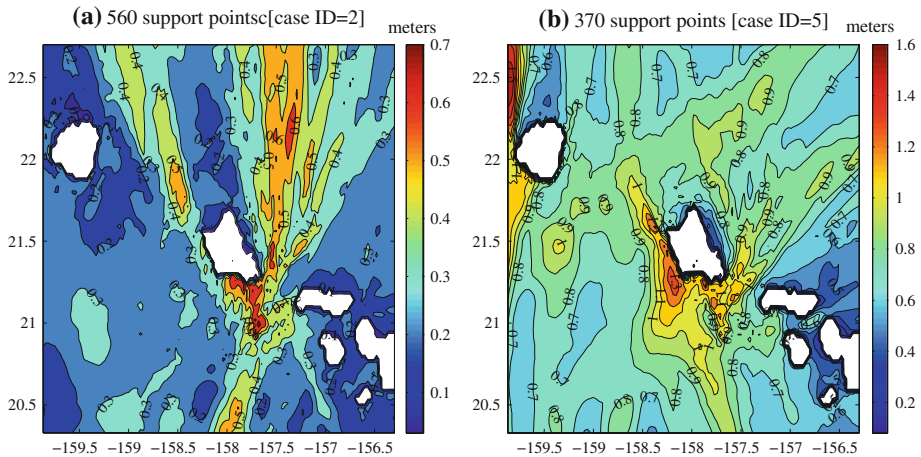


Fig. 6 Mean error (ME) for optimal linear, surrogate model for **a** 560 [case ID = 2] and **b** 370 support points [case ID = 5]

heights in this case) that have higher values and thus higher relative errors. The most important result coming, now, from Table 2 is the significance of performing an explicit optimization for the characteristics of the RSA model; the performance improvement for the optimal case over the case for no optimization is significant. It also shows that optimizing for the exact basis functions can provide a considerable improvement when the number of support points is smaller, whereas the explicit optimization for r_i 's, which ultimately dictate the moving character of the surrogate model within the different direction in the model parameter space \mathbf{x} , also has a significant impact, especially for a smaller number of support points. This discussion clearly shows that the optimization for the characteristics of the response surface approximation can provide important accuracy benefits.

Finally, comparison of Figs. 6 and 5 does not indicate any correlation between the errors of the surrogate model and the statistics of the initial database of observation points. Comparison between the two cases presented in Fig. 6 shows that though the regional trends are similar, there also exist significant differences, that is, errors in some regions relatively increase when the fewer support points are used. This further demonstrates that the distribution of the landfall location for the support points can have an impact on the spatial variability of the errors established through the response surface approximation (since the difference in the surrogate models considered is ultimately with respect to the distribution of the landfall locations of the support points).

5.5 Storm surge

For the storm surge, the database is initially augmented following the discussion in Sect. 3.3. Close to 30 % of the database points required updating using the proposed approach. Tables 3 and 4 present the optimization results, in a similar format as Tables 1 and 2. In this case, an additional measure is introduced to evaluate the efficiency of response surface approximation; the percentage of misclassified locations with respect to inundation, that is, the percentage of points that were classified as inundated when they were actually dry or classified as dry when they were inundated. A subset of the locations is also created,

Table 3 Fit for the optimal response surface for the storm surge level for different number of support points and optimization cases (transformation of response variables)

Number of support points	Optimization case	ID number	AR ²	AME (%)	AS (m)	ANS	Misclassification (%)
560	Linear	12	0.930 (0.974)	5.23 (3.38)	0.078 (0.048)	0.213 (0.146)	2.56 (0.002)
	Linear _n	13	0.930 (0.974)	5.17 (3.32)	0.078 (0.047)	0.212 (0.145)	2.55 (0.001)
	Log _n	14	0.931 (0.975)	5.06 (3.24)	0.077 (0.047)	0.210 (0.143)	2.52 (0.001)
370	Linear	15	0.816 (0.891)	9.18 (6.48)	0.148 (0.103)	0.398 (0.315)	3.42 (0.054)
	Linear _n	16	0.823 (0.895)	9.11 (6.42)	0.146 (0.101)	0.395 (0.312)	3.38 (0.052)
	Log _n	17	0.816 (0.888)	9.39 (6.95)	0.158 (0.105)	0.392 (0.327)	3.52 (0.018)

considering the points that were wet for all simulations; these correspond to 15,867 nearshore locations in our grid. These are locations for which no modification was required in our initial database for developing the surrogate model (as discussed in Sect. 3.3). In parenthesis in both tables, the statistics for this subset are presented. Finally, Fig. 7 presents a comparison between predicted and exact storm surge for all 20 validation scenarios for 50 locations (out of the 77,175) randomly selected.

The results in Tables 3 and 4 indicate the same trends with Tables 1 and 2; the fit when the entire database is used is very good (Table 3), with the optimization for the exact type of basis functions and the values, or r_i 's having an important role, especially for the surrogate model informed by a reduced number of support points (Table 4). Though the errors increase here compared to the ones for the wave height, when the evaluation is constrained to the 15,867 nearshore locations in the grid that always remain wet, this difference is very small. Thus, the larger errors for the surge can be attributed to the challenges associated with the inland locations that may remain dry. Even for those cases, though, the surrogate model provides an overall adequate fit when the entire database is used, with a small rate of misclassifications (below 3 %). The quality of this fit is further demonstrated by the results in Fig. 7.

5.6 Combined output

A final study is performed by combining the different type of outputs (wave height and surge) into a single response vector \mathbf{z} and performing the optimization for the response surface approximation. Table 5 presents the results by evaluating the fit separately for the wave height or surge (only AME and ANS fit measures are reported due to space limitations); it also presents the results for the optimal cases established when developing surrogate models only for the wave height or surge response outputs (as described in the previous two sections). The results show that there is indeed a reduction in accuracy for either of the response quantities when the optimization does not explicitly have it as a target. That reduction is though not large. The combined optimization yields results that are closer to the optimization for surge, which is expected since the number of surge points used were larger (ratio of almost 9:1), and thus, relative contribution towards the total error is larger. This discussion indicates that the response surface surrogate model can perform adequately for simultaneously predicting both the response quantities of interest here, but also that a surrogate model that targets a specific response quantity can offer an improved accuracy. The later demonstrates that, if possible (in terms of computational cost), separate

Table 4 Fit for different optimization sub-problems for storm surge level for different number of support points for linear, optimization case

Number of support points	Optimization sub-problem	ID number	AR ²	AME (%)	AS (m)	ANS	Misclassification (%)
560	Optimal basis— c_p, v_f, R_m linear	13	0.930 (0.974)	5.17 (3.32)	0.078 (0.047)	0.212 (0.145)	2.55 (0.001)
	Full quadratic basis c, k, r_i 's	18	0.930 (0.973)	5.25 (3.43)	0.078 (0.048)	0.212 (0.148)	2.50 (0.003)
	c, k	19	0.911 (0.95)	6.93 (5.22)	0.096 (0.068)	0.266 (0.214)	2.81 (0.008)
370	None	20	0.903 (0.955)	6.79 (4.80)	0.097 (0.064)	0.269 (0.203)	2.70 (0.005)
	Optimal basis— c_p, v_f, R_m linear	16	0.823 (0.895)	9.11 (6.42)	0.146 (0.101)	0.395 (0.312)	3.38 (0.052)
	Full quadratic basis c, k, r_i 's	21	0.698 (0.777)	14.01 (11.03)	0.203 (0.151)	0.542 (0.489)	4.55 (0.051)
	c, k	22	0.699 (0.781)	13.81 (10.72)	0.197 (0.150)	0.542 (0.467)	4.30 (0.077)
	None	23	0.681 (0.805)	14.41 (10.66)	0.202 (0.138)	0.552 (0.429)	4.63 (0.084)

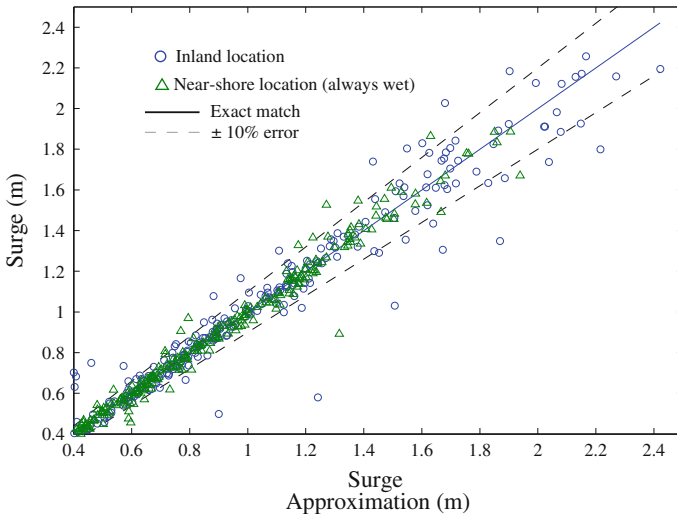


Fig. 7 Comparison between approximated and exact storm surge elevation for 50 locations

Table 5 Comparison between the response surface optimal models established for response vector corresponding to the storm surge or wave height or for them combined for linear_n optimization case

Number of support points	Optimization for	ID number	Wave height		Surge		
			AME (%)	ANS	AME (%)	ANS	Misclassification (%)
560	Surge	13	3.12	0.120	5.17	0.212	2.55
	Wave height	2	2.37	0.102	6.47	0.293	2.74
	Combined	24	3.10	0.120	5.19	0.213	2.57
370	Surge	16	6.11	0.258	9.11	0.395	3.38
	Wave height	5	6.37	0.276	10.83	0.449	4.01
	Combined	25	6.10	0.258	9.16	0.398	3.41

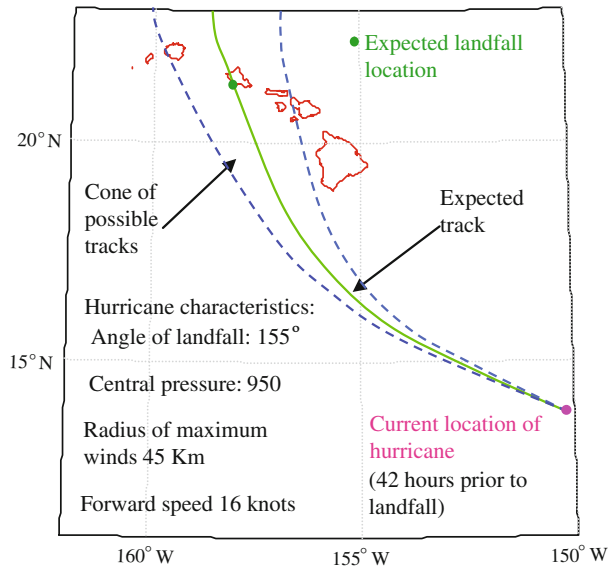
surrogate models should be developed and implemented for each of the response quantities.

Finally, looking collectively (extending to all cases considered in the previous three sections) at the results in Table 7 of Appendix 2, it can be inferred that the landfall location and angle of final approach (especially the former) (1) have in general a greater importance in affecting hurricane responses and that (2) the responses exhibit larger variability with respect to these two parameters. These two attributes are indicated, respectively, by the fact that (1) larger optimal values for r_i 's are identified for them and that (2) in no case is a linear basis function identified as optimal for them.

5.7 Example of risk assessment implementation

To demonstrate a possible implementation of the proposed surrogate modelling approach, the optimal response surface surrogate model is used to estimate hurricane risk, based on the theoretical and computational framework discussed in Appendix 1, for the hurricane

Fig. 8 Hurricane track and characteristics used in the illustrative risk assessment implementation



illustrated in Fig. 8 approaching landfall to Oahu. This corresponds to the real-time hurricane risk assessment implementation that has been advocated in (Taflanidis et al. 2012). The *estimated* mean values for hurricane parameters [landfall longitude, heading at landfall in degrees measured clockwise from south; central pressure in mbar; forward speed in knots; radius of maximum winds in km] are chosen as

$$\mathbf{x}_{\text{mean}} = [158.12^\circ \quad 155^\circ \quad 950 \text{ mbar} \quad 16 \text{ knots} \quad 45 \text{ km}]^T \quad (23)$$

For defining $p(\mathbf{x})$, these parameters are assumed to follow independent Gaussian distributions with mean values given in Eq. (23) and standard deviations selected here as [based on *estimation* errors for a time until landfall equal to 42 h (Taflanidis et al. 2012)]

$$\sigma_{\mathbf{x}} = [0.28^\circ \quad 17.5^\circ \quad 10.5 \text{ mbar} \quad 3.5 \text{ knots} \quad 2.5 \text{ km}]^T \quad (24)$$

Hurricane risk is evaluated through Eq. (29) using $N = 2,000$ samples. Figure 9 shows [part (a)] the probability that the significant wave height will exceed 9 m and [part (b)] the wave height with probability of being exceeded 5 %. Then Fig. 10 shows the probability that the surge will exceed threshold β for two locations, the first with coordinates 21.2999°N, 158.1101°W (near Pearl Harbor) and the second with coordinates 21.3101°N, 157.9302°W (close to the old runway in the airport). Results with and without including the prediction error of Eq. (2) (due to the surrogate modelling) are shown in this figure. The comparison indicates that the prediction error can have a significant impact on the calculated risk (though not always), and it will lead to more conservative estimates for rare events (with small probabilities of occurrence). This demonstrates that it is important to explicitly incorporate it in the risk estimation framework. These trends agree with the results published recently by Resio et al. (2012) within the context of storm surge risk estimation. Then, Fig. 11 presents similar results for the wave height; it shows the probability that the wave height will exceed threshold β for two locations, the first with coordinates 21.29°N, 157.92°W (close to the coast) and the second with coordinates 21.17°N, 158.34°W (in the open sea). In this case, the higher accuracy of the surrogate model (smaller prediction error variance as discussed in previous sections) leads to smaller

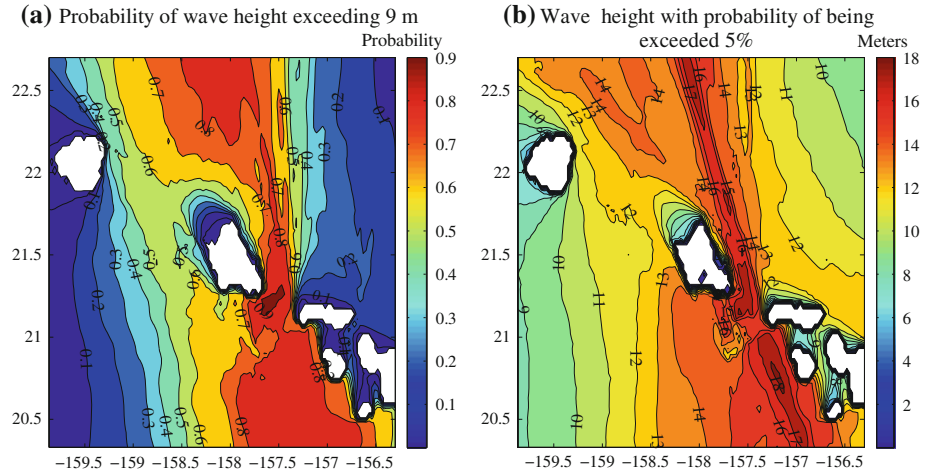


Fig. 9 **a** Probability contours that the significant wave height will exceed 9 m and **b** wave height with probability of being exceeded 5 %

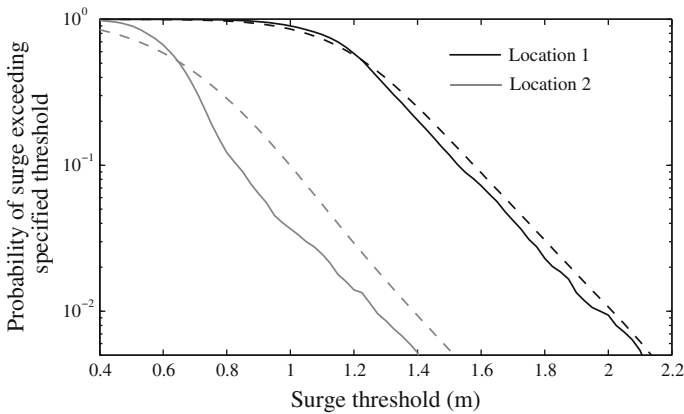


Fig. 10 Probability that surge will exceed threshold β for two locations in Oahu with coordinates 21.2999°N, 158.1101°W (location 1) and 21.3101°N, 157.9302°W (location 2). Results with (dashed lines) and without (solid lines) the prediction error of the surrogate model are shown

(relatively to Fig. 10) impact of the prediction error on the calculated exceedance probabilities. It is important to note that, like in the case of Fig. 10, the impact of including the prediction error is higher when looking at rare events, that is, for small probabilities of exceedance.

The total time needed for this risk assessment is 2 min on a 3.2 GHz single core processor with 4 GB of RAM. This corresponds to a tremendous reduction of computational time compared to the high-fidelity model, which required over 1500 hours for analysing a single hurricane track. Thus, a risk computation requiring independent evaluations of the surrogate model for $N = 2,000$ different scenarios is still approximately 10^4 times faster than evaluating a single hurricane scenario using the high-fidelity model SWAN + ADCIRC. Note that ability to use a large number of evaluations in the stochastic

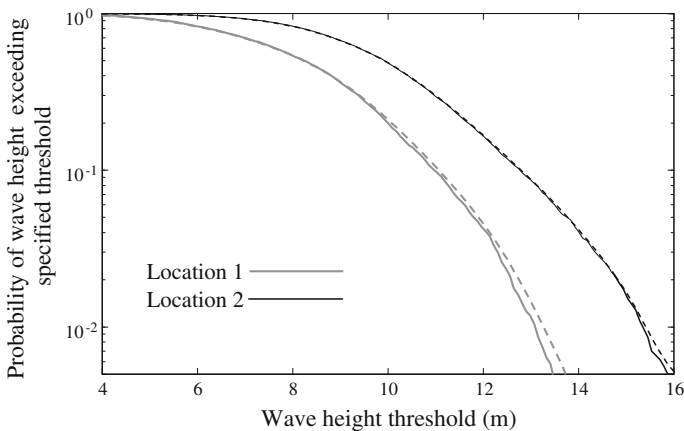


Fig. 11 Probability that significant wave height will exceed threshold β for two locations in Oahu with coordinates 21.29°N, 157.92°W (location 1) and 21.17°N, 158.34°W (location 2). Results with (*dashed lines*) and without (*solid lines*) the prediction error of the surrogate model are shown

simulation-based evaluation (29) of the risk integral (25) contribute to improved accuracy [reduced coefficient of variation (30)] for that estimate. The computational simplicity of the stochastic simulation-based estimation (29) along with the computational efficiency of the developed surrogate model also promote the development of automated risk assessment tools for hurricane/storm risk predictions, as advocated in (Taflanidis et al. 2012) and (Kijewski-Correa et al. 2012). These arguments illustrate the efficiency of the proposed theoretical and computational scheme, based on an optimized response surface approximation.

6 Conclusions

The systematic implementation and optimization of MLS response surface approximations for hurricane wave and surge response prediction were discussed in this paper. Using an existing database of high-fidelity simulations, the surrogate model can efficiently predict hurricane impact for any hurricane scenario of interest (within the bounds of the database), for any quantity of interest and for a large coastal region that might involve thousands of locations. The optimal selection of the basis functions for the response surface and of the parameters of the MLS character of the approximation were addressed. The impact of the number of high-fidelity simulations informing the surrogate modelling was investigated and different hurricane response characteristics (surge and significant wave height) were also analysed. Special attention was given to predicting the surge in inland locations for which the possibility of the location remaining dry needs to be additionally addressed. An approximate methodology was proposed here for these cases, based on a modification of the initial database; the storm surge for any location that remained dry was correlated to the closest inundated location. Though this ultimately has an impact on the accuracy of the response surface approximation, the established percentage of misclassifications (wet nodes that are identified as dry and vice versa) was small.

The results for the case study considered (Oahu Island) demonstrate that MLS response surface models can provide accurate approximations for both the wave and surge responses

and can ultimately facilitate a highly efficient estimation of hurricane risk. The explicit optimization of the characteristics of the response surface model provide considerable improvements in accuracy; this pertains to both the order of basis functions used as well as to the parameters that dictate the moving character of the surrogate model. The number of support points also has an impact on the results (as is expected in any surrogate model), with the landfall locations for the support points playing a key role in the case of storm responses. Though this result will definitely depend on the topographical characteristics of the coastal region in which this methodology is implemented, it was shown in the case study for Hawaii (focusing on Oahu) that very high accuracy was established when the landfall location for the support points had an average distribution of 0.3° , with a moderate accuracy decrease when this distribution was increased to 0.5° . Overall, the landfall location and angle of approach were the two model parameters influencing the hurricane impact more, and thus the accuracy of the surrogate model. Though the proposed response surface approximation can provide an overall good fit for multiple response quantities, it was also shown that improvement in accuracy can be established when the surrogate model is optimized with specific quantities of interest in mind. This shows that the optimization of separate surrogate models should be considered when the quality of the approximation needs to be improved for specific responses of interest (for example, for waves compared to surge or for surge for different regions within the coast compared to the surge for the remaining regions). Finally, it was shown that within a risk estimation framework, the prediction error due to surrogate model approximation can have a significant impact on the calculated risk, especially for rare events (small probabilities of occurrence) for which it will typically lead to more conservative estimates.

Acknowledgments This research effort is partially supported by US Army Corps of Engineering grant W912HZ-09-C-0086 under the Surge and Wave Island Modeling Studies, Coastal Field Data Collection Program. This support is greatly appreciated. Permission to publish this work was granted by the Chief of Engineers, US Army Corps of Engineers. Also the contribution of Professor Joannes Westerink and graduate student Mark Hope, both at the University of Notre Dame, in developing and performing the high-fidelity simulations for the illustrative example is greatly appreciated.

Appendix 1: Implementation in coastal risk assessment

Within the JPM setting, hurricane/storm coastal risk may be quantified in terms of the response $\hat{\mathbf{z}}(\mathbf{x})$ provided by the surrogate model and a *probability density function* $p(\mathbf{x})$ describing the uncertainty in the input hurricane parameters \mathbf{x} . For real-time risk evaluation, that is, when predicting the risk due to an approaching hurricane, $p(\mathbf{x})$ may be constructed through the estimates provided by the National Hurricane Center (<http://www.nhc.noaa.gov>); each component of \mathbf{x} can be selected to follow an independent Gaussian distribution with mean equal to the forecast quantities and standard deviation equal to the associated statistical estimation error (Taflanidis et al. 2012). For long-term hurricane risk evaluation for a region, $p(\mathbf{x})$ is selected based on statistical data, and it further incorporates information on occurrence rates for hurricanes, not just on relative plausibility of the model parameters (Resio et al. 2007).

Risk is ultimately expressed as some desired statistic of the response \mathbf{z} , for example the probability that the wave height will exceed some specific threshold or the median surge. The exact selection used for these statistics leads to definition of the risk consequence measure $h_i[\cdot]$. Ultimately, for any component z_i of the response vector, the risk, denoted H_i , is provided by the multi-dimensional probabilistic integral (Taflanidis et al. 2012)

$$H_i = \int_X h_i[\hat{z}_i(\mathbf{x})] p(\mathbf{x}) d\mathbf{x} \tag{25}$$

where X corresponds to the region of possible values for \mathbf{x} . The risk consequence measure depends on the definition for H_i , and it additionally addresses the prediction error ε_i . Through its appropriate selection different potential hurricane risk quantifications can be addressed. For example, if H_i corresponds to the expected (mean) value for some z_i , $H_i = E[z_i]$ (where $E[\cdot]$ denotes expectation), then (Taflanidis et al. 2012)

$$h_i[\hat{z}_i(\mathbf{x})] = \hat{z}_i(\mathbf{x}) \tag{26}$$

and the model prediction error has no impact on the risk consequence measure. If alternatively, H_i corresponds to the probability that some z_i will exceed some threshold β_i , $H_i = P[z_i > \beta_i]$ (where $P[\cdot]$ denotes probability), then (Taflanidis et al. 2012)

$$h_i[\hat{z}_i(\mathbf{x})] = F_i(\hat{z}_i - \beta_i) \tag{27}$$

where F_i corresponds to the cumulative distribution function for the model prediction error for z_i . In this case, the statistics of the prediction error do have an impact on the risk quantification. This simplifies to

$$h_i[\hat{z}_i(\mathbf{x})] = \Phi\left[\frac{\hat{z}_i(\mathbf{x}) - \beta_i}{\sigma_{\varepsilon_i}}\right] \tag{28}$$

for the proposed case of Gaussian distribution for the model prediction error, where $\Phi[\cdot]$ denotes the standard Gaussian cumulative distribution function.

Once risk for z_i has been quantified by the proper selection of the consequence measure (dependent only on \hat{z}_i), the probabilistic integral in Eq. (25) can be estimated by stochastic simulation (Robert and Casella 2004). For the simplest approach (direct Monte Carlo), and using N samples of \mathbf{x} randomly selected from $p(\mathbf{x})$, the estimate for H_i is given by

$$\tilde{H}_i = \frac{1}{N} \sum_{m=1}^N h_i[\hat{z}_i(\mathbf{x}^m)] \tag{29}$$

where vector \mathbf{x}^m denotes the sample of the uncertain parameters used in the m th simulation. The quality of this estimate is assessed through its coefficient of variation, δ obtained by

$$\delta \approx \frac{1}{\sqrt{N}} \sqrt{\frac{\frac{1}{N} \sum_{m=1}^N (h_i[\hat{z}_i(\mathbf{x}^m)])^2}{\tilde{H}_i^2} - 1} \tag{30}$$

which decreases (i.e. estimation improves) proportionally to \sqrt{N} . Thus, estimation of risk may be efficiently and accurately performed using the established surrogate model, as evaluation of $\hat{z}_i(\mathbf{x})$ requires minimal computational effort [thus a large number of samples N can be used for (29)].

Appendix 2: Summary of cases considered for response surfaces and corresponding optimal parameters

The following two tables present a summary of all cases considered (Table 6) in the case study along with the optimal parameters, that is, type of basis functions, values for c and

Table 6 Summary of cases considered for response surfaces

ID number	Response quantity for optimization	Optimization problem solved	Transformation of response and objective function	Number of support points
1	Wave height	Full	Linear	560
2	Wave height	Full (same for only c, k, r_i 's)	Linear _n	560
3	Wave height	Full	Log _n	560
4	Wave height	Full	Linear	370
5	Wave height	Full	Linear _n	370
6	Wave height	Full	Log _n	370
7	Wave height	Only c, k	Linear _n	560
8	Wave height	None	Linear _n	560
9	Wave height	Only c, k, r_i 's	Linear _n	370
10	Wave height	Only c, k	Linear _n	370
11	Wave height	None	Linear _n	370
12	Storm surge	Full	Linear	560
13	Storm surge	Full	Linear _n	560
14	Storm surge	Full	Log _n	560
15	Storm surge	Full	Linear	370
16	Storm surge	Full	Linear _n	370
17	Storm surge	Full	Log _n	370
18	Storm surge	Only c, k, r_i 's	Linear _n	560
19	Storm surge	Only c, k	Linear _n	560
20	Storm surge	None	Linear _n	560
21	Storm surge	Only c, k, r_i 's	Linear _n	370
22	Storm surge	Only c, k	Linear _n	370
23	Storm surge	None	Linear _n	370
24	Combined	Full	Linear _n	560
25	Combined	Full	Linear _n	370

Table 7 Optimal values for model parameters of response surfaces

ID number	Basis functions	c	k	r_1	r_2	r_3	r_4	r_5
1	Full quadratic	0.159	0.50	1	1.405	0.050	0.204	0.289
2	Full quadratic	0.158	0.50	1	1.417	0.050	0.205	0.289
3	R_m linear	0.507	1.61	1	1.169	0.063	0.305	0.668
4	c_p, v_f, R_m linear	0.449	1.99	1	0.498	0.248	0.311	0.148
5	c_p, v_f, R_m linear	0.466	1.99	1	0.543	0.277	0.344	0.164
6	R_m linear	0.334	0.83	1	0.817	0.272	0.164	0.051
7	Full quadratic	0.759	1.89	1	1	1	1	1
8, 11, 20, 23	Full quadratic	0.400	1.00	1	1	1	1	1
9	Full quadratic	0.299	0.94	1	0.652	0.346	0.299	0.134
10	Full quadratic	0.817	1.97	1	1	1	1	1
12	c_p, v_f, R_m linear	0.233	0.58	1	0.996	0.423	0.489	0.435
13	c_p, v_f, R_m linear	0.149	0.50	1	0.988	0.427	0.487	0.441

Table 7 continued

ID number	Basis functions	c	k	r_1	r_2	r_3	r_4	r_5
14	c_p, v_{fs}, R_m linear	0.157	0.50	1	0.987	0.429	0.486	0.444
15	c_p, v_{fs}, R_m linear	0.524	1.75	1	0.482	0.453	0.511	0.284
16	c_p, v_{fs}, R_m linear	0.513	1.66	1	0.483	0.449	0.506	0.277
17	c_p, v_{fs}, R_m linear	0.278	0.53	1	0.453	0.453	0.453	0.292
18	Full quadratic	0.375	0.50	1	0.778	0.464	0.266	0.262
19	Full quadratic	0.946	1.99	1	1	1	1	1
21	Full quadratic	0.907	0.60	1	0.739	0.649	0.560	0.077
22	Full quadratic	0.994	1.95	1	1	1	1	1
24	c_p, v_{fs}, R_m linear	0.150	0.50	1	0.994	0.419	0.491	0.446
25	c_p, v_{fs}, R_m linear	0.506	1.64	1	0.496	0.442	0.501	0.259

k and for weights r_k for them (Table 7). These correspond to all unique cases that were examined in Sects. 5.4, 5.5 and 5.6, by considering the full- and sub-optimization problems for the different outputs and the different normalization of response and objective function selections (more information on the selection is provided in Sect. 5.3). Each case is referenced here by its identification index, ID. Note that for parameters in Table 7 for which their basis function are not explicitly referenced as *linear*, the chosen polynomial degree is *quadratic*.

References

- Breitkopf P, Naceur H, Rassineux A, Villon P (2005) Moving least squares response surface approximation: formulation and metal forming applications. *Comput Struct* 83(17–18):1411–1428
- Burges CJC (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disc* 2:121–167
- Choi KK, Youn B, Yang RJ (2001) Moving least squares method for reliability-based design optimization. In: 4th world congress of structural and multidisciplinary optimization
- Condon AP, Sheng YP (2012) Evaluation of coastal inundation hazard for present and future climates. *Nat Hazards* 62:345–373
- Das HS, Jung H, Ebersole B, Wamsley T, Whalin RW (2010) An efficient storm surge forecasting tool for coastal Mississippi. In: 32nd international coastal engineering conference, Shanghai, China
- Dietrich JC, Bunya S, Westerink JJ, Ebersole BA, Smith JM, Atkinson JH, Jensen R, Resio DT, Luettich RA, Dawson C, Cardone VJ, Cox AT, Powell MD, Westerink HJ, Roberts HJ (2010) A high resolution coupled riverine flow, tide, wind, wind wave and storm surge model for southern Louisiana and Mississippi: part II—synoptic description and analyses of hurricanes Katrina and Rita. *Mon Weather Rev* 138(2):378–404
- Grimmett G, Stirzaker D (2001) Probability and random processes. Oxford University Press, Oxford
- Ho FP, Myers VA (1975) Joint probability method of tide frequency analysis applied to Apalachicola Bay and St. George Sound. NOAA technical report NWS 38, Washington, DC, USA
- Holmstrom K, Goran AO, Edvall MM (2009) User's guide for TOMLAB 7. Tomlab Optimization, San Diego. www.tomopt.com/tomlab/
- Irish J, Resio D, Cialone M (2009) A surge response function approach to coastal hazard assessment. Part 2: quantification of spatial attributes of response functions. *Nat Hazards* 51(1):183–205
- Kennedy AB, Gravois U, Zachry B (2011a). Observations of landfalling wave spectra during Hurricane Ike. *J Waterw Port Coastal Ocean Eng* 142–145. doi:10.1061/(ASCE)WW.1943-5460.0000081
- Kennedy AB, Gravois U, Zachry BC, Westerink JJ, Hope ME, Dietrich JC, Powell MD, Cox AT, Luettich RL, Dean RG (2011b). Origin of the Hurricane Ike forerunner surge. *Geophys Res Lett* L08805. doi:10.1029/2011GL047090

- Kennedy AB, Westerink JJ, Smith J, Taflanidis AA, Hope M, Hartman M, Tanaka S, Westerink H, Cheung KF, Smith T, Hamman M, Minamide M, Ota A (2012) Tropical cyclone inundation potential on the Hawaiian islands of Oahu and Kauai. *Ocean Model* 52–53:54–68
- Kijewski-Correa T, Taflanidis AA, Kennedy AB, Kareem, Westerink JJ (2012) CYBER-EYE: integrated cyber-infrastructure to support hurricane risk-assessment. In: ATC and SEI advances in hurricane engineering conference
- Loweth EL, De Boer GN, Toropov VV (2010). Practical recommendations on the use of moving least squares metamodel building. In: Thirteenth international conference on civil, structural and environmental engineering computing, Crete, Greece
- Myers VA (1975) Storm tide frequencies on the South Carolina coast. NOAA technical report NWS-16
- Myers RH, Montgomery DC (2002) Response surface methodology. Wiley, New York
- Niedoroda AW, Resio DT, Toro GR, Divoky D, Reed C (2010) Analysis of the coastal Mississippi storm surge hazard. *Ocean Eng* 37(1):82–90
- Resio DT, Westerink JJ (2008) Modeling of the physics of storm surges. *Phys Today* 61(9):33–38
- Resio DT, Boc SJ, Borgman L, Cardone V, Cox A, Dally WR, Dean RG, Divoky D, Hirsh E, Irish JL, Levinson D, Niedoroda A, Powell MD, Ratcliff JJ, Stutts V, Suhada J, Toro GR, Vickery PJ (2007) White paper on estimating hurricane inundation probabilities. Consulting report prepared by USACE for FEMA
- Resio D, Irish J, Cialone M (2009) A surge response function approach to coastal hazard assessment—part 1: basic concepts. *Nat Hazards* 51(1):163–182
- Resio DT, Irish JL, Westering JJ, Powell NJ (2012) The effect of uncertainty on estimates of hurricane surge hazards. *Nat Hazards*. doi:[10.1007/s11069-012-0315-1](https://doi.org/10.1007/s11069-012-0315-1)
- Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, New York
- Song YK, Irish JL, Udoh IE (2012) Regional attributes of hurricane surge response functions for hazard assessment. *Nat Hazards*. doi:[10.1007/s11069-012-0309-z](https://doi.org/10.1007/s11069-012-0309-z)
- Taflanidis AA (2012) Stochastic subset optimization incorporating moving least squares response surface methodologies for stochastic sampling. *Adv Eng Softw* 44(1):3–14
- Taflanidis AA, Beck JL (2010) Reliability-based design using two-stage stochastic optimization with a treatment of model prediction errors. *J Eng Mech* 136(12):1460–1473
- Taflanidis AA, Kennedy AB, Westerink JJ, Smith J, Cheung KF, Hope M, Tanaka S (2012) Rapid assessment of wave and surge risk during landfalling hurricanes; a probabilistic approach. *ASCE J Waterw Coast Port Auth*. doi:[10.1061/\(ASCE\)WW.1943-5460.0000178](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000178)
- Toro GB, Niedoroda AW, Reed C (2007). Approaches for the efficient probabilistic calculation of surge hazard. In: 10th international workshop on wave hindcasting and forecasting and coastal hazards
- Toro GR, Niedoroda AW, Reed CW, Divoky D (2010a) Quadrature-based approach for the efficient evaluation of surge hazard. *Ocean Eng* 37:114–124
- Toro GR, Resio DT, Divoky D, Niedoroda A, Reed C (2010b) Efficient joint-probability methods for hurricane surge frequency analysis. *Ocean Eng* 37:125–134
- Udoh LE, Irish JL (2011). Improvements in hurricane surge response functions: incorporating the effects of forward speed, approach angle, and sea level rise. In: International conference on vulnerability and risk analysis and management/fifth international symposium on uncertainty modeling and analysis
- Westerink JJ, Luettich RA, Feyen JC, Atkinson JH, Dawson C, Roberts HJ, Powell MD, Dunion JP, Kubatko EJ, Pourtaheri H (2008) A basin- to channel-scale unstructured grid hurricane storm surge model applied to southern Louisiana. *Mon Weather Rev* 136(3):833–864. doi:[10.1175/2007MWR1946.1](https://doi.org/10.1175/2007MWR1946.1)