Re-analysis of extreme sea state events modelling using a data-driven technique

Cédric Goeury, Fouquet Thierry, Bagnis Tristan, Benoit Michel, Teles Maria

Re-analysis of extreme sea state events modelling using a data-driven technique

Cedric GOEURY[1,2], Thierry FOUQUET[1], Tristan BAGNIS[1], Michel BENOIT[1,2], Maria TELES[1]

[1] National Laboratory for Hydraulics and Environment (LNHE), EDF R&D, Chatou, France

[2] LHSV, Ecole des Ponts, EDF R&D, Chatou, France

email: cedric.goeury@edf.fr, thierry.fouquet@edf.fr

ABSTRACT

With the objective of better answering marine and coastal engineering needs, sea state databases are created by hindcast simulations over long periods. The wave hindcast ANEMOC-3 database, built using the spectral model TOMAWAC for wave generation and propagation and the hydrodynamic circulation model TELEMAC-2D for tidal water levels and currents calculation, covers a period of 45 years from 1979 to 2023. During the calibration of the database, the numerical results must be consistent with past observational data (*in situ* or satellite measurements). Among other things, this process relies on calibration to determine "empirical adequacy" (Oreskes et al., 1994). In particular, the calibration aims at simulating a series of reference events by adjusting some uncertain physically-based model parameters until the agreement with available data reaches a satisfactory level. The objective of this work is to implement an efficient calibration algorithm, capable of processing measurements optimally, and of estimating the partially known parameters in the numerical model.

1. Introduction

Many problems in science and engineering require the estimation of unknown or uncertain model's parameters that will produce a solution that best fits a finite set of measurements. The modelling of marine flows and waves is no exception. Hindcast wave databases are sophisticated computer-based simulations providing complete time series of wave parameters and/or full wave spectra with a fine spatial resolution. They are powerful tools for a range of applications, including marine engineering, coastal infrastructure design, offshore operations, wave power assessment and environmental studies. By providing a detailed understanding of past sea states over several decades, they help researchers and professionals make informed decisions and predictions related to ocean environments and their impact on various activities. In this framework, EDF R&D LNHE together with the Cerema developed the hindcast wave database ANEMOC about 16 years ago with the main interest devoted to the French coastal areas of the Atlantic Ocean, English Channel and North Sea (Benoit et al., 2008). With the significant advances in the means of observing continental waters, the ability to exchange data as computational results has become a growing need. Data assimilation algorithms for integrating observation data into real cases are now increasingly applied to hydraulic problems with two main objectives: optimizing model parameters and improving hydraulic and wave simulations and forecasting. The objective of this work is to implement an efficient calibration algorithm, based on data assimilation, to best estimate partially known parameters of the source and sink terms in the wave model (namely wind generation and white-capping dissipation coefficients).

This paper is organized as follows. Section 2 deals with the methods used in this work. Since it is essential to understand, in depth, the relationship between the calibration of modelling parameters and the simulated sea state variables which are compared to the observations, a sensitivity analysis is associated to define the parameters to be calibrated, which is also subject of Section 2. Section 3 includes a description of the hindcast wave database ANEMOC. Section 4 presents the results. Section 5 offers some conclusions and outlooks.

2. Materials and methods

2.1. Optimal Calibration using variational data assimilation 3D-VAR

The inverse problem of calibration can be understood as the computation of the *posterior* distribution $\pi(\mathbf{X}|\mathbf{Y})$, where model parameters constitute the *p*-components of the parameter control vector $\mathbf{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$ composed of independent variables defined on some probability space, $\mathbf{Y} \in \mathbb{R}^m$ is the observation vector, also defined on a probability space, around the unknown parameter vector $\mathbf{X} \in \mathbb{R}^p$ and $\mathbb{R}^m$ the observation space is defined as:

$$\mathbf{Y} = G(\mathbf{X}) + \epsilon_o \tag{1}$$

where $G: \mathbb{R}^p \longrightarrow \mathbb{R}^m$ is a vector-valued function of vector $\mathbf{X}$ and $\epsilon_o \in \mathbb{R}^m$ is an observable measurement noise such as $\mathbb{E}(\epsilon_o) = 0$ and $\mathcal{R} = cov(\epsilon_o) = \mathbb{E}(\epsilon_o {\epsilon_o}^T) \in \mathbb{R}^{m \times m}$, and identified as a multivariate normal distribution, $\epsilon_o \sim \mathcal{N}(0, \mathcal{R})$.

The *posterior* distribution $\pi(\mathbf{X}|\mathbf{Y})$ can be determined through the well-known Bayes rule:

$$\pi(\mathbf{X}|\mathbf{Y}) \propto \mathcal{L}(\mathbf{Y}|\mathbf{X})\pi(\mathbf{X}) \tag{2}$$

The term $\mathcal{L}(\mathbf{Y}|\mathbf{X})$, called the likelihood, can be interpreted as the probability density function of the observed data, conditional upon a set of parameter values (considered as random variables), and as a measure of the information provided by the observations on the parameter values. From Eq. (1), the likelihood is expressed as $\mathcal{L}(\mathbf{Y}|\mathbf{X}) \propto exp\left(-\frac{1}{2}\big(\mathbf{Y} - G(\mathbf{X})\big)\mathcal{R}^{-1}\big(\mathbf{Y} - G(\mathbf{X})\big)^T\right)$.

The term $\pi(\mathbf{X})$ represents *a priori* knowledge of the unknown parameters $\mathbf{X}$. This term is classically taken as a multivariate normal distribution with known mean $\mathbf{X}_0$ (derived from measurement data or previous computation) and covariance matrix $\mathcal{B} \in \mathbb{R}^{p \times p}$ positive definite such as $\pi(\mathbf{X}) \propto exp\left(-\frac{1}{2}(\mathbf{X} - \mathbf{X}_0)\mathcal{B}^{-1}(\mathbf{X} - \mathbf{X}_0)^T\right)$. From the previous expressions of *a priori* and likelihood terms, the *posterior* distribution is given by the following equation:

$$\pi(\mathbf{X}|\mathbf{Y}) \propto \exp\left(-\frac{1}{2}\left[\big(\mathbf{Y} - \mathrm{G}(\mathbf{X})\big)\mathcal{R}^{-1}\big(\mathbf{Y} - \mathrm{G}(\mathbf{X})\big)^{\mathrm{T}} + (\mathbf{X} - \mathbf{X}_0)\mathcal{B}^{-1}(\mathbf{X} - \mathbf{X}_0)^{\mathrm{T}}\right]\right) \tag{3}$$

The maximum *a posteriori* (MAP) is equivalent to the formulation of the optimal search of control vector $\mathbf{X}$, which must satisfy the *a priori* error statistics $J_b = \frac{1}{2}(\mathbf{X} - \mathbf{X_0})\mathcal{B}^{-1}(\mathbf{X} - \mathbf{X_0})^T$ and the equivalent observation error statistics $J_o = \frac{1}{2}\left[\big(\mathbf{Y} - \mathrm{G}(\mathbf{X})\big)\mathcal{R}^{-1}\big(\mathbf{Y} - \mathrm{G}(\mathbf{X})\big)^{\mathrm{T}} + (\mathbf{X} - \mathbf{X}_0)\mathcal{B}^{-1}(\mathbf{X} - \mathbf{X}_0)^{\mathrm{T}}\right]$. This is known as the traditional variational data assimilation cost function, called 3D-VAR (Carrassi *et al.*, 2018).

Mathematical methods can be used to solve optimization problems. The former can vary significantly according to the form of the cost function (convex, quadratic, nonlinear, etc.), its regularity, and the dimension of the space. Many deterministic optimization methods are known as gradient descent methods, among which the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method used in this work. The chosen optimization method involves computing the adjoint of the observation operator G (or the partial derivatives of the operator with respect to its input parameters). In this work, the partial derivatives are approximated using a classical finite difference method. As the inverse problem is defined, a relevant question is: what are the effects of the modelling calibration parameters $\mathbf{X}$ on the simulated sea state variables $\mathrm{G}(\mathbf{X})$ which are compared to the observations? This question can be addressed by a sensitivity analysis.

2.2. Sensitivity analysis

The sensitivity analysis aims at quantifying the impact of uncertainty in input parameters on the accuracy of the model output variables. Conventional approaches to Global Sensitivity Analysis (GSA) imply the stochastic estimation of statistical moments. The variance-based methods aim at decomposing the variance of the output. The sensitivity analysis intends to quantify the relative importance of each input parameter of a model to the model output variance. Generally, these techniques compute sensitivity indices called Sobol' indices (Sobol', 1993). The definition of the Sobol sensitivity indices is a result of the ANOVA (Analysis Of VAriance) variance decomposition constructed on the hypothesis that the input variables are independent and the model is a square-integrable function of the input parameters. The ANOVA variance decomposition is given by the following equation:

$$\begin{cases} Var[G(\mathbf{X})] = \sum_{i=1}^{p} V_i(G(\mathbf{X})) + \sum_{i<j} V_{ij}(G(\mathbf{X})) + \cdots + V_{12\ldots P}(G(\mathbf{X})) \\ V_i(G(\mathbf{X})) = Var[\mathbb{E}(G(\mathbf{X})|X_i)] \\ V_{ij}(G(\mathbf{X})) = Var\left[\mathbb{E}\left(G(\mathbf{X})|X_i, X_j\right)\right] - V_i(Y) - V_j(Y) \\ \ldots \end{cases} \quad (4)$$

The term $\mathbb{E}(G(\mathbf{X})|X_i)$ represents the conditional expectation of the output $G(\mathbf{X})$ under the assumption that the uncertain variable X_i remains constant. From the resulting decomposition of the variance, the first order S_i and the total S_{Ti} Sobol sensitivity indices are defined by:

$$\begin{cases} S_i = Var(\mathbb{E}[G(\mathbf{X})|X_i])/Var(G(\mathbf{X})) \\ S_{Ti} = Var(\mathbb{E}[G(\mathbf{X})|X_{-i}])/Var(G(\mathbf{X})) \end{cases} \quad (5)$$

where X_{-i} refers to the set of uncertain input factors excluding X_i.

In this work, a polynomial chaos expansion has been carried out to estimate Sobol's sensitivity indices (Sudret, 2008).

3. Application to ANEMOC hindcast wave numerical database

The hindcast database ANEMOC (in French "Atlas Numérique d'Etats de Mer Océaniques et Côtiers) is built from numerical simulations and provides sea state conditions over different geographic domains. The subject of the present work is the ANEMOC-3 version built over the Atlantic Ocean, English Channel and North Sea French coast covering the period from 1979 to 2023. It is built with the help of two nested meshes (Fig. 1). The first mesh, called "oceanic", covers the whole Atlantic Ocean. It has a relatively coarse spatial resolution with a minimum element size of approximately 20 km along the European coastline. Its main purpose is to compute wave generation and propagation for providing boundary conditions to the other (finer) mesh, called the "coastal" mesh. This latter one has an element size of under 1 km along the French coastline. Only the oceanic mesh is subject of the present work.

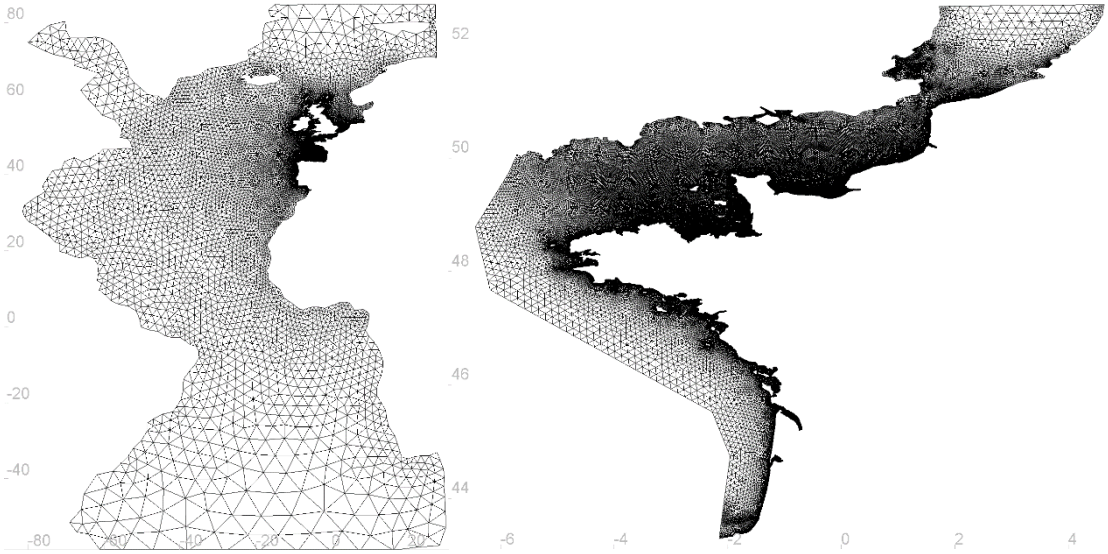

Fig. 1. ANEMOC-3 oceanic (left panel) and coastal (right panel) domains.

Raoult *et al.* (2018) included tidal water levels and currents effects on the wave propagation over the coastal domain of ANEMOC-3. In order to do so, chained simulations were used aligning the hydrodynamic 2D model, TELEMAC-2D, and the third-generation spectral wave model TOMAWAC model (Benoit et al., 1996), which are part of the TELEMAC-MASCARET numerical platform (**www.opentelemac.org**). While TOMAWAC is run on both oceanic and coastal domains, TELEMAC-2D is run on the coastal domain only, such that tidal water levels and currents are computed and updated in TOMAWAC coastal domain every 15 min. ANEMOC-3 gives 30 min period output of wave parameters such as spectral significant wave height H_{m0}, mean wave periods $T_{m-1,0}$ and T_{m02} , peak period T_p , directional spreading σ, mean wave direction Dir_m and mean wave power (per meter of length crest). The results presented hereafter concern the oceanic domain only. More details concerning the hindcast database ANEMOC-3 are described in Teles *et al.* (2022).

3.1. The wave action conservation equation

TOMAWAC solves the wave action balance equation with energy source/dissipation processes modelled with semi-empirical parameterizations. The relevant variable for describing the sea state is the directional spectrum of wave energy which is also known as wave directional spectrum of energy and will henceforth be denoted as E. From the wave directional spectrum, the wave action can be expressed as follows:

$$N = E/\rho g\sigma = F/\sigma \tag{6}$$

where σ denotes the relative or intrinsic angular frequency, i.e. the angular frequency being observed in a coordinate system moving at the velocity of current, ρ the water density, g the gravitational acceleration and F the directional variance spectrum.

TOMAWAC solves the following action flux conservation or balance equation:

$$\frac{\partial N}{\partial t} + \frac{\partial(\dot{x}N)}{\partial x} + \frac{\partial(\dot{y}N)}{\partial y} + \frac{\partial(\dot{k_x}N)}{\partial k_x} + \frac{\partial(\dot{k_y}N)}{\partial k_y} = Q(k_x, k_y, x, y, t) \tag{7}$$

where x and y are the horizontal Cartesian coordinates, t is time, k_x and k_y the wave number for directional spectrum discretization along x and y coordinates respectively and $Q(k_x, k_y, x, y, t)$ are the source, transfer and sink terms.

The source and sink terms that compose $Q(k_x, k_y, x, y, t)$ in the right-hand members of Eq. (7) gather the contributions from the physical processes listed in Fig. 2.

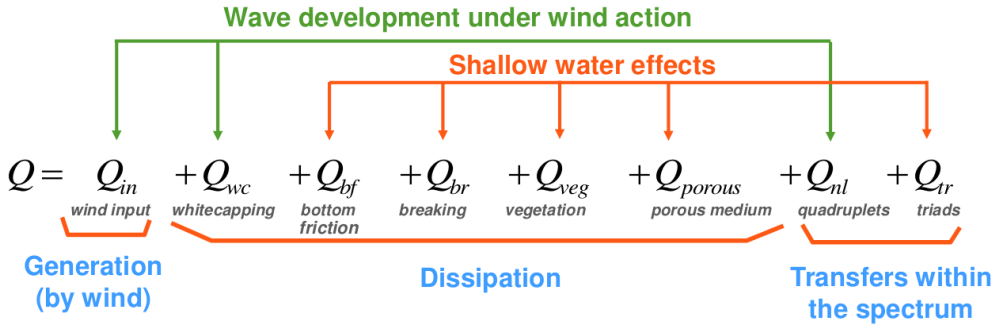

Fig. 2. Overview of TOMAWAC source and sink terms

In this work, the wind generation source term (Q_{in}) based on Janssen's model (Janssen, 1991) and white-capping dissipation term (Q_{wc}) from the van der Westhuysen et al. (2007) model are studied. Within the oceanic mesh, only these two terms together with the non-linear quadruplets term are activated in the calculations.

3.1.1 Wind-driven wave generation

Janssen's input model, coming from a quasi-linear theory for modelling the ocean/atmosphere interactions, is function of the directional variance spectrum $F(f, \theta)$:

$$Q_{in} = \sigma\epsilon\beta\left(\left[\frac{u_*}{C} + z_\alpha\right] max[cos(\theta - \theta_w); 0]\right)^2 F(f, \theta) \tag{8}$$

with ϵ the ratio of air and water densities, $C = \sigma/k$ the wave phase velocity, θ_w the local wind direction (direction where it blows), u_* the friction velocity, being linked to the surface stress, z_α a constant allowing to offset the growth curve, β a function such as $\beta = (\beta_m/\kappa^2)\mu \ln^4 \mu$ where β_m denotes a coefficient set to 1.2 by Janssen (1991), κ is the Von Karman's constant and μ denotes the non-dimensional critical height such as $\mu = min\left[\frac{gz_0}{C^2} exp\left(\kappa/\left[\frac{u_*}{C} + z_\alpha\right] cos(\theta - \theta_w)\right); 1.0\right]$ with $z_0 = \alpha \frac{{u_*}^2}{g\sqrt{1-\tau_w/\tau_s}}$ where α denotes the Charnock constant set to 0.01 in Janssen (1991) and τ_s and τ_w are respectively the surface and wave stress.

To conclude, the wind-driven wave generation model contains two input parameters: the Charnock constant α and the coefficient β_m. The Charnock constant α and the coefficient β_m are assumed to be contained in intervals set to respectively $[0.007; 0.02]$and $[0.8; 1.6]$.

3.1.2 White-capping-induced energy dissipation

In this study, the white-capping or the wave steepness induced breaking is taken from van der Westhuysen (2008). This model is based on a saturation-based model formulation, which defines the Q_{wc} term as depending on the saturation threshold B_r. Its formulation combines the white-capping dissipation parameterization from the Komen *et al.* (1984) model with that of van der Westhuysen *et al.* (2007) as follows:

$$Q_{wc} = f_{br}(f)Q_{wc}^{W} + \left(1 - f_{br}(f)\right)Q_{wc}^{K} \tag{9}$$

with $f_{br}(f) = \frac{1}{2} + \frac{1}{2}\tanh\left\{10\left[\left(\frac{B(k)}{B_r}\right)^{1/2} - 1\right]\right\}$ where the saturation threshold B_r is a model parameter, Q_{wc}^{W} and Q_{wc}^{K} are respectively the dissipation parametrizations from van der Westhuysen *et al.* (2007) and Komen *et al.* (1984).

The expression proposed by van der Westhuysen *et al.*, (2007) is:

$$Q_{wc}^{W} = -C_{dis,break}\left(\frac{B(k)}{B_r}\right)^{p_0/2} g^{1/2}k^{1/2}F(f,\theta) \tag{10}$$

where $B(k) = C_g k^3 \frac{E}{2\pi}$ with $C_g$the wave group speed, $p_0 = 3 + \tanh\left\{w\left(\frac{u_*}{C} - 0.1\right)\right\}$ with w set to 25 in van der Westhuysen *et al.* (2007) and $C_{dis,break}$ a coefficient parameter.

The expression proposed by Komen *et al.*, (1984) is:

$$Q_{wc}^{K} = -C_{dis}\bar{\sigma}\bar{k}^4 m_0^2\left(\delta\frac{k}{\bar{k}} + (1-\delta)\left(\frac{k}{\bar{k}}\right)^2\right)F(f,\theta) \tag{11}$$

with m_0 the total variance, $\bar{\sigma}$ the average intrinsic frequency and $\bar{k}$ the average wave number. C_{dis} and δ two parameters of the dissipation model.

At the end, the white-capping-induced energy dissipation contains four constants assumed here to be calibration parameters: the parameters $C_{dis,break}$ and C_{dis}, the white-capping weighting coefficient δ in Komen's relationship and the saturation threshold B_r. The parameter variation intervals are set such as: $C_{dis,break} \in [0.1; 1.] \times 10^{-4}$, $B_r \in [0.5; 2.5] \times 10^{-3}$, $C_{dis} \in [1.; 6.]$ and $\delta \in [0.; 1.]$.

3.1.3 Summary of parameter range variation

All the model input parameters and their associated probability distribution are summarized in Table 1.

Table 1. Input parameters and associated probabilistic models for ANEMOC-3 application.

Variable name	Nature	Variation interval	Probability distribution
Charnock constant α	Real scalar	$[0.007; 0.02]$	Uniform
Wind-driven coefficient β_m	Real scalar	$[0.8; 1.6]$	Uniform
Van der Westhuysen dissipation coefficient $C_{dis,break}$	Real scalar	$[0.1; 1.] \times 10^{-4}$	Uniform
Saturation threshold B_r	Real scalar	$[0.5; 2.5] \times 10^{-3}$	Uniform
White-capping dissipation coefficient C_{dis}	Real scalar	$[1.; 6.]$	Uniform
White-capping weighting coefficient δ	Real scalar	$[0.; 1.]$	Uniform

3.2. Available data

The capability of ANEMOC to model mean and extreme sea states is here evaluated based on buoy measurements during the one-month period of February 2014. Several very intense storms occurred over the northern part of the Atlantic Ocean within this time frame. The measurements are collected from both offshore (deep waters) and coastal (intermediate waters) wave buoys from the CANDHIS (France) and UKMO networks, with water depths ranging from 20 m to 4300 m (Fig. 3).

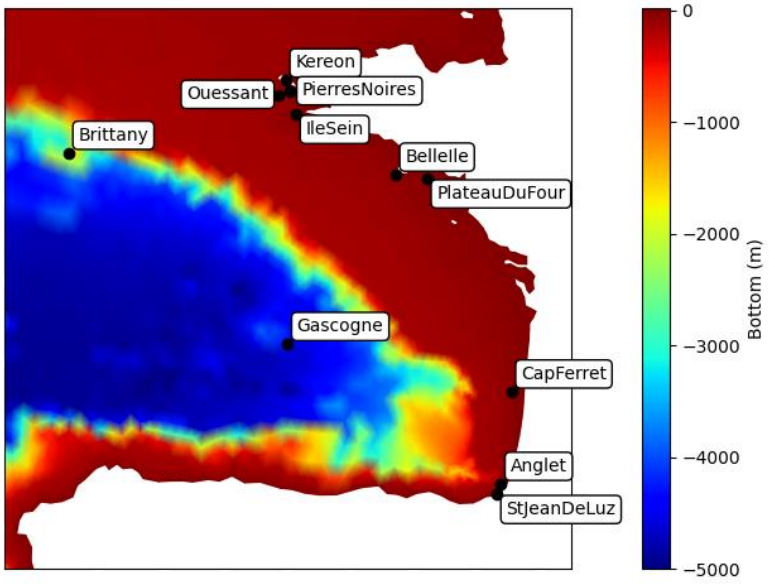

Fig. 3. Bottom elevation and location of buoys considered over the ANEMOC-3 oceanic model.

The spectral significant wave height $H_{m0}$was continuously measured from CANDHIS (every 30 min) and UKMO (every 60 min) wave buoys. As an example, Fig. 4 shows the time series of H_{m0}, during January and February 2014, recorded at the Brittany buoy (UKMO 62163), located offshore Brittany in the Atlantic Ocean. It can be noticed that, in the first 15 days of February 2014, H_{m0} remained higher than 4 m, with high peak values reaching 13-14 m. In this 2-month period January-February 2014 alone, 6 storms have their peak H_{m0}values exceeding 10 m at this location.

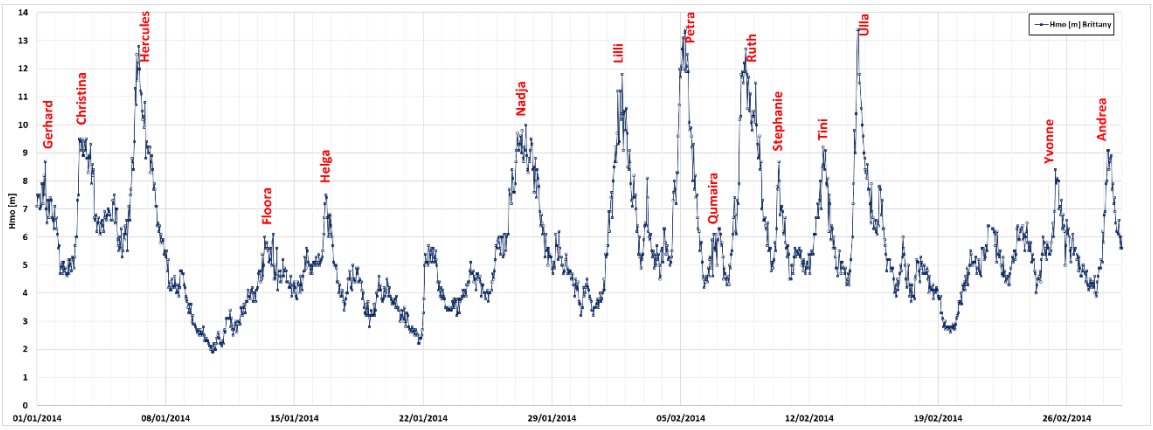

Fig. 4. Significant wave height evolution during the months of January and February 2014 measured at Brittany buoy (UKMO 62163).

4. Results

This study has been carried out following the Application Programming Interface (API) framework described by Goeury *et al.* (2022). In this framework, the sensitivity analysis and the calibration algorithm are performed by coupling TOMAWAC respectively with the open-source library for uncertainty treatment named "OpenTURNS", standing for "Open source initiative to Treat Uncertainties, Risks'N Statistics" (www.openturns.org) (Baudin *et al.*, 2015) and the data assimilation library ADAO (A module for Data Assimilation and Optimization) (https://pypi.org/project/adao) (Argaud, 2019).

4.1. Sensitivity analysis

To handle dynamic system behaviour under parameter uncertainty, a set of sample configurations is generated using random sampling. In this study, a design of experiment of size 1000 is constructed for the Polynomial Chaos Expansion (PCE) learning step. This number of model evaluations was determined based on a convergence study carried out the sensitivity analysis. In the sampling procedure, the parameter uncertainties are taken in a uniform distribution whose limits are defined by the minimum/maximum values of the variation range of each parameter (Tab. 1). The solver TOMAWAC ensures the relationship between a configuration

vector of model uncertain inputs $\mathbf{X}^j = (\mathrm{X}_{j,1}, \mathrm{X}_{j,2} \cdots \mathrm{X}_{j,p})$ and the output quantity of interest $G(\mathbf{x})$ composed of scalar output given at discrete time $t \in [1, \ldots, \mathrm{T}]$ for each point of interest corresponding to buoy locations. As an example, Fig. 5 shows the significant wave height evolution measured and resulting from Monte Carlo computations at Ouessant buoy during February 2014.

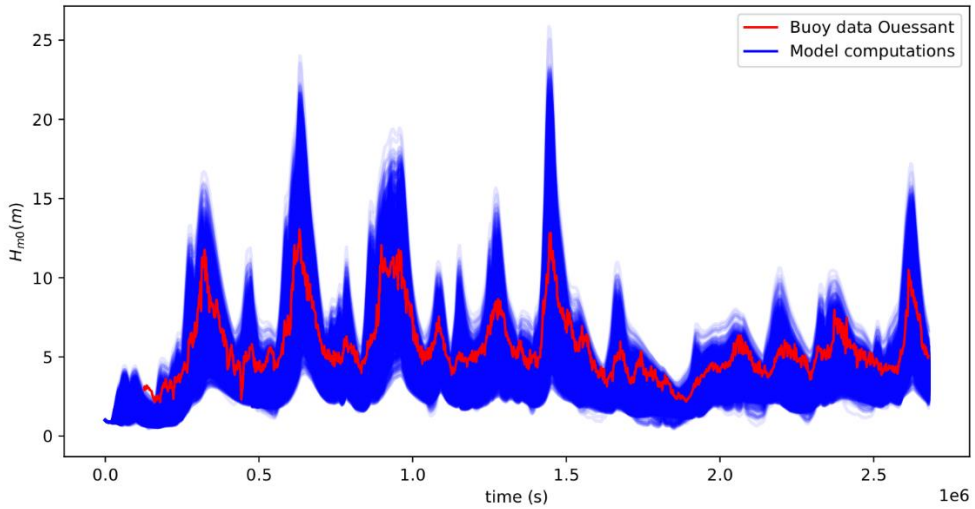

Fig. 5. Significant wave height evolution measured and resulting from Monte Carlo computations at Ouessant buoy during February 2014.

The construction of the PCE is carried out based on Least Angle Regression Stagewise (LARS) method to construct an adaptive sparse PCE. In this approach, a collection of possible PCEs, ordered by sparsity, is provided and an optimum one can be chosen with an accuracy estimate such as corrected leave-one-out error used in this study. Then, the Sobol's indices are obtained based on post-treatment of the constructed PCE as explained in Sudret (2008). Fig. 6 displays the time evolution of the sensitivity analysis estimate at Belle Ile buoy.

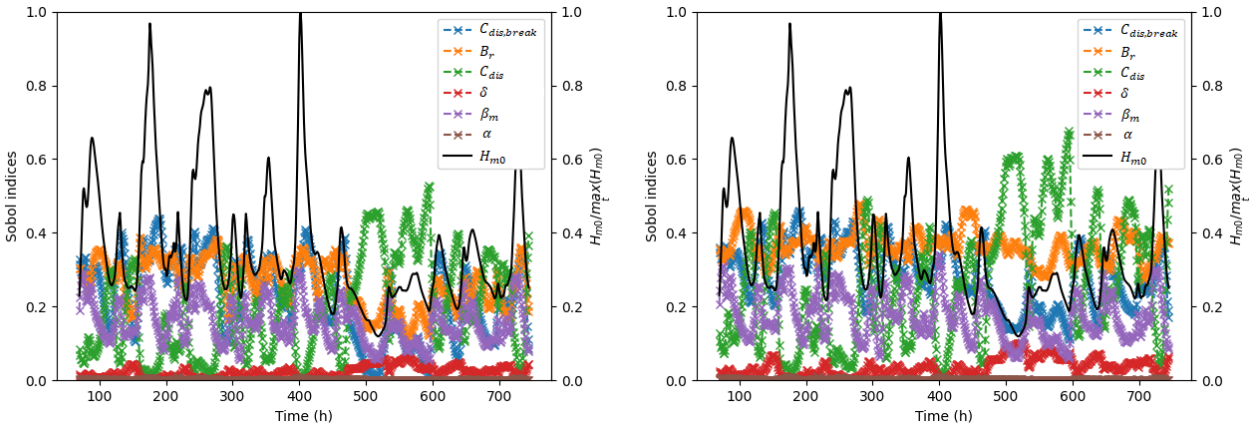

Fig. 6. Time evolution of Sobol' sensitivity indices at Belle Ile buoy: First (left) and total (right) order indices.

As shown by the sensitivity analysis, the most influent variables on H_{m0}, are the van der Westhuysen dissipation coefficient $C_{dis,break}$, the white-capping dissipation coefficient C_{dis}, the saturation threshold B_r and the wind-driven coefficient β_m. In fact, the sensitivity to the white-capping weighting coefficient δ and the Charnock constant α can be considered negligible in comparison. It can be noticed that there is no significant difference between the first and the total Sobol' indices for the wind-driven coefficient β_m. Since the variance part explained by variable interactions of the input factor with the other uncertain parameters is determined by substracting the total Sobol sensitivity index and the first order Sobol index, this means that the interactions between the wind-driven coefficient β_m with the other parameters are negligible. On the contrary, the interactions of the white-capping-induced energy dissipation variables have an impact on the spectral significant wave height.

To conclude, the calibration problem initially composed of six input variables can be reduced to four (van der Westhuysen dissipation coefficient $C_{dis,break}$, white-capping dissipation coefficient C_{dis}, the saturation threshold B_r and the wind-driven coefficient β_m) after considering the results of sensitivity analysis.

4.2. Calibration results

As shown by Eq. (3), the optimal search for the control vector **X** takes a minimization form of an objective or cost function. This minimization process, equivalent to the maximum *a posteriori* search, is carried out using the 3D-VAR algorithm. The control parameter is composed of the four most influential variables identified by the sensitivity analysis performed previously. The initial guess $\mathbf{X}_0$ is set to random values inside the constrained search space (Tab. 1). The observation vector **Y** is the spectral significant wave height extracted every 60 min at Ouessant, Belle Ile and Cap Ferret buoys from midnight February 2 to midnight March 1, 2014. The chosen optimization method involves computing the partial derivatives of the observation operator G with respect to **X**, a classical finite difference method with a differential increment set to 10^{-3} is used here. The error background and observation covariance matrices respectively identified by $\mathcal{B}$ and $\mathcal{R}$ are taken to be diagonal, meaning they have no error correlations. A small variance value for $\mathcal{R}$, justifying great confidence in the observation value, is considered. On the contrary, little confidence is given to the prior knowledge with a high variance value for $\mathcal{B}$. After repeating 10 times the calibration methodology with different initial guesses $\mathbf{X}_0$, the optimal value giving the best cost function result at the end of the optimization process is selected. The calibration algorithm finds an optimal solution in about 17 iterations with the following set of parameters $\mathbf{X}_{\mathrm{MAP}} = \left(C_{dis,break} = 5.59 \times 10^{-5};\ B_r = 2.15 \times 10^{-3}; C_{dis} = 1.08\ ; \beta_m = 1.05\right)$. Fig. 7 displays the results of the calibration process over the computation period at Cap Ferret buoy. The simulation obtained from the optimal set of parameters $\mathbf{X}_{\mathrm{MAP}}$ is compared with that obtained from the parameter default values in TOMAWAC solver ($\mathbf{X}_{\mathrm{PDV}} = \left(C_{dis,break} = 5 \times 10^{-5};\ B_r = 1.75 \times 10^{-3}; C_{dis} = 3.29\ ; \beta_m = 1.2\right)$).

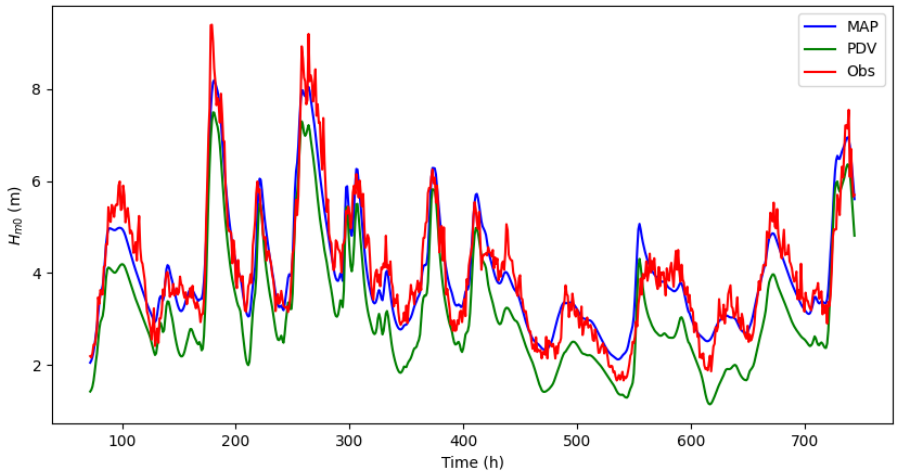

Fig. 7. Comparison of the spectral significant wave height time evolution with and without calibration (respectively - MAP and - PDV) with respect to the buoy measurements at Cap Ferret (- Obs) during February 2014.

As expected, the spectral significant wave heights, calculated with the calibrated parameter configuration, are much closer to measurements than the ones computed with the parameter default values. Hereafter, Fig. 8 shows the root mean square error (RMSE) performance of the re-analysis methodology on different buoys.

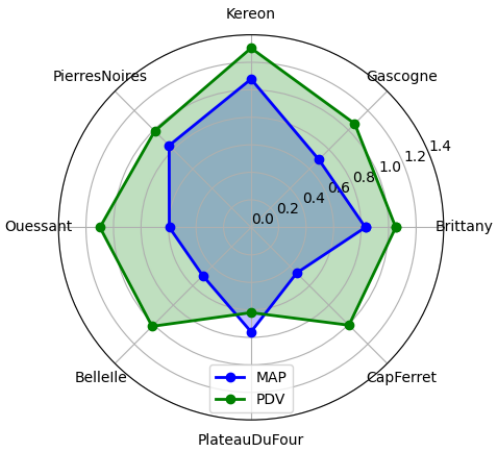

Fig. 8. Comparison of the RMSE on spectral significant wave height evolution with and without calibration (respectively - MAP and - PDV) with respect to the buoy measurements.

The result emphasizes the efficiency of the automatic calibration tool in the framework of ANEMOC-3 application. In fact, after the re-analysis procedure, the RMSE decreases significantly for almost all buoys, indicating that the model after calibration fits the data better. A particular behaviour can be noticed for Plateau du Four buoy with a deterioration in results after calibration. Finally, the re-analysis results are compared to the ANEMOC-3 configuration where the white-capping-induced energy dissipation is taken from Komen *et al*. (1984) (with dissipation coefficient $C_{dis} = 2.1$ and weighting coefficient $\delta = 0.4$). Fig. 9 and Fig. 10 show the scatter plot between observed versus modelled significant wave height results and the superposed Q-Q (quantile-quantile) plot (in black) for the ANEMOC-3 (left) and for re-analysis (right) configurations respectively at Belle-Ile and Cap Ferret buoys. As expected, a general good agreement is found, and the re-analysis configuration presents a slight improvement in simulated results compared to ANEMOC-3 ones.

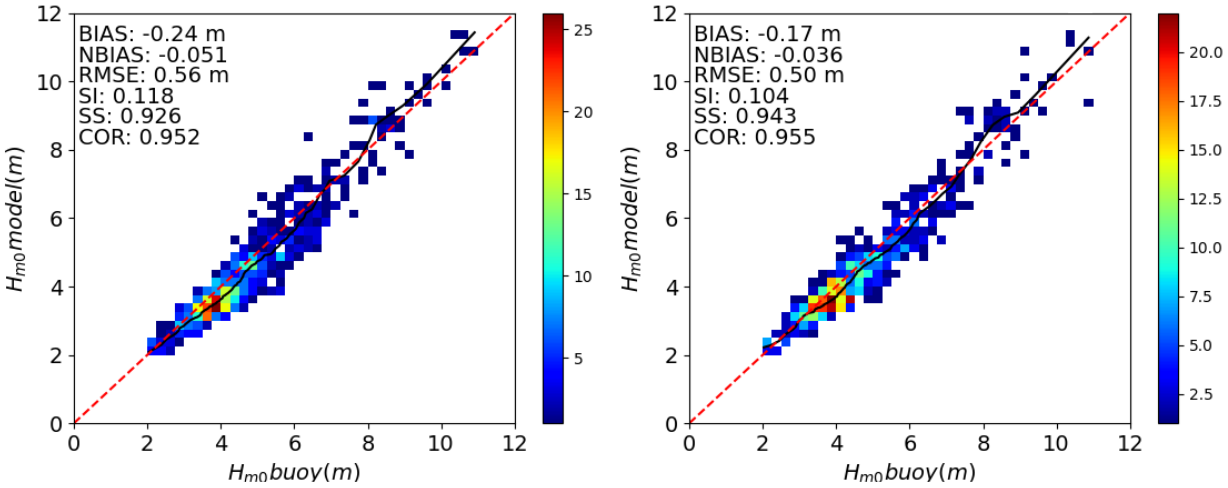

Fig. 9. Scatter plot and Q-Q plot of modelled versus measured H_{m0} domains at Belle-Ile buoy for ANEMOC-3 (left) and calibrated (right) configurations.

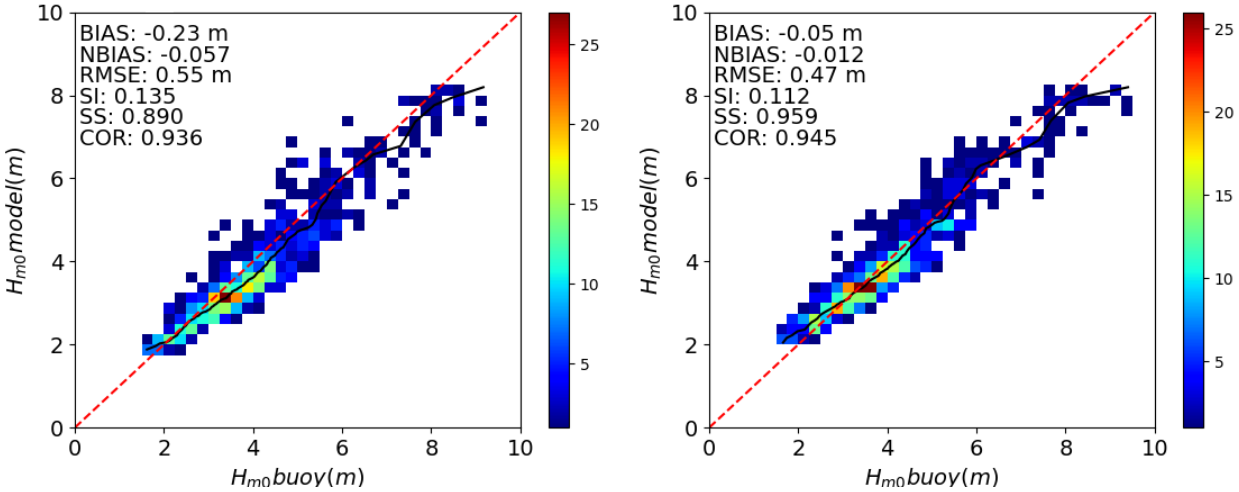

Fig. 10. Scatter plot and Q-Q plot of modelled versus measured H_{m0} at Cap Ferret buoy for ANEMOC-3 (right) and calibrated (left) configurations.

5. Conclusions

With the significant advances in the means of observing continental waters, the ability to exchange data as computational results has become a growing need. The objective of this work is to implement an efficient calibration algorithm, based on data assimilation, to best estimate partially known parameters of the source and sink terms (namely wind generation and white-capping dissipation coefficients) in the ANEMOC-3 hindcast wave numerical database application. This case deals with a serie of reference events by adjusting some uncertain physically based parameters until the comparison with observations achieves sufficient accuracy. If performed manually, the model calibration is time-consuming. Fortunately, the process can be largely automated to significantly reduce human workload, as shown in this paper. The efficiency of the re-analysis has been demonstrated over a selected period of one month including several intense storms. In fact, calculated statistical errors confirm a good fit with buoy measurement and the comparison with ANEMOC-3

configuration shows a slight improvement in simulated results. From this work, different axes of research are identified. The data-driven algorithm deployed here could be extended to (i) different numerical models (larger and/or more refined area covering for instance the European coastline), (ii) test alternative parameterizations of the physical sink and source terms in the wave model to better reproduce extreme events, (iii) integrate more observation data (remote sensing altimeter data, for instance), or (iv) correct the simulated wave spectra using observations.

Acknowledgements

The authors gratefully acknowledge contributions from the open-source community, especially that of OpenTURNS (Open source initiative for the Treatment of Uncertainties, Risks'N Statistics) and ADAO (a module for Data Assimilation and Optimization). We would like to thank Angélique Ponçot, Jean-Philippe Argaud and OpenTURNS team from EDF R&D for their constructive discussions.

References

Argaud, J. P. (2019) User documentation, in the SALOME 9.3 platform, of the ADAO module for "Data Assimilation and Optimization". Technical report 6125-1106-2019-01935-EN, EDF/R&D.

Baudin, M., Dutfoy, A., Iooss, B., Popelin, A. L. (2015) Open TURNS: An industrial software for uncertainty quantification in simulation. arXiv preprint arXiv:1501.05242.

Benoit M., Marcos F., Becq F. (1996) Development of a third generation shallow water wave model with unstructured spatial meshing, In Proc. 25th Int. Conf. on Coastal Eng. (ICCE'1996), Orlando (FL, USA), pp. 465-478.

Benoit M., Lafon F., Goasguen G. (2008) Constitution et exploitation d'une base de données d'états de mer le long des côtes françaises par simulation numérique sur 23 ans. Base ANEMOC en Atlantique – Manche – Mer du Nord. European Journal of Environmental and Civil Engineering, Vol. 12/1-2, pp 35-50. http://dx.doi.org/10.1080/19648189.2008.9692994

Carrassi A., Bocquet M., Bertino L., Evensen G. (2018) Data assimilation in the geosciences: An overview of methods, issues, and perspectives. WIREs Clim Change, 9:e535. https://doi.org/10.1002/wcc.535

Goeury C., Audouin Y., Zaoui F. (2022) Interoperability and computational framework for simulating open channel hydraulics: Application to sensitivity analysis and calibration of Gironde Estuary model, Environmental Modelling & Software, Vol. 148, 105243.

Janssen P.A.E.M (1991) Quasi-linear theory of wind-wave generation applied to wave forecasting, Journal of Physical Oceanography, Vol 21, pp. 1631-1642.

Komen G.J., Hasselmann S., Hasselmann K. (1984) On the existence of a fully developed wind-sea spectrum. Journal of Physical Oceanography, Vol. 14, pp 1271–1285.

Oreskes N., Shrader-Frechette K., Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263(5147), 641-646.

Raoult, C., Joly, A., Andreevsky, M., Joly-Laugel, A. (2018) ANEMOC-3: Improving the ANEMOC-2 Sea state database by adding tide effects. In Actes des 16èmes Journées de l'Hydrodynamique, 27-29 Novembre 2018, Marseille, France.

Sobol' I.M. (1993) Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiment, 1(4):407–414.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. Reliability engineering & system safety, 93(7), 964-979.

Teles M., Weiss M., Benoit M. (2022) Assessment of the ANEMOC-3 sea state hindcast database for modelling a series of energetic winter storms along the French coast, XVIIèmes Journées Nationales Génie Côtier – Génie Civil, Chatou, France, pp 171-182. DOI:10.5150/jngcgc.2022.019

Van der Westhuysen A.J., Zijlema M., and Battjes J.A. (2007) Nonlinear saturation-based white-capping dissipation in swan for deep and shallow water. Coastal Eng., 54: 151–170.

Van der Westhuysen A.J. (2008) Advances in the spectral modelling of wind waves in the nearshore, PhD thesis, University of Technology Delft, The Netherlands.