Robust Ensemble-Based Evolutionary Calibration of the Numerical Wind Wave Model

Pavel Vychuzhanin[(✉)], Nikolay O. Nikitin[(✉)], and Anna V. Kalyuzhnaya

ITMO University, 49 Kronverksky Pr., St. Petersburg 197101, Russian Federation
pavel.vychuzhanin@gmail.com, nikolay.o.nikitin@gmail.com

Abstract. The adaptation of numerical wind wave models to the local time-spatial conditions is a problem that can be solved by using various calibration techniques. However, the obtained sets of physical parameters become over-tuned to specific events if there is a lack of observations. In this paper, we propose a robust evolutionary calibration approach that allows to build the stochastic ensemble of perturbed models and use it to achieve the trade-off between quality and robustness of the target model. The implemented robust ensemble-based evolutionary calibration (REBEC) approach was compared to the baseline SPEA2 algorithm in a set of experiments with the SWAN wind wave model configuration for the Kara Sea domain. Provided metrics for the set of scenarios confirm the effectiveness of the REBEC approach for the majority of calibration scenarios.

Keywords: Evolutionary algorithm · SWAN wind wave model · Ensemble modelling · Robust optimisation · Model calibration

1 Introduction

The various tasks of offshore development and coastal shipping make it necessary to use the regional configurations of the numerical wind wave models to reproduce historical extreme events and predict potential hazards. To obtain the forecasts and hindcasts of desired quality, the suitable physical parameters of models should be identified for the specific simulation conditions.

The numerical model calibration of ocean wind wave model involves the fitting of simulation results with the in-situ and satellite wave measurements. The purpose of calibration is the identification of the physical parameters set that allows minimising the discrepancy between the model and observations.

However, it is a sophisticated task to calibrate the model manually even with the metocean experts' involvement. The modern wind wave models are computationally intensive, and each simulation run can take hours to compute. Also, a dramatically low time-spatial coverage of the available historical wave measurements and low quality of atmospheric reanalyses in some regions (like

the Arctic seas, in particular the Kara Sea region described in the paper) makes it hard to validate the parameter set reliably. The obtained parameters with minimal discrepancy can be very specialised in case of over-fitting to the low number of observed data points and can actually decrease the quality of long-term simulation results in non-observed locations or time ranges [3].

There are many well-known optimisation approaches that can be applied to automate the parameters' tuning for environmental models as well as [8]. Despite this, in the paper a task-specific robust evolutionary algorithm is proposed. It allows to make reliable calibration decisions in situations with high environmental uncertainty and tries to ensure a tolerable solution identification.

At the moment, the modern atmospheric reanalysis still has quality issues in the Arctic region [7]. We proposed an algorithm that establishes artificial diversity for wind velocity fields. It was used to generate the probabilistic ensemble of input wind fields to take the impact of the surface forcing uncertainty into account. Then, the multi-objective fitness function was used to achieve the trade-off between robustness and performance of the optimised model.

We conducted a set of experiments to verify the effectiveness of the proposed approach against the baseline SPEA2 algorithm using the Kara Sea domain and the SWAN (Simulating WAves Nearshore) [2] model as the case study. The nine spatially distributed points were chosen to analyse the performance and robustness of the model's configurations obtained after calibration in one-month training runs of the model. The several configurations with different subsets of calibration and validation points were compared to estimate the statistical metrics of optimisation effectiveness for both algorithms.

This paper is structured as follows. Section 2 describes the problem statement and mathematical formalisation of the robust optimisation task. Section 3 provides an overview of various calibration approaches and their applicability for the problem. Section 4 contains a detailed description of the baseline SPEA2 algorithm and the proposed robust algorithm. Section 5 is dedicated to the experimental studies (model configuration, datasets, results and metrics). Section 6 summarises the obtained results and highlights of the key findings.

2   Problem Statement

As it was noted in the introduction, coverage of observed met-ocean data (especially oceanic observations) is extremely sparse. Although, reliable information about met-ocean characteristics is needed in many regions (e.g. Arctic seas). That's why during last decades it became a common practice to obtain the information about met-ocean events and processes from forecasting or hindcasting (retrospective) simulation results from numerical hydrodynamic models. Nevertheless, for solving such task the numerical models should be fitted (through model parameters) to the certain water area. Taking into account few spatial points and small sizes of datasets with observations, there is a serious risk of model overfitting when model fits to specific features of observed data instead of fitting to common features of the target region. Description of the solution to this problem is the main goal of this article.

Hydrodynamic model fitting through the tuning of model parameters (or model calibration) can be formulated as an optimisation task. For this purpose, it is reasonable to present the simulation process in a general mathematical notation (1).

$$Y = \{Y_1, Y_2, ..., Y_k\} = M(\xi \mid \theta), \tag{1}$$

where $Y = \{Y_1, Y_2, ..., Y_k\}$ denotes multivariate output data (simulated fields, e.g. wave heights), $M(\bullet)$ is the model operator, ξ is the input data (boundary and initial conditions), θ is the set of model parameters.

With that, the tuning of model parameters (or model calibration) can be formalized in terms of multi-objective optimisation in the model parameter space and written as:

$$\begin{aligned} \theta_{opt} &= \arg \min_\theta F(\theta), \\ F(\theta) &= \mathcal{G}(f_i(\theta, Y, \{x, y\})), \end{aligned} \tag{2}$$

where $\mathcal{G}(\bullet)$ is an operator for multiobjective transformation to F, f_i is the objective function, $i = 1 \dots n$, $\{x, y\}$ are spatial coordinates of a point-of-interest.

In a case of wind waves hindcasting, the poor time and spatial coverage of observations make the model optimisation much harder. The over-fitting of the solution to the specific events represented in small data samples can cause a non-optimal model configuration with lower quality under different external conditions. One of the ways to improve the robustness of optimisation results is to enlarge training dataset with new instances with relatively small artificial disturbances. This issue makes it necessary to take the simulation uncertainty factors into account.

The uncertainty in the wind wave model can be represented not only by disturbances in design variables [18]. There are deviations in the environment variables that can be represented through input data sets diversity (for the SWAN model the wind forcing obtained from atmospheric reanalysis is most important). In this case input data ξ should be transformed to ensemble realisation $\{\xi\}_n = \{\xi_1, ..., \xi_n\}$ by addition of artificial disturbance (or noise) and Eq. (2) transforms into Eq. (3). A detailed description of the ensemble procedure is given in Sect. 4.3.

$$\begin{aligned} \theta_{rob} &= \arg \min_\theta \tilde{F}(\theta \mid \{\xi\}_n), \\ \tilde{F}(\theta \mid \{\xi\}_n) &= \mathcal{G}(\tilde{f}_i(\theta \mid \{\xi\}_n, Y, \{x, y\})). \end{aligned} \tag{3}$$

An ensemble objective function $\tilde{f}_i$ defines landscape of objective function over the space of parameters considering ensemble of input states $\{\xi\}_n$. As an example, ensemble fitness function can be represented by the expected function for the ensemble of runs with small disturbances in input data (shown in Eq. (4)). This approach can be used to produce better solutions for the set of diverse environmental scenarios and increase the expected performance.

$$\tilde{f}(\theta \mid \{\xi\}_n) = \int_{-\infty}^{\infty} f(\mathbf{x}, \xi + \boldsymbol{\delta}) \cdot p(\boldsymbol{\delta}) d\boldsymbol{\delta} \tag{4}$$

As an example of the hydrodynamic model for experimental studies, third-generation wind wave model SWAN [2] was chosen. The wind waves are surface waves in the oceans and seas that caused by the interaction between water masses and sea-level wind. Wind waves models of third-generation (e.g. SWAN) allow to simulate the wave spectra and to reconstruct characteristics of waves (e.g. heights, periods, directions). The SWAN model can be described with the action balance equation (5).

$$\frac{\partial}{\partial t}N + \frac{\partial}{\partial x}c_x N + \frac{\partial}{\partial y}c_y N + \frac{\partial}{\partial \sigma}c_\sigma N + \frac{\partial}{\partial \theta}c_\theta N = \frac{S}{\sigma}, \qquad (5)$$

where on the left-hand side $N = \frac{E}{\sigma}$ denotes the wave action density and E is an energy of wave spectrum, σ is the relative frequency, θ is the group wave direction, c is the group velocity in corresponding space. The right-hand side represents the source and sink term in a form Eq. (6).

$$S = S_{in} + S_{ds} + S_{nl}, \qquad (6)$$

where S_{in} is the input energy obtained by wind, S_{ds} is the energy of dissipation and S_{nl} denotes the energy of wave-wave nonlinear interaction.

These three terms represent the genesis of wave energy sources/sinks and are a powerful handle for wave model fitting. From this point of view, it is convenient to express energy sources through model parameters. Wind energy is characterised by the drag function (DRG), wave dissipation—by the wave breaking (STMP) and bottom friction (CFW) functions. Energy flow from nonlinear interactions is relatively small and wasn't taken into account in the current paper.

In the frame of this article, the experimental study (Sect. 5) was provided to assess the practical effectiveness of the proposed robust calibration method in comparison with the general-purpose calibration algorithms. The SWAN model configuration for the Kara Sea was chosen as a case study because of the importance of this region for offshore industrial development and extremely low density of sensors in areas of interest.

3   Related Work

Model calibration or tuning is a subject with extensive literature [8,25]. The conservative approach is to estimate the parameters in an expert way [10,16]. It includes the development of several candidate sets of parameters based on previous simulation experience and manual individual adjustment of each parameter. The quality metrics for the model quality assessment are calculated with the comparison of model time series and historical values obtained from the reanalyses and observations.

Since the manual "trials-and-errors" method is time-consuming and gives solution only for particular model setup, the automatic calibration of models is widely used for different aspects of environmental simulations like atmospheric [6] and ocean [26] forecasting tasks. As a basic approach, the space-filling

design for the parameter space can be used [24] for model calibration. However, the high-resolution configurations of wind wave models of 3rd generation are computationally-intensive and require a lot of time to process the appropriate date range and spatial domain. This problem makes it necessary to reduce the number of runs required for calibration.

There are many well-known optimisation methods applied to environmental models like derivative-free optimisation [22], various Bayesian optimisation methods [5] and surrogate-assisted methods [9].

However, the evolutionary (genetic) algorithms are efficient enough to perform a robust solution search [14] in complex parameter space with a lack of historical data for quality assessment [20]. The applicability of evolutionary algorithms for SWAN wave model calibration is demonstrated in [12].

The robust optimal design approaches have a lot of applications in many fields [27]. They are often based on Monte Carlo methods that allow representing the uncertainty from different sources [4]. The perturbation-based ensemble allows sampling the modelling uncertainty in a more systematic way [19]. A set of simulation with small differences induced by stochastic modifications allows to increase the variability of the calibration dataset and improve the quality of models [17].

Nevertheless, the discrepancy usually simulated as additional noise in model output and observations [1] without taking the actual sources of external uncertainty (e.g. wind forcing for reanalysis) into account. The task of a reliable calibration of a wave model for a specific domain with poor observational coverage makes it necessary to implement the approach that combines the ensemble-based diversity of environmental variables with multi-objective evolutionary optimisation.

4   Evolutionary Algorithms for Models Fitting

We compared the robust wave model calibration with a baseline solution—the multi-objective evolutionary algorithm that estimates the most suitable solution without taking uncertainty into account. The other approach is based on the same algorithm with modified fitness functions—it estimates the performance and robustness of the solution with the ensemble of forecasts obtained from several model's runs with noised inputs. The source code of both algorithms was implemented in Python and available in [23].

4.1   Baseline Approach

The commonly used SPEA2 multi-objective optimisation algorithm [28] was chosen as a baseline solution for the calibration task. In terms of evolutionary algorithms, in our case, each individual corresponds to a genotype represented by a certain set of model parameter and the phenotype (values of the objective function) are the errors of the model predictions, corresponding to these parameters. At each iteration of evolution, the Pareto-optimal set of individuals is

selected according to the values of the fitness function, and all non-dominated solutions are saved in the archive. Then the mating pool is filled with a binary tournament selection and recombination and a mutation operator are applying for each individual. The resulting mating pool becomes a new population at the next iteration of the algorithm.

Despite the fact that some modern evolutionary algorithms outperform SPEA2 in some synthetic tasks [13], we decided to base the experiments on a well-studied [12] algorithm to separate the impact from the proposed ensemble-based modifications from other features' influence.

4.2 Robust Ensemble-Based Evolutionary Calibration (REBEC) Approach

The main disadvantage of the baseline algorithm is that the model variables optimise exactly for the specific conditions that were used for the fitness function evaluation. It allows to maximise the performance for the observed case, but the solution found can be unstable even after small changes in external conditions. The lack of the time-spatial coverage of observational data for wave parameters in target regions makes it complicated to take the different external uncertainties (e.g. forcing-induced, resolution-induced, etc.) into account.

The more robust approach to model parameters optimisation can be implemented using the ensemble of wave models configured using different input data sets. We can form the stochastic ensemble of wave models with the perturbed wind forcings and search for more robust model parameters using this ensemble instead of a single model with certain forcing.

For this purpose, we can adapt the baseline SPEA2 algorithm (that was introduced above) by changing the fitness assignment strategy: for a given genotype, the set of phenotypes corresponding to the elements of the ensemble is estimated and based on its values the robust metric is calculated. The flowchart of the proposed algorithm is presented in Fig. 2.

It is important to find a compromise between performance and robustness of the obtained solution [11], so the fitness function for the algorithm is based on the composite estimation robustness and performance metrics. The performance can be calculated as a vector of root-mean-square errors (RMSE) against observations for a set of target points, and the robustness can be simulated in various ways [15]. Figure 1 depicts the set of the ensemble error surfaces that are used for metric calculation.

We tried to use the mean-variance as a robust metric, but it causes the domination of the solutions with low wind drag and, consequently, near-zero wind-induced variability. So, the ensemble mean was chosen as a trade-off metric.
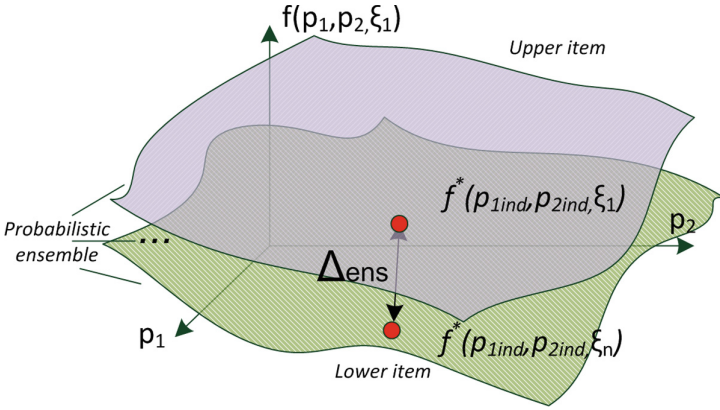
Fig. 1. The landscape of an objective function for a probabilistic ensemble

The pseudocode of the final implementation of the robust algorithm is presented in Algorithm 1.

Data: Initialised ensemble, populationSize,
archiveSize, crossoverRate, mutationRate
Result: best individual from archive
pop ← InitPopulation(*populationSize*)
archive ← ∅
while *not ConvergenceCriterion()* **do**
    for individ *in* pop **do**
        ensObjectives ← ∅
        for model *in* ensemble **do**
           | ensObjectives [model] ← CalculateObj(individ, model)
        end
        bestByObjectives ← TakeBestByMean(ensObjectives, *ensAmount*)
        individ.objectives ← Mean(bestByObjectives)
    end
    union ← archive + pop
    for individ *in* union **do**
        | individ.fitness ← CalculateFitness(individ)
    end
    archive ← TakeNonDominated(union, *archiveSize*)
    matingPool ← BinaryTournamentSelection(archive, *populationSize*)
    pop ← CrossoverAndMutation(matingPool, *crossoverRate*,
    mutationRate)
end
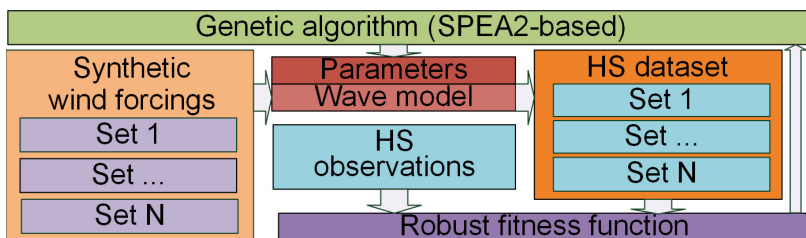**Algorithm 1.** The pseudocode of the implemented REBEC algorithm

Fig. 2. The main logical blocks and interconnections of proposed robust evolutionary algorithm

4.3   Synthetic Input Data Generation with Artificial Noise

To implement the proposed probabilistic optimisation method, we developed the supplementary algorithm that allows to add specific noise to wind velocity variables—U (eastward) and V (northward) vector components from atmospheric reanalysis data that is used by the wave model as an external forcing.

The algorithm starts from uniform scattering of the randomly-located sources of artificial noise in the gridded data. To obtain the realistic wind field after the application of noise, the time-spatial correlation terms are added to control the noise spreading from the source.

The noise function for the wind vector component U produced by one noise point can be written as:

$$f^*(j,t) = N(0,\sigma) \cdot corr(U_j, V_j) \cdot corr(U_t, U_{t-1}) \tag{7}$$

where j is the spatial index of source points, t is the time step index, σ is the standard deviation parameter of the Gaussian distribution, U is the matrix of wind U-components.

Then, the aggregated noise from N source points for specific data point induced by all source points can be obtained as:

$$f(i,t) = \sum_{j=1}^{N} f^*(j,t) \cdot corr(U_i, U_j) \tag{8}$$

where i is the spatial index of the data point, j is the spatial index of the noise point, t is the time step index, N is the number of noise points, U is the matrix of wind U-components.

The example of the wind field augmented with noise by the described method (with σ equal to 25% of basic value magnitude) is presented in Fig. 3.

It can be seen that the common wind patterns are similar but some wind speed variability exists. The additional post-processing procedure was applied to the perturbed model runs output to suppress the non-realistic wind height peaks in the observed calm periods. However, the near-peaks variability was preserved.

In this way, the ten wind data sets augmented with artificial noise were generated and used in an experimental study.
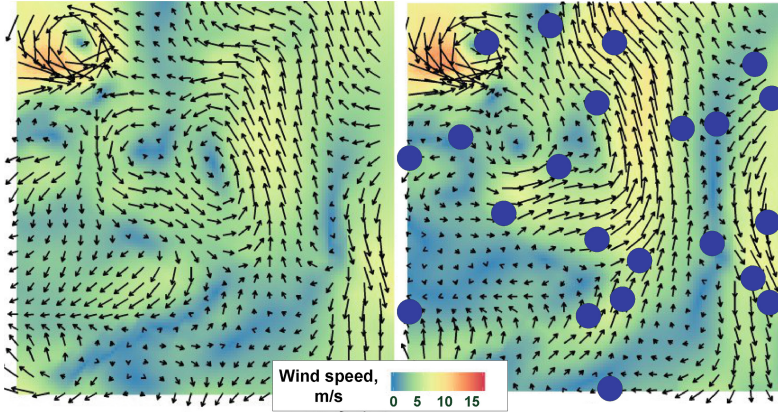
Fig. 3. The example of comparison of (a) basic ERA-Interim for Kara Sea region and (b) wind field augmented with noise for the same region. The blue marks depicts the noise source points locations. (Color figure online)

5   Experimental Study

The case study for the calibration task is based on the SWAN model configuration for the Kara Sea region. The significant wave height (Hsig) variable was chosen as a target variable. Moreover, the results in nine representative points were analyzed (P1–P9 presented in Fig. 4) to take into account possible spatial variability of the optimal solution.

5.1   Synthetic Data for Wave Observations

The wave observations data are required for the validation of the model quality and calibration algorithm effectiveness. However, such data often cannot be obtained from open data sources. To perform a reproducible experiment with Kara Sea configuration, we used the simulation results from the high-resolution WaveWatch III model [21] configuration. The systematic biases of synthetic observations against model were removed. Then we analysed the error metrics for the significant wave height variable against real observations in points 1–3 (RMSE is 0.29 m and MAE is 0.21 m). We accept the quality of the WaveWatch III output as sufficient to be used as the reference dataset for the optimisation algorithms' evaluation.

To maintain the variability of experimental scenarios, we prepared 18 subsets of synthetic observational points to be used for calibration. They consist of observation points located in various spatial areas with different depths and distances to the coast. To reproduce various scenarios, the calibration subsets were initialised with random point groups of a certain size: from a single-point situation to the all-points-instead-one case. For each subset, the points that were not used during the calibration were assigned to validation sets. It allows

to compare the effectiveness of the robust algorithm with baseline SPEA2 and analyse the dependency of results from selected observation points set.

5.2   Model Configuration

The SWAN model was configured with the regular curvilinear grid in cartesian coordinates. The initial conditions were obtained for a preliminary monthly spin-up run. The boundary conditions were not set (since the control points were distanced from the grid boundaries, see Fig. 4). The simulation dates range was set from 20140814.120000 to 20140915.000000. The time step for integration was defined as 120 min and output time step is 3 h. The parameterisations GEN3, COLLINS, QUADRUPL, TRIAD and DIFFRACtion were enabled. The output was configured to obtain the significant wave height (HS) values in 9 spatial points. Their locations are specified in Fig. 4.
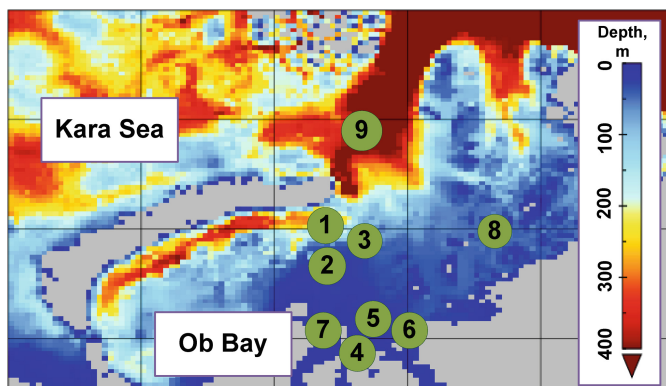

Fig. 4. The part of the bathymetry of the simulation domain: the Kara Sea and Ob bay. The land cells are shaded with a grey mask. The locations of observation points and their indices are specified with green marks. (Color figure online)

5.3   Sensitivity Analysis

The sensitivity analysis of model parameters described in Sect. 2 was performed to estimate their significance. We ran the set of experiments with every parameter independently modified by additive noise with Gaussian distribution with $\sigma/\mu = 0.25$ assumption. The results are obtained from 50 experiments with sequential noising of each variable from the chosen set.

The wind drag was identified as the most sensitive parameter with a high relative output and input variance ratios, the wave steepness is the second one and the sensitivity bottom friction coefficient is quite low for most of the comparison points. In can be concluded that the SWAN model error function has a wide "plateau" with similar error values and many local minimums in the "valley" area that can affect the algorithm's convergence and robustness.

5.4   Validation of REBEC Approach

A set of experiments was conducted to compare the results of optimisation experiments. The initial population for both calibration approaches was produced using Latin hypercube sampling (LHS) in the parameter space. The parameters for both calibration algorithms was chosen as: population size is set to 20 individuals, the number of generations—60, the archive size—5 individuals, the probability of mutation and crossover—0.2.

Two objective functions were chosen for model results quality assessment: the mean absolute error (MAE) and root mean square error (RMSE).

The calibration for every scenario was repeated 100 times to obtain the distribution of the relative improvement of RMSE and MAE against the model configuration with default parameter values (DRF = 1.0, CFW = 0.015, STPM = 0.00302). Also, the mean relative standard deviation of the calibrated parameters set is provided. The boxplots for the results with different scenarios are presented in Fig. 5.
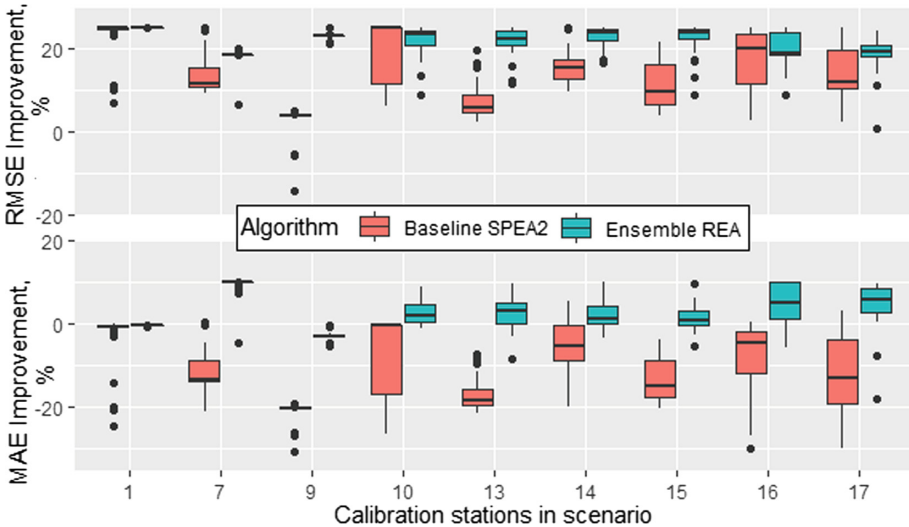

Fig. 5. The comparison of the baseline and robust algorithms' performance on the validation set of stations in all scenarios. The RMSE and MAE metrics are presented as an improvement against the corresponding values for the default configuration.

It can be seen that the variance of metrics for the robust algorithm is lower and the quality is better. The detailed metrics for all scenarios and stations sets are provided in Table 1.

As can be seen, the robust approach provides a better or equivalent improvement of model performance for the validation points in all groups of scenarios. The standard deviation for both model parameters and relative improvement

Table 1. Error metrics for the baseline and robust algorithms. The "test" block contains the metrics for the verification points. The "train" block contains the metrics for the calibration points. The boldface numbers indicate the best metrics for all station sets (the higher improvement and lower standard deviation is better)

Scenario	Algorithm	Validation points							Calibration points						
		Improvement, %						Par. SD	Improvement, %						Par. SD
		RMSE			MAE				RMSE			MAE			
		Mean	Max	SD	Mean	Max	SD		Mean	Max	SD	Mean	Max	SD	
1–9	BL	**11**	**22.1**	7.8	−9.2	1.8	7.1	3.3	6.6	**26.4**	13.7	2.1	13.6	6.8	3.3
	RB	11	15	**2.7**	**4**	**6.9**	**2**	**2.7**	**11.3**	16.4	**3**	**2.5**	**6.5**	**3**	**2.7**
10–14	BL	16.6	**23.6**	4.9	−7.2	0.7	5.2	2.7	**24.4**	**29.1**	4.1	7.9	12.8	2.7	2.7
	RB	**18.1**	21	**2.8**	**4.1**	**9.4**	**2.5**	**2.4**	22.9	26.9	**3.9**	**8.7**	**14.4**	**2.2**	**2.4**
15–18	BL	17.6	**25.1**	5.7	−6.6	2.8	6.5	3.1	**27.2**	**34.2**	6.2	**11.2**	**17.1**	**2.8**	3.1
	RB	**18.7**	24.7	**4.1**	**5.4**	**10**	**4.1**	**3**	23.4	33.4	**5.1**	11.2	17.9	3.2	**3**
All	BL	14.0	**23**	6.6	−8.4	1.7	6.4	3.2	14.5	**27.9**	9.7	4.7	**13.5**	4.9	3.2
	RB	**14.6**	18.3	**3.1**	**4.6**	**8.7**	**2.7**	**2.8**	**15.8**	21.6	**3.9**	**5.2**	10.5	**2.9**	**2.8**

values are also lower than the baseline. In can be concluded that the optimal algorithm choice for validation points varies in different scenarios. The scenarios 1–9 operate with a single-point calibration set. The performance of the robust algorithm for this group of validation points is similar to baseline RMSE (but outperforms it for the MAE metric and calibration points metrics). For the other scenarios, the gain is near 1–2% RMSE and 10% MAE against the baseline.

Also, the calibration set quality averaged for all scenarios for the robust approach also outperforms the baseline. The standard deviation of the obtained metrics is smaller for all scenarios, as well as the mean standard deviation for model parameters. We can claim that a robust approach is effective for the cases with several spatially scattered points that can be applied for calibration. It is important to notice that the calibration points' quality is not affected in a negative way.

6   Conclusion

In the paper, the practical approach to the calibration of numerical wave models under data quality and availability constraints was proposed. The algorithm for the simulation of artificial data diversity was implemented and applied to the ERA-Interim reanalysis wind data. The regional configuration of the SWAN model was used as a case study for the parameters tuning algorithm effectiveness evaluation.

The proposed REBEC approach was compared with the baseline SPEA2 algorithm in a set of experiments. The lower variability and better performance metrics for the spatially distributed calibration and verification points were obtained. It confirms the effectiveness of the robust calibration approach for

the simulation domains with a small number and poor coverage of real observations. However, the negative impact of the proposed approach for computational performance (several simulations should be performed for each candidate parameters set) makes the robust optimisation potentially non-preferable for the model configurations with the sufficient spatial coverage of observations and high-quality atmospheric reanalyses.

The source code of the algorithms for calibration, pre- and post- processing as well as the configuration files for SWAN are available in an open repository [23].

References

1. Bhat, K.S., Haran, M., Goes, M., Chen, M.: Computer model calibration with multivariate spatial output: a case study. In: Chen, M.-H., Müller, P., Sun, D., Ye, K., Dey, D. (eds.) Frontiers of Statistical Decision Making and Bayesian Analysis, pp. 168–184. Springer, New York (2010). https://doi.org/10.1007/978-1-4419-6944-6
2. Booij, N., Ris, R.C., Holthuijsen, L.H.: A third-generation wave model for coastal regions: 1. Model description and validation. J. Geophys. Res. Ocean. **104**(C4), 7649–7666 (1999)
3. Brynjarsdóttir, J., OHagan, A.: Learning about physical parameters: the importance of model discrepancy. Inverse Probl. **30**(11), 114007 (2014)
4. Che, J., Wang, J., Li, K.: A Monte Carlo based robustness optimization method in new product design process: a case study. Am. J. Ind. Bus. Manag. **4**(07), 360 (2014)
5. Cornejo-Bueno, L., Garrido-Merchán, E.C., Hernández-Lobato, D., Salcedo-Sanz, S.: Bayesian optimization of a hybrid system for robust ocean wave features prediction. Neurocomputing **275**, 818–828 (2018)
6. Duan, Q., et al.: Automatic model calibration: a new way to improve numerical weather forecasting. Bull. Am. Meteorol. Soc. **98**(5), 959–970 (2017)
7. Fredriksen, L.E.: An evaluation of the reanalyses ERA-Interim and ERA5 in the Arctic using N-ICE2015 data. Master's thesis, UiT The Arctic University of Norway (2018)
8. Hourdin, F., et al.: The art and science of climate model tuning. Bull. Am. Meteorol. Soc. **98**(3), 589–602 (2017)
9. James, S.C., Zhang, Y., O'Donncha, F.: A machine learning framework to forecast wave conditions. Coast. Eng. **137**, 1–10 (2018)
10. Jin, K.R., Ji, Z.G.: Calibration and verification of a spectral wind-wave model for Lake Okeechobee. Ocean Eng. **28**(5), 571–584 (2001)
11. Jin, Y., Sendhoff, B.: Trade-off between performance and robustness: an evolutionary multiobjective approach. In: Fonseca, C.M., Fleming, P.J., Zitzler, E., Thiele, L., Deb, K. (eds.) EMO 2003. LNCS, vol. 2632, pp. 237–251. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36970-8_17
12. Kovalchuk, S.V., et al.: A conceptual approach to complex model management with generalized modelling patterns and evolutionary identification. Complexity **2018**, 15 pages (2018)

13. Li, M., Yao, X.: An empirical investigation of the optimality and monotonicity properties of multiobjective archiving methods. In: Deb, K., Goodman, E., Coello Coello, C.A., Klamroth, K., Miettinen, K., Mostaghim, S., Reed, P. (eds.) EMO 2019. LNCS, vol. 11411, pp. 15–26. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12598-1_2

14. Liu, Y., Khu, S.T.: Automatic calibration of numerical models using fast optimisation by fitness approximation. In: International Joint Conference on Neural Networks, IJCNN 2007, pp. 1073–1078. IEEE (2007)

15. McPhail, C., Maier, H., Kwakkel, J., Giuliani, M., Castelletti, A., Westra, S.: Robustness metrics: how are they calculated, when should they be used and why do they give different results? Earth's Future **6**(2), 169–191 (2018)

16. Mortlock, T.R., Goodwin, I.D., Turner, I.L.: Calibration and sensitivities of a nearshore SWAN model to measured and modelled wave forcing at Wamberal, New South Wales, Australia (2014). https://doi.org/10.7158/C14-016.2014.12.1

17. O'Donncha, F., Zhang, Y., Chen, B., et al.: Ensemble model aggregation using a computationally lightweight machine-learning model to forecast ocean waves. arXiv preprint arXiv:1812.00511 (2018)

18. Paenke, I., Branke, J., Jin, Y.: Efficient search for robust solutions by means of evolutionary algorithms and fitness approximation. IEEE Trans. Evol. Comput. **10**(4), 405–420 (2006)

19. Rougier, J., Sexton, D.M., Murphy, J.M., Stainforth, D.: Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. J. Clim. **22**(13), 3540–3557 (2009)

20. Schmitt, C., Rey-Coyrehourcq, S., Reuillon, R., Pumain, D.: Half a billion simulations: evolutionary algorithms and distributed computing for calibrating the simpoplocal geographical model. Environ. Plan. B Plan. Des. **42**(2), 300–315 (2015)

21. Tolman, H.L., et al.: User manual and system documentation of WaveWatch III TM version 3.14. Technical note, MMAB Contribution 276, 220 (2009)

22. van Vledder, G.P.: Calibration of SWAN 40.20 for field cases Petten, Slotermeer and Westerschelde (2003)

23. Vychuzhanin, P., Nikitin, N., Deeva, I.: The source code of the robust evolutionary algorithm for SWAN model calibration (2019). https://github.com/ITMO-NSS-team/SwanEvolution

24. Wainwright, J., Mulligan, M.: Environmental Modelling: Finding Simplicity in Complexity. Wiley, Chichester (2005)

25. Williams, J.J., Esteves, L.S.: Guidance on setup, calibration, and validation of hydrodynamic, wave, and sediment models for shelf seas and estuaries. Adv. Civil Eng. **2017**, 25 pages (2017)

26. Williamson, D.B., Blaker, A.T., Sinha, B.: Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model. Geosci. Model Dev. **10**(4), 1789–1816 (2017)

27. Zang, C., Friswell, M., Mottershead, J.: A review of robust optimal design and its application in dynamics. Comput. Struct. **83**(4–5), 315–326 (2005)

28. Zitzler, E., Laumanns, M., Thiele, L.: SPEA 2: Improving the strength Pareto evolutionary algorithm. TIK-Report 103 (2001)