Article

Prediction of Potential Seagrass Habitat Using Remote Sensing Big Data and Machine Learning

Bohao He[1,†], Yanghe Zhao[1,†], Siyu Liu[1], Shahid Ahmad[1], Mao Wei[1,*]

1 College of Ecology and Environment, Hainan University, Hainan 570100, China; 20095132210054@hainanu.edu.cn (B.H.); zyh15525152637@163.com (Y.Z.)
* Correspondence: 994290@hainanu.edu.cn (M.W.)
† These authors contributed equally to this work.

Abstract: Globally, seagrass meadows provide critical ecosystem services. However, seagrasses are globally degraded at an accelerated rate. The lack of information on seagrass spatial distribution and seagrass health status seriously hinders seagrass conservation and management. Therefore, this study proposes to combine remote sensing big data with a new hybrid machine learning model (RF-SWOA) to predict potential seagrass habitats. The multivariate remote sensing data is used to train the machine learning model, which can improve the prediction accuracy of the model. This study shows that a hybrid machine learning model (RF-SWOA) can predict potential seagrass habitats more accurately and effectively than traditional models. At the same time, it has been shown that the most important factors influencing the potential habitat of seagrass in the Hainan region were the distance from land (38.2%) and the depth of the ocean (25.9%). This paper provides a more accurate machine learning model approach for predicting the distribution of marine species, which can help develop seagrass conservation strategies to restore healthy seagrass ecosystems.

Keywords: seagrass; remote sensing; machine learning; species distribution model (SDM); hybrid model; habitat suitability; niches; meta-heuristic optimization

1. Introduction

Seagrass is one of the extremely important marine resources [1,2]. Globally, seagrass habitats are rapidly degrading, and the loss of seagrass habitats will lead to multiple risks such as increased global climate change, shoreline destruction, and declining biodiversity [3-6]. Accurate knowledge of seagrass habitat and understanding of what factors limit or even threaten seagrass growth has become an urgent issue [4,7]. Unfortunately, many seagrass habitats around the world do not have clear spatial information [8,9], which seriously hinders marine environmental management and seagrass conservation. Traditional experimental methods for mapping seagrass distribution require large-scale field surveys, but such methods are costly and inefficient. In recent years, due to the development of remote sensing technology, more and more data and methods have been applied to marine predictive modeling, such as satellite data, unmanned aerial vehicles (UAV), acoustic surveys, and Geographic Information Systems (GIS) [10-13].

Species distribution models (SDM) are used to predict the regional distribution maps of the study species [14,15] and to assess the degree of habitat suitability of study species [16-18]. As SDM has been intensively studied, more and more studies have chosen to use machine learning for SDM modeling and have produced excellent results [19,20]. Downie, *et al.* [21] used the GAM and MaxEnt models to predict seagrass distribution, and the results showed that machine learning could accurately predict seagrass distribution. However, Bittner, Roesler and Barnes [14] found differences in the relative importance of environmental factors in different models in predicting seagrass distribution and therefore concluded that a more practical machine learning model should be selected for prediction. Among the SDM modeling methods, random forest models are widely used [20,22,23].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

Because the random forest is a very representative tree modeling algorithm that is applied in species distribution modeling due to its high accuracy [24-36].

In recent years, with the development of machine learning, hybrid machine learning models have been widely used [37,38]. Meta-heuristic algorithms have been found to improve the classification accuracy of models significantly [39,40]. In addition, population-based hybrid optimization algorithms can improve the search capability of global variables, moving from many individuals previously to collaborative group-sharing solutions, thus dramatically increasing the speed and power of the algorithm. The excellent performance and optimal solutions of metaheuristic algorithms solve the puzzles of multidisciplinary research, ranging from engineering and social sciences to ecology. This led to the widespread use of metaheuristics in many studies [41-51] (e.g., the whale optimization algorithm (WOA) [52-54].

Some applications have demonstrated the usability of hybrid machine learning models in providing insight into various knowledge domains. Still, few have explored the use of hybrid machine learning models to predict species suitability distributions. Therefore, this paper uses remote sensing big data with a hybrid machine learning model (RF-SWOA) to map seagrass habitat suitability areas. Furthermore, an attempt is made to apply hybrid machine learning to SDM and evaluate the performance of its model. Thus, the objectives of this study are: (1) to develop a hybrid machine learning model for predicting potential seagrass habitat; (2) to explore the effects of environmental variables on seagrass habitat; and (3) to evaluate the predictive advantages and limitations of the hybrid machine learning model.

2. Materials and Methods

2.1. Seagrass occurrence data

Hainan Province, located in the southernmost part of China, is the largest province in China in terms of land area (land plus sea). The climate of Hainan Island belongs to the monsoon tropical climate, which is between the two temperature zones of the tropics and subtropics. The weather is hot and humid with a long summer without a winter. The annual average temperature is 24 °C. Heat is abundant. Hainan Island is rich in plant and animal resources, of which seagrass is one of the main aquatic seed plant resources.

Hainan Island accounts for 64% of China's total seagrass area [55]. Therefore, this study conducted a field survey to determine the distribution of seagrass on Hainan Island from March to August 2021 (Figure 1). The presence of seagrass was marked with latitude and longitude, and samples were collected to identify seagrass species according to the method advocated by international seagrass researchers [56]. The literature and other relevant data were also combined to form the known distribution of seagrass beds on Hainan Island. We identified the boundaries of the seagrass beds by walking around the perimeter of the seagrass beds and recording the specific distribution of the seagrass beds when the study site was exposed at low tide.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

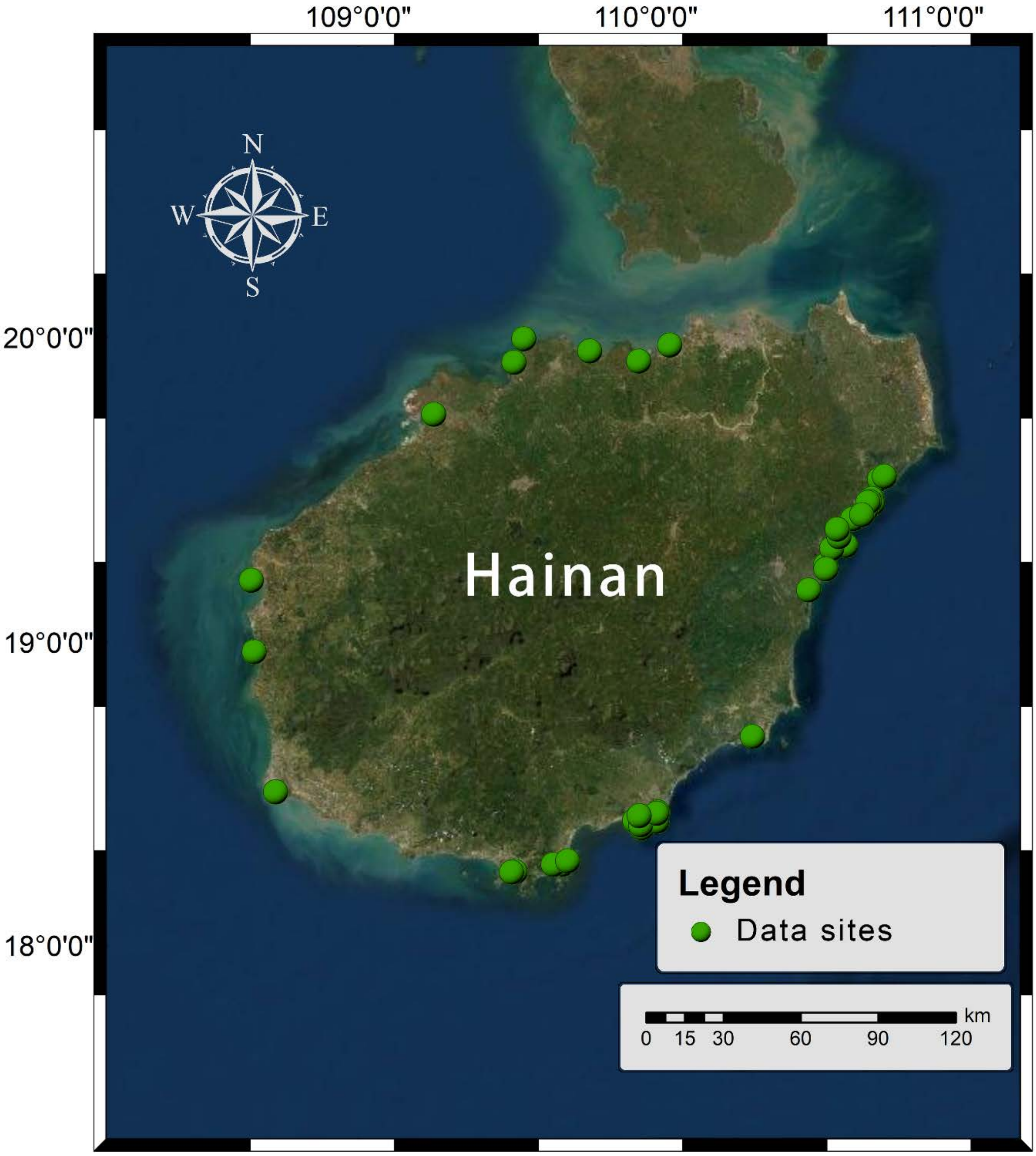

Figure 1. Study area and seagrass field distribution location sites.

2.2. Environmental data

To study potential seagrass habitats, 15 environmental variables were obtained for modeling (Table 1). Temperature, salinity, velocity, nitrate, phosphate, silicate, phytoplankton, calcite, pH, and attenuation was obtained by Bio-ORACLE 2.2 version (https://www.bio-oracle.org/index.php, accessed on February 5, 2022). Ocean slope data from GMED 2.0 version (https://gmed.auckland.ac.nz/index.html, accessed on February 6, 2022). Ocean chlorophyll-*a* concentration data from Google Earth Engine (https://earthengine.google.com/, accessed on February 5, 2022). Photosynthetically active radiation data from the MODIS aqua sensor (https://oceancolor.gsfc.nasa.gov/data/aqua/, accessed on February 6, 2022). Distance to nearest-shore data from NASA's Ocean Biology Pro-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

cessing Group (https://oceancolor.gsfc.nasa.gov/docs/distfromcoast/, accessed on February 6, 2022). Bathymetric dataset from GEBCO global network (https://www.gebco.net/, accessed on February 6, 2022). Table 2 shows the minimum (MIN), maximum (MAX), mean (MEAN) and standard deviation (STD) of the 15 different environmental data. Data from different environmental variables showed different trend changes and were able to determine the range of environments under the most suitable growth conditions for seagrass (Figure A1). All environmental variables were interpolated to 1 km spatial resolution using kriging interpolation in the ArcGIS 10.8 version of Geostatistical analysis. The strong covariance between environmental variables was fully considered, so the spdep R package [57] was used to calculate the spatial autocorrelation matrix.

Table 1. Environmental variables were used in this study.

Notation	Description	Units
silicate	Ocean silicate concentration	mol.m-3
attenuation	Diffuse attenuation	m-1
calcite	Constituent minerals in the ocean	mol.m-3
chlorophyll	Ocean chlorophyll-a concentration	mol.m-3
depth	Ocean depth	F
land	Distance from land	F
nitrate	Ocean nitrate concentration	mol.m-3
par	Photosynthetically active radiation	E.m-2.day-1
pH	Hydrogen ion concentration	1
phosphate	Ocean phosphate concentration	mol.m-3
phytoplankton	Phytoplankton in the ocean	umol.m-3
salinity	Ocean salinity	PSS
slope	Ocean slope	F
temperature	Ocean surface temperature	ºC
current	Currents velocity	m-1

Table 2. Statistical analysis results for different environmental variables.

Notation	MIN	MAX	MEAN	STD
silicate	5.87	13.12	8.13	1.96
attenuation	0.04	0.27	0.15	0.06
calcite	0.00	0.04	0.01	0.01
chlorophyll	0.12	0.56	0.25	0.10
depth	-100.09	-1.11	-30.82	18.70
land	0.02	0.28	0.14	0.05
nitrate	0.02	1.93	0.52	0.51
par	36.04	42.37	39.21	1.54
pH	8.18	8.19	8.19	0.00
phosphate	0.00	0.10	0.05	0.03
phytoplankton	0.92	2.97	1.53	0.45
salinity	32.94	33.31	33.18	0.08
slope	0.02	0.19	0.09	0.04
temperature	24.97	27.13	26.20	0.67
current	0.12	0.55	0.24	0.09

2.3. Machine learning models and evaluation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

2.3.1. Random forest model

The Random Forest (RF) algorithm is an extension of Bagging [58,59], in which the base learners are fixed as decision trees and the forest is made up of multiple trees (Figure 2). RF also uses sample perturbations added with put-back sampling when training the base learners, and it also introduces an attribute perturbation. Compared to bagging integration of decision trees, RF has poor starting performance, but as the number of base learners increases, RF tends to converge to a lower generalization error. Also, unlike bagging, in which the decision tree selects the optimal division attributes from all attribute sets, RF selects the division attributes in only a subset of the attribute set and thus is more efficient to train.

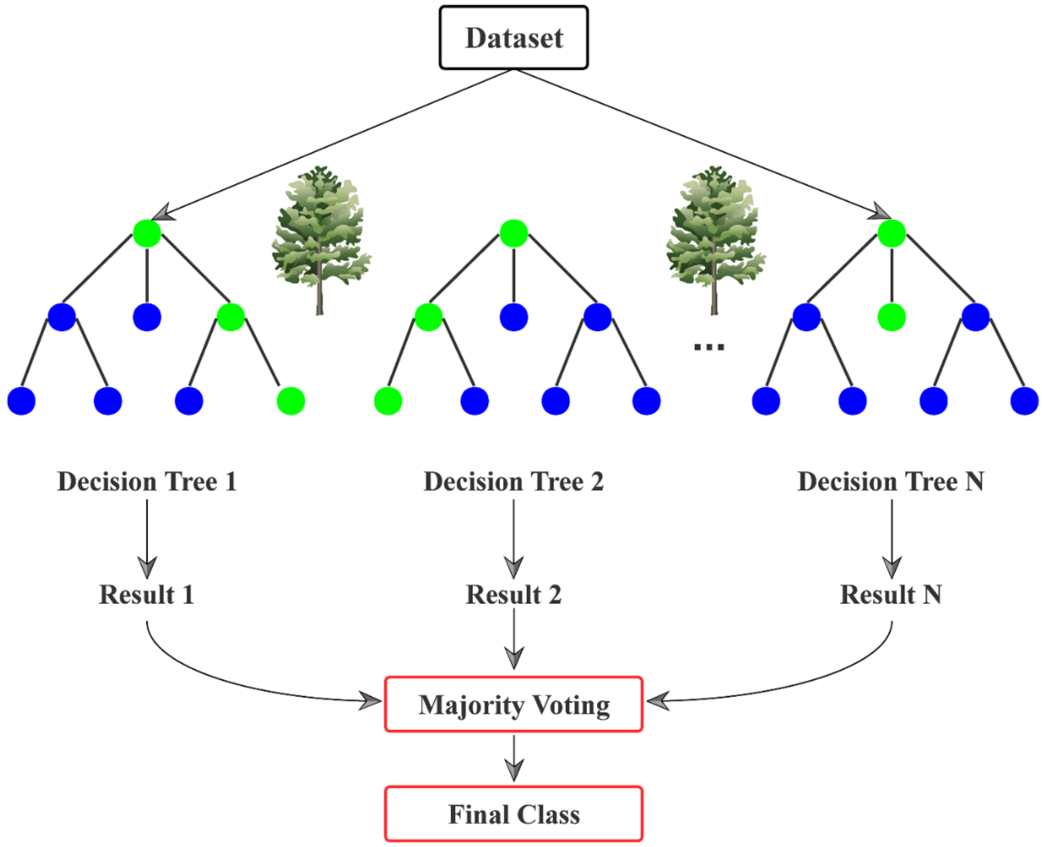

Figure 2. Random forest model structure.

2.3.2. Hybrid model

The Whale Optimization Algorithm (WOA) was introduced by Mirjalili in 2016 [53]. Inspired by the way whales hunt, the predation behavior is organized into three mathematical models: prey encirclement, bubble net attack and prey search [53,60]. The whale encircles the prey while locating the best search position with increasing number of iterations, while updating in real time. The mathematical expression of this behavior is

$$D = |C * X_L(t) - X(t)|,$$
$$X(t+1) = X_L(t) - A * D, \tag{1}$$

where A and C are the coefficient vectors, t indicates the current iteration, X_L is the position vector of the best solution obtained so far, X is the position vector, $|\ \ |$ is the absolute value, and * is an element-wise multiplication. Calculate A and C as follows:

$$A = 2a * r - a$$
$$C = 2 * r \tag{2}$$

A new position must be defined between the initial search position and the optimal search position to adjust the parameters. In this case, it is described as follows:

$$X(t+1) = \mathrm{D} * e^{bl} \cdot \cos(2\pi l) + X_L(t), \tag{3}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

where: b is a constant coefficient and l is a random vector between [0, 1]. The whale contraction or spiral model approach is selected based on a 50% probability. Based on the mathematical model, the whale's prey is simulated in a spiral circle, as follows:

$$X(t+1)=\begin{cases}X_L(t)-A*D & \text{if } p<0.5\\ D*e^{bl}*\cos(2\pi l)+X_L(t) & \text{if } p\geqslant 0.5\end{cases} \tag{4}$$

Contraction envelope and spiral position updates are performed simultaneously, with contraction according to p and spiral wandering according to $1-p$, where $p\in[0,1]$.

As the whale searches for prey, it moves toward the local optimal location point while expanding the global optimal point search, and this phase can be described as

$$D=| C * X_{rand} - X(t) |,$$
$$X(t+1)=X_{rand}-A*D, \tag{5}$$

where X_{rand} is a vector of random locations. A more detailed explanation of the WOA algorithm can be found in Mirjalili and Lewis [53].

The WOA is modified by adding a chaotic map to optimize global search capabilities. SWOA is mathematically described as follows:

$$p_{k+1}=ap_k^2\sin(\pi p_k),p_0\in[0,1],0<a\leqslant 4, \tag{6}$$

where k is the number of iterations and a is the description parameter within $0<a\leqslant 4$. For more information on the SWOA algorithm, see [54].

The model randomly selects 80% of the seagrass occurrence data for training and the remaining 20% for testing. The RF and RF-SWOA models were developed in Python 3.8 [61].

2.3.3. Model evaluation

A comprehensive evaluation of the model was conducted using six evaluation metrics. They are AUC, Omission rate, correct classification rate, Sensitivity, Specificity, Kappa.

$$AUC=\frac{1+\dfrac{True\ positive}{True\ positive\ +\ False\ negative}-\dfrac{False\ positive}{False\ positive+True\ negative}}{True\ number\ ^*\ False\ number} \tag{7}$$

$$Omission\ rate=\frac{False\ negative}{False\ negative+True\ negative} \tag{8}$$

$$Correct\ classification\ rate=\frac{True\ number}{Total\ sample} \tag{9}$$

$$Sensitivity=\frac{True\ positive}{(True\ positive+False\ negative)} \tag{10}$$

$$Specificity=\frac{TN}{(False\ positive+True\ negative)} \tag{11}$$

$$Kappa=\frac{\dfrac{True\ positive}{Total\ sample}-\dfrac{a1\times b1+a2\times b2+\cdots+aC\times bC}{Total\ sample*Total\ sample}}{1-\dfrac{a1\times b1+a2\times b2+\cdots+aC\times bC}{Total\ sample*Total\ sample}} \tag{12}$$

3. Results

3.1. Correlation analysis between environments

A high spatial autocorrelation between variables will seriously affect the prediction results of the species distribution model [62-64], therefore environmental variables with high correlation coefficients (r > 0.7) were excluded. The results of the study clearly show the spatial autocorrelation between all environmental variables (Figure 3). Therefore, the phosphate, phytoplankton, par, and attenuation environmental variables are removed, and the rest of the environmental variables are brought into model training.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

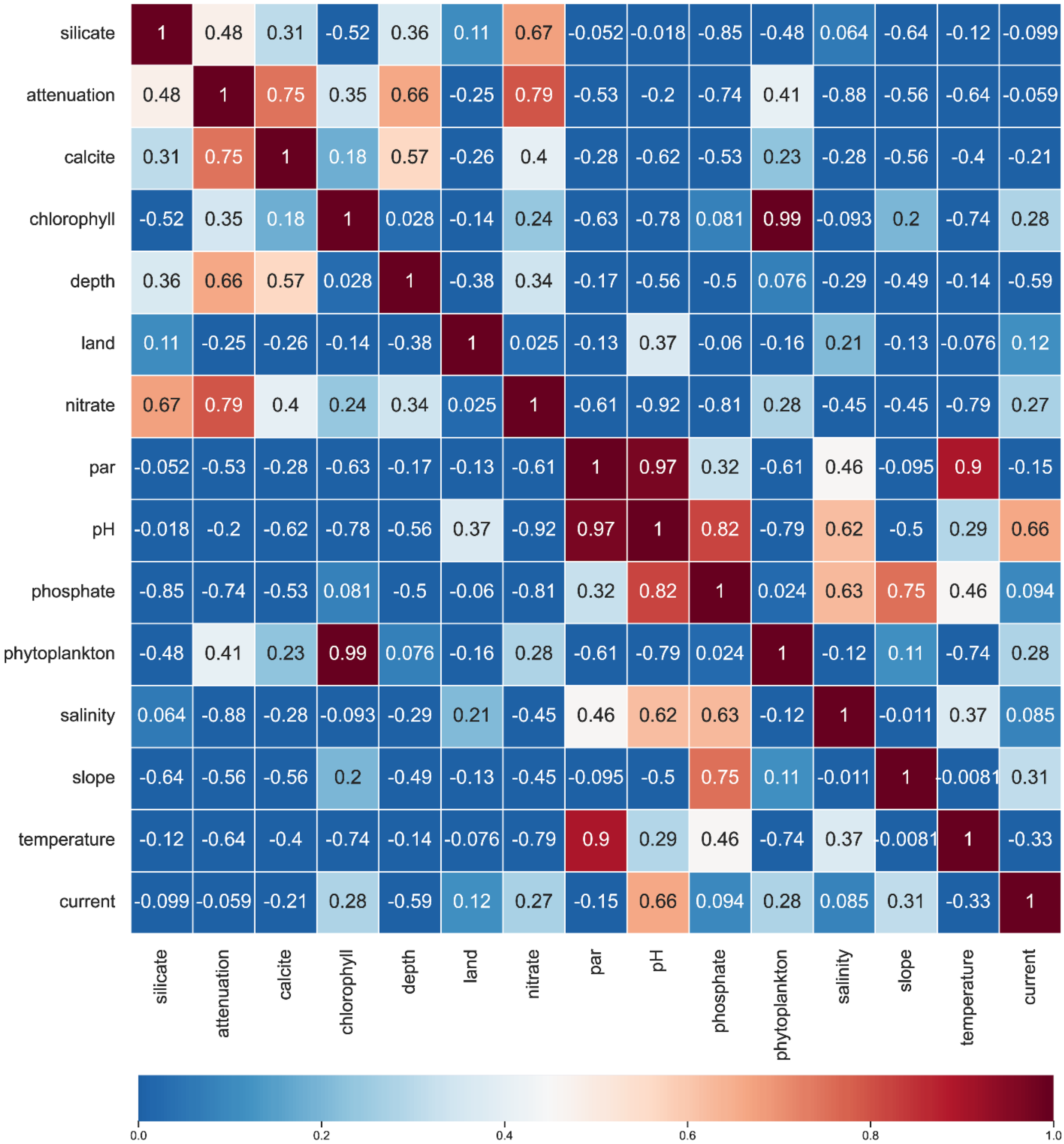

Figure 3. Correlation analysis matrix for different environmental variables.

3.2. Importance of environment features

The results of the importance of environmental characteristics showed that the most important environmental characteristics to predict the potential habitat of seagrass were the distance to land (38.2%) and the depth of the ocean (25.9%). The rest of the environmental variables showed small contribution values (<6%) to the prediction of the potential habitat of seagrass (Figure 4).

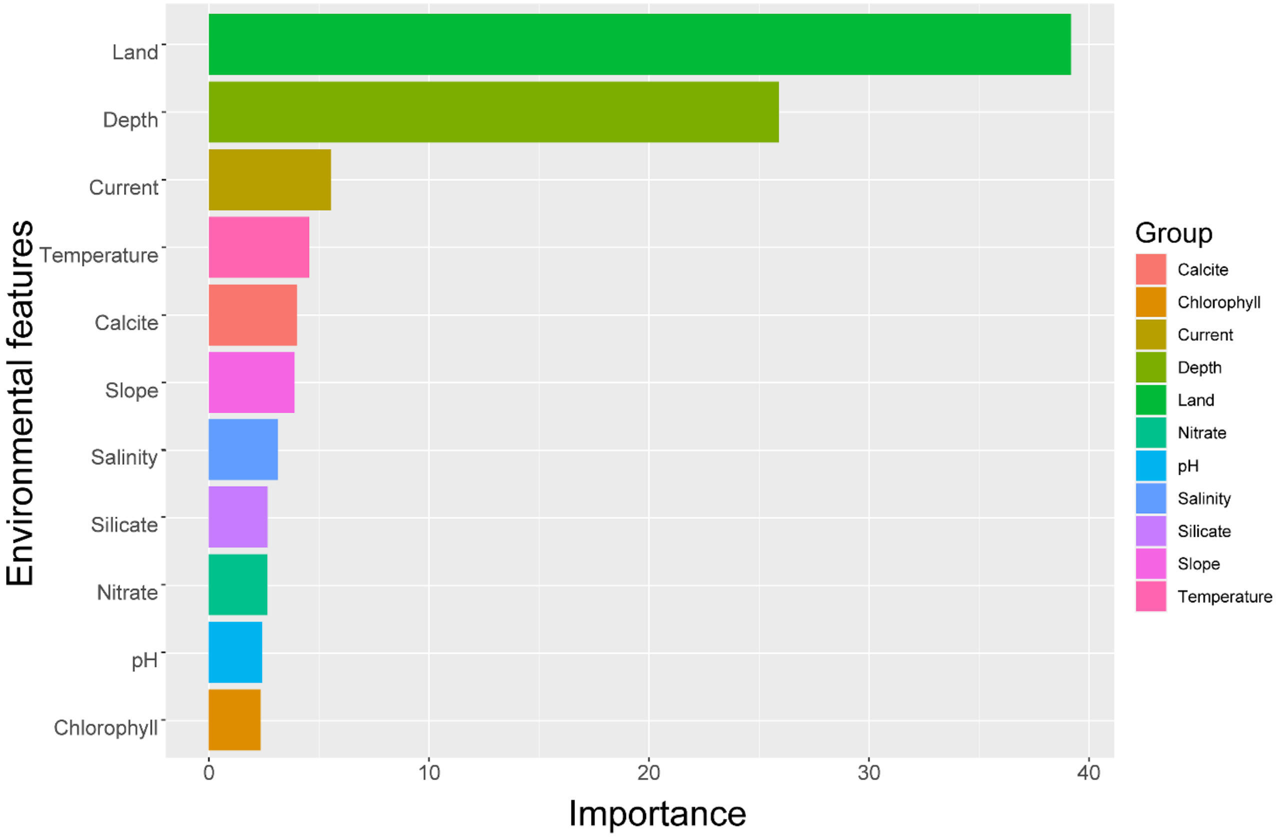

Figure 4. Importance analysis of 11 environmental features.

3.3. Potential seagrass habitat

Both models (RF and RF-SWOA) mapped potential seagrass habitat areas (Figure 5). It is clear that the RF model severely overestimates the potential habitat of seagrass and makes a more optimistic prediction, but this is not consistent with actual observations. In contrast, the potential seagrass habitat area given by the RF-SWOA model is closer to the actual observation site. From Figure 5, it can be found that the further the potential seagrass habitat is from land, the less likely it is to exist. This is reflected in both models.

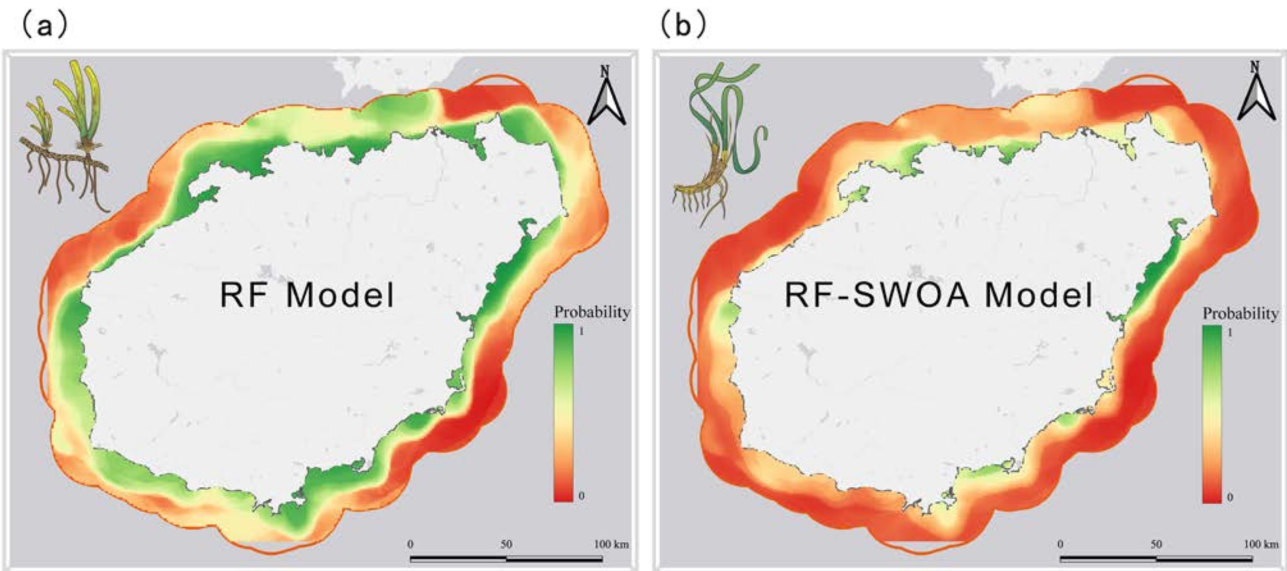

Figure 5. Potential habitat areas (Predicted by a. RF model and b. RF-SWOA model).

3.4. Model performance evaluation

RF-SWOA and RF models are compared in Figure 6. The results show that RF-SWOA has a higher AUC, correct classification rate, Kappa, and lower omission rate than the RF model. RF-SWOA produced a more accurate and stable prediction of seagrass habitat than RF alone. Figure 7 shows the results of sensitivity and specificity tests. As a result, the sensitivity and specificity of the proposed model (RF-SWOA) are better than those of the RF model. Hybrid machine learning algorithms with higher sensitivity and specificity in prediction can reduce errors in the potential distribution of seagrass, making the results more reliable.

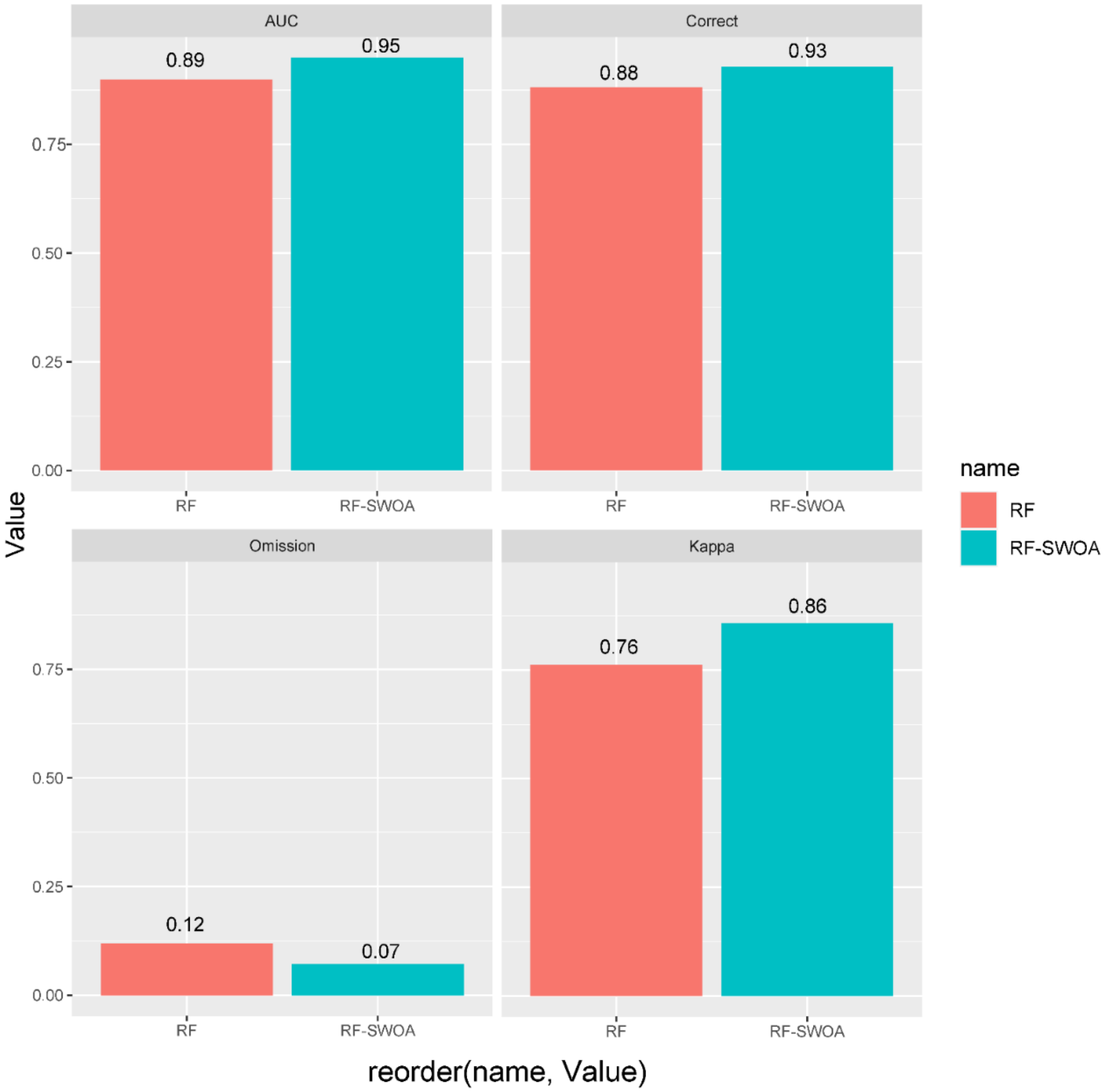

Figure 6. RF and RF-SWOA model performance evaluation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

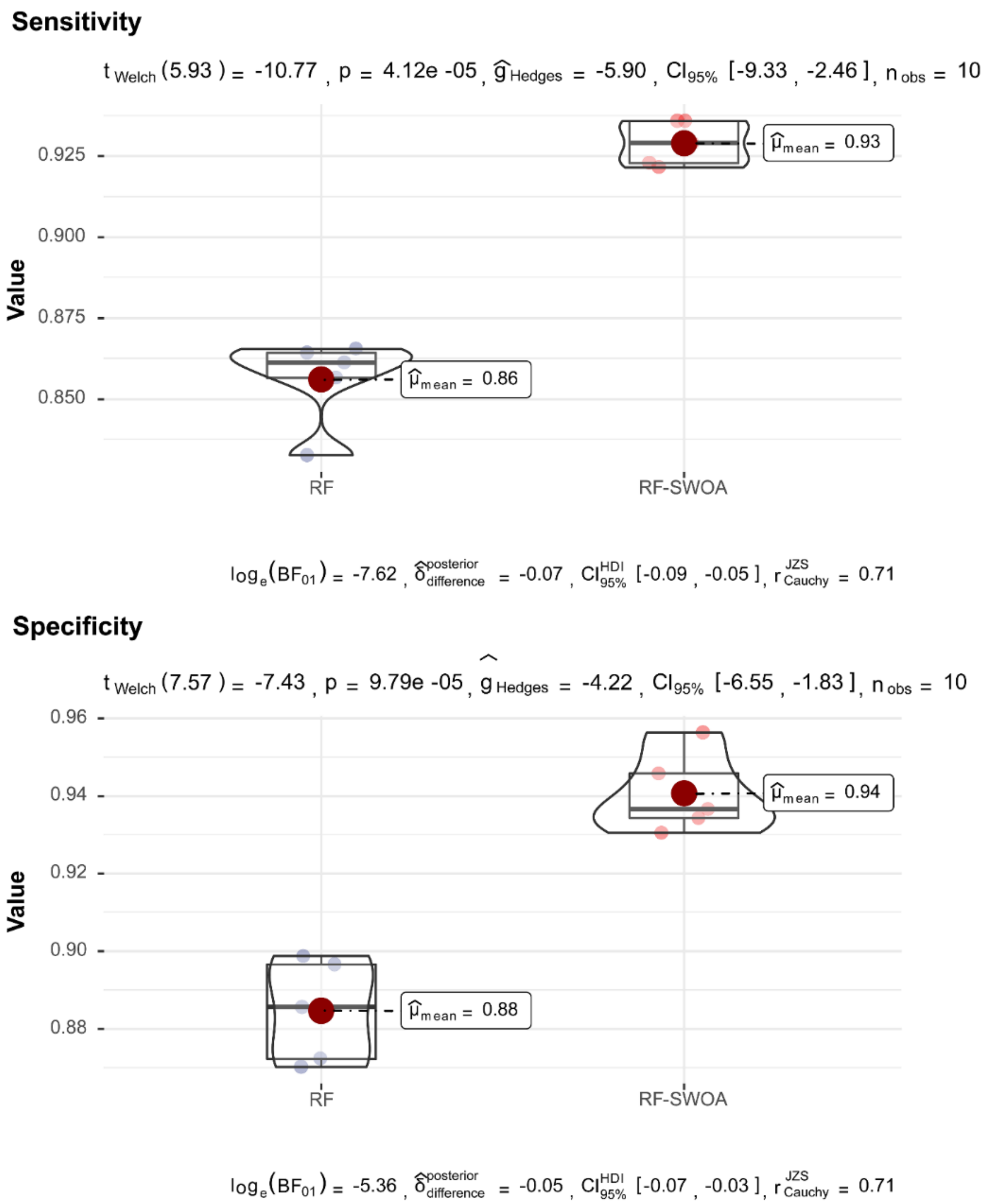

Figure 7. Sensitivity and specificity tests of RF and RF-SWOA models. The upper part of the panel shows the statistical test results of frequentist analysis, and the lower part of the panel shows the statistical test results of Bayesian analysis. The results follow the gold standard of statistical reporting [65].

4. Discussion

4.1. SWOA hybrid model evaluation

Intelligent optimization algorithms are widely used in various engineering practices [66-68], and also simple operation is one of the advantages of the WOA algorithm. It has excellent optimization capabilities and few parameters, which can dramatically increase the accuracy of the solution and convergence speed in the process of optimizing machine learning functions [69,70]. Although WOA has obvious advantages compared with other

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

intelligent algorithms, it has similar problems to other intelligent algorithms, such as being easy to get trapped in a local optimum and thus unable to jump out of the loop, and other problems. The SWOA algorithm proposed in this paper can update its position according to the adaptive parameter strategy while updating the optimal individual to achieve the ability to optimize the global search. This study further verified the performance of the SWOA algorithm through simulation experiments. Four standard test functions (Table 3) are used to assess the performance of the SWOA algorithm. In this test, the F1 and F2 functions are used to determine whether the SWOA algorithm can find the optimal value quickly and efficiently. To test whether the algorithm is capable of jumping out of its local optimal value, the F3 and F4 test functions are used. Each simulation test is to solve the performance of the 1000-dimensional test function. By testing the performance of SWOA and WOA algorithms through simulation, it is seen that the SWOA algorithm has better global convergence and robustness (Figure 8). Based on this finding, the optimized model is capable of improving marine species distribution predictions (e.g., seagrasses).

Table 3. Four simulation test functions.

Simulation function expression	Function name	Search space	Global optimum	Characteristic
$F_1 = 1 + \sum_{i=1}^{n} \frac{x_i^2}{4000} - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) \sqrt{b^2 - 4ac}$	Griewank	[-600,600]	0	Unimodal function
$F_2 = \sum_{i=1}^{n} \lvert x_i \rvert$	Schwefel 2.20	[-100,100]	0	Unimodal function
$F_3 = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n} \cos(2\pi x_i)\right) + 20 + \exp(1)$	Ackley	[-32,32]	0	Bimodal function
$F_4 = 10n + \sum_{i=1}^{n} (x_i^2 - 10\cos(2\pi x_i))$	Rastrigin	[-5.12,5.12]	0	Bimodal function

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

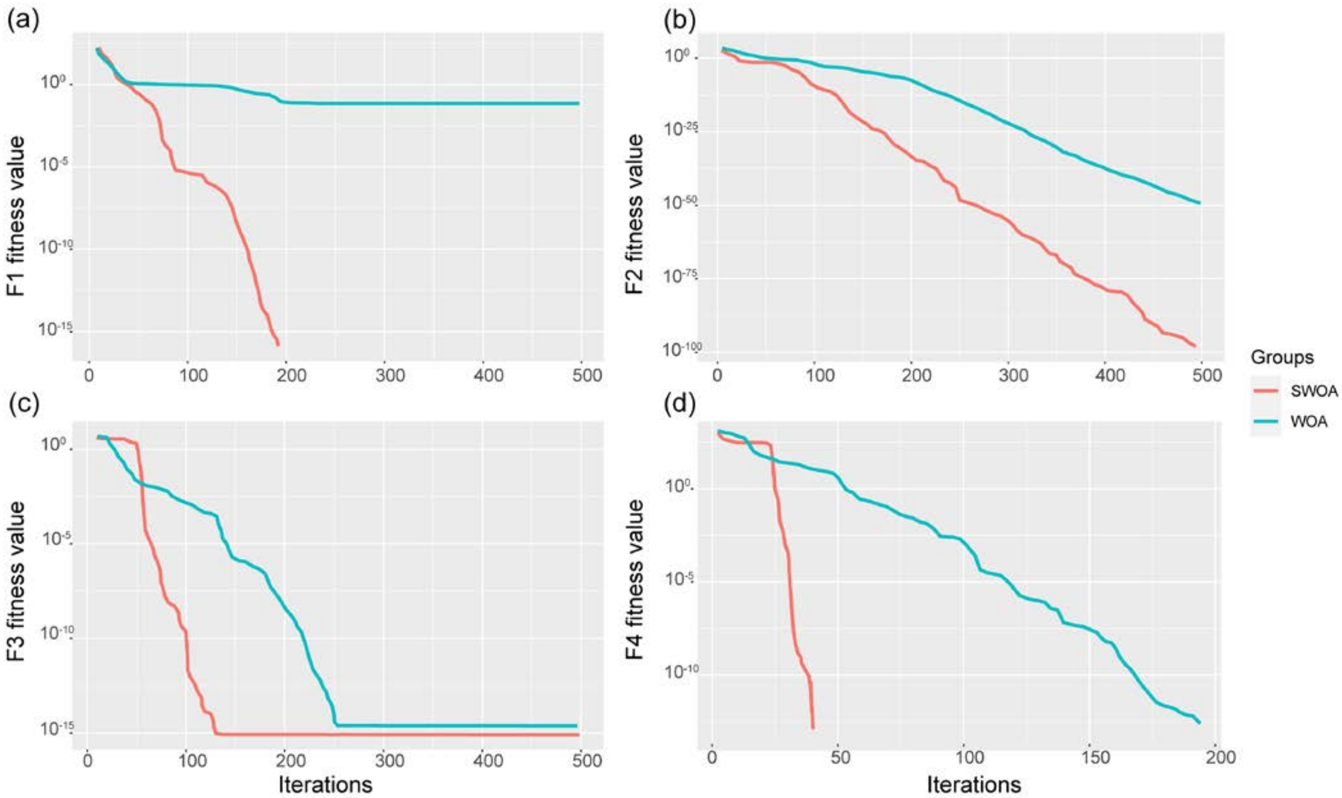

Figure 8. RF and RF-SWOA model performance evaluation.

4.2. Environmental drivers of seagrass habitat

Seagrass potential habitats are subject to the combined effects of multiple environmental variables. In this study, environmental variables that can affect seagrass growth were combined to model potential seagrass habitats. Our results show that the most critical environmental factors affecting seagrass habitat are the distance from land, ocean depth, and current velocity. This reflects the particular importance of physical environmental variables for seagrass habitats. However, this does not mean that chemical and biological types of environmental variables do not affect seagrass survival. In contrast to previous studies, the modeling of seagrass distribution was found to be influenced by different environmental drivers in different regions. A global model showed that the temperature of the sea surface and the distance to the land were the most important environmental variables to predict the distribution of seagrass [71]. On a regional scale, surface nitrate concentration and availability of benthic light became the most important environmental variables for predicting seagrass distribution in a model of seagrass species distribution in the US Gulf of Mexico, while in another sea area, distance to sandy shore and depth were the most important environmental drivers [14,72]. In summary, at the global level, the ocean temperature is particularly important for global seagrass growth, while at the small-scale distance to land, it becomes the most important factor. Therefore, we propose to establish a seagrass habitat simulation in the study area to identify which environmental factors will lead to seagrass growth limitation in order to better target seagrass conservation and restoration [73]. Although the model proposed in this study made relatively accurate predictions of potential seagrass habitats in the study area, the lack of high-resolution environmental data layers resulted in the inability to discern local spatial heterogeneity, which greatly limited the generalizability of the model. Meanwhile, there is still room for improvement in the classification accuracy of the model, and the model should be further optimized in subsequent studies, for example, by trying to use deep learning models to predict the potential seagrass habitats. In future studies, we will investigate how environmental variables affect seagrass distribution, for example, by introducing SDM models into interpretable machine learning models to investigate the response within environmental variables and the effect of multiple environmental variables in combination on seagrass distribution.

5. Conclusions

This study proposed a new hybrid machine learning model (RF-SWOA) to predict potential seagrass habitats accurately. We integrated multivariate remote sensing big data to develop a new model to map the extent of the potential seagrass habitat of Hainan Island. The results of this study indicated that the RF-SWOA model could effectively be applied to model seagrass distribution. The most important environmental factors affecting seagrass distribution were the distance from land, ocean depth, and current velocity. Therefore, the potential habitat map developed based on the RF-SWOA model can contribute to the adequate protection and restoration of seagrasses and provide scientific guidance for the planning of seagrass areas. In addition, despite the high predictive potential of the new method, the global interpretability and interaction analysis of the model are limited. So, future studies should use larger spatial scales and more seagrass samples to better understand the suitable distribution areas of seagrasses and the severe threats they face.

Author Contributions: Conceptualization, B.H. and M.W.; formal analysis, B.H. and Y.Z.; funding acquisition, B.H. and M.W.; investigation, B.H., Y.Z. and M.W.; methodology, B.H.; resources, M.W.; supervision, M.W., S.A; visualization, B.H.; writing—original draft, B.H. and Y.Z.; writing—review and editing, B.H., Y.Z., S.L., S.A. and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Major Science and Technology Plan of Hainan Province (RZ2100005452), the National Natural Science Foundation of China (42276235), the Start-up funding from Hainan University (kyqd20035), the innovative research projects for graduate students in Hainan Province (Qhys2021-16).

Data Availability Statement: Temperature, salinity, velocity, nitrate, phosphate, silicate, phytoplankton, calcite, pH, and attenuation was obtained by Bio-ORACLE 2.2 version (https://www.bio-oracle.org/index.php, accessed on February 5, 2022). Ocean slope data from GMED 2.0 version (https://gmed.auckland.ac.nz/index.html, accessed on February 6, 2022). Ocean chlorophyll-*a* concentration data from Google Earth Engine (https://earthengine.google.com/, accessed on February 5, 2022). Photosynthetically active radiation data from MODIS aqua sensor (https://oceancolor.gsfc.nasa.gov/data/aqua/, accessed on February 6, 2022). Distance to nearest-shore data from NASA's Ocean Biology Processing Group (https://oceancolor.gsfc.nasa.gov/docs/distfromcoast/, accessed on February 6, 2022). Bathymetric dataset from GEBCO global network (https://www.gebco.net/, accessed on February 6, 2022). The seagrass distribution data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank the members of the 609 offices of the College of Ecology and Environment at Hainan University for their help with research ideas. We are grateful to Biying Jia, Peng Wang, and Yingshuai Liu for their help with data acquisition and geographic information system. The authors thank university of Maryland's integration and application network.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2022 doi:10.20944/preprints202209.0221.v1

Figure A1. Covariance matrix of 15 environmental variables.

References

1. Fourqurean, J.W.; Duarte, C.M.; Kennedy, H.; Marbà, N.; Holmer, M.; Mateo, M.A.; Apostolaki, E.T.; Kendrick, G.A.; Krause-Jensen, D.; McGlathery, K.J. Seagrass ecosystems as a globally significant carbon stock. *Nature geoscience* **2012**, *5*, 505-509.
2. Cullen-Unsworth, L.C.; Unsworth, R. A call for seagrass protection. *Science* **2018**, *361*, 446-448.
3. Orth, R.J.; Carruthers, T.J.; Dennison, W.C.; Duarte, C.M.; Fourqurean, J.W.; Heck, K.L.; Hughes, A.R.; Kendrick, G.A.; Kenworthy, W.J.; Olyarnik, S. A global crisis for seagrass ecosystems. *Bioscience* **2006**, *56*, 987-996.
4. Waycott, M.; Duarte, C.M.; Carruthers, T.J.; Orth, R.J.; Dennison, W.C.; Olyarnik, S.; Calladine, A.; Fourqurean, J.W.; Heck Jr, K.L.; Hughes, A.R. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the national academy of sciences* **2009**, *106*, 12377-12381.

5. Kendrick, G.A.; Nowicki, R.J.; Olsen, Y.S.; Strydom, S.; Fraser, M.W.; Sinclair, E.A.; Statton, J.; Hovey, R.K.; Thomson, J.A.; Burkholder, D.A. A systematic review of how multiple stressors from an extreme event drove ecosystem-wide loss of resilience in an iconic seagrass community. *Frontiers in Marine Science* **2019**, *6*, 455.
6. Moksnes, P.O.; Röhr, M.E.; Holmer, M.; Eklöf, J.S.; Eriander, L.; Infantes, E.; Boström, C. Major impacts and societal costs of seagrass loss on sediment carbon and nitrogen stocks. *Ecosphere* **2021**, *12*, e03658.
7. Short, F.T.; Wyllie-Echeverria, S. Natural and human-induced disturbance of seagrasses. *Environmental conservation* **1996**, *23*, 17-27.
8. Short, F.; Carruthers, T.; Dennison, W.; Waycott, M. Global seagrass distribution and diversity: a bioregional model. *Journal of experimental marine biology and ecology* **2007**, *350*, 3-20.
9. McKenzie, L.J.; Finkbeiner, M.A.; Kirkman, H. Methods for mapping seagrass distribution. *Global seagrass research methods* **2001**, 101-121.
10. Fingas, M. Remote sensing for marine management. In *World seas: An environmental evaluation*; Elsevier: 2019; pp. 103-119.
11. Ouellette, W.; Getinet, W. Remote sensing for marine spatial planning and integrated coastal areas management: Achievements, challenges, opportunities and future prospects. *Remote Sensing Applications: Society and Environment* **2016**, *4*, 138-157.
12. Belkin, I.M. Remote sensing of ocean fronts in marine ecology and fisheries. *Remote Sensing* **2021**, *13*, 883.
13. Picart, S.S.; Sathyendranath, S.; Dowell, M.; Moore, T.; Platt, T. Remote sensing of assimilation number for marine phytoplankton. *Remote sensing of environment* **2014**, *146*, 87-96.
14. Bittner, R.E.; Roesler, E.L.; Barnes, M.A. Using species distribution models to guide seagrass management. *Estuarine, Coastal and Shelf Science* **2020**, *240*, 106790, doi:https://doi.org/10.1016/j.ecss.2020.106790.
15. Gonzalez-Irusta, J.M.; Gonzalez-Porto, M.; Sarralde, R.; Arrese, B.; Almon, B.; Martin-Sosa, P. Comparing species distribution models: a case study of four deep sea urchin species. *Hydrobiologia* **2015**, *745*, 43-57, doi:10.1007/s10750-014-2090-3.
16. Pollock, L.J.; Tingley, R.; Morris, W.K.; Golding, N.; O'Hara, R.B.; Parris, K.M.; Vesk, P.A.; McCarthy, M.A. Understanding co‐occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* **2014**, *5*, 397-406.
17. Zimmermann, N.E.; Edwards Jr, T.C.; Graham, C.H.; Pearman, P.B.; Svenning, J.C. New trends in species distribution modelling. *Ecography* **2010**, *33*, 985-989.
18. Miller, J. Species distribution modeling. *Geography Compass* **2010**, *4*, 490-509.
19. Li, X.; Wang, Y. Applying various algorithms for species distribution modelling. *Integrative zoology* **2013**, *8*, 124-135.
20. Evans, J.S.; Murphy, M.A.; Holden, Z.A.; Cushman, S.A. Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology*; Springer: 2011; pp. 139-159.
21. Downie, A.-L.; von Numers, M.; Boström, C. Influence of model selection on the predicted distribution of the seagrass Zostera marina. *Estuarine, Coastal and Shelf Science* **2013**, *121*, 8-19.
22. Bradter, U.; Kunin, W.E.; Altringham, J.D.; Thom, T.J.; Benton, T.G. Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm. *Methods in Ecology and Evolution* **2013**, *4*, 167-174.
23. Mi, C.; Huettmann, F.; Guo, Y.; Han, X.; Wen, L. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* **2017**, *5*, e2849.
24. Attorre, F.; Alfo, M.; De Sanctis, M.; Francesconi, F.; Valenti, R.; Vitale, M.; Bruno, F. Evaluating the effects of climate change on tree species abundance and distribution in the Italian peninsula. *Applied Vegetation Science* **2011**, *14*, 242-255, doi:10.1111/j.1654-109X.2010.01114.x.

25. Freeman, E.A.; Moisen, G.G.; Frescino, T.S. Evaluating effectiveness of down-sampling for stratified designs and unbalanced prevalence in Random Forest models of tree species distributions in Nevada. *Ecological Modelling* **2012**, *233*, 1-10, doi:10.1016/j.ecolmodel.2012.03.007.
26. Hanberry, B.B.; He, H.S.; Palik, B.J. Comparing Predicted Historical Distributions of Tree Species Using Two Tree-based Ensemble Classification Methods. *American Midland Naturalist* **2012**, *168*, 443-455.
27. Kosicki, J.Z. Should topographic metrics be considered when predicting species density of birds on a large geographical scale? A case of Random Forest approach. *Ecological Modelling* **2017**, *349*, 76-85, doi:10.1016/j.ecolmodel.2017.01.024.
28. Kosicki, J.Z. Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. *Environmental and Ecological Statistics* **2020**, *27*, 273-292, doi:10.1007/s10651-020-00445-5.
29. Luan, J.; Zhang, C.; Xu, B.; Xue, Y.; Ren, Y. The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research* **2020**, *227*, doi:10.1016/j.fishres.2020.105534.
30. Mi, C.; Huettmann, F.; Guo, Y.; Han, X.; Wen, L. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *Peerj* **2017**, *5*, doi:10.7717/peerj.2849.
31. Miller, K.B.; Huettmann, F.; Norcross, B.L. Efficient spatial models for predicting the occurrence of subarctic estuarine-associated fishes: implications for management. *Fisheries Management and Ecology* **2015**, *22*, 501-517, doi:10.1111/fme.12148.
32. Mouton, A.M.; Alcaraz-Hernandez, J.D.; De Baets, B.; Goethals, P.L.M.; Martinez-Capel, F. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. *Environmental Modelling & Software* **2011**, *26*, 615-622, doi:10.1016/j.envsoft.2010.12.001.
33. Saranya, K.R.L.; Lakshmi, T.V.; Reddy, C.S. Predicting the potential sites of Chromolaena odorata and Lantana camara in forest landscape of Eastern Ghats using habitat suitability models. *Ecological Informatics* **2021**, *66*, doi:10.1016/j.ecoinf.2021.101455.
34. Siaulys, A.; Bucas, M. Species distribution modelling of benthic invertebrates in the south-eastern Baltic Sea. *Baltica* **2012**, *25*, 163-170, doi:10.5200/baltica.2012.25.16.
35. Wang, A.; Goslee, S.C.; Miller, D.A.; Sanderson, M.A.; Gonet, J.M. Topographic variables improve climatic models of forage species abundance in the northeastern United States. *Applied Vegetation Science* **2017**, *20*, 84-93, doi:10.1111/avsc.12284.
36. Williams, J.N.; Seo, C.; Thorne, J.; Nelson, J.K.; Erwin, S.; O'Brien, J.M.; Schwartz, M.W. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* **2009**, *15*, 565-576, doi:10.1111/j.1472-4642.2009.00567.x.
37. Ardabili, S.; Mosavi, A.; Várkonyi-Kóczy, A.R. Advances in machine learning modeling reviewing hybrid and ensemble methods. In Proceedings of the International Conference on Global Research and Education, 2019; pp. 215-227.
38. Bies, R.R.; Muldoon, M.F.; Pollock, B.G.; Manuck, S.; Smith, G.; Sale, M.E. A genetic algorithm-based, hybrid machine learning approach to model selection. *Journal of pharmacokinetics and pharmacodynamics* **2006**, *33*, 195.
39. Beheshti, Z.; Shamsuddin, S.M.H. A review of population-based meta-heuristic algorithms. *Int. J. Adv. Soft Comput. Appl* **2013**, *5*, 1-35.
40. Singh, P.; Kottath, R. An ensemble approach to meta-heuristic algorithms: comparative analysis and its applications. *Computers & Industrial Engineering* **2021**, *162*, 107739.
41. Boussaid, I.; Lepagnot, J.; Siarry, P. A survey on optimization metaheuristics. *Information Sciences* **2013**, *237*, 82-117, doi:10.1016/j.ins.2013.02.041.
42. Cruz-Duarte, J.M.; Amaya, I.; Ortiz-Bayliss, J.C.; Conant-Pablos, S.E.; Terashima-Marin, H.; Shi, Y. Hyper-Heuristics to customise metaheuristics for continuous optimisation. *Swarm and Evolutionary Computation* **2021**, *66*, doi:10.1016/j.swevo.2021.100935.

43. de Melo, V.V.; Carosio, G.L.C. Investigating Multi-View Differential Evolution for solving constrained engineering design problems. *Expert Systems with Applications* **2013**, *40*, 3370-3377, doi:10.1016/j.eswa.2012.12.045.
44. Dokeroglu, T.; Sevinc, E.; Kucukyilmaz, T.; Cosar, A. A survey on new generation metaheuristic algorithms. *Computers & Industrial Engineering* **2019**, *137*, doi:10.1016/j.cie.2019.106040.
45. Hassan, A.; Pillay, N. Hybrid metaheuristics: An automated approach. *Expert Systems with Applications* **2019**, *130*, 132-144, doi:10.1016/j.eswa.2019.04.027.
46. Herran, A.; de la Cruz, J.M.; de Andres, B. Global Search Metaheuristics for planning transportation of multiple petroleum products in a multi-pipeline system. *Computers & Chemical Engineering* **2012**, *37*, 248-261, doi:10.1016/j.compchemeng.2011.10.003.
47. Marichelvam, M.K. An improved hybrid Cuckoo Search (IHCS) metaheuristics algorithm for permutation flow shop scheduling problems. *International Journal of Bio-Inspired Computation* **2012**, *4*, 200-205, doi:10.1504/ijbic.2012.048061.
48. Memmah, M.-M.; Lescourret, F.; Yao, X.; Lavigne, C. Metaheuristics for agricultural land use optimization. A review. *Agronomy for Sustainable Development* **2015**, *35*, 975-998, doi:10.1007/s13593-015-0303-4.
49. Moya, I.; Bermejo, E.; Chica, M.; Cordon, O. Coral reefs optimization algorithms for agent-based model calibration. *Engineering Applications of Artificial Intelligence* **2021**, *100*, doi:10.1016/j.engappai.2021.104170.
50. Talbi, E.-G. Combining metaheuristics with mathematical programming, constraint programming and machine learning. *Annals of Operations Research* **2016**, *240*, 171-215, doi:10.1007/s10479-015-2034-y.
51. Xiong, N.; Molina, D.; Ortiz, M.L.; Herrera, F. A Walk into Metaheuristics for Engineering Optimization: Principles, Methods and Recent Trends. *International Journal of Computational Intelligence Systems* **2015**, *8*, 606-636, doi:10.1080/18756891.2015.1046324.
52. Kaur, G.; Arora, S. Chaotic whale optimization algorithm. *Journal of Computational Design and Engineering* **2018**, *5*, 275-284, doi:10.1016/j.jcde.2017.12.006.
53. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Advances in engineering software* **2016**, *95*, 51-67.
54. He, B.; Jia, B.; Zhao, Y.; Wang, X.; Wei, M.; Dietzel, R. Estimate soil moisture of maize by combining support vector machine and chaotic whale optimization algorithm. *Agricultural Water Management* **2022**, *267*, 107618, doi:https://doi.org/10.1016/j.agwat.2022.107618.
55. Zheng, F.; Qiu, G.; Fan, H.; Zhang, W. Diversity, distribution and conservation of Chinese seagrass species. *Biodiversity Science* **2013**, *21*, 517-526.
56. Kuo, J.; Den Hartog, C. Seagrass taxonomy and identification key. *Global seagrass research methods* **2001**, *33*, 31-58.
57. Bivand, R.; Altman, M.; Anselin, L.; Assunção, R.; Berke, O.; Bernat, A.; Blanchet, G. Package 'spdep'. *The Comprehensive R Archive Network* **2015**.
58. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5-32.
59. Breiman, L. Bagging predictors. *Machine learning* **1996**, *24*, 123-140.
60. Aljarah, I.; Faris, H.; Mirjalili, S. Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Computing* **2018**, *22*, 1-15.
61. Python, W. Python. *Python Releases for Windows* **2021**, *24*.
62. Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* **1993**, *74*, 1659-1673.
63. Koenig, W.D. Spatial autocorrelation of ecological phenomena. *Trends in Ecology & Evolution* **1999**, *14*, 22-26.
64. F. Dormann, C.; M. McPherson, J.; B. Araújo, M.; Bivand, R.; Bolliger, J.; Carl, G.; G. Davies, R.; Hirzel, A.; Jetz, W.; Daniel Kissling, W. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **2007**, *30*, 609-628.
65. Patil, I. Visualizations with statistical details: The'ggstatsplot'approach. *Journal of Open Source Software* **2021**, *6*, 3167.

 doi:10.20944/preprints202209.0221.v1

66. Su, Z.; Wang, J.; Lu, H.; Zhao, G. A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting. *Energy conversion and management* **2014**, *85*, 443-452.
67. Li, W.; Wang, G.-G.; Gandomi, A.H. A survey of learning-based intelligent optimization algorithms. *Archives of Computational Methods in Engineering* **2021**, *28*, 3781-3799.
68. Wang, J.; Du, P.; Hao, Y.; Ma, X.; Niu, T.; Yang, W. An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting. *Journal of environmental management* **2020**, *255*, 109855.
69. Chakraborty, S.; Saha, A.K.; Sharma, S.; Mirjalili, S.; Chakraborty, R. A novel enhanced whale optimization algorithm for global optimization. *Computers & Industrial Engineering* **2021**, *153*, 107086.
70. Sun, Y.; Wang, X.; Chen, Y.; Liu, Z. A modified whale optimization algorithm for large-scale global optimization problems. *Expert Systems with Applications* **2018**, *114*, 563-577.
71. Jayathilake, D.R.; Costello, M.J. A modelled global distribution of the seagrass biome. *Biological Conservation* **2018**, *226*, 120-126.
72. Downie, A.-L.; von Numers, M.; Boström, C. Influence of model selection on the predicted distribution of the seagrass Zostera marina. *Estuarine, Coastal and Shelf Science* **2013**, *121-122*, 8-19, doi:https://doi.org/10.1016/j.ecss.2012.12.020.
73. Mao, W.; Zhao, Y.; He, B.; Jia, B.; Li, W. Review on degradation mechanism and restoration strategies of seagrass ecosystem. *Journal of Desert Research* **2022**, *42*, 87.