Response to Reviewers' comments to manuscript ESSD-2025-264

"Spatial Patterns of Sandy Beaches in China and Risk Analysis of Human Infrastructure Squeeze Based on Multi-Source Data and Ensemble Learning"

Dear Reviewers:

Thank you very much for your thoughtful and detailed review. Your suggestions have provided us with important and constructive insights, which have significantly improved the manuscript. We have carefully considered all of your comments and have made substantial revisions to the manuscript based on your feedback. We have done our best to enhance the manuscript and hope that the revised version will meet your approval. A point-by-point response to the outstanding comments is attached to this manuscript. The major revisions are summarized as follows:

Response to Comments by Reviewer 1:

1. *I have carefully read the manuscript entitled « Spatial Patterns of Sandy Beaches in China and Risk Analysis of Human Infrastructure Squeeze Based on Multi-Source Data and Ensemble Learning » by Jie Meng et al. and tested associated datasets available in Zenodo. It presents original information on sandy beach locations throughout China, taking the form of a shape file produced using ensemble learning and multi-source data, from which various beach spatial characteristics (e.g., number, width, area) and risk analysis of coastal squeeze can be assessed and analyzed. It is not the first time sandy beaches were mapped in China using satellite imagery - with already existing datasets mentioned and assessed in comparison by the authors - yet it is the first time ensemble learning is used, which according to the authors improved detection accuracy. Besides, when previous studies used Sentinel-2 data only, Sentinel 1 and Google Earth imagery were also included in this study. I have found the manuscript well written and illustrated, although additional careful proofreading by the authors could have avoided several mistakes such as identical section and figure titles, repetitions and typos. Figure captions in particular must be improved. I attach an annotated version of the manuscript, listing a number of technical corrections and specific comments, that the authors should take into account while preparing their revision.*

Response:

We would like to express our gratitude to the reviewer for their valuable comments. Based on the suggested revisions in the attached document, we have made the following adjustments to the manuscript format:

1.L53: We have changed "**technologies**" to "**techniques**" in the manuscript.

2.L57: Small grayscale differences refer to situations where the variation in brightness (grayscale

value) between different regions or pixels in the image is minimal. For instance, certain areas may display nearly identical grayscale values, which can lead to poor performance of segmentation algorithms in these regions. This typically happens in low-contrast images, regardless of whether the image is grayscale or color.

3.L62: We have changed "**rough**" to "**exploratory**" in the manuscript.

4.L70: We have changed "**tidal image interference**" to "**the impact of tidal variations on sandy beach extraction from remote sensing images**" in the manuscript.

5.L72: We have changed "**is**" to "**are**" in the manuscript.

6.L76: We have changed "**contributing over**" to "**contributing to over**" in the manuscript.

7.L80: We have modified the sentence from: "However, despite the rapid economic development in these regions, there is a lack of nationwide dynamic monitoring tools for sandy beaches, and the risks posed by human infrastructure squeeze are not well understood." to: "However, despite the rapid economic development in these regions, there is a lack of nationwide dynamic monitoring tools for sandy beaches, and the risks posed by human infrastructure squeeze—particularly due to urban development and coastal expansion—are not well understood."

8.L86: We have modified the sentence from: "Multi-temporal data from 2016 to 2023 are used to build an annual representative beach dataset, reducing tidal fluctuation impacts and supporting precise mapping and monitoring;" to: "Multi-year sandy beach data extracted using ensemble learning from 2016 to 2024 are merged to construct an annual representative sandy beach dataset, reducing tidal fluctuation impacts and supporting precise mapping and monitoring."

9.L95: Sandy beaches are typically located along the coastline. To ensure the study area includes as many relevant beaches as possible, while aligning with the recognized coastal zone boundaries defined by experts, we established a buffer zone. Specifically, the study area incorporates a buffer zone extending 10 km inland and 20 km offshore. The primary goal of this design is to maintain the integrity of the sandy beach area, preventing the buffer zone from being too narrow, which could lead to certain beach areas being truncated or excluded from the study region. Such a design would result in an underestimation of China's sandy beach area, not accurately reflecting the true extent of the coastline.

10.L101: We have changed "**Figure 1: Location of China's coastal zone and distribution of partial test points for result verification**" to "**Figure 1: Location of China's coastal zone and distribution of partial typical land feature points for results verification.**"

11.L119: We have changed "**Figure 2: Spatial distribution of Sentinel-1 and Sentinel-2 images used in this study: (a) Image Count of Sentinel-1, (b) Image Count of Sentinel-2**" to "**Figure 2: Spatial distribution of Sentinel-1 and Sentinel-2 images used in this study from 2016 to 2024: (a) Image Count of Sentinel-1, (b) Image Count of Sentinel-2.**"

12.L125: Regarding the issue you raised, this is indeed the case. The phrase "**this is the range of pixel grayscale values retained**" refers to the valid range of grayscale values we retained during the processing. This approach ensures the validity and consistency of the image data while also minimizing potential noise interference.

13.L142: We have changed "**Li et al., 2022**" to "**Miao et al., 2022**" in the manuscript.

14.L145: We have changed "**Study area and materials**" to "**Methodology**" in the manuscript.

15.L150: Regarding the repetition you mentioned, I included the sentence "**To accurately monitor the current status of sandy beaches in China, this study integrates multi-source data from 2016 to 2024**" to make the technical process appear more complete and help readers better understand my methodology. This section aims to clearly outline the background and data sources of the study, making the subsequent research methods and steps easier to comprehend.

16.Regarding Figure 3: First, the field surveys mentioned in **L164** were used to determine a subset of training and testing points for beach extraction. Additionally, the "**Public data**" in the figure was updated to include beach data from the OSM database for comparison. Finally, we have changed the caption of **Figure 3** from "**Spatial distribution of Sentinel-1 and Sentinel-2 images used in this study: (a) Image Count of Sentinel-1, (b) Image Count of Sentinel-2**" to "**Figure 3: The technical framework of the study.**"

17.L164: The field surveys refer to a simple process where we go to the field to check if the area is a sandy beach. If it is, we record the GPS coordinates (latitude and longitude). It is a straightforward operation, so it was not described in detail here.

18.L165: This section just establishes a sample library of test points for subsequent validation. These are point data used to verify the minimum pixel at the location of the point in the classification results (which is 10m). The specific points differ each year, and the number of points for each year is shown in Figure 4.

19.L170: Since I have previously conducted similar work in Fujian Province, and based on the features referenced in the literature, I selected these features as the input variables for the model. The paper "**Meng, J., Xu, D., Tao, Z., and Ge, Q.: Sandy Beach Extraction Method Based on Multi-Source Data and Feature Optimization: A Case in Fujian Province, China, Remote Sens., 17(16), 2754, doi:10.3390/rs17162754, 2025**" discusses in detail how each feature affects the model. Additionally, we included the importance analysis results in **L245**, which show the contribution of each input feature to the results each year.

20.**L179**: Due to the large number of input features, providing a detailed description would take up a lot of space. As many studies have already used these features, we have provided the calculation methods for each input feature along with the corresponding references for readers to consult.

21.L190: The parameters here are set to default values. We have replaced them with better base learner models for beach extraction. The reason for listing the parameters is to help readers effectively replicate and use this method.

22.L193: We have changed the manuscript from "**Pixel-based classification algorithms inevitably produce salt-and-pepper noise, and some small patches—such as buildings and bare land—are difficult to distinguish from environmentally influenced sandy beaches (Mattson et al., 2024)**" to "**Pixel-based classification algorithms inevitably produce salt-and-pepper noise because sandy beaches share similar spectral characteristics with certain areas, such as buildings and bare land, making them difficult to distinguish (Mattson et al., 2024).**"

23.L207: Due to the large number of equations and input parameters, providing a detailed description would take up a lot of space. Since many studies have already used these evaluation metrics for accuracy assessment, we have abbreviated the formulas and cited the corresponding articles. Readers can refer to these articles for more detailed formulas for accuracy evaluation.

24.L215: Sandy beaches are generally stable features, with minimal changes over time. However, tidal influences can cause some beaches to be misclassified as water bodies due to water level fluctuations, resulting in variations in the extracted beaches from year to year. Therefore, we used multi-year beach extraction results and combined them to create a merged multi-year beach dataset. This approach significantly reduces the impact of tidal fluctuations on beach extraction, providing a more accurate reflection of the real situation regarding the distribution, data, and area of sandy beaches nationwide. Based on this, we used impermeable surfaces from the land-use data of 1990-2024 to create a 100-meter buffer zone for clipping, producing the human infrastructure squeeze risk areas for 1990-2024.

25.L217: The detailed explanation of Sen's slope would require considerable space. Since this method is widely used and has been employed in many studies, we have cited the relevant references for readers to consult.

26.L219: We have changed "**risk areas across regions**" to "**risk zones across different regions**" in the manuscript.

27.L222: We have changed the manuscript from "**The performance of the ensemble learning algorithm was validated by calculating PA, UA, F1-score, OA, and Kappa coefficient, and we obtained the validation results (Table 5)**" to "**The ensemble learning algorithm's performance was evaluated using key metrics such as Accuracy, Precision, Recall, sandy beach F1-score, and AUC, with the results presented in Table 5.**"

28.L225: In the initial version, I used a relatively simple set of metrics, which led to minimal differences in the results. Based on your suggestions, I have updated the metrics to include Accuracy, Precision, Recall, sandy beach F1-score, and AUC. These updated metrics now provide more distinct

results, helping to better assess the model's performance.

29.L228: The "merged data" refers to the combined beach extraction results from 2016 to 2024. Tidal influences led to some beaches being misclassified as water bodies, causing the beach area to be underestimated. Since sandy beaches are relatively stable features, merging multi-year data helps reflect the current state of the beaches, significantly reducing the impact of tidal fluctuations on beach extraction.

30.L236: The inconsistency between Figure 5 and its corresponding conclusion was due to an error in drawing the figure. I have corrected this now. Additionally, due to the large longitudinal extent of Guangdong, multiple bars need to be stacked together for comparison.

31.L236: The data presented here are based on the merged beach data from 2016 to 2024. The width represents the average width of each province, while the perimeter and area reflect the total perimeter and area of each province.

32.L246: We have changed the manuscript from "**Figure 6: Spatial distribution of sandy beaches in China: (a) Spatial distribution of sandy beach numbers, (b) Spatial distribution of sandy beach length, (c) Spatial distribution of sandy beach width, (d) Spatial distribution of sandy beach area**" to "**Figure 6: Spatial distribution of sandy beaches in China: (a) sandy beach numbers, (b) sandy beach length, (c) sandy beach width, and (d) sandy beach area.**"

33.L263: We have replaced "**reference**" with "**published**" and made corresponding changes throughout the manuscript. Additionally, we have made the necessary adjustments to Figure 8 as well.

34.L270: Previously, the manuscript compared various datasets. We have now changed it to a comparison of models. It was not reasonable to equate datasets using the accuracy evaluation method I established, so we removed the accuracy comparison between datasets, as it was meaningless. Instead, we have focused on comparing popular models to highlight the advantages of our model and demonstrate the effectiveness of our sandy beach extraction. However, comparisons of beach distribution and sandy beach attributes across different datasets have still been retained.

35.L277: We have changed "**a clear advantage**" to "**provides a certain supplementary advantage.**"

36.L293: Dataset 3 shows significantly higher values because a large number of other land cover types were misclassified as sandy beaches. Even with other scales, the results remain hard to read. We will include a table in the appendix listing the number of beaches in each province for every dataset.

37.L295: We have changed the manuscript from "**in the risk area of human infrastructure squeeze in the study area**" to "**in the risk areas of human infrastructure squeeze in the study area.**"

38.L304: The unit for the y-axis has been moved closer to the y-axis title.

39.L311: We have changed the manuscript from "**high coastal development intensity**" to "**have high coastal development density.**"

40.L321: We have changed the manuscript from "**Figure 10: Spatial changes of human infrastructure squeeze risk area: (a) Current risk area of human infrastructure squeeze, (b) Spatial changes of human infrastructure squeeze risk area from 1990 to 2023**" to "**Figure 11: Spatial changes of human infrastructure squeeze risk area: (a) current risk area, (b) changes in the risk area from 1990 to 2024.**"

41.L340: Regarding the texture features of sandy beaches, at the image scale, certain characteristics behave as expected, but this pattern may not always hold true in different contexts or broader analyses. However, at the 10m scale in this study, this pattern is consistent.

42.Figure 12: We have modified Figure 12 by retaining only one value.

43.L368: We have changed the manuscript from "**strong applicability**" to "**robust generalization.**"

44.L381: We have changed the manuscript from "**Overall, the distribution of sandy beaches in China reflects the combined effects of sediment supply, coastal type, and hydrodynamic conditions, resulting in more sandy beaches in the north and south, and fewer on the central coast**" to "**Overall, the distribution of sandy beaches in China is shaped by sediment supply, coastal type, and hydrodynamic conditions, with more beaches in the north and south, and fewer along the central coast.**"

45. We have revised the manuscript to change "**Figure 13: Spatial changes of various factors from 1990 to 2023: (a) Spatial change of per capita GDP, (b) Spatial change of resident population, (c) Spatial change of built-up area, (d) Spatial change of road area**" to "**Figure 14: Spatial changes of various factors from 1990 to 2024: (a) per capita GDP change, (b) resident population change, (c) built-up area change, (d) road area change.**"

2. *I have carefully read the manuscript entitled « Spatial Patterns of Sandy Beaches in China and Risk Analysis of Human Infrastructure Squeeze Based on Multi-Source Data and Ensemble Learning » by Jie Meng et al. and tested associated datasets available in Zenodo. It presents original information on sandy beach locations throughout China, taking the form of a shape file produced using ensemble learning and multi-source data, from which various beach spatial characteristics (e.g., number, width, area) and risk analysis of coastal squeeze can be assessed and analyzed. It is not the first time sandy beaches were mapped in China using satellite imagery - with already existing datasets mentioned and assessed in comparison by the authors - yet it is the*

first time ensemble learning is used, which according to the authors improved detection accuracy. Besides, when previous studies used Sentinel-2 data only, Sentinel 1 and Google Earth imagery were also included in this study. I have found the manuscript well written and illustrated, although additional careful proofreading by the authors could have avoided several mistakes such as identical section and figure titles, repetitions and typos. Figure captions in particular must be improved. I attach an annotated version of the manuscript, listing a number of technical corrections and specific comments, that the authors should take into account while preparing their revision.

Response:

Thank you for your valuable feedback; it has been extremely helpful. I will now respond to each of your comments:

1.Firstly, regarding the test set, we constructed a sample set through field surveys and visual interpretation using Sentinel-2 and Google images. Once the sample set was obtained, we randomly divided it into a 7:3 ratio for training and testing. Of course, it is essential to ensure that the proportion of sandy and non-sandy beach samples in both the training and test sets, as well as the sample distribution across provinces, remain consistent.

2.Since our product is first generated using ensemble learning followed by post-processing and visual interpretation corrections, the workload is minimal, and it better compensates for the sandy beach areas, avoiding the issue of underestimating the sandy beach area due to missed identification during visual interpretation. In contrast, Dataset 1 was labeled without any prior knowledge, resulting in many missing sandy beach areas. You are correct in pointing out that relying solely on Dataset 1 seems unreasonable, so we replaced it with a more reliable source. The current dataset is derived from the **OpenStreetMap (OSM)** database, and we have made updates in **Figure 3**, Table 1, and **Line 138** of the manuscript. The revised text now reads: "**In this study, we used three datasets to evaluate our identified sandy beach dataset (Table 1): (1) The China sandy beach dataset, data directly obtained from the OpenStreetMap (OSM) database; (2) The 2022 China 10m sandy beach dataset identified by Ni et al. (Ni et al, 2024) using a support vector machine method based on Sentinel-2 imagery; (3) The 2020 China coastal land use dataset at 10m resolution, identified by Miao et al. (Miao et al, 2022) using an object-oriented classification method based on Sentinel-2 imagery.**" The updates can be seen in the revised figures and tables. Additionally, we have modified the comparison between subsequent datasets **(L262, L275**). The updated text reads: "**To further assess the accuracy and reliability of the dataset, this study compared three published datasets in selected areas of Fujian, Shandong, and Guangdong (Fig. 8). Dataset 1, directly obtained from the OpenStreetMap (OSM) database, is highly subjective and tends to misclassify non-beach areas as sandy beaches, while also missing some actual sandy beach areas. Dataset 2, constructed using a support vector machine on Sentinel-2 imagery, shows high consistency with our dataset but still**

misses some sandy beach areas. Dataset 3, created using an object-oriented approach, demonstrates high accuracy for other land cover types but faces significant misclassification issues with bare land and urban areas. The results show that our dataset provides higher accuracy in sandy beach classification, significantly reducing misclassification." "**According to the comparison results, our dataset shows significant advantages over published datasets 1, 2, and 3 in several key metrics, particularly in terms of sandy beach area, perimeter, width, and number (Fig. 9). Overall, our dataset provides a supplementary advantage in sandy beach area coverage, with larger areas in all regions: Fujian (54.57 km²), Guangdong (78.88 km²), and Taiwan (46.60 km²), significantly surpassing published datasets 1 (30.35, 20.59, and 19.07 km²) and 2 (29.17, 58.35, and 25.51 km²). Regarding perimeter, our dataset closely matches actual sandy beach boundaries: Fujian (1435.89 km), Guangdong (2849.39 km), and Taiwan (1324.98 km), compared to published datasets 1 (581.95, 826.40, and 509.22 km) and 2 (756.92, 1856.62, and 906.63 km). In terms of width, our dataset also outperforms published datasets 1 (52.21, 20.94, and 48.42 m) and 2 (45.18, 32.58, and 27.56 m), with values of 54.91 m in Fujian, 38.92 m in Guangdong, and 57.17 m in Taiwan. The number of identified sandy beaches in our dataset is higher than in published datasets 1 and 2, further highlighting the reduced misclassification and noise in our results (Fig. 9). Moreover, published dataset 3 has significantly higher area, perimeter, width, and number values than the other datasets and actual values, leading to many non-sandy beaches being incorrectly identified as sandy beaches.**" These changes are reflected in **Figures 8 and 9**.

3.Additionally, we realized that it was unreasonable to compare different products using the test set we constructed. Therefore, this part has been removed from the manuscript and replaced with a comparison between the current mainstream models and our ensemble learning model. In this updated version, we replaced the base learner for the ensemble learning model with tree-based models, which are less sensitive to parameters and better suited for handling large datasets and wider areas. The revised text is as follows (**L183**): " **Stacking is a powerful ensemble learning method that uses predictions from multiple base learners as inputs to a meta-learner for final prediction. It combines the strengths of different models to overcome individual limitations, enhancing accuracy, stability, and generalization (Chen et al., 2024). In this study, Random Forest (RF), Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGB), and Light Gradient Boosting Machine (LGBM) were selected as base learners. Their output classification probabilities were used as input features for the meta-learner, which adopted Logistic Regression (LR) to integrate probabilities, calculate distances to target classes, and produce final results (Table 3). For performance comparison, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), RF, Classification and Regression Tree (CART), and Convolutional Neural Network (CNN) models were also employed as benchmark models.**" Furthermore, the parameters for the base learners were also updated, and the revised parameters are reflected in **Table 3**. **Table 6** has been updated to show a comparison between different models, where our ensemble learning model yields the best results (2016-2024 average). The accuracy, precision, recall, F1-score, and AUC are 0.9335, 0.9160, 0.9014, 0.9084,

and 0.9802, respectively, outperforming the combined results of SVM, KNN, RF, CART, and CNN models. In addition, we have added the corresponding discussion in **L257**. The updated text reads: "**We evaluated the performance of our model and several other machine learning algorithms (SVM, KNN, RF, CART, and CNN) using accuracy, precision, recall, F1-score, and AUC on the validation set (Table 6). Our model outperformed all others, achieving the highest accuracy (0.9335), precision (0.9160), and AUC (0.9802). It showed significantly better classification performance, especially in reducing misclassification. In comparison, the SVM model had good precision but lower recall and accuracy, while KNN performed the worst across all metrics**." These results are reflected in **Table 6**.

4.Previously, there were indeed only 14,694 points used for testing between 2016 and 2023, covering 5 land cover categories. Each year, the dataset was redefined with less than 400 sandy beach points per year, which was clearly insufficient. Now, we have updated the study to cover a 9-year period from 2016 to 2024, as reflected throughout the manuscript. The time span mentioned in L15 and other sections referring to the years has been updated from 2016-2023 to 2016-2024. Additionally, the number of Sentinel-1 and Sentinel-2 images used has been revised accordingly, with the updated numbers being **23,859** and **80,759**, respectively, as reflected in **L108**, **L111**, and also shown in **Figure 2**. Furthermore, we have updated the land use data to cover the period from 1990 to 2024 (**L212**), and we revised the method for establishing buffer zones. The updated approach creates a buffer zone around sandy beach data and then performs the clipping, which results in more accurate sandy beach extraction than clipping based on land use data. The revised text now reads: "**Therefore, this study established a 100-meter buffer zone around the Chinese land use dataset and conducted an overlay analysis using the obtained sandy beach data. The analysis generated annual human infrastructure squeeze risk areas to evaluate the squeeze effects of infrastructure on sandy beaches from 1990 to 2024.**"

Regarding the issue of sample size, we recognized that the previous sample set was insufficient. Therefore, we changed the problem to a binary classification task, merging the other four land cover classes into one, with an emphasis on labeling sandy beach data. A total of **341,057** points were labeled from 2016 to 2024, with **249,098** points for the training set and **49,1959** points for the testing set. The breakdown for sandy beach and non-sandy beach samples for both the training and testing sets from 2016 to 2024 is as follows:

- 2016-2024 training set (sandy beach): **6776, 7308, 7841, 8231, 8722, 9549, 10632, 11716, 14130**
- 2016-2024 testing set (non-sandy beach): **19290, 19431, 18822, 18568, 18510, 17981, 18138, 18136, 16470**
- 2016-2024 testing set (sandy beach): **2901, 3128, 3358, 3525, 3735, 4088, 4550, 5017, 6054**

- 2016-2024 testing set (non-sandy beach): **4175, 4200, 4797, 4756, 4738, 5352, 5399, 5398, 7056**

This has been reflected in **Figure 4**. Due to the changes in both the years and the sample points, some data in the results section of the article will be updated accordingly. For example, in **L236**, the sandy beach area, perimeter, width, and number have changed. The revised text now reads: "**From a provincial perspective, Guangdong has the most sandy beaches, with 1,096 sandy beaches, and also ranks first in both total length (2,849.39 km) and total area (78.88 km²). In terms of sandy beach width, Hebei has the widest sandy beaches, with an average width of 68.77 m. Other regions, such as Fujian (sandy beach area 54.57 km², total length 1,435.89 km, width 54.91 m) and Hainan (sandy beach area 51.65 km², total length 1,977.96 km, width 41.02 m), also show significant sandy beach resources.**"

3. *Concerning quality metrics, why were they chosen, and are they really complementary to each other? Tables 5 and 6 suggest similar interpretations can be made for all metrics as results do not differ significantly. Additional information should be provided to guide readers in how to interpret these quality metrics and the results obtained.*

Response:

Thank you for the reviewer's valuable feedback. Indeed, the five metrics (PA, UA, F1-score, OA, and Kappa coefficient) previously selected in Tables 5 and 6 did not show significant differences in results, and thus were not effective in distinguishing the performance of different models. However, these metrics are commonly used in multi-class classification tasks. To better evaluate the performance of our binary classification model, I have converted the problem into a binary classification task. As a result, we replaced these metrics with more suitable evaluation metrics for binary classification: Accuracy, Precision, Recall, Sandy Beach F1-score, and AUC (Area Under the Curve). These metrics are more effective in reflecting the model's performance in binary classification tasks. In the revised Tables 5 and 6, the differences between the newly selected five metrics can be observed (see **L227** and **L270**). Due to space limitations, we have only referenced the relevant literature in the manuscript without providing a detailed introduction to these evaluation metrics, as they are mainstream and commonly used in the field.

1.Accuracy is the most basic evaluation metric, measuring the proportion of correct predictions among all predictions. It is calculated by dividing the number of correct predictions by the total number of samples. Although accuracy is straightforward and intuitive, it may not fully reflect model performance, especially in cases of imbalanced data. In tasks like sandy beach extraction, where the ratio between positive (sandy beach) and negative (non-sandy beach) samples may vary greatly, accuracy may not be an adequate measure of performance.

2.Precision measures the proportion of actual positive samples among those predicted as positive by the model. It reflects the confidence of the model when predicting positive class instances. Precision is crucial in sandy beach extraction tasks, as it helps assess how well the model avoids misclassifying non-beach areas as beaches. A low precision indicates that the model is making many false positives, which could lead to errors in beach management and protection efforts.

3.Recall is the proportion of correctly predicted positive samples among all actual positive samples. A higher recall indicates that the model is better at identifying positive samples (i.e., sandy beach regions). However, this could lead to an increase in false positives, as non-beach regions might be misclassified as beaches. Therefore, recall needs to be evaluated together with precision to assess the model's performance comprehensively.

4.F1-score is the harmonic mean of precision and recall, providing a balance between the two. F1-score is especially useful for tasks with imbalanced class distributions. In sandy beach extraction applications, F1-score helps us balance precision and recall, ensuring the model minimizes misclassification while identifying as many sandy beach areas as possible.

5.AUC (Area Under the Curve) is a critical metric for evaluating a binary classification model's discriminative ability. It is derived from the ROC (Receiver Operating Characteristic) curve and measures the model's ability to distinguish between positive and negative samples. A higher AUC value indicates that the model is better at distinguishing between the classes. AUC is particularly robust in imbalanced datasets and provides an overall performance evaluation, especially when dealing with multiple threshold values.

With these updated evaluation metrics, we are now able to assess the performance of the binary classification model more accurately, especially in the context of sandy beach extraction. These metrics not only assist in more reliable identification of sandy beach regions but also help reduce misclassifications, providing more trustworthy results for the model.

4. *Likewise, the authors evaluate other (independent) datasets (some produced by the authors themselves, eg. Dataset 1) which they name reference datasets. I think « reference » is misused in this context as it generally implies that the data were used for validation (truth data), which is not the case here.*

Response:

Thank you for your valuable feedback. The dataset 1 is no longer the one I created; instead, we have replaced it with a sandy beach dataset from the OpenStreetMap (OSM) database. Additionally, we believe that the term "reference dataset" is not appropriate, so we have changed it to "published dataset."

It is important to note that these three datasets are only used for comparison with our own dataset, not for accuracy validation.

5. *Additional information could be provided and discussed in regard to the liability and capacities of the machine learning method used. What is the minimum beach size that can be detected? Looking at the dataset, it seems small pocket beaches can remain undetected. In contrast, elongated features (some artificial) may be erroneously detected as sandy beaches. What is the impact of tide range (potentially variable across China) and having images obtained at different tide levels on data consistency and how can this be improved? It is said using annual averages (composite images) helps mitigate this issues, but it is not clear how, particularly as Figure 2 shows an heterogeneous spatial distribution of satellite images and certainly, with regular revisits, this means images are obtained at different tide levels throughout the country. From this could arise systematic biases in the spatial characteristics deduced from the dataset (eg., beach width and area). Was this, and how, mitigated for this study?*

Response:

Thank you for your valuable feedback. Indeed, the impact of tides on beach extraction is a complex issue, especially when dealing with beach data on a national scale. The differences in tidal times across various regions can cause dynamic changes in beach areas, which in turn affect the accuracy of the extraction. To address this, we have synthesized the results of beach extractions over multiple years to minimize the influence of tidal fluctuations and construct a more representative dataset reflecting the actual beach distribution. Firstly, regarding the minimum beach size, we have set the threshold at **1,000 square meters**. This threshold helps avoid retaining small areas below this size during post-processing, as these small areas are often misidentified as beaches due to noise or fluctuations. This setting ensures the reliability of the beach extraction and improves the overall quality of the dataset. However, the effect of tides cannot be ignored, particularly in areas where there is a clear transition between beaches and water bodies. Each year, due to tidal changes, certain beach areas may be submerged by water, leading to misclassification as water bodies in the imagery. Although beaches themselves are relatively stable as natural features and do not exhibit significant expansion or contraction, tidal influence causes yearly variations in the beach extraction results, especially between high and low tides. Additionally, due to varying tidal times in different regions, satellite images are obtained at different times, further complicating data consistency. For example, some regions may show beaches covered by water in satellite images obtained at one time, while others may show exposed beaches during the same period.

To better address this issue, we decided to merge the beach extraction results over multiple years (from 2016 to 2024) to create a more representative beach dataset. Although annual composite images may be affected by tidal fluctuations, combining these results helps capture the widest extent of beach coverage each year. This method not only reduces misclassification due to tidal changes but also

provides a more stable dataset that better reflects beach distribution, particularly one that is closer to the low tide beach distribution. Since low tide exposes the largest area of the beach, combining multiple years of data can effectively capture the beach area during low tide, avoiding omissions and misclassification that may occur when relying on data from only one year. Furthermore, using multiple years of data helps mitigate the systematic bias caused by tidal fluctuations in any single year. For example, if high tide in a given year causes part of the beach to be submerged and misclassified as water, using only data from that year could result in an underestimation of the beach area. However, by combining data from multiple years, other years' images might show these areas as beaches, thus helping to reconstruct the actual beach extent. This approach allows us to create a more accurate beach distribution map and reduces errors in beach extraction. It is important to note that while merging data from multiple years can minimize the impact of tidal fluctuations on beach extraction, this method is not perfect. Due to differences in tidal times between regions, composite images may still show some local discrepancies. In regions where the transition between high and low tide occurs, beach boundaries may still be inaccurately represented due to the timing of the satellite images. Therefore, we adopted the approach of synthesizing the maximum beach coverage from each year's extraction results to generate a representative beach distribution map. This method helps minimize errors caused by tidal fluctuations and provides the best possible estimate of beach distribution. Through this approach, we aim to obtain a more stable and reliable beach distribution map, especially during low tide, ensuring that the data better reflects the actual beach situation. This not only helps in accurately estimating beach area, perimeter, and width, but also provides reliable foundational data for beach management and protection. For example, many beach conservation efforts require monitoring and evaluation during low tide, so obtaining accurate low tide beach data is crucial for implementing effective protection measures. In conclusion, while tidal fluctuations do have an impact on beach extraction, combining multi-year beach extraction results allows us to produce a more stable and accurate dataset. This approach effectively alleviates errors caused by tidal changes and provides a more accurate basis for long-term monitoring and protection of beaches.

6. *The temporal change in coastal squeeze is assessed over the period 1990-2024, yet beach distribution data outside 2016-2024 are not available. Thus, which data did you use for sandy beach spatial coverage? Currently, this is not explained in the text.*

Response:

We sincerely appreciate the reviewer's valuable feedback on our study. Regarding the accuracy of beach extraction, we fully agree that beaches are relatively stable landforms that typically do not undergo drastic changes. However, due to the influence of tides, especially in areas with significant tidal fluctuations, the exposure of beaches can change. These tidal variations often result in beach areas being incorrectly identified as water bodies in the yearly extraction results, leading to an

underestimation of the actual beach area. The impact of tides not only introduces errors in the single-year beach extraction results but also exacerbates the errors due to the timing differences of satellite image acquisitions, further complicating the tidal effects in different regions.

To address this issue, we employed a multi-year synthesis approach, combining the yearly beach extraction results to derive the maximum beach extent. This method effectively reduces errors caused by tidal fluctuations, as the beach area extracted at different tidal levels each year varies. The combined result provides a more comprehensive distribution of the beaches, especially reflecting the maximum exposure during low tide. We believe that by synthesizing multi-year beach data, we can more accurately represent the actual distribution of beaches, avoiding errors caused by tidal fluctuations in single-year data, particularly for beach areas that may be covered by water during high tide.

Furthermore, to further improve the accuracy of the results, we also used land use data from 1990 to 2024 for a compression analysis. By creating a 100-meter buffer zone, we were able to identify the impact of land use on beach areas, particularly how the expansion of human infrastructure affects beach distribution.