ESSD-2025-514: Reply to comments from Referee
(Reviewer comments in bold, author responses in blue)

This is my second review of this manuscript. I abandoned my first review after reading the first 200+ lines because I found the presentation to be confusing, incomplete, and in places incorrect, making it difficult to follow the authors' reasoning. I provide examples of each of these in my review. The authors have made a serious effort to address my concerns but, unfortunately, the revision falls short. There are two classes of problems: 1) questions about the procedures and 2) the clarity of the writing (organizational/editorial issues). I deal with each of these below:

We sincerely thank the reviewer for their detailed second review and for acknowledging our previous efforts; we fully accept the criticism that the revision still falls short regarding procedural clarity and writing organization. We are particularly grateful for the specific examples and concrete rewriting suggestions provided, which have given us a clear roadmap for improvement.

Questions about Procedures

1) The abstract argues that no publicly "*available, longterm ocean front dataset currently exists, and the existing detection methods often rely on time-consuming manual labelling or traditional algorithms with limited accuracy in complex frontal regions*." This is true, but there have been a significant number of papers describing the global frontal field but, again, none that are publicly available. The important point is that the vast majority of previous work has been with fronts detected in L2 or L3 SST fields, not L4 fields but the dataset produced as part of this work is from L4 model reanalyses and, yet this is not mentioned in the abstract. I tend not to use L4 fields because of lost resolution and because of fronts resulting from artifacts in the field. This is an issue that is really not addressed in this manuscript.

We sincerely thank the reviewer for this insightful comment and fully agree with your perspective. In response to this valuable feedback, we have revised the manuscript as follows:

Clarification of data sources: We have revised the Abstract and Data section to explicitly state that our model was trained utilizing both L3 remote sensing satellite SST data and the GLORYS12V1 L4 reanalysis SST product. The specific revision reads: "*The model was trained utilizing both L3 remote sensing satellite SST data and the GLORYS12V1 L4 reanalysis SST product*." This addition clarifies the ambiguity regarding our data sources.

We acknowledge that while L4 reanalysis data may entail some loss of resolution compared to raw satellite observations, it provides the spatial and temporal continuity essential for training a robust deep learning model on a global scale. Moreover, by incorporating L3 products as supplementary data, we have enriched the data samples to further enhance the model's representativeness.

2) It was not always clear to me what data you were using—critical to really understand what you did. In Section 2 you say that you will use the daily average sea-surface temperature (SST) output from GLORYS12v1. That's clear although it would be good to provide a DOI for the specific data you used as well as a date you downloaded it. However, in Section 3.2 you say "*To expand the dataset, satellite remote sensing images*…". This sentence sounds like you are using raw satellite SST fields, not the GLORYS12v1 fields. If that is the case, what is the provenance of these fields? If not, this sentence should be rewritten. When you mention the preprocessing step (lines 174-175), it sounds like you are using raw satellite SST data, not the L4 GLORYS12v1 product. And in Section 4.5.1 Data source errors you say "*The accuracy of ocean front detection is influenced by data quality. If the data include noise or missing values or the data source is inconsistent, the model will learn inaccurate features and introduce errors. The reliance on satellite imagery as the primary data source for the proposed method may limit its applicability in certain situations. In such cases, complementary data sources, such as in situ measurements or marine numerical model results, may enhance the accuracy and robustness of the mode*l." I'm confused here. Didn't you use a L4 model derived product not original satellite imagery? If so, what sort of noise are you referring to here and what gives rise to missing data. This section is important but it's really not clear where you're going with it. Are you suggesting the problems that might arise if one were to apply your algorithm to L3 SST data or are you discussing something else? This brings me back to the lack of clarity in places with regard to the data you are using. I think that the manuscript would benefit if you were more explicit.

We thank the reviewer for this important and detailed comment. We agree that the data sources were not consistently described throughout the manuscript, which may have caused confusion. We have substantially revised the relevant sections to provide a clear and consistent description of the data used in this study.

Section 2 (Data sources): We have clarified that the primary data source is the GLORYS12V1 L4 reanalysis product (1993–2023, DOI: 10.48670/moi-00021). To augment the training dataset, we additionally utilized the ODYSSEA Global Ocean L3 multi-sensor SST product (2021–2022, DOI: 10.48670/moi-00164). Both DOIs and download periods are now specified.

Section 3.2 (Data labels): We have revised this section to clearly state that the annotated SST gradient maps were derived from the data sources described in Section 2. The ambiguous phrase "satellite remote sensing images" has been removed. We now explicitly define each annotated sample as one SST gradient map with its corresponding JSON file, specify that multiple fronts can be annotated per map, and describe the JSON output format.

Preprocessing steps: We have revised the preprocessing description to clarify that the L3 satellite data were interpolated to a 1/12° grid and screened for cloud cover, while the L4 reanalysis data (GLORYS12V1) are already quality-controlled and did not require additional preprocessing.

Section 4.5.1 (Data source limitations): We have rewritten this section to more

accurately reflect the context. The discussion now focuses on the model's generalizability and potential limitations when applied to lower-quality data sources (e.g., single-sensor L2 satellite observations with noise or missing values), rather than implying issues with the training data themselves. We clarify that the model was developed using high-quality L4 and L3 products, and we discuss how adverse weather conditions could affect performance in other deployment scenarios.

We believe these revisions provide a clear and consistent description of the data used throughout the manuscript and address the reviewer's concerns. We appreciate the reviewer's careful reading and constructive feedback, which significantly improved the clarity of our work.

3) Lines 174-175: "*2) Pretreatment: Preprocessing steps for the SST image data were conducted to enhance the quality and remove any noise or artefacts*." What do you mean by noise or artifacts here? You're using L4 fields right? What kind of noise did you detect in them and how did you remove this noise or these artifacts?

We sincerely thank the reviewer for raising this critical point, which has allowed us to clarify the description of our data sources and preprocessing methods, thereby enhancing the reproducibility of our work. We have revised the relevant sections as follows:

Clarification of Data Sources: In the "Data Acquisition" section, we have now explicitly stated that the SST data used in this study comprise two sources: the GLORYS12V1 L4 reanalysis product (primary data source) and supplementary Level-3 (L3) satellite observational data.

Revision of Preprocessing Description: In response to the vague term "noise/artifacts," we have entirely rewritten the pretreatment subsection. The key steps are:Spatial Resolution Unification: The L3 data were interpolated to a uniform 1/12° grid to match the resolution of the GLORYS12V1 L4 reanalysis data. Handling Cloud Cover: We did not interpolate missing values due to clouds. Instead, our approach was to screen the dataset to identify and select only the valid (cloud-free) L3 satellite observation samples within the study area, thereby ensuring data integrity for subsequent analysis.

These revisions have been integrated into the "Data Acquisition" and "Pretreatment" subsections of the manuscript. We believe that by clearly defining the data composition and the specific steps of our data handling, these sections now provide the necessary clarity and reproducibility. We appreciate the reviewer's feedback in helping us refine this part of the manuscript.

4) In your discussion of training dataset size (lines 242-258 and Table 3), it sounds like you are varying the test dataset. Varying the train/test split is fine for studying data-efficiency, but the test set should not be used to select the split (or any other setting). Please clarify whether a separate validation set (or nested CV) was used for selecting hyperparameters/split ratio, and whether the reported test set was held fixed and used only once for final evaluation. Also clarify what is meant by 'adding test data'

improving generalization.

We thank the reviewer for this important methodological comment. We agree that the original wording was ambiguous and could be misinterpreted as using the test set for model selection. To clarify, the experiments in Table 3 were designed to study data efficiency, and the data splits correspond to training and validation sets, not a test set. The validation set was used to evaluate model performance under different training data proportions and to guide the selection of the training/validation ratio. Hyperparameters (learning rate, batch size, number of iterations) were fixed across these experiments.

For final evaluation, we used a separate, independent test set consisting of SST data from 2023, which was not involved in any training or validation processes. This test set was held fixed and used only once to evaluate the final model performance.

We have revised the text to clarify these points. Specifically, we now:

1.Refer to the non-training portion in these experiments as the validation set rather than the test set;

2.Explicitly state that the training/validation split was selected based on validation performance, and hyperparameters were tuned using this validation set;

3.Clarify that a separate test set (2023 data) was used only for final evaluation.

We have also removed the inaccurate statement that "adding test data improves generalization," as generalization is influenced solely by training data and model design. The revised text and Table 3 caption have been updated accordingly in the manuscript. We appreciate the reviewer's careful reading and constructive feedback, which helped us improve the methodological clarity of our work.

5) Lines 296-298: "*Compared with traditional methods, the ocean fronts extracted by the deep learning model have a higher degree of fit, and in addition, the deep learning model can better reflect the characteristics of ocean fronts*." I'm confused here. The gradient of the input data is shown on the left in the Fig. 8 and the gradient of the deep learning image is shown on the right. Isn't the gradient from the input image the actual correct value? The deep learning image is simulating a gradient, but it seems that you want that gradient to be close to the one observed from the real data but not necessarily equal to it. You need to clarify why the deep learning values are better if that's what you mean.

We sincerely thank the reviewer for raising this critical point. We fully agree that the original wording was logically flawed. It incorrectly implied that the deep learning results were "better" than the traditional method's results, which themselves served as the comparative benchmark. This was indeed confusing and unjustified.

In response to your comment, we have thoroughly revised the wording of the relevant paragraph (please see the corresponding lines in the revised manuscript). The key amendments are as follows:

1.Clarified the logic of comparison: We have corrected the description to explicitly state that the results from the traditional gradient method serve as the "reference ocean front maps" against which the deep learning model's results are compared and validated.

This establishes the correct evaluation framework.

2.Adopted objective, descriptive language: We have removed the subjective and conclusive terms such as "a higher degree of fit," "better," and "more accurate." Instead, we now objectively describe the observation from Fig. 8: "the ocean fronts extracted by the deep learning model are highly consistent with those derived from the traditional gradient method."

3.Drew a reasonable inference from the observation: From this objective observation of "high consistency," we then draw a reasonable and supported conclusion: this "demonstrates the model's effectiveness in capturing frontal structures."

We believe the revised text resolves the original logical inconsistency and presents the argument in a more scientific and rigorous manner. We greatly appreciate the reviewer's help in improving the quality of this key section.

6) Line 307: You say that the width is defined as "distance from the centerline to the boundary," but it is not clear how the centerline could be defined and the distance to the boundary computed. Please add one or two sentences clarifying how the "centerline" is defined/computed from the front mask and how the distance to the boundary is measured (e.g., skeleton/medial axis + Euclidean distance transform, and whether the reported width is an average over centerline points). This would make the width metric fully reproducible.

We thank the reviewer for this valuable suggestion, which significantly improves the clarity and reproducibility of our width metric definition. As suggested, we have added a detailed clarification in the manuscript. The added text explicitly states that: The centerline is obtained by extracting the morphological skeleton of the detected front mask. The width at each skeleton point is computed as twice the Euclidean distance transform value at that point. The final reported width is the average of these per-point widths over the entire skeleton. This revision ensures that our width measurement methodology is now fully unambiguous and reproducible.

Clarity of the Writing

In my first review I suggested that the authors use a chatbot to help them with the English, specifically to address editorial/rhetorical problems. The authors responded that they used an editing service for this revision but this editing service does not appear to have addressed structure/clarity issues, which I think still plague the manuscript. Some examples:

We sincerely thank the reviewer for taking the time to re-evaluate our manuscript and for raising the crucial concerns regarding its structure and clarity of expression. We fully understand and acknowledge the issues you have pointed out. We are committed to implementing substantial revisions to the manuscript in order to thoroughly address the structural and clarity-related problems you have identified.

1) Section 3.1 Gradient Calculation Method. First, it is not clear to me why the authors refer to this as the gradient calculation 'method', why not just the gradient calculation. Second, the paragraph contains extraneous material, which doesn't add much if anything to the story. Specifically, this section can be reduced to a few precise sentences defining the centered-difference derivatives and the gradient magnitude.
For example, "We estimate the SST gradient magnitude G from the gridded SST field T(i,j) using centered finite differences:
D_x(i,j)=\frac{T(i,j+1)-T(i,j-1)}{2\Delta x},\qquad
D_y(i,j)=\frac{T(i+1,j)-T(i-1,j)}{2\Delta y},
G(i,j)=\sqrt{D_x(i,j)^2 + D_y(i,j)^2}.
Here \Delta x and \Delta y are the grid spacings (in km), and (i,j) denotes the grid indices."
Note that this is much shorter but, more importantly, it is to the point and adds clarity to the description with the specification that the spacing is in kms, assuming that it is. Also, note that I have fixed the incorrect subscript in Eq. 3.

We sincerely thank the reviewer for the detailed and constructive feedback on the description of the gradient calculation. We agree that the original presentation was unnecessarily verbose and contained extraneous information. We have revised Section 3.1 entirely in accordance with the reviewer's suggestion. The revised section is now concise and to the point, precisely defining the centered-difference scheme and the gradient magnitude calculation as suggested. Specifically, the section now reads:

"The SST gradient magnitude, G, is estimated from the gridded SST field, T(i,j), using centered finite differences:

$D\_x(i,j) = [T(i,j+1) - T(i,j-1)] / (2\Delta x)$,

$D\_y(i,j) = [T(i+1,j) - T(i-1,j)] / (2\Delta y)$,

$G(i,j) = \sqrt{[D\_x(i,j)^2 + D\_y(i,j)^2]}$.

Here, Δx and Δy are the zonal and meridional grid spacings (in kilometers), and (i, j) denotes the grid indices."

We have also corrected the subscript in the corresponding equation (formerly Eq. 3) in the manuscript. We believe this revision significantly improves the clarity and precision of the method description.

2) Section 3.2 Data Labels. This section is currently too vague to be reproducible. It introduces Labelme before explaining what it is, and it is unclear what constitutes one "annotation dataset." Please define the unit of annotation (one image/field? one tile?), specify whether multiple fronts can be annotated per image, and clarify the source of the "satellite remote sensing images" used for dataset expansion (are these the same fields described in Section 2, or an additional data source?). Finally, state the output label format and what each JSON file corresponds to.
Here's a suggested revision of the section:

3.2 Data labels
Ocean fronts were manually delineated using Labelme (an open-source image-

annotation tool) by drawing polygon boundaries around visually continuous front features in each SST-gradient map. Each annotated map was saved as a Labelme JSON file containing one or more polygons (multiple fronts may occur in a single map). We assembled a labeled dataset of 5,000 annotated SST-gradient samples (each sample = one SST-gradient map plus its corresponding JSON annotation file). To increase diversity, the samples were selected across different regions and seasons.

Of course, this is my understanding of what you have done so it may not be quite correct. If additional satellite remote-sensing products were used for this augmentation, the specific data sources and preprocessing steps should either be specified here or in another section, which you reference. Also, there are at least two versions of Labelme software so you should reference the one you use; e.g., if you use the Python tool, you might reference Wada, K. labelme: Image Polygonal Annotation with Python (computer software). Zenodo. https://doi.org/10.5281/zenodo.5711226

```
@software{wada_labelme_2021,
   author = {Wada, Kentaro},
   title = {labelme: Image Polygonal Annotation with Python},
   year = {2021},
   publisher = {Zenodo},
   doi = {10.5281/zenodo.5711226},
}
```

But you may be using the MIT CSAIL LabelMe project/tool?

These are just examples of significant rhetorical modifications that could be made to improve the readability of the document. I used ChatGPT to help me structure these and, if the Journal allows you to use an LLM to improve your manuscript, again, I suggest that you do so.

We thank the reviewer for the detailed and constructive suggestions. We have substantially revised Section 3.2 to improve clarity and reproducibility. Specifically, we now: (1) introduce Labelme as a Python-based open-source image annotation tool with a proper citation (Wada, 2021); (2) define each annotated sample as one SST gradient map with its corresponding JSON file; (3) clarify that multiple polygons (i.e., multiple fronts) can be annotated per map; (4) specify that the annotated data are derived from the GLORYS12V1 L4 reanalysis and L3 satellite SST data described in Section 2; and (5) describe the output format (JSON) and its content (polygon coordinates and metadata). The revised section is provided in the manuscript. We appreciate the reviewer's thoughtful feedback, which has helped us significantly improve this section.

Here are other editorial issues I noticed while reading the document:

3) Line 55: "with an mean Dice Similarity Coefficient (mDSC)" —> "with [a] mean..." and a reference to 'Dice Similarity Coefficient' would be helpful.

We thank the reviewer for pointing out this error and for the helpful suggestion.

We have corrected the article from "an" to "a" and added a reference to the original Dice Similarity Coefficient (Dice, 1945). The revised text now reads: *"with a mean Dice Similarity Coefficient (mDSC) (Dice, 1945)"*. We appreciate the reviewer's careful reading and constructive feedback.

4) Lines 126-128. I found the use of colour to be confusing here: "*However, detecting and characterizing these fronts is challenging due to their complex and dynamic nature. In particular, the visual similarities among different fronts can make it difficult to distinguish them based on colour and shape alone*." Please use consistent terminology—you use grayscale elsewhere in the manuscript—distinguishing the data (SST field/gradient field/raster) from visualizations (color-mapped images). If inputs are single-channel, avoid 'colour' and instead refer to 'intensity/value/gradient magnitude' and spatial pattern/texture. For example, you might want something more like: "In particular, the visual similarity of different fronts can make them difficult to distinguish based on gradient magnitude patterns and shape alone." if single channel gradients or "In particular, the visual similarity of different fronts can make them difficult to distinguish based on colormap intensity patterns and shape alone (noting that color is a visualization choice rather than a physical variable)." if colormap. And, if you truly use RGB composites (unlikely from what you've shown) then you need to state what the channels are.

We thank the reviewer for this important terminological clarification. We agree that the use of "colour" is inappropriate given that our input data are single-channel gradient fields. To maintain consistency and accuracy, we have revised the sentence to refer to gradient magnitude patterns and spatial shape rather than color. The revised text now reads: *"However, detecting and characterizing these fronts is challenging due to their complex and dynamic nature. In particular, the visual similarities among different fronts can make them difficult to distinguish based on gradient magnitude patterns and shape alone."*We appreciate the reviewer's careful attention to terminological precision.

5) Line 194:"First, the accuracy of the border was represented..." What do you mean by the border, is it the border of the image or the front you have found? It's actually pretty clear that it is not the edge of the image but up until now you have used front so, if that is what you mean, why not use it?

We thank the reviewer for this clarification. We agree that the term "border" is ambiguous in this context. We intended to refer to the boundaries of the detected ocean fronts. To maintain terminological consistency throughout the manuscript, we have revised the text to use "front" instead of "border." The revised text now reads: *"First, the accuracy of the front was represented..."*. We appreciate the reviewer's careful reading and helpful suggestion.

6) Line 220: You mention 'COCO-style instance labels'. I'm not familiar with these. You need at a minimum a parenthetical description accompanied with a reference.

We thank the reviewer for this helpful suggestion. We have added a brief parenthetical description clarifying that COCO-style instance labels refer to JSON-formatted annotations containing polygon coordinates and instance-level segmentation information. We have also included a reference to the Microsoft COCO dataset (Lin et al., 2014) to provide proper context. The revised text now reads:"These masks were then converted into COCO-style instance labels (i.e., JSON-formatted annotations containing polygon coordinates and instance-level segmentation information) to train the Mask R-CNN model (Lin et al., 2014)." We appreciate the reviewer's careful reading and constructive feedback.

7) Lines 229-230: "*Notably, to ensure the accuracy of the evaluation, the test set data do not need to participate in the training process*." What do you mean by the test step not needing to participate? Either they do or they don't. If the latter, which I'm assuming is the case, just say so: "…*the test set data do not participate in the training process*."

We thank the reviewer for this clarification. We agree that the original wording was ambiguous. We have revised the sentence to clearly state that the test set data do not participate in the training process. The revised text now reads: *"Notably, to ensure the accuracy of the evaluation, the test set data do not participate in the training process."* We appreciate the reviewer's careful reading and constructive suggestion.

8) Lines 265-280: The first paragraph here is just a verbalization of what's in Table 4. You should only verbalize the parts of the table, which you emphasize in subsequent discussion. In fact, the 2nd paragraph discusses the table in more detail. These two paragraphs can be combined and shortened.

We thank the reviewer for this helpful suggestion. We agree that the original text contained redundant line-by-line descriptions of Table 4. Following the reviewer's advice, we have significantly shortened the first paragraph and integrated the key data points into the second paragraph. The revised text now presents the essential trends from Table 4 in a concise manner. We believe this revision improves conciseness and readability. The updated text has been incorporated into the manuscript.

9) Line 324: "*Spring and summer have relatively fewer ocean fronts, with the*..." Fig. 9 doesn't show numbers but rather gradients. Yes, there are places where there are not gradients in spring and summer so there are less but, still, I would have an image with color being counts if the number is really important.

We thank the reviewer for this important observation. We agree that Fig. 9 displays gradient magnitudes rather than frontal counts, and the original wording referring to "number of ocean fronts" was not directly supported by the figure. To address this, we have revised the text to align with what is actually shown in Fig. 9, focusing on gradient intensity and the spatial extent of frontal features rather than explicit counts. The revised

text now reads: *"Spring and summer exhibit relatively weaker gradient magnitudes and less extensive frontal structures, with the weakest signals observed in summer. In terms of frontal activity inferred from gradient intensity, the order, from weakest to strongest, is summer < spring < autumn < winter."* We appreciate the reviewer's careful reading and constructive suggestion to improve the consistency between the text and the figure.

10) Lines 326-327 reference locations in Fig. 9. It would help for readers not very familiar with the region to label the locations mentioned in the figure.

We thank the reviewer for this helpful suggestion. To improve clarity for readers who may not be familiar with the region, we have added labels for the mentioned locations directly in Fig. 9. The revised figure now clearly indicates the relevant geographic locations, making it easier to follow the corresponding discussion in the text. We appreciate the reviewer's thoughtful advice to enhance the accessibility of our work.

11) Lines 354-356: I wouldn't use 'hydrological' here. In most scientific writing, hydrological points readers toward hydrology on land (precipitation, runoff, rivers, groundwater, watershed processes) and "the water cycle," not the ocean's water-mass properties. If you mean T/S/density and water masses, you are talking about hydrographic conditions. If you mean currents/front dynamics specifically, then use oceanographic or physical oceanographic conditions. I would make it ; e.g., "…*Rapid changes in temperature and salinity gradients may affect*…" if that is appropriate..

We thank the reviewer for this important terminological clarification. We agree that "hydrological" is not appropriate in this oceanographic context. Following the reviewer's suggestion, we have revised the sentence to directly refer to the relevant physical parameters. The text now reads:"*Rapid changes in temperature and salinity gradients may affect the spatial structure and position of the fronts*."We believe this revision accurately reflects the intended meaning and aligns with standard scientific terminology. We appreciate the reviewer's careful attention to language precision.