Response to Reviewer 2:

Dear Reviewer:
We would like to thank you for your patience in reading the paper in detail and your valuable comments. We sincerely think their detailed comments have helped us to improve the manuscript. Below, we present our point-by-point response (text in blue denotes our replies). We hope the manuscript is now acceptable following our revisions and explanations.

Major comments:

Comments 1: It is not clear what goal this AI SWH model would like to achieve. Is it to make good 0-10 days SWH predictions similar to traditional numerical wave models? If yes, we should use predicted wind fields from GFS (or IFS) as the forcing instead of the ERA5 reanalysis wind fields. In the current manuscript, one year of SWH series was generated by the AI SWH model. But it essentially functions like post-processing the ERA5 wind analysis to 0-6h short-term SWH forecasts. This is not the regular 0-10-day SWH forecasts we normally expect.

Response 1: Thank you for your comments. Regarding the question of whether our model is solely designed for wave prediction, the answer is "No. " While one application of our AI SWH model is wave prediction, it is essential to emphasize that wave models—whether numerical or statistical—are not limited to forecasting wave conditions over a few days. Wave models and weather forecasting models fundamentally differ in purpose and methodology. Weather forecasting is a typical initial value problem, focusing on accurate predictions of atmospheric conditions over short periods, with outputs highly sensitive to initial conditions. In contrast, phase-averaged wave models, such as Numerical Wave Models (NWMs), address a forcing problem where wave evolution is primarily driven by external forcings, such as wind fields. These wave models are used not only for forecasts but also for hindcasts and projections.

A robust wave model should perform well when driven by high-quality forcing fields and exhibit errors proportional to the quality of those fields. Therefore, wave models are typically evaluated using high-quality reanalysis forcing fields to isolate errors originating from the model itself rather than from the forcing fields. For this reason, we did not use forecasted wind fields (e.g., GFS or IFS forecasts) as forcing inputs for our model, as their larger errors could compromise the training and calibration of the wave model.

The primary goal of our study was to develop an AI-based SWH model that mirrors the input-output structure of NWMs but excludes wave spectra to reduce computational complexity. We anticipated that the absence of directional wave spectra might lead to error divergence in rolling predictions, even during hindcasts. However, the rate of divergence was unclear. Our findings demonstrate that the error stabilizes after approximately 200 hours, with stabilized errors comparable to those of state-of-the-art NWMs, particularly in wind-sea-dominated regions. This highlights the potential effectiveness of our AI-based wave model surrogate.

Traditional NWMs often downplay data assimilation's importance due to the forcing-driven nature of directional wave spectra evolution. However, the observed error divergence in our AI SWH model underscores the significance of initial conditions. Thus, the second focus of our study was to assess whether direct data assimilation of altimeter-derived SWH (bypassing wave spectra) could mitigate error divergence. Our results confirm that assimilating altimeter data effectively reduces error divergence and enhances the model's reliability.

Furthermore, the error divergence of our AI SWH model, even during hindcasts, highlights its resemblance to an initial value problem, unlike traditional NWMs. Consequently, data assimilation becomes particularly beneficial, especially in swell-dominated regions. Without data assimilation, as shown in Figure 4, the model exhibits poorer performance in "swell pools." To our knowledge, this study is the first to address error divergence in AI-based SWH rolling models and propose altimeter data assimilation as a solution. Additionally, the absence of directional wave spectra simplifies SWH assimilation by eliminating the need for spectral adjustments—an advantage not previously reported in the literature.

Comments 2: Section 2.4 and other relevant parts: (1) data assimilation will improve initial conditions, but it is not related to the establishment of this AI SWH model and it is suggested to remove this part from this manuscript; (2) the data assimilation method here is too simple. Nowadays, we would expect at least a 2DVAR method that considers the uncertainties of the background and the observations.

Response 2: We appreciate the concerns raised by the reviewers in the data assimilation section.

For comment (1), As previously mentioned, the AI SWH model developed in this study exhibits error divergence, even during hindcasts, due to the absence of directional wave spectra. This error divergence suggests that the AI SWH model shares more characteristics with an initial value problem than traditional NWMs. Consequently, data assimilation becomes an essential component of the AI model, particularly in swell-dominated regions. Without data assimilation, as demonstrated in Figure 4, the model performs poorly in "swell pools. " This represents a significant limitation. To our knowledge, this study is the first to identify and address error divergence in AI-based SWH rolling models by proposing the assimilation of altimeter data as a solution. Additionally, the absence of directional wave spectra simplifies the assimilation of SWH, eliminating the need for spectral adjustments. To the best of our knowledge, this straightforward assimilation capability of the AI SWH model has not been reported in previous studies.

For comments (2), there are two primary practical and computational reasons for our preference for objective analysis (OA)—a form of optimum interpolation (OI)—over variational methods:

1. OA is significantly less computationally demanding than variational methods. One of the AI model’s greatest advantages is its efficiency and lightweight nature. Introducing variational methods for assimilation would increase computational demands by several orders of magnitude, rendering the AI model inefficient and impractical. Furthermore, OI enables incremental assimilation of observations, allowing for continuous updates as new data becomes available. In contrast,

variational methods typically require a complete assimilation cycle, which may not be feasible for fast-paced AI applications.

2. To our knowledge, there is no clear evidence that variational methods outperform OI in wave modeling. For instance, major operational centers such as ECMWF continue to use OI for wave modeling (IFS DOCUMENTATION – Cy43r1 –PART VII: ECMWF WAVE MODEL, available at: https://www.ecmwf.int/sites/default/files/elibrary/2016/79992-ifs-documentation-cy43r1-part-vii-ecmwf-wave-model_1.pdf), although there is still debate on whether NWMs need data assimilation. Therefore, our lightweight assimilation scheme is sufficient for the AI SWH model, as supported by our results.

We have also added some of the above explanation to our revised manuscript.

Specific comments:

Lines 48-49: "However, …" The statement is not accurate. The SWH prediction is both a forcing and an initial value problem. When focusing on the short-term (a few hours to a day) forecasts, the initial value will have a large impact. Zhou et al., 2021 and Ouyang et al., 2023 targeted 24h and 3 days forecasts respectively.

Thank you for pointing out this issue. We acknowledge that SWH prediction involves both a forcing and an initial value problem, especially when focusing on short-term forecasts. We will revise the manuscript to clarify this distinction. We will make the following changes in the revised manuscript: "Consequently, some studies have already explored AI applications in wave modeling. Some have attempted to replicate the AI weather forecasting approach by treating wave modeling as a purely nonlinear auto-regression problem of spatio-temporal series (e.g., Zhou et al., 2021; Ouyang et al., 2023). However, this approach overlooks the fact that phase-averaged wave modeling should not be treated as an initial value problem. Without a wind field driving the model, it is physically impossible to accurately simulate waves directly from past wave evolution alone. While initial conditions do play a role in short-term prediction, these auto-regression models cannot even run without the initial conditions provided by an NWM. "

Additionally, we emphasize that while initial values are essential for short-term wave forecasts, they can be readily derived from past forcing fields. This characteristic underscores why wave modeling is often viewed more as a forcing problem rather than an initial value problem.

Line 124-125: "Particularly, if a long series…" Could you clarify which part of Song and Jiang, 2023 drew this conclusion?

Song and Jiang (2023) successfully modeled the directional wave spectrum at a single point using only historical wind field data, without relying on any initial wave field conditions. Their approach was based on the premise that waves are either generated by local winds or influenced by remote historical winds. In their models, the only inputs were wind fields from the preceding 240 hours, with no wave data used as inputs, yet they achieved good results. This conclusion is further supported by a recent study by Wang and Jiang (2024) (Physics-guided deep learning for skillful wind-wave modeling, Science Advances, https://www.science.org/doi/10.1126/sciadv.adr3559). In this study, global SWH was modeled using only wind fields from the past 240 hours, without incorporating initial SWH field data, also yielding good results.

Lines 133-134: "We believe…": Evidence, instead of a subjective "belief", is expected here to demonstrate why U-Net is suited for this work.

While there may be various reasons for selecting a specific AI model architecture, determining whether a model is genuinely well-suited to a particular problem requires thorough testing. The rationale for choosing U-Net is briefly outlined in the manuscript: "The encoder progressively extracts features from the input through convolution and pooling, while the decoder reconstructs spatial resolution using de-convolution and up-sampling. Skip connections link corresponding layers of the encoder and decoder, preserving high-resolution details". Admittedly, this initial selection was based on a subjective belief regarding the model's suitability. However, our results provide strong evidence that U-Net is indeed well-suited for this task.

To present this explanation more objectively, we have revised the text as follows: "Such a CNN-based deep learning model is well-suited for wave statistical modeling using our input-output structure, and the effectiveness of U-Net in in wave modelling has been shown in previous studies (e.g., Gao and Jiang, 2023, Wang and Jiang, 2024). "

Lines 135-136: "The processes of both…": In AI for NWP practices, lots of literature has demonstrated that AI struggles to resolve different scales at the same time (for example, smoothing to get better medium-range forecasts and hence not able to resolve smaller scales features). So how can it be assured that "The processes of both local wave generation by wind and wave propagation in space can be captured by convolutional kernels at different scales? More discussions on this are needed.

The "smoothing problem" observed in AI weather models is not an issue for AI wave models. The wave response to wind inherently acts as a low-pass filter for wind fields, as waves are essentially the integral of wind forces. This "smoothing effect" aligns well with the requirements for modeling wave generation, particularly for SWH. While for wave propagation, the propagation of swells can be regarded as a linear progress, which can also be captured be convolutional kernels at different scales. In addition, in AI weather models, the "smoothing effect" poses a significant challenge as it inhibits the model's ability to run independently without initialization from numerical models. However, this "smoothing effect" does not impact the rolling modeling of waves, as wave models, being forcing-driven, can even operate in a "cold start" mode without relying on any initial conditions.

Lines 139-142: It looks like there is a problem with the "-190 degree to 190 degree" trick: How do waves at 180 degrees propagate to -179 degrees in this method?

The method is indeed feasible, although it may not have been clearly described in the manuscript.

In the revised version, we will clarify this implementation to enhance understanding: "To handle the wraparound at the -180° and 180° longitude boundary, we employed an engineering trick of extending the input fields from -180° to 180° (720 longitudes) to -190° to 190° (760 longitudes). Specifically, the data from -180° to -170° were duplicated and appended to 180° to 190°, and a similar treatment was applied to the opposite boundary. This effectively connecting the two boundaries and avoiding discontinuities during the modeling process." This approach ensures that waves can propagate seamlessly from 180° (-180° during computation) to -179°.

Line 161-168: (1) How does the epoch ensemble method work? Need more theoretical discussions here. Generally, we would like to use all available data to train the best AI model (instead of splitting them into different epochs). (2) We can add different perturbations to generate ensembles or use ensemble wind forcings (such as from GEFS) (3) The reduction of RMSE from the 4 ensembles shown in the manuscript may come from the smoothing effect.

Thank you for your thoughtful comment. The epoch ensemble method is a commonly used engineering trick in deep learning, distinct from the concept of ensemble forecasting in NWP. Specifically, during model training, multiple models that have already converged on the validation set are saved, and their mean output is taken as the final result. Ideally, obtaining the best AI model would be preferable. However, practical limitations, such as insufficient training data, constraints in time and cost, model hyper-parameter tuning, and model complexity, often make it challenging to achieve the optimal model. This challenge is further compounded when the loss function represents the (weighted) sum of multiple components. During training, some regions may be overestimated while others are underestimated; in subsequent iterations, these regions may reverse. Achieving convergence across all locations without over-fitting can be particularly difficult. In such cases, the epoch ensemble method can provide a straightforward solution to balance these loss discrepancies.

For Comment (2), we agree that ensemble wave outputs can be generated by using ensemble wind forcings. This represents a significant potential application of our wave model surrogate. The AI model's computational efficiency allows it to process more ensemble members of wind forcing compared to NWMs, reducing time and computational costs. However, this application is beyond the scope of the discussion in this section.

Regarding Comment (3), again, we need to mention that smoothing effect is not an issue for wave modelling. Additionally, it is important to note that the epoch ensemble method involves "smoothing" across ensemble members rather than spatial or temporal smoothing of the data itself.

Fig 2: It looks like the "AI model with data assimilation" results are NOT free forecasts but analyses performed every 6 hours along with short-term (0-6h) forecasts.

Yes, the assimilation is performed every 6 hours, and the results are not forecasts but hindcasts (with ERA5 wind as inputs). Also, the time interval for assimilation can be freely configured in the provided python codes.

Line 220: "This suggests that the simple AI model can function independently, at least, in certain scenarios." What does this mean and what scenarios it refers to? Need clarifications.

Many previous AI SWH models, such as Zhou et al. (2021) and Ouyang et al. (2023), have to be used with the help of NWMs (providing the initial field), so that they cannot be used independently. While this model does not need any information from a NWM so we call it "function independently". After realizing that this might be a bit confusing, this sentence is revised to: "This suggests that the simple AI model can work without the assimilation of observation and the information from NWMs, at least, in some applications such as modelling the SWH in wind-sea dominated regions"

Fig. 3, 4, 5: Are the results in these figures from the AI model with or without data assimilation?

The results shown in Figs. 3, 4, and 5 are from the AI model without data assimilation. It is worth noting that even without assimilation, the proposed AI rolling model achieves accuracy comparable to numerical models when evaluated against the CCI altimeter dataset in wind-sea-dominated regions. To avoid potential misunderstandings, we will revise the figure captions in the revised manuscript as follows:

Figure 3: The variation of global overall error metrics between the AI SWH model outputs and ERA5 with simulation time: (a) CC, (b) bias, (c) RMSE, and (d) SI. The orange lines and shaded areas are the same as those in Fig. 2, but no epoch ensemble is used. The blue lines and shaded area are the corresponding results for the cold start with an initial field of zero SWH. These results do not use data assimilation.

Figure 4: Comparison of SWHs from the AI model (without data assimilation) at 240-h hindcast time (when the errors are stable) with ERA5 for the year 2020. (a) Scatter plot between the SWHs from the two datasets. (b-e) Global spatial distributions of CC, bias, RMSE, and SI, respectively.

Figure 5: The same as Fig. 4, but the comparison is between the 240-h SWH hindcasts of the AI model and the CCI-Sea State dataset.

Line 378-379: "why the AI model can slightly outperform the NWM in these areas." If we want to draw this conclusion, we will need to run both the NWM and AI models with the same settings and compare them directly.

We have to admit that it is difficult for the NWM and AI models to have the same setting because they are fundamentally different approaches. The settings used for AI models cannot be directly implemented in NWMs, and vice versa. However, we have provided a comparison between a state-of-the-art NWM hindcast dataset, WW3-ST6, and altimeter observations in Figure S4 of the Supplementary Materials. We put this figure in the Supplementary Materials to save space, because it is only used as a reference. This figure can be directly compared with Figure 5 in the main text. The results indicate that the AI model slightly outperforms the NWMs in certain regions, such as the westerlies in the Southern Hemisphere.

Lines 386-390: The initial value problem has a predictability limit (around 10 days for the wave forecasts). So one would not expect the impact of initial SWH will last beyond ~10 days.

We agree with the reviewer's observation, but it appears we may not have fully grasped the main point of this comment. Wave modeling, fundamentally, is not an initial value problem. The 10-day limit of wave forecasts primarily stems from the 10-day predictability limit of weather (wind) forecasts. If high-quality wind data are available, such as in hindcast scenarios, it is feasible to model waves over longer periods for both our AI model and NWMs.

Line 423: "An important advantage of the AI SWH model proposed here is its low computational cost compared to traditional NWMs.": It will be good if we can give concrete examples of the computational resources needed by the AI SWH model and the traditional NWMs.

We have incorporated the corresponding description into the revised manuscript according to your comment: " For example, on a personal laptop equipped with a single GTX 3080 GPU, the AI model can perform a 1-year global SWH rolling simulation at a resolution of 0.5° × 0.5° × 1h in just 10 minutes. In contrast, traditional NWMs, such as the WAVEWATCH III model, typically require several days to complete a simulation with the same output, even on supercomputing facilities."

Lines 445-446: "While training such a model would be challenging, it is not an impossible task, and the rapid advancements in AI may make this goal more achievable in the future": This sentence is too subjective, Consider revising.

We will revise the sentence to make it more objective:
"Training such a model would be challenging due to the complexity of the task, but ongoing advancements in AI methodologies, particularly in deep learning, are continuously improving the possibility of achieving this goal."

Edit:

Line 304: "clearly evident" -> "evident"

Revised.

Thank you again for your comments and suggestions. We will update the manuscript accordingly.