CC 2: Comment from Ping Chen:

Reply:

Thanks for the comments! During the reply, we found the page numbers and the reference numbers are not consistency to the manuscript, then based on the specific issues mentioned in these comments, we response as the following to clarify related data processing and analysis issues in the comments.

1. **On page 22, in the third step, what is the specific process of data screening based on the gamma parameter? How are Hs and kp in the empirical spectrum obtained? Are they directly calculated from the SWIM spectrum data?**

   The Hs data of the SWIM spectrum is derived from the Hs at the nadir point.

   The SWIM spectrum data exhibits obvious spurious peaks and the surfbeat effect under low to medium sea states. Note that these are two different non - linear effects. Therefore, the spectrum data under such sea states is incorrect and cannot be used as reference spectrum data. Or the wave parameters kp and other conclusions obtained using these spectra as reference data are unreliable.

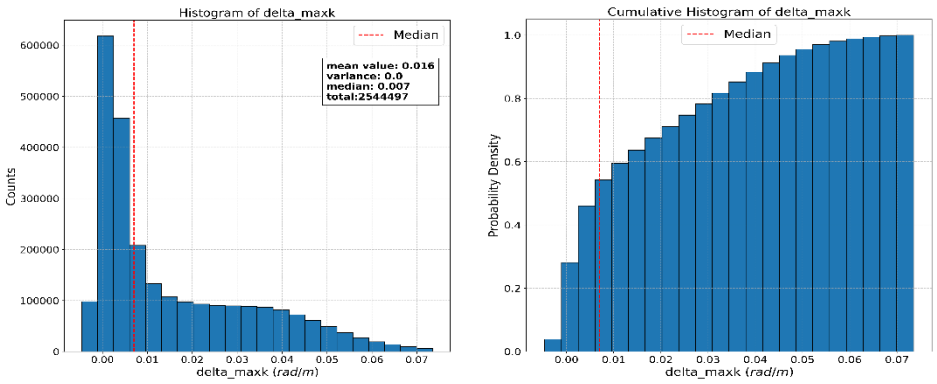
   Reply: We appreciate the reviewer's valuable comments regarding the data screening process based on the gamma parameter. During the preprocessing SWIM L2 product data, we employed a 0.5%-99.5% confidence interval to eliminate abnormal samples of peak enhancement factors outside this range. Statistical analysis shows that this confidence interval corresponds to peak enhancement factor values ranging approximately from 1 to 12, so we ultimately selected wave units with peak enhancement factors between 1-12 as our research data.

   Regarding the construction of the Combined spectrum (C spectrum), we utilized both significant wave height $H_{\frac{1}{3}}$ and peak wavenumber k_p information from SWIM L2 data.

   The significant wave height $H_{\frac{1}{3}}$ is directly provided by the SWIM L2 product, while the peak wavenumber k_p can be calculated from the dominant wavelength λ_p data in the SWIM L2 product.

   We consider that spurious peaks and the surfbeat effect may cause the observed spectral peak to correspond to low-wavenumber false peaks, leading to an underestimated maximum peak wavenumber k_{max} and affecting data reliability. For this, we implemented quality control by calculating the difference $\Delta = k_p - k_{max}$ and setting a reasonable threshold to exclude samples with excessively large Δ values. This has been explained at the content around the last 3 lines in part 3.1.1. Specifically, in practice, k_p is directly obtained from SWIM L2 data, while k_{max} is determined by identifying the wavenumber corresponding to the maximum peak in the observed spectrum. Since false

peak data typically exhibit significant discrepancies between k_p and k_{max}, whereas valid data should show close agreement between these values, Δ should theoretically be as small as possible. In this way, the data with those effects are screened out. Statistical analysis of the Δ distribution guarantees and verifies this quality control procedure, and shown specifically in the following figure.



(a) (b)

Figure Re-CC2-1 statistic on Δ

The figure above (Figure Re-CC2-1) shows the frequency histogram and cumulative probability density histogram of Δ, showing that over 99.5% of the data samples are positive, indicating that most k_p values lie to the right of k_{max}. The median Δ is 0.007, with a mean value of 0.016. Panel (a) reveals that the data are primarily concentrated between 0-0.007, and panel (b) demonstrates that this range accounts for half of the total samples (approximately 1.5 million boxes). To minimize data errors, we selected the median Δ as the threshold for screening, ensuring that all obtained data fall within this threshold.

To address this clearly, we update part 3.1.1 for description of SWIM data. We have expanded the content regarding data screening using the Δ threshold of the revised manuscript. The content has been modified from "Subsequently, to suppress abnormally high values at low wave numbers of SWIM caused by parasitic peaks (Merle et al., 2021; Xu et al., 2022), " to "To suppress abnormally high values at low wave numbers of SWIM caused by spurious peaks and surfbeat effects (Merle et al., 2021; Xu et al., 2022), Data screening was implemented by calculating $\Delta = k_p - k_{max}$ and setting a reasonable threshold to exclude samples with excessing large Δ values. k_p represents the SWIM-observed peak wavenumber and k_{max} corresponds to the wavenumber of maximum spectral energy. Since false peak data typically exhibit significant discrepancies between k_p and k_{max}, whereas valid data should show close agreement between these values, Δ should theoretically be as small as possible. As shown in Fig. 1, the frequency histogram in panel A reveals that the data are primarily concentrated below 0.007, and the cumulative distribution in panel B demonstrates half of the total samples fall within this range for $\Delta \leq 0.07$. After balancing the trade-off between the sufficiency of retained samples and screening effectiveness, 0.007 was selected as the threshold value. Subsequently, according to the histogram of γ, the boxes for γ ranging from 1 to 12 are selected as the research data that depicts the sea surface well. In

total, 1 251 067 boxes are found as the experiment data in 2021 for modelling, and 1 307 256 boxes in 2022 are obtained for validation." And Figure 1 below is newly included:

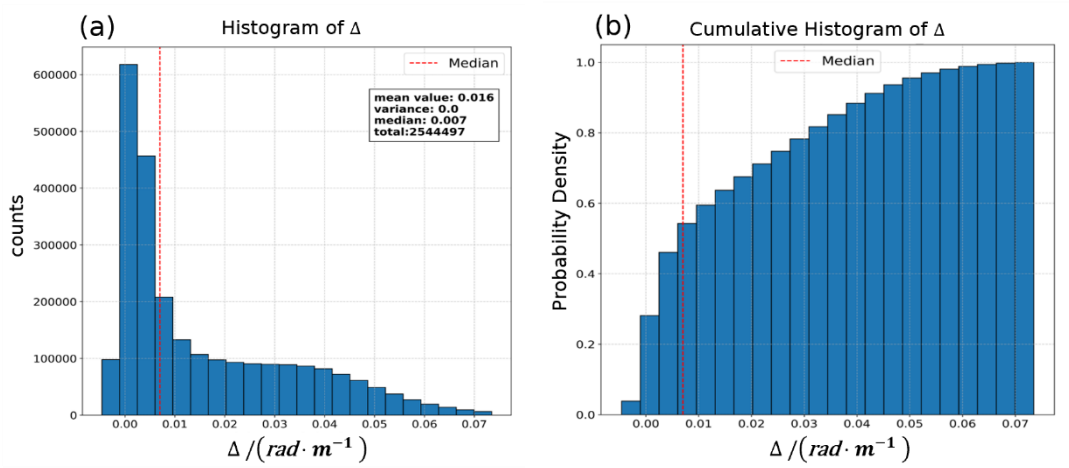


Figure 1: The histogram and Cumulative Histogram of Δ

2. **On page 33, in Figure 3 - 9, is the SWIM spectrum the average of multiple samples? Due to the statistical fluctuations of the SWIM spectrum, even under the same sea surface conditions, the kp corresponding to the SWIM - measured spectrum still varies. Directly averaging the spectra will lead to a decrease in the spectral peak and an increase in the spectral width. Therefore, the common practice is to perform wavenumber - normalized spectral averaging. A detailed description of this method can be found in Ref. [52].**

   Reply: Thanks to the reviewer for pointing the issue of the potential for direct averaging required for SWIM spectra. Spectral averaging under the same sea surface conditions may lead to reduced peak intensity and increased spectral width due to statistical fluctuations in SWIM spectra. Though we didn't demonstrate the SWIM spectra in figures, the averaging issue is a good point for SWIM data applications. Then we would like to clarify that for the development of the Combined spectrum. In this manuscript, we perform parameter fitting using combination of bunches of individual observation samples. Thus, instead of focusing on individual expression of the spectrum, we focus on the statistics features, this has been explained in the manuscript in part 3.1.1 and section 3.2 for data amount and data applied specifically in the fitting details respectively, and the concerned spectral averaging does not affect our results.

3. **Because the SWIM spectrum data has its own problems under low to medium sea states, it is not suitable to use SWIM data as reference data to establish an empirical spectrum model.**

   Reply: We appreciate the reviewer's comment regarding the limitations of SWIM spectrum data under low to medium sea states and its suitability as reference data for establishing an empirical spectrum model. We fully acknowledge that SWIM wave spectrum data may be affected by nonlinear effects such as spurious peaks and the surfbeat effect under these conditions. To minimize this adverse effect, we implemented

data clean primarily based on the difference between peak wavenumber k_p and the maximum peak wavenumber k_{max}, as detailed in our reply to your first comment (CC2 Comment 1), to exclude abnormal wave box data as much as possible with mainly data expressing the surface features obtained for our further procedures. Moreover, for validation of our results, in addition to using SWIM observation data, we also employed independent validation with NDBC buoy data, as in part B of section 3.4.1. Evaluation results using both DI and R^2 metrics demonstrate that compared to classical wave spectrum models such as Elfouhaily and Goda spectra, the Combined spectrum can more accurately characterize real sea conditions. This also in another aspect, validated the method proposed. With the response to the first comment, we updated part 3.1.1, by inclusion of specific details for date selection, the specific update please refer to the reply to your comment 1 (CC2 Comment 1).