Response to Reviewer #2

Manuscript ID: ESSD-2025-694 Earth System Science Data

Title: A global gridded dataset of significant wave height via fusion of multi-mission altimetry and numerical hindcast

Authors: Haoyu Jiang (Haoyujiang@szu.edu.cn), Hao Su

Owing to lack of data used as a benchmark in global wave modeling authors of the current paper look at the proper approach of correction satellite altimetry products and their conclusion is quite clear. Their research focuses on the important question of how to create trustful dataset of significant wave height (Hs) by fusing multi-mission satellite altimeter observations with numerical wave models and understand its effects on precision. From technical point of view the proposed fusion method employs a space-time optimal interpolation that incorporates spatial, temporal, and Hs differences, as well as a correlation-based threshold, to optimally blend model and observational data. Furthermore, the size of the period taken into consideration is relatively small (2010-2016) to draw a reliable conclusion. The standard operational baseline period designated by the World Meteorological Organization (WMO) is 1991–2020. Besides this fact modeled wave climate is not fully Hs climate (due to wave systems partition method, e.g PTM), however multi-mission satellite wave climate is purely Hs climate (nrt, inter-calibrated, calibrated by buoys), as aptly pointed out by the authors in the main body of the text when they describe discrepancies in observations. It is not entirely clear whether we are improving the model or the observations.

I have carefully read the paper and I believe that it is an interesting study worth publishing after minor revision.

We thank the reviewer for the positive assessment of the manuscript and for raising these important points. We agree that the original description was not sufficiently precise regarding the validation period, the meaning of wave climate in the context of this study, and the respective roles of the numerical hindcast and satellite observations. We have therefore clarified these issues throughout the revised manuscript.

First, the period 2010–2016 refers only to the independent validation against the selected NDBC buoy observations. It does not represent the temporal coverage of the fused dataset or the period used to define a climatological baseline. The two-satellite fused product intended for climate-related applications covers October 1992–December 2023. The 30-year period recommended by WMO is the current standard period for calculating climatological normals. However, the purpose of the 2010-2016 analysis in this study is to quantify the accuracy of instantaneous SWH estimates against independent buoy measurements, rather than to calculate a climatological normal or derive long-term climate trends. We have revised the manuscript to make this distinction explicit and have avoided drawing climatological conclusions directly from the validation period.

Second, we agree that SWH alone does not provide a complete description of wave climate, which may also include wave period, direction, spectral shape, and partitioned wind-sea and swell components. We have therefore replaced several general references to "wave climate" with the more precise terms "SWH-based wave climate". Nevertheless, the quantities fused in the present study are physically consistent because both datasets represent total SWH.

Third, the objective of the proposed method is neither to recalibrate the original satellite observations nor to modify the physics or numerical integration of the wave model. Instead, the method constructs an observation-constrained gridded SWH dataset. The numerical hindcast provides a spatially and temporally continuous background field, while the jointly calibrated altimeter observations provide correction increments. The original altimeter measurements and the WAVEWATCH III model configuration remain unchanged. Therefore, the resulting product is therefore neither an observation-only gridded product nor an improved numerical simulation, but a fused SWH analysis constrained by both sources.

These clarifications have been added to the manuscript:

"It is noted that these buoy data is used only for independent validation against the selected NDBC buoy observations and does not represent the temporal coverage of the fused dataset or a climatological reference period."

"To address the limitations outlined above, this study presents a dataset that fuses jointly calibrated satellite altimeter observations with NWM hindcast outputs. The proposed fusion framework addresses above limitations in different ways. First, it uses the jointly calibrated CCI Sea State altimeter record as the observational reference, thereby avoiding calibration against the model background. Second, because the fusion is performed offline after the hindcast has been generated, both past and future altimeter observations can be used to constrain the SWH field at a given analysis time. Third, this retrospective constraint also reduces the dependence of the reconstruction on the forward spectral propagation in the wave model. Therefore, no adjustment of spectral shape is required, and the associated uncertainty is therefore avoided. The resulting dataset is therefore neither an observation-only gridded product nor an improved numerical simulation, but an offline fused SWH analysis constrained by both sources. This dataset effectively mitigates the undersampling

issue inherent to altimetry while preserving the wave variability signals captured by the altimeters, making it well suited for SWH-based wave-climate studies."

Minor comments

Lines 43-45: "...satellite based wave sensors and wave model of SAR or scatterometer..." sentencem probably need to be rephrased. Current understanding in this particular sentence is that "wave model" can be defined in multiple meaning, e.g. Generattion/type/empirical method to retrieve. Can authors write more precisely what did they mean using term of wave model sensor.

We thank the reviewer for pointing out this ambiguous wording. The term "wave model" was a typo and should have been "wave mode," referring to the wave-imaging mode of synthetic aperture radar. We also agree that the original term "wave scatterometers" was insufficiently precise, because conventional satellite scatterometers are primarily designed to retrieve ocean surface winds rather than wave parameters. We have therefore revised the sentence to explicitly refer to synthetic aperture radar operated in wave mode and spaceborne wave spectrometers, such as CFOSAT/SWIM:

"However, satellite altimeters and other spaceborne wave-observing instruments, such as synthetic aperture radars operated in wave mode and wave spectrometers (e.g., Surface Waves Investigation and Monitoring instrument onboard China-France Oceanography Satellite), sample the ocean only along ground tracks or at spatially discrete footprints, resulting in sparse spatiotemporal coverage."

Lines 64-65 "...higher accuracy of satellite altimeter observations." Another question for auhtors: what is the measurement error according to the type and the manner by which the data is collected or stored?

We thank the reviewer for this comment. Validation of calibrated and denoised altimeter SWH products against offshore buoy measurements generally reports RMSEs of approximately 0.2 m. We have added this error information to clarify the stated accuracy of satellite altimeter observations:

"A more promising approach for generating a high-quality SWH dataset is to combine the seamless spatiotemporal coverage of NWM hindcasts with the higher accuracy of satellite altimeter observations, which generally have an RMSE of approximately 0.2 m when validated against offshore buoy measurements (Jiang, 2023)."

Lines 67:70: Besides the fact that ERA5 wave model assimilation of along track data with optimum interpoaltion method (of 5 satellite missions out of 10 that are in operational use nowdays) there is WAVERYS (MERCATOR) https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_WAV_001_032/description that also assimilates satellite altimetry products (such as family of missions named after Sentinel and SAR). Do you agree with the statetment that for the purpose of validation numerical models versus observations: already assimilated satellite measurements could not be used for postrpocessing output as it consists of data fusion (model+observations)?

Wave reanalysis products such as ERA5 and WAVERYS, which have already assimilated satellite altimeter observations, are not suitable as background fields for the present offline fusion with altimeter data. Otherwise, the same or closely related observational information would be used twice, making it difficult to isolate the contribution of the proposed fusion and potentially introducing dependence between the background field and the observations. For this reason, we use a WAVEWATCH III hindcast that does not assimilate wave observations as the background field. We have clarified this point and added WAVERYS as another example of an assimilative wave reanalysis:

"Another example of wave reanalysis dataset is the WAVERYS, which also assimilates SWH observations from historical altimeter missions (Law-Chune et al. 2021). However, the "online" assimilation methodology in ERA5 and WAVERYS presents several limitations when used for historical SWH data reconstruction"

"Assimilative wave reanalyses such as ERA5 are not suitable as background fields for the present study, because subsequent fusion with altimeter observations would lead to repeated use of the same observational information. Therefore, the SWH data from a WAVEWATCH III (WW3) hindcast based on the source term package 6 (ST6), without wave data assimilation, is adopted as the background field."

Fig 4, 6 should be rearanged and colormaps should be enhanced especially to repeat the same colormap for RMSE, CC. Bias on Figs. 4-7 is not meaningful, better enhance the limits.

We thank the reviewer for this suggestion. We assume that "Figs. 4 and 6" refers to Figs. 4 and 5, because Fig. 6 does not contain RMSE or CC panels. Within each figure, identical colorbar limits are used for the corresponding panels to allow direct comparison before and after fusion. However, we think the colorbar limits are not necessarily identical between Figs. 4 and 5 because the two figures represent different comparisons. Figure 4 evaluates the original and fused fields against independent altimeter observations, whereas Fig. 5 compares results obtained using two different numerical hindcasts as background fields. The magnitudes and distributions of the corresponding errors are therefore different. Using limits adapted to the data range in each figure provides better visual contrast and more clearly reveals the spatial error patterns. The same colormap type is retained for each metric to ensure visual consistency.

We are also not entirely certain how to interpret the statement that "Bias on Figs. 4–7 is not meaningful" and the suggestion to "enhance the limits". We respectfully consider the bias information in these figures to be meaningful for different purposes. In Fig. 4, the comparison of bias before and after fusion demonstrates the substantial reduction in systematic errors achieved by the proposed method. In Fig. 5, the bias between the results obtained using two different background hindcasts is used to evaluate the sensitivity of the fused product to the choice of background field. In Fig. 6, the differences between the fused and observation-only gridded products illustrate that satellite undersampling can introduce substantial systematic deviations. In Fig. 7, bias is reported together with the other standard error metrics and confirms that the remaining systematic error is small.

We have therefore retained the bias results in these figures. We would appreciate further clarification if our interpretation of the reviewer's comment is not correct.