Summary:

This study constructed an AI model for predicting the significant wave height (SWH) parameter globally using a convolution neural network with the U-Net architecture. The AI SWH model is trained on 18 years of ERA5 reanalysis by using the SWH and the 10-m surface wind vector fields at two consecutive times (i.e., rolling prediction strategy). Therefore, the AI model "simulates" SWH in a manner similar to the numerical wave models with an initial SWH field and the forecasted 10-m wind fields. Evaluation of AI SWH model performance in 2020 shows that this AI SWH model performs as good as the WaveWatch III model with the ST6 physics. The global error patterns against ERA5 SWH and CCI-Sea State analysis product further show that the AI-SWH model produces more reliable SWH prediction in wind-sea conditions than in swell-dominant conditions. The authors conclude that this AI SWH model can be a more efficient approach to produce global forecast of significant wave height than traditional numerical wave models.

General Comments:

Introduction:

My impression is that the introduction somewhat overstated the powerfulness of AI model or AI SWH model. It is true that the numerical wave models have limitations in parameterizations of the wind input term and the dissipation term that govern the spectral evolutions. But I don't think the AI model are completely free from these limitations since it learns from ERA5 and inherently adopts those limitations the authors stated. I suggest the authors toning down a bit this aspect when writing about the advantages of the AI model and not giving an impression that the AI model alone could overcome the physical limitations of the numerical wave models.

Thinking about the results from a more physical perspective:

It is quite interesting that the AI model is skillful in predicting the SWH associated with wind-seas. I am just curious if this means that the AI SWH model has learned some physics of the wave evolution. Could the authors comment on whether this AI model be run in an idealized setup to produce the SWH of fetch-dependent wind waves under constant and uniform wind forcings at different wind speeds? Would the relationship between SWH and U_{10} in this AI model (i.e., $(SWH\text{-}U_{10})_{AI}$) behave similar to some empirical relations between $U_{10,}$ fetch, and SWH? For example, for fully developed seas, I think the authors can compute the SWH associated with the Pierson-Moskowitz spectrum at different wind speeds and obtain a SWH-U_{10} relationship predicted by the Pierson-Moskowitz spectrum (i.e., $(SWH\text{-}U_{10})_{PM}$). For fetch-dependent seas similarly, $(SWH\text{-}U_{10})_{JONSWAP}$ can be found for different fetches.

Specific Comments:

Methods:

1. How long does it take to train this AI Model on 18 years of data? Would it be fair to mention this training time as well?
2. Would the results changes if testing was conducted using data from 2018, 2020, and 2021 together? Have the authors tested how sensitive this model is to different ratios of the training data, the evaluation data, and the model testing data? Can authors provide some answers to these questions in the method?
3. Did the authors perform some model tuning based on the evaluation dataset? if so, it would be great if the authors document what parameters have been tuned using the validation set from 2022.
4. Also, it is not very obvious to me how or why choosing 2022 for validation can prevent over-fitting. Could the authors demonstrate that this AI model is not overfitting in some way?

Results:

1. Figure 2: With data assimilation, why do the time series of the 4 error metrics have a zig-zag pattern?
2. Figure 4: Do the spatial distributions of the 4 error metrics change in different seasons?
3. By focusing on analyzing results after the errors stabilize, do the authors imply that this AI SWH model is more suitable for wave forecast beyond 10 days (240 hrs) without data assimilation and beyond 3-4 days with data assimilation?
4. Although the authors acknowledged that this paper does not compare with in-situ observations, to showcase the effectiveness of this AI model, I think it can still be worthwhile to compare the AI SWH model, WW3-ST6 hindcast, and ERA5 reanalysis, against a few in-situ buoy observations in the manner of a short time series at some key locations (e.g., some key swell-dominated locations versus wind-sea dominated locations) or weather conditions (e.g., westerlies or more uniform wind conditions versus tropical or extra-tropical cyclones).

Discussion:

It will be helpful if the authors can be more specific about the suitable applications with the AI SWH model. (e.g., time scales of the operational wave forecast, locations, seasons etc.)