Dear Dr. Ciccon,

Thank you very much for your review of my paper. I really appreciate your work and I am thankful that you took the time to read and comment on my work.

Below are concise, point-by-point replies to this review. First, my general ones are:

- We respectfully disagree that Hmax from models and from buoys cannot be compared. It is standard in the literature to compare buoy zero-crossing Hmax with the model's expected maximum envelope height ⟨Hmax⟩, provided the definitional caveat is stated; we now do so in Methods and figure captions, and we cite prior work using this comparison. Other work that has done this include:

  Barbariol, F., Bidlot, J-R., Cavaleri, L., Sclavo, M., Thomson, J., & Benetazzo, A. (2019). *Maximum wave heights from global model reanalysis.* Progress in Oceanography, 175, 139–160

  Wang, J., et al. (2024). *Performance of WWIII in simulating the ratio of maximum to significant wave height in the China Sea.* Ocean Engineering.

  Benetazzo, A., et al. (2021). *Towards a unified framework for extreme sea waves from space-time extremes.* Ocean Engineering.

  Davison, S., et al. (2024). *Characterization of extreme wave fields during Mediterranean tropical-like cyclones.* Frontiers in Marine Science.

  Cavaleri, L., et al. (2022). *The 2015 exceptional swell in the Southern Pacific: generation, advection, forecast and implied extremes.* Progress in Oceanography, 206, 102840.

- We will include more information on the inverted r peaks' scope and false positives.

Please know that following Dr. Bidlot's guidance I've done a new version (haven't sent it yet) with the changes below:

- I've reprocessed the ECMWF dataset and re-did figures 2, 3 and 4. They were corrected and look very different than before.
- I've applied the same Hm0 > 1 m cutoff to ERA5/ECMWF.
- I've added explicit notes in figure legends about FOWD filtering.
- I've added a section/paragraph on different Hmax definitions (zero-crossing vs envelope).

Below are the line-specific comments responses:

1. In 22 (2Hs vs ~2.2Hs). Agreed—2Hs is conservative and departures from Rayleigh are often noted around ~2.2Hs; however, this is the most common literature definition of rogue wave used, and that is why we used that. Some examples are as follows:

   Dysthe, K.; Krogstad, H.E.; Müller, P. Oceanic RogueWaves. Annu. Rev. Fluid Mech. 2008, 40, 287–310.

Baschek, B.; Imai, J. RogueWave Observations Off the US West Coast. Oceanography 2011, 24, 158–165.

Clauss, G.F.; Schmittner, C.E.; Hennig, J. Systematically varied rogue wave sequences for the experimental investigation of extreme structure behavior. J. Offshore Mech. Arct. Eng.-Trans. ASME 2008, 130, 021009.

Garrett, C.; Gemmrich, J. Unexpected Waves. J. Phys. Oceanogr. 2008, 38, 2330–2336.

2. ln 91 (skewness, kurtosis, BFI not viable predictors). Agreed. We now treat these as descriptive diagnostics of non-Gaussianity rather than as predictive variables for real-world rogues, and we will tighten the wording and add citations noting their limited skill.
3. ln 121 (FOWD name). Correct—Free Ocean Wave Dataset. We will correct everywhere.
4. ln 166 (“smaller” → “smallest”). Thanks—will fix.
5. Fig 1 (colormap + define seasons). We used this colorbar because it had more options of colors. I will check different ones and see if they make it easier to see.
6. ln 265+ (winter bias vs Hm0>1 m filter). Point taken. To address potential sampling bias in the next revision
7. ln 270 (“…to nonlinear interactions that form rogue waves”). We will rephrase to avoid implying modulational instability as the sole cause; we will attribute winter dominance to high-energy, steep, multi-modal seas and group dynamics without asserting a specific nonlinear mechanism.
8. ln 281 (“unstable sea conditions… favourable”). We will define “unstable” (high steepness, rapidly evolving/multi-modal spectra, shifting directionality) and add supporting citations; if needed we’ll soften the claim to “consistent with conditions often associated with extreme crests.”
9. ln 331–345 (whiskers vs rogue probability). Agreed—the whisker plots do not by themselves give rogue probability; we will confine probability statements to Hmax/Hs and explicit occurrence metrics.
10. ln 349 (symbol r used twice). We will reserve r for crest–trough correlation and use ρ for Pearson correlation in scatter plots (or label “corr.” explicitly).
11. ln 438 (add citations for asymmetry/peakedness preceding rogues). Agreed—we will add citations supporting that elevated skewness/kurtosis are consistent with steeper, more asymmetric groups in observed rogue environments.
12. ln 483 (FOWD vs raw CDIP discrepancies). Good catch—we will re-audit time-window alignment and clarify that FOWD uses multi-window processing (10/30 min + dynamic), which can make rogue signals more/less pronounced relative to single-window raw reconstructions. We will correct the text to match Fig. 5 and footnote any residual differences.

13. In 535 (no fixed r=0.6 threshold). Agreed. We will avoid implying a fixed threshold; we emphasize instead the dynamic r drop-and-rebound as our proposed early-warning signal, consistent with probabilistic framing.
14. In 596 (“correlation-based” → “crest-trough correlation-based”). Will change phrasing accordingly.
15. In 597 (do not compare 0.72 predictive power to our 75.3%). Agreed—these are not directly comparable (different definitions/targets). We will remove any numerical comparison, recast 75.3% as a recall statistic for our detector, and report precision/recall with false-positive rates.

Best Wishes,

Laura