Review of “Recent changes in extreme wave events in the Southwestern South Atlantic”

This study investigates the wave mean and extreme climate in the South West South Atlantic ocean (SWSA) and the potential associated coastal hazards from 1993 to 2021, focusing on a portion of the Brazilian coast. To do so, the author first used the outputs of regional wave datasets from the CMEMS and a historical coastal hazards dataset from the The Baixada Santista Coastal Hazards database (BDe-BS). The regional wave climate is investigated using several parameters (Significant wave height H_s, wave energy period Te, wave power, number of extreme events and intensity) for both mean regime and extreme wave climate (Hs 95th percentile). The seasonality is also investigated with a particular attention on the winter period during which most intense wave events occur. Despite the limited period of study (29 years), significant and valuable results on trends are found for both the wave climate and coastal hazards, especially for the number of extreme events.

Overall, I found the general idea and workflow coherent, and the use of 2 different dataset very valuable. This study provides valuable results promoting the need of further research for coastal disaster prevention. Though, before publication, a significant number of clarifications on the methods is needed, extra efforts to link the wave climate to costal hazards is expected, along with an exhaustive review of the whole manuscript concerning Figures (both format and captions) and general writing.

For the following reasons, I reckon that the manuscript should be considered for publication after undergoing major revisions.

General comments:

- Although not a native speaker, I suggest checking for the whole manuscript for syntax and English writing. Also, many times there is a lack of connection and logic between statements (that alone are generally very true) within the same sentence. Though the article is fully understandable, the high number of technical corrections needed results in a major revision. Related comments are enumerated below in the technical and small corrections part (corrections carefully noted only until line 160 of the article).

- In general, there is lack of clarity on the CMEMS data used and on the methods used to compute for instance Hs 95th percentiles (from hourly Hs time series or independent local events?), yearly time series of spatial average of specific areas (results showed in Figures 3 and 4), or trends from Sen’s slope and parametric linear regression… Also, the comparison of the number of extreme events obtained from CMEMS database and BDe-BS (Figure 4c) is not very coherent, another analysis could be done, as further proposed. Related comments are chronologically presented in the Specific comments section. In the case all clarifications solve my doubts or misunderstandings, this would be a minor comment, otherwise a major revision is expected.

Specific comments:

1. Line 55-71, 2.1 Datasets: What is the point using these 2 CMEMS dataset? Is it because WAVERYS misses the year 2021? You state, "so a more consistent analysis can be achieved despite using different sources", and further compute statistical indicators (such as the 95th percentiles) with no explanation of the actual data used (Figure 1, caption: "based on the CMEMS hindcasts", what data are you plotting?). Furthermore, combining data from a 1) global reanalysis and a 2) global analysis and forecast from the same service (CMEMS) is a bit redundant.

2. It is not clear if you are using Hs independent events or hourly Hs time series when you compute H_s 95th percentiles and the number of extreme events above the 95th percentiles. Line 75: "… the Hs peaks …", it could be easier to define a Hs^{peaks} to further in the article refer to Hs independent events in an efficient way (e.g., line 77, "Hs distribution" is confusing), because percentiles computed from H_s time series or H_s peaks can give very different results… Another way would be to specify that further in the article, H_s 95th values will refer to percentiles computed from H_s independent events, while mean H_s corresponds to average computed from hourly H_s time series.

3. Lines 76-80: This part is not very clear, maybe you could give more details of the method used to eventually obtain the H_s 95th values used in Figures 1,2 and 3. From what I understood, you first compute the monthly H_s^{peaks} 95th percentile at every grid point of the study domain. From these values, the average seasonal and annual percentiles are computed (here, you could precise how you defined your seasons, SON, DJF…). Finally, at every grid point, your time series (1993-2021, 29 values) are the yearly percentiles computed independently of the seasonal variability, and yearly percentiles for each season.

4. Lines 83-89: I understand the method using Sen's slope to find trends and if they are statistically significant with a Mann-Kendall test. You have a trend +- error, and that is significant or not. Then, it is not very clear why and what are doing by computing a parametric linear regression analysis. First, as you mention lines 197-198, it assumes a normal distribution, which is not true for extreme wave event distribution. Then, you are for instance applying a bootstrap method (n = 1000) to a set of 29 yearly values of Hs 95th percentiles. How do you compute de confidence intervals (CIs)? Figure 3 shows constants errors but Cis not constant, with a narrowing at mid-term of the whole period.

5. Line 97: Here, you explain how the duration of each event was computed, following *Weisse and Günther (2007)*. Line 48, in the introduction, you also mention the duration, but you never further investigate this characteristic!

6. Lines 98-100: You just defined the duration of each event, and the intensity associated to this event, that is the difference between the SET value and the maximum value during this event. Following *Weisse and Günther (2007)*, you should clarify that you will then use the mean intensity, i.e., the average of the intensity of all individual events (above SET) within a year, both independently of seasonal variability and for each season.

7. Line 182-185: “The extreme event … warning subareas (Fig. 3).” I assume that you want to point out the benefits and limitation of such approach to perform coastal hazards assessment, to introduce the complementary focus on specific subareas (A, B, C, D) where you have a “terrain” database from the BDe-BS. Instead of “Due to that”, I think it should be “Therefore”. Moreover, it is not because of reanalyses derived-sparse results that you analyse the trends in A, B, C, D (previous results do not show higher statistical significance there), but because you have data there!! Finally, for each area, I assume that you work with spatially averaged trends (you should also explain it along with the associated spatially averaged errors), but what do you mean by “most relevant trends”? You compute spatial trends for each area and parameter and then find if they are statistically significant, but ALL trends are analysed. It is therefore quite hard to understand the reasoning here and I recommend modifying these lines to be clearer.

8. Lines 213-231: The trends calculated (from the reanalyses and the BDe-BS) are indeed similar in the subareas C. But as you say, yearly number of events from both databases often differ with a clearly higher number of wave event for the reanalyses-derived data than coastal hazards from the BDe-BS. This result raises the same concern than the comment 35 (“because percentiles computed from H_s time series or H_s peaks can give very different results”). Let’s assume that for the number of extreme wave events over the Hs 95th percentile, you used the distribution independent events (2 days’ time-window), H_s^{peaks}. Therefore, in theory, the maximum number of independent events within one year is roughly 183. This is the distribution of wave independents events. However, you are working with EXTREMES wave events corresponding to events with Hs values higher than the SET value (Defined by the 95th percentile). So, from 183 events, 5% are above de 95th percentile, ≈ 9 extreme event per year at most, when Figure 4c shows that the number of extreme wave events ranges from 23 to 36. So, it means that, though you used the independent events distribution (H_s^{peaks}) for the 95th percentile H_s values, the number of extreme events was computed from the 95th percentile of hourly Hs time series.
   So, to conclude: 1) I think that it is confusing and, in agreement with comment 35, you should clarify whether Hs hourly time series are used or H_s^{peaks}. 2) except for trends, the number of extreme events obtained from the 2 the reanalyses and BDe-Bs are not comparable with this method. For instance, if you would chose the 99th percentile instead of the 95th, the number of extreme wave events detected would be lower and more like the number of extreme events given by the BDe-BS (you actually say that you could use any percentile between 90 and 99, Lines 94-95). In my opinion, I would only compare trends with this method. However, it could be interesting to assess the capacity of the reanalyses to reproduce extreme wave events that are documented by the BDe-Bs, giving an idea of the typical wave forcing leading to coastal hazards in the subarea C. I am aware of the extra work that it represents, but it is an opportunity to make the link between “the regional wave climate and with coastal hazards” (Line 201).

9. Lines 223-231: On the assumptions of differences observed between reanalyses-derived extreme wave events and coastal hazards from the BDe-BS. “Maybe these

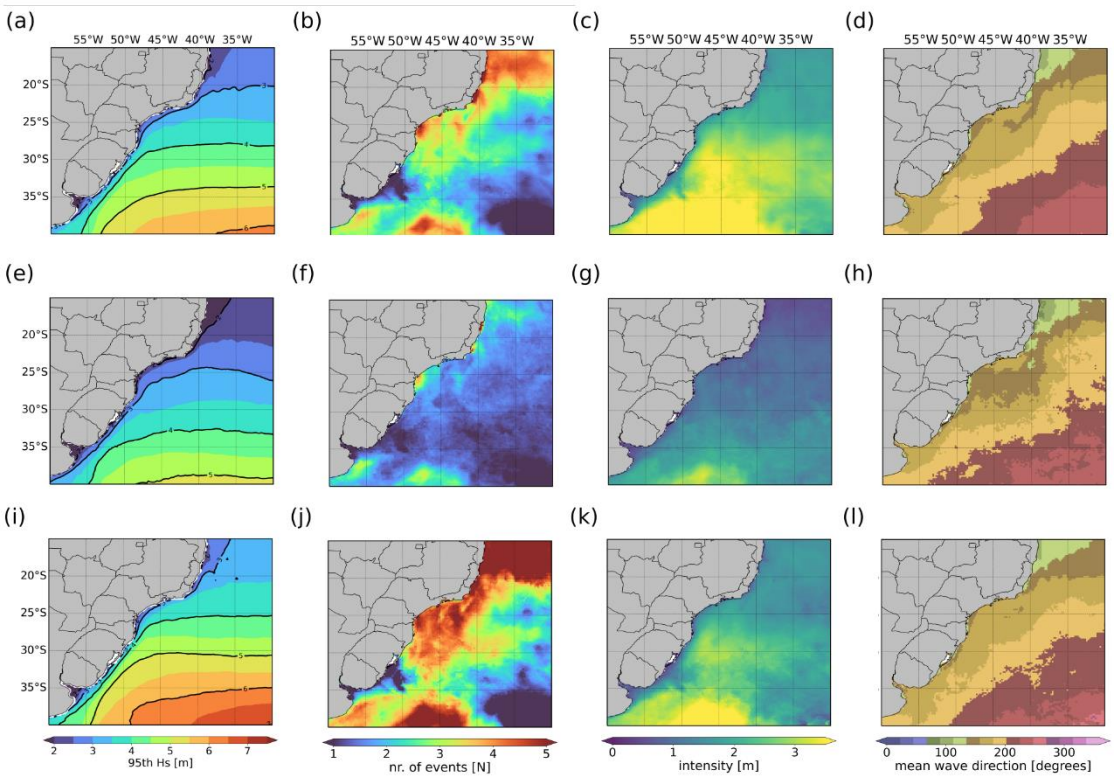
wave events would not become a hazard if the local sea level rises did not allow waves to reach further into the continental area": this is indeed a potential consequence of sea level rise. However, the hypothesis of sea level rise explaining the difference observed only from the absence of peaks in 2002 and 2009 in the reanalyses-derived number of extreme wave events is off topic. It is a long and large-scale effect that has a low impact one year to another. Also, it occurs over the whole period, not only during two isolated years, and the provided data here to make such assumption is simply not enough. Finally, the number of extreme events is not necessarily related to the intensity of these events. A year with a low number of extreme wave events but powerful ones can lead to a higher number observed extreme event at coast than a year with a high number of relatively weaker extreme wave events.

10. The results show significant and high trends. Line 220: for the subarea C, maybe you could take advantage of giving an example of the consequence over 29 years, in order to give more weigh to your findings. For instance, 0.20 event/year increase looks small, but over 29 years it represents 5-6 more events per year in the subarea C that shows between 24-34 events per yaer (~20% increase !!).

Figures comments:

Overall, all figures need to be modified. First, I suggest removing Figure 4.a, and adding as a first Figure a plot that describes the study area. For instance, you could provide a Figure showing the whole area where you are using CMEMS data, and within this plot add a zoom on the A, B, C, D zones. It is a bit disturbing to provide a figure of the area of study at the end of the paper. Further comments on figures:

1. Figures 1,2,3: Overall, you have a high number of subplots. To gain space, when the x,y label and ticks are common to several subplots, only show it on the edges, same for the colorbars. This way you can narrow down the space between subplots. You may then increase the fontsize of labels and ticks (with same size for all, e.g., coordinates have not the same size). Also, putting directly on the figure "whole period, summer, winter" make it easier. For instance, you could have something like the following figure:



2. The colorbar of Mean wave direction (Figure 1) should be the circular (e.g. “hsv”)
3. I suggest to modify the legend or Figure 1 “~~Annual/Seasonal mean~~ Mean of annual ~~(a,e,i)~~ 95th-percentile Hs (m) values (a,e,i) ~~(m)~~, extreme wave event (b,f,j) number, (c,g,k) mean intensity (Hs - 95th Hs) (m) (d,h,l) mean direction (degrees) ~~in~~ for the (a-d) whole period of study (1993-2021), … ”
4. Line 185: 1) Figure 3 misses 2 subplots (95th percentile). 2) For the whole paragraph describing results from Figure 3, I suggest adding “(see Fig Xx)” for every observation, e.g., Line 187 “In the winter, the A and C subareas presented significant trends in the number of events per year (see Fig. 3b).” 3) Hs 95th trends are not described, probably because the corresponding figures are missing.

Technical and small corrections:

This is not an exhaustive list of needed corrections. Because of not being a native speaker, some comments/corrections might not be adequate, but still the manuscript should be carefully reviewed.

1. Line 4-5: “... the occurrence of ~~wave~~ extreme wave events but also other extreme wave indicators that may impact the ~~off-shore~~ offshore and coastal areas.” Extreme wave indicators don’t “impact” offshore and coastal areas, extreme waves do. Rephrase with something like “… but also extreme wave indicators that characterise the potential wave impact/hazards on offshore and coastal areas.”
2. Line 5-8: “For a more direct application … impact the coastal zone”. More direct than what? I suggest moving this sentence after the part describing the CMEMS data you

are using, since you first perform a regional assessment of the wave climate, and then link your findings with coastal hazards referenced by the BDe-BS.

3. Line 5-6: need to rephrase “For a more direct … a more focused …” because it is repetitive.
4. Line 9: it would be nicer to write H_s instead of Hs for the whole paper, as well as 95^{th} instead of 95th.
5. Line 10: “… statistics are ~~used~~ performed to analyse ~~the~~ wave mean and extreme climate patterns …”. I would rather write “mean and extreme wave climate”, “mean wave climate”, “extreme wave climate” here and then.
6. Line 13: if H_s values, then wave period values, wave power values … For instance: “… in mean values of Hs and wave period, and consequently wave power.
7. Line 24: I suggest removing “thus reflecting directly on hazards along the coast” because it is a bit a standard phrase that is not needed.
8. Line 25: extreme wave ~~occurrence~~ and storm surges occurrence?
9. Line 26-27: Rephrase “the SWSA … exploration fields”. “Which is” refers to the SWSA, as if the ocean is responsible for the transportation … For instance, rephrase: “… where 755 million tons of goods were transported…”
10. Line 30: order chronologically the references (ICMBio, 2018; Pereira-Filho et al., 2021); Check the whole manuscript (e.g., Line 243).
11. Lines 33-35: “The difficulties remain mostly in the still-limited knowledge and understanding of the local physical processes and climate variabilities …” What unknown local physical processes can prevent from assessing wave trends? What climate variabilities? Following “for several reasons.”, the reader expects a clear explanation or enumeration, but the reasons given are very vague.
12. Line 36: remove “In general” or “most”
13. Line 37: “… this ~~increase~~ tendency”?
14. Line 39: “… present ~~a~~ larger uncertainty~~y~~ies.“
15. Line 40: In addition to understanding the significant wave height (H_s) trends
16. Line 42: “… at both ~~the~~ regional …”
17. Line 43-44: Specify the period you are mentioning. Do you mean the changes between the present climate conditions and future scenarios? Or it is the changes that already occurred under the present climate (1993-2021), and other changes in the future (e.g., towards the end of the century)?
18. Line 45: “… coastal risk assessments …”. The connection between the two parts of the sentence is not very clear to me and vague. What do you mean by linking the regional to local wave extremes? Regional scale would be the SWSA and local scale an area like you define later (A to D)? or beach scale? “… future scenarios … demand special efforts …” Why, knowing the future wave climate, would it be harder to link the regional and local wave extremes than for the present climate?
19. Line 47: “the recent wave climate”, specify the exact period of study
20. Line 50-51: same comment than for 3.
21. Line 55: “global ~~hindcast~~ reanalysis “, you have the wave reanalysis WAVERYS and a global analysis and forecast product (GLO-NRT)
22. Line 55-57: For the reader, it is a good idea to sum up the products used in a table. Though, too little information is given. I suggest including the main characteristics of the CMEMS product, such as temporal and spatial resolution, temporal coverage, wind forcing, and available data for the BDe-BS (the table indicates the period ~~1993-2021~~ 1928-2021)

23. Line 57: write 1/5º instead of 0.20º to compare with the following 1/12º resolution.
24. Line :60 “Both products are produced ~~by the wave model called~~ using the Météo France Wave Model”
25. Line 61: “WAVERYS is forced by hourly? surface winds and sea ice fraction fields …”
26. Line 65: “… western South Atlantic ~~American~~ wave climate…” as said in Crespo etal. (2022).
27. Line 65: “The~~se~~ authors compared ~~the~~ Hs …”
28. Line 67: “… against wave buoy measurements ~~in~~ at three locations…”
29. Line 68-69: “The ERA5 performance in representing the winds is also relevant once this field dominates the wave generation process”, I don’t get the point of this sentence, ocean surface wave generation is always initially driven by winds. Do you mean that the ERA5 performance is relevant for young-sea state while not crucial for swells that travelled away from their generation area?
30. Line 70: I suggest removing “even”, because it suggests that it would be okay if it did not reproduce storm variability, or that is something not expected.
31. Line 78: “… next section …”, aren’t you referring to section 2.4 instead of the next section (2.3)?
32. Line 103, suggestion: change the section title to “Extreme wave event analysis”
33. Lines 106: Wave power: you are using the mean wave period T_e directly available from CMEMS; In Figure 2, peak period (T_p)trends are showed; You should use the same. If you are using the T_p, the wave energy flux P (equation 1) is a bit different (see *Guillou, 2019, https://doi.org/10.1016/j.renene.2020.03.124*). For instance, using T_p, you should replace in equation (1) T_e by αT_p, where $\alpha = 0.9$ for a standard JONSWAP wave spectrum.
34. Line 106: Same comment as for 5., Hs,Te --> H_s, T_e.
35. Line 114: Specify that MetArea V corresponds to Atlantic waters west of 20ºW from 35º50’S to 7ºN.
36. Line 117: the results found here to improve future monitoring and warning system development ~~and improvement~~.
37. Line 120: I suggest changing “the hemerographic method” for “a hemerographic literature review”.
38. Line 122: Remove “Thus …”, what is said in the previous phrase does not imply the definition of a coastal hazard.
39. Line 123: ~~This later~~, the latter.
40. Line 137: “… controlled by the storm track due to the strongest winds associated with the cyclones”, I suggest rephrasing “… controlled by cyclones storm tracks and associated strong winds”
41. Line 138: “The main storm track position, between … characterises”, I suggest rephrasing “The storm tracks that are mainly confined between … characterise”
42. “… wave height distribution ~~by the~~ with a Hs gradient towards the south …”
43. Line 142: “storm tracks”.
44. Lines 140-144: I think you should first present the results and then provide an explanation about the differences observed, otherwise it is a bit confusing. 1) what differences are observed between winter and summer (direction, 95th Hs, nr of events and intensity) and 2) why? (Southward shift of storm tracks).
45. Line 145: correction needed “… while spring (SON) remains the winter pattern”
46. Line 146: need to rephrase “For these reasons we are going to focus on the analysis of the winter since it heads the top list of extreme wave events in the study domains”

47. Line 153: Fig 1~~1~~ b,j
48. Line 160: I suggest removing “Starting from the basic extreme statistics …”
49. Line 160: “present~~ed~~s”, use the same tense for the whole paper. You could sometimes use “show” for instance instead of always using “present”
50. Line 168: same as line 160.
51. Line 175: As you did earlier for the Hs trends, and since you observe a small but significant wave power increase, you could give the magnitude of a significant wave power increase.
52. Lines 180-181: aren’t statistics more robust when the trend is significant? Whether for the Hs 95th or the number of events plots, the trend is overall significant at a higher of location for the whole period than for the winter season.
53. Line 190 “… high interannual variability”
54. Line 197-198: “… a normal distribution required by the parametric test.” Then again, what is the point using parametric linear regression analysis?
55. Line 203-204: the percentage values are not coherent “… 48% (78) … 30%(~~79~~49) ...”
56. Line 205: What do you mean by “… the same proportion (35%) of total events”, same than for the whole period?
57. Line 206: “High ~~wave generation~~ wave events frequency”
58. Line 207: “Table 1 presents …” move this sentence before starting to mention data from the Table 1 (before Line 203)
59. Line 219-220: “~~moreover, it is possible~~ …” remove this because you directly compare both trends! Also, since you have the 0.20 positive trend from Fig. 3a results, you can indicate the trend value for coastal hazards data.
60. Lines 243-244: First and second parts of the sentence are understandable alone, but lacks connection, I suggest rephrasing.
61. Line 248: You did not elect the winter season as representative of more extreme wave climate, you focused on winter because it shows the most extreme wave patterns!