Thank you for the thoughtful comments from each of the two reviewers. We have carefully considered each comment and below outline the changes we have made. We believe these changes have strengthened the manuscript and so thank each reviewer for their time.

REVIEWER #1 COMMENTS

1. Figure 1: change "model location, with location of" to "model domain, with locations of". Also, I think there is a typo "at ~ 250 km depth", it should be "at ~250 m depth".

A1. Thank you, we have made this change.

2. Table 1 should be moved to Section 2.1 for discussion. Also the title of Table 1 indicates values reported are for tidal constitutes, which constitutes?

A2. Thank you for spotting this discrepancy in the caption; we have now corrected it to read '*Validation of sea levels output by the model'*. We did not in the end move Table 1 to Section 2 (The hydrodynamic model), but instead inserted it into the newly created 'Data and Methods' section, where it is now first mentioned.

3. I have some concerns on the value reported in the Table, it is very questionable that the "mean tidal range" is greater than 8.5 m for Xiamen, 7 m for Beihai, 5.9 m for Vong Tau. Figure 3 clearly showed that the max tidal range at Vong Tau is less than 4 m. Please check and clarify.

A3. We thank the reviewer for pointing out this inconsistency and we shall examine these calculations more closely. We decided however, for this paper, to remove this column entirely from the table as it wasn't strictly adding a great deal towards the core message.

4. The model resolution of 11 km along mostly the coastal, with finest of ~2.3 km along the coastline of Vietnam, is not sufficient to represent the details of the shoreline and bathymetric feathers, which are critical for storm surge simulations. Discussion on model accuracy in estuaries and bays is necessary to clarify that the model results should be treated with care in those areas.

A4. Thank you for your comments and we understand your concerns. We do agree this model resolution is not adequate to represent storm surge detail around estuaries and bays and as such have added text to the discussion section to acknowledge that this resolution does not capture the level of detail in these areas. The discussion section now states, '*It should be noted that there is potential for sub-optimal accuracy in storm surge levels, particularly within small coastal features such as inlets, bays or estuaries, where coastal resolution is insufficient to capture features in detail. For example, Bertin et al. (2015) showed that within small seas wave radiation can induce set up that transforms storm surge levels along exposed coastlines, with even the small waves entering bays and inlets affecting water levels*.'

There were a couple of reasons why we did not model at higher resolution along these shorelines. Firstly, we wanted to simulate 10,000 years of Tropical Cyclone (TC) behaviour, and to do this twice-over (once for the present day and once for the future scenarios). Each synthetic year normally also having multiple TC events, within our region of interest. The execution of all these model simulations would be both computationally expensive and time-consuming to get through. Therefore, we aimed to strike a balance between model resolution and duration run time. Given our regional focus, we felt 2.3 km was a good compromise that enabled us to capture as much of the TC information as possible. In the event, even with use of a High Performance Computer at the university, we spent around ~6 months to just run all the various models.

There is also the fact that this project was supported by UK and Vietnam Government funding. Our Vietnam colleagues do not have access to a High Performance Computer, and they wanted to be able to use the model for other applications. Hence, our second reason was our aim to keep the model resolution at a point that our project partners could run it on a standard desktop computer after the project ends.

5. Section 2.3 and Figure 4: storm surge validation is important for this study. It is not sufficient to use Figure 4 to demonstrate the Holland Model is better than ERA5. Model-data comparison at more stations and error statistics should be provided.

A5. A good point and thank you for highlighting this omission. We used TC Ketsana as an example because the event was really nicely captured as it passed fairly close to a tide gauge. As discussed within the paper a common problem is that TC events are difficult to capture on record, due to proximity or timing. However, a small handful of historic TC events were similarly close enough to a recording device to be used to determine the best methodology and validate model, and these additional TC events are now shown within the supplementary section. One such TC event is Typhoon Sally (aka Typhoon Maring) which struck Zhapo, China in September 1996 the figure of which shown in the reproduced in the below Amended Figures and Tables section (Figure S1).

A table showing the validation RMSE between observed and modelled results, for these total four historic TC events, is also now provided within the supplementary section (also copied in the attached Figures and Tables document). Furthermore, the text of the main paper has been updated to replace ‘not shown’ in Line 263 with ‘*shown in the supplementary section’*, to point to this new information.

6. Figure 5. Are the values shown in the figure the number of cyclones? Please provide unit in the legend.

A6. We are happy to update the legend of Figure 5 to make it clearer. The below Amended Figures and Tables section shows this update. Yes the values are the number of cyclone tracks passing through each gridded location, within the ~35 year assessment period. Figure 5c shows a heatmap of where activity is estimated to increase most intensely over time, under the CNRM 2015-2050 climate projection.

7. Line 449: statement “…that tides intensify storm surge hazard” implies the nonlinear effect of tide and surge interaction but in ~Line 336 the authors showed the nonlinear effect is negligible and therefore the total water level can be calculated by the sum of two separate runs: tide and storm surge. Please clarify.

A7. This is now corrected and reworded in Lines 449 and 499 - thank you for finding this inconsistency in the document. Using ‘tides intensify storm surge hazard’ was a poor choice of words since we proved that the nonlinear effect is negligible along the shorelines of interest.

8. How tidal elevation is added to storm surge to obtain the total water level since the synthetic TC does not have a realistic time and date?

A8. Thank you for your comment. It is correct to say that the database of synthetically generated TCs does not assign literal dates. However, there is a synthetic year (between year 1 and year 10,000) and a month associated with each generation. Our approach was therefore to (1) generate tide levels for the model domain, covering a recent 19-year period, to encompass the full 18.6 year nodal cycle. Then (2) assign each synthetic TC a new random year, day and time, but leaving it with the same month as generated. Then (3) combine the modelled storm surge and tide together taking care that the calendar month of the TC and tide were the same - ensuring the storm surge occurs (synthetically) in the correctly aligned season/time of year. This 19-year reproduction of tides and storm surges was afterwards used to generate statistics for total return period levels that are illustrated in Figure 9.

We have clarified this in the final paragraphs of Section 2.2.3 of the paper which now reads, *'… we also estimated total still sea level return periods for each OCP. Because non-linear interactions between tide and non-tidal components were determined to not be an issue for this region, we do this by adding surge levels to a semi-randomly selected tide. The first step to calculate total still sea levels was to run a tide-only model simulation (for the year 2009 where we had already obtained OTIS tide data) and to save modelled tide levels at each of the OCPs at 10-minute intervals. We then secondly input this year of detailed tide levels into the MATLAB T-Tide script (Pawlowicz et al., 2002) to obtain the tidal constituents and predict the tides over a longer recent 19-year period (2003 to 2021). A full 19-year period was targeted because it encompasses a complete 8.85-year cycle of lunar perigee and covers the 18.6-year nodal astronomical tidal cycle, both of which can influence extreme sea levels (Baranes et al., 2020; Peng et al., 2019; Haigh et al., 2011). The third and final step was to select a semi-random date from this 19-years of tide data that would match with a TC in the baseline and future datasets. The original past/present and future STORM datasets has TCs develop largely between May and November, as occurs in the natural record for this region. Our sub-sets of baseline and future TC data, derived from STORM, all have a simulated month and synthetic year assigned to them. Because of this it was possible to match each TC (surge) to the correct month in the tide data, preserving TC seasonality...'*

REVIEWER #2 COMMENTS

2 Major comments

2.1 Writing and the manuscript structure

In my opinion there is a large scope to improve the overall writing of the manuscript. Following is a list of comments -

2.1.1. Introduction was rather hard to follow. For example, the §1 opened with the future surge hazard, then immediately switched to current surge hazard in §2, and then revert back to future surge hazard in §3. My suggestion would be to reorganize it into §2 → §1 → §3.

2.1.2. Two method discussed at L75 and L80 are not same thing it seems. One of them extends data over century, another extends only on decade. The sequence and presentation of these paragraph reads odd.

2.1.3. L91-105: This paragraph is troublesome as it opens with synthetic storm, follows by reanalysis forcing, followed by synthetic data again and STORM dataset etc.

2.1.4. L135-140 should go to a dedicated Dataset section, (Beginning of section 3?) and their properties should be discussed. It is again a bit odd to read about data and method at the end of the Introduction section. Here I was expecting how you are going to structure the manuscript.

2.1.5. The data/model/discussion section also has similar structure-related issues. I stopped suggesting explicit updates from this point on. I suggest the authors to carefully revise the manuscript with readability (and story) in mind.

A2.1.(all). Thank you for the valuable feedback for the above 5 points relating to the paper's structure and writing style. We have reorganised the text and structure of the manuscript to improve its readability bearing the above comments in mind. We have added extra headings to distinguish the data and methods sections more clearly (points 2.1.4 and 2.1.5). We have also reorganised sentences and paragraphs throughout the document to make it easier to read, improve clarity and flow more logically (points 2.1.1 – 2.1.3).

2.2 Modelling and validation

2.2.1. L146: Is this model already published? If so, please provide past reference.

A2.2.1. The model was originally created several years ago by our project partners the Southern Institute of Water Resources Research (SIWRR) by our co-investigators Le Quan Quan, Nguyen Nghia Hung and Tran Ba Hoang. The model has been used in various applications by SIWRR but has not previously been published in an English Language International Journal. We also considerably changed the original version of the model by re-configuring the model grid, boundary condition and set up and conducting a more extensive quantitative validation exercise. Hence, this current paper under review here, is the first reference that describes the current model.

2.2.2. How did you choose the model resolution? For a regional study, a resolution of 2.3km (finest) seems rather low, as GSTM 2020 model discussed in the introduction has 2.5km at the coast. Additionally, in most cases the resolution is set to 5-7 km, which is twice as coarser than GSTM.

A2.2.2. Thank you for your comments – this was an issue raised also by Reviewer #1. The model resolution was determined as a compromise between computational efficiency, model accuracy and simulation run-time. While certainly the GSTM 2020 model is global and would have a greater computational cost at 2.5 km grid resolution than our regional model, we took the position that because we were running almost one hundred thousand simulations of TC movements (when including both baseline and future scenarios), that this range of grid-resolution was the best solution for this task. Additionally, it was a consideration for us that the MIKE 21 FM model should continue to be used by our partners SIWRR in Vietnam after project completion, on a regular desktop computer.

We have added a few sentences in the discussion section to acknowledge that this resolution does not capture shoreline detail in these locations. The full paragraph reads, *'The model has a variable triangular grid resolution, with greater detail along coastlines. It should be noted that there is potential for sub-optimal accuracy in storm surge levels, particularly within small coastal features such as inlets, bays or estuaries, where coastal resolution is insufficient to capture features in detail. For example, Bertin et al. (2015) showed that within small seas wave radiation can induce set up that transforms storm surge levels along exposed coastlines, with even the small waves entering bays and inlets affecting water levels. Unfortunately, the number of TC simulations entailed in this study meant that there had to be a trade-off between rendering coastal detail and reasonable computation timescales. Future work in such locations, looking at local sections of coastline, would require detailed modelling to estimate extreme sea levels due to storm surge. Additionally, currents and wave action is specifically not incorporated in the modelling as it was outside of the project scope. Such wave models will be a valuable addition to the scientific discussion, as their high spatial resolution at the coastline would ensure that nearshore wave dynamics, such as wave setup, are adequately resolved (Hinkel et al., 2021; Saulter et al., 2017)*'.

The model configuration section has also been updated with the prefix, '*To create a model which could be run around one hundred thousand times, but still offer fair accuracy, we selected grid resolutions at the model's coastal boundaries of …*'.

2.2.3. Is any wave model is incorporated? If so which one and how? The resolution issue is further problematic as Wave models needs metric (10s of meter) resolution to properly model the wave setup [1]. In case wave model is not coupled, then wave setup is not computed which can further augment the storm surges [2]. My understanding is that no wave model is coupled. Please provide discussion and adds in the limitations.

A2.2.3. No wave modelling was done. We can confirm that our focus was on still water levels to ensure the computational effort and runtimes remained manageable with the scope of the funded project. We have added a sentence to the model configuration section of the manuscript to emphasise this, it says, '*No wave modelling was carried out in this analysis either, since the focus of this paper is to be on still water level*'. Additionally, we have added text to the discussion section to note this limitation. The full paragraph is copied into the response to Comment 2.2.2 above.

2.2.4. What do you mean by updated bathymetry in L161?

A2.2.4. The bathymetry in the model was updated from the original one provided by our project partners. We utilised a newer SRTM15+ bathymetry which had recently been released by Scripps Institution of Oceanography on their website. Since this is unclear the old sentence '*For bathymetry, an updated 15 arcseconds resolution global dataset from SRTM15+ (v2) was used*' has been replaced with, '*For bathymetry, we replaced the data in the original model with a 15 arcseconds resolution global dataset from SRTM15+ (v2)'*.

2.2.5. What does it mean by '2016 tidal model solution' in L167?

A2.2.5. The Oregon State University Tidal Inversion Software provides astronomical tidal components for this region. We downloaded the latest output model (dated 2016) for the China seas and Indochina, and used this to create tidal levels for our model boundary locations using their proprietary TMD MATLAB toolbox.

Since it is unclear, the text around this has been modified within the manuscript as follows: '*Tides were only separately modelled for validation and to estimate total sea levels. To simulate tides we generated the astronomical tidal component within the domain using tide data obtained from the Oregon State University Tidal Inversion Software (OTIS, Martin et al., 2009; Egbert & Erofeeva 2002). The harmonic constituents were downloaded from the OTIS web site (http://volkov.oce.orst.edu/tides/) for the seven model open sea boundaries of our model. The data are provided for eight primary (M2, S2, N2, K2, K1, O1, P1, Q1), two long period (Mf, Mm) and three non-linear (M4, MS4, MN4) harmonic constituents. With these data, the tide was then predicted, for each boundary grid point, using the Tidal Model Driver (TMD) MATLAB toolbox (http://polaris.esr. org/ptm_index.html) with the 'China Seas and Indochina region (2016)' tidal model option.'*

2.2.6. L181-184: Annual mean sea level values fluctuate, and linearly increase. How do you make it equivalent to model datum (EGM96) by removing a fluctuating value?

A2.2.6. Thank you for rising this valid point. Indeed, annual mean sea levels do fluctuate from year to year and have increased over time. However, the data lengths and periods varied greatly among the 27 tide gauge records that we had access to in our region. Some sites only had a year or two of data. Hence, it was not possible to use a long epoch (for example 18.6 years covering the full nodal cycle) to compute a similar vertical datum at each site, and make this exactly equivalent to the model datum. As a compromise we found that the year covered by the most amount of tide gauge records was 2019. Hence, we simply extracted the annual mean of the tide gauge record from 2019 at each site. Where this year was not available, we took the next nearest year. In the latest version of GTSM (Muis et al., 2020), water levels are referenced to the present-day mean sea level by removing the annual average sea level for each year, and subsequently also by subtracting the mean over the last 19 years from the (de-trended) time series. Ideally, we would have followed the same approach, but as discussed above, the length of several tide gauge records in this region are short. There is a small (few cm) difference between the annual mean of 2019 compared to the mean over the last 19 years. This may cause a small bias in the validation statistics, but at least the bias is consistent among sites. In conclusion we felt this was the best approach, considering the sea level record data length limitations.

2.2.7. L184-185: Tidal analysis does not remove meteorological influences. At daily scale, for example, S1 tide is forced by diurnal atmospheric pressure loading. At seasonal scale, for example, Sa and SSa components are often influenced by long-term meteorology, freshwater input etc [3].

A2.2.7. Thank you for raising this point. We acknowledge that the tidal analysis does not remove all of the meteorological influences, such as the daily scale S1 tide, but it does remove the vast majority of meteorological forcing factors. Importantly, we apply exactly the same tidal analysis procedures to our measured tide from tide gauge records, to the time-series of total water level predicted by the model. Hence, at least these small biases are consistently compared. Furthermore, in the paper we focus mainly on the validation of the main constituents (M2, S2, O1, K1).

2.2.8. L185-187: If possible please compare for a consistent year.

A2.2.8. As we discussed above in regard to point 2.2.6, the 27 records do not all cover the same period, and at some sites, the data only covers a year or two. We choose to focus on January 2019, because this period coincided with when we had the greatest number of complete datasets available across all 27 tide gauge sites. Note, no single year had all 27 records available, so there was always going to be a compromise. Within the data availability, we do the best we can. Where the data for 2019 was not available, we considered the closest year.

2.2.9. Modelled tide and tide gauge data are inherently correlated (L200), hence validation using MAE and correlation fundamentally does not capture the error in the various 'waves' (e.g. tidal constituents). Please use complex error for validation and update Table 2 with complex error instead of MAE. You might also provide a map of the station with Complex Error for the 4 main constituents noted in L214. See for example Tanchant et al 2021 [4].

A2.2.9. Thank you for raising this point, we respectfully would prefer to stick with using MAE as this validation method has been used in a wide range of past papers (e.g. Muis et al., 2020 & 2016; Cavaleri et al., 2018; Wahl et al., 2017; Martin et al., 2009). The reviewer mentions that MAE does not capture the error in various tidal constituents; this is why we deliberately validated the total time series, in addition to considering the four main tidal constituents. We therefore don't feel that it is necessary to include complex error as well.

2.2.10. L239: is 0.25 degree enough resolution? Why it was chosen?

A2.2.10. The 0.25 degree resolution of the imported wind and pressure data from IBTrACS was originally decided on to match the ERA5 data it was to be compared with (Lines 243-245), in the validation exercise. The same template was carried over into the full simulations also (Line 321-322). Increasing the resolution had ramifications that it significantly increased the file sizes of the meteorological forcing datasets, across the 10,000-year storm datasets. We conducted a series of sensitivity tests and found that a 0.25 degree resolution appeared appropriate to capture the meteorological fields. Moreover, within MIKE 21 FM, the meteorological forcing is interpolated (within the software) onto the hydrodynamic flexible mesh, and so it doesn't matter that much what the resolution of the forcing files are, as it is then interpolated to the higher resolution of the hydrodynamic mesh. Additional text has been inserted into the beginning of this section to make this clearer. This states of the output cyclone wind files generated in MIKE 21 FM; '*This spatial resolution was sufficient to resolve the TC within these forcing files, especially as the wind/pressure files would be further interpolated in the MIKE 21 FM software to the higher resolution of the model mesh as the cyclone traverses through the model domain*'.

2.2.11. L231-235: A map of these tracks would be useful. This can be added to, for example, figure 2b.

A2.2.11. The track of Typhoon Ketsana is shown in the insert in Figure 4b. We totally agree with this comment, and have therefore included the tracks of the other TC events that were used for validation in an inset window within each of the new Supplementary Material figures S1-S3 too.

2.2.12. L264: Can you show the landfall time in Figure 4? How about other stations? Can those be added as supplementary materials?

A2.2.12. Thank you for this comment. We have added a green vertical line to both panels in Figure 4, to indicate when TC Ketsana made landfall (and repeated the same for the other TCs used for validation, in the Supplementary Material figures S1-S3, for other tide gauge locations). Only a handful of TC data could be used for validation as many stations only had short sea level records and captured relatively few landfalling TCs in reality. TC surge levels are generally poorly detected, even between neighbouring tide gauge locations, and so additional station data simply was difficult to obtain for the manuscript.

2.2.13. I could not find what was the model timestep used for tide, or surge simulation. Please provide this vital modelling detail.

A2.2.13. The MIKE 21 FM model set up uses a variable time-step. We have added new text to section 2.2.2 'Hydrodynamic model implementation' to say this. This reads; '*These steering files define the model solution technique, to integrate the time and space variables within the shallow water equations via an explicit scheme, utilising a variable time step interval in the calculation*'.

2.2.14. I also could not find any detail regarding the model friction formulation. Is it manning? Chazy? roughness length? What is the distribution? Regionally uniform or different? How it was chosen?

A2.2.14. Thank you for your comment on this omission. The answer is that a default Manning number was used over the whole domain, with constant value 32 $m^{1/3} s^{-1}$. We continued the text in 2.2.2 ‘Hydrodynamic model implementation’ to now read; ‘*The recommended default Manning number (i.e., 32 m1/3s-1) was used to define bed resistance over the entire domain*’.

2.2.15. Please update figure 2b, with a better colorbar resolving the nearshore bathymetry, which is very critical for storm surge modelling [5].

A2.2.15. Thank you for this comment, Figure 2b has been updated with contour lines and a different colorbar – with more detail in the nearshore bathymetry. A comparison between the ‘before’ and ‘after’ versions of Figure 2b, is shown in the below Amended Figures and Tables section, for reference.

2.3 STORM dataset

2.3.1. At several places, Bloemendaal et al. (in review) were referred, and at L292 it was explicitly asked to see a under review paper. The paper was however not accessible. Is it in a closed review journal? If so, I would ask the authors to add substantiating details in the current manuscript. For example, explain in L291 the delta approach used in Bloemendaal et al (under review). Please feel free to use supplementary materials to add further details.

A2.3.1. This is an important point and we are therefore pleased to say that the referred-to Bloemendaal et al., paper was accepted for publication on 26th April 2022, in the journal ‘Science Advances’. The reference has now been updated in the main text and references list. Here is the link to the new manuscript:

https://www.science.org/doi/full/10.1126/sciadv.abm8438.

2.3.2. It is also of interest to know if the seasonal distribution of the STORM dataset follows the observed distribution. Also, if the storm category distribution matches the observed? - or how well they matches.

A2.3.2. Bloemendaal et al., (2022) can now provide the detail required, however in response we can confirm that the seasonal distribution of STORM does match the observed distribution. Also, it is shown in Bloemendaal et al., (2022) that the storm category distribution is consistent with observed data in the baseline STORM dataset, whereas within the future dataset there is an increase in the frequency of intense (category 4 and category 5) TCs.

2.3.3. L304-309: I do not follow why CNRM-CM6 model was chosen? Some linkage with previous study is missing here. Where is this decreasing trend of frequency is found?

A2.3.3. There was no particular reason to prefer the CNRM-CM6 Global Climate Model (GCM) over other available GCMs. However, we elected to utilise just one future climate projection in our modelling - it was simply not possible to run all four future GCMs each with 10,000 years of TC activity, within the time-scales of our funded project. In the future we would like to try and run the other three datasets. The fact that the number of more intense TCs (category 4 and category 5) are neither too high or too low in the CNRM GCM, compared to, for example, the extremes of CMCC (Table 3: 9,424 category 5 TCs) or HadGEM (Table 3: 21,661 category 5 TCs) climate models, within the wider Western North Pacific region, was an influencing factor in an otherwise random selection.

Apologies this was not clear, we have updated the text in L304-309 to now read, '*In order to avoid the large computational cost of simulating the storm surge conditions associated with the equivalent of 10,000 years of cyclone activity four times over, we elected to only use data from a single climate model projection. The future STORM dataset based on the CNRM-CM6-1 climate model was selected as - in the range of available climate model options - it ranked in-between extremes in the number of most intense TCs (i.e. category 4 and category 5, Table 3) in the domain, and thus suggested a middle pathway from all the four climate models options'*.

The text relating to decreasing trend of TC frequency was not really helpful in the paragraph alluded to and was creating confusion, therefore it has been removed for clarity. Thank you for this feedback. However to answer, as stated in Bloemendaal et al., (2022), in the future there is projected to be a '*global decrease in annual TC genesis frequency compared to the baseline*' (baseline = present day genesis frequency).

2.3.4. L310: does that mean there is an increase in frequency along Vietnam coast? and the TC frequency is decreasing in WNP? L343 indicates that for current around 40 year period there is 30,800 cyclones, and for future 35 year period there is 63300 cyclones? The statements in L304 about decreasing frequency, does not make sense with the number presented here. Please explain and clarify. Perhaps it is the selected cyclones, not 'frequency' in the data?

A2.3.4. We concur that the statement on TC frequency decreasing in the projected future WNP (Lines 304-306) is confusing in this paragraph and therefore, as also stated in answer to point 2.3.3 above, we have removed it to improve clarity. The point of this section was simply to state that, yes, when comparing similar ~35 year timescales, the projected future TC numbers exceed present day numbers in our model domain and this effects the coastline of Vietnam (and the Philippines) most of all. TC track distributions are also seemingly altered in this region, as illustrated in the Figure 5 hotspot map. This suggests that, by 2050, there is projected to be an uptick in TC activity along all coastlines, but the greatest difference, over this timescale, is around the coastline of central Vietnam (between Hainan Island and southern Vietnam). This section of text has been updated to hopefully clarify this point.

2.4 Statistical analysis (Section 3.3)

2.4.1. I do not understand what does it mean by "annual maxima approach" was selected? As you explain in Appendix, the STORM dataset has only month information in it, no date, time, or year - is it the case? If so, then how do you decide which storm fall in which year? Needs a clear explanation and logic behind the method chosen here.

A2.4.1. The STORM dataset has a synthetic month (weighted by the observed seasonal dataset – see Bloemendaal et al., 2022) and year associated with each created TC. The years are numbered between 1 and 10,000, and while this has no calendar meaning, these years were used to separate a 'year' of TC activity for this statistical analysis. Hence, once we had run all the TC through the hydrodynamic model, we were able to identify the maximum storm surge or water level within each coastal grid cell, for the TC associated with year 1, then year 2 and so on. With excess of 10,000 'years' of data the annual maxima approach is a viable statistical choice, considering that there was no information whether algorithm-generated TCs in any given year-month were independent events (i.e. occurring far enough apart, spatio-temporally, that sequential wind and pressure fields had no residual dependence). We have updated the text in the manuscript to make this clearer. At the start of section 3.3 the text now reads: '*In estimating these RPLs, we employed the following methodology: (i) the annual surge maxima was found for every one of the 10,000 years of the synthetic record, for each OCP (since each TC in the STORM database has a given synthetic month and year);'*. And in the supplementary section, *'Simulating tides adds the complexity that each cyclone must be randomly assigned a specific day and start time, but STORM only assigns each cyclone to a year and month*'.

2.4.2. In Figure 8. please provide estimate of confidence interval. For example, bootstrap method can be used for such large dataset. That will also tell us if presenting 10000 year return period value is meaningful.

A2.4.2. Thank you for your comment. We do not think it is appropriate to include a confidence interval in Figure 8, as it would make the detail in the three lines difficult to make out. Clearly the uncertainty will - and does - increase as you move to higher return periods, which is maximum at the 1000 year return period level. The approach we used to calculate confidence intervals is described in the last paragraph of section 2.2.3 'Computation of return periods'. For each model grid point, we randomly combined the storm surge with a different astronomical tide, repeating the process 100 times. Hence, for each grid point we have 100 different return level estimates, and in the paper we focus on the mean.

2.5 Discussions

2.5.1. A lot of discussion was put forward on bath-tub model, and flooding, but those were neither shown nor discussed in the results section (or appendix, supp.). Detail will be appreciated. For example, sea dykes are discussed in L506, but no information, height, location is presented. Please consider presenting those information - otherwise it is hard to check this claim.

A2.5.1. Thank you for this useful feedback. We are looking into this in greater detail in a second paper and for this reason we feel that this reference to a bath-tub flood area might not be a useful contribution to the discussion section for this paper after all. Exactly for the reasons that this comment describes. We have therefore removed this text from the manuscript. We are currently undertaking simulations of inundation in the Mekong Delta using DHI's MIKE 11 hydrodynamic model, using our MIKE 21 FM storm surge data at the coastal boundaries, and so will include discussion of this in a future 'following-on' paper.

2.5.2. How the storm surge timing differs/similar to river flood timing? Does river discharge has any impact? Does inland flooding can module surge level? Please discuss? Please also indicate that inland flooding is not considered directly in the model.

A2.5.2. This is certainly a very interesting point worthy of examination, but unfortunately it is outside of the scope of this study. We are aiming to assess this in the aforementioned second paper, discussed briefly above, coming soon.

2.5.3. Please discuss the model resolution issue. (Elaborated above).

A2.5.3. This is valuable feedback, thank you. A paragraph of discussion on model resolution and its impact on storm surge levels within small coastal features such as inlets and bays has been added into the discussion section of the paper. This full paragraph has already been copied into the answer for comment 2.2.2 above.

3. Minor comments

- **L24: What do you mean by "± tides"?**

With and without tides added. The authors have updated this text around Line 24.

- **L25-26: Repetition of the previous line "storm surge hazard" increases, is essentially stated again in the beginning of the next line. This can be shortened. See comment about "regional amplification" in the Major comments.**

These two sentences have been amended and shortened.

- **L34: This line is a bit jarring to read. I suggest to drop "Because of projected sea-level rise" completely from this line, as you have again said it in L39. Please also re-refer to Kirezci et al. at L39.**

Amended as recommended.

- **L41: The statement that there has been an assumption of stationarity of storm surge statistics needs to be backed up by references. Do you mean the "storm surge" statistics were assumed stationary, or the "storm" statistics were assumed stationary? Needs clarification.**

Line 41 text has been updated and references inserted. Previously there has been an assumption that the properties of storm surges will not change over time, but this has been challenged in recent work. The new text starts with some new references; '*Most past studies have assumed to date that storm surge extreme behaviour has been, and will continue to be, stationary and that the extreme wave climate will change little over large ocean regions (Vitousek et al, 2017; Hinkel et al., 2014). But with projections of a changed climate by the end of this century, this hypothesis has been challenged in recent studies (e.g. Tadesse and Wahl, 2021; Lin-Ye et al., 2020). For European coastlines by 2100, modelling shows that extreme storm surge levels may augment relative sea-level rise by over 30%, under the mean SSP5-8.5 climate projections (Vousdoukas et al., 2016). More recently Calafat et al. (2022) examined 1960-2018 tide gauge observations for north-western European seas and discovered changed trends in surge extremes due to climate variability and anthropogenic forcing*'.

- **L47: "these TC. . . " which one you are indicating? I believe all the TC, not the low-probability ones? is it?**

Yes indeed, all TCs but he more intense TCs are most damaging. The paper has been updated to clarify this; the text now reads, '*It is estimated that currently almost 230 million people around the world are directly exposed to some level of storm surge hazard from either tropical or extra-tropical cyclone activity, based on SwissRe global models (SwissRe, 2017). The populations most acutely at risk from storm surge induced extreme sea levels are those located on low-lying coastlines within tropical zones associated with intense cyclone activity (Dullaart et al., 2021; Edmonds et al., 2020; Kirezci et al., 2020; Woodruff et al., 2013; McGranahan et al., 2007; see also supplementary material section 1)*'

- **L49-51: It is a bit odd to read that 3% of the global population is currently exposed to storm surge hazard, while 2.4-4.1% of global population will be exposed to 1/100 year flood. Are these both quantify the same thing? Does that mean the estimate by one of these two author was over/underestimated? Or do they fall into the error bar of the estimation? This statement also hurt the claim that "it is an increase of 52% compared to today" in L36.**

Thank you for this comment - we can see now how this does seem inconsistent. It is correct that the two sources are quantifying slightly different things and they also estimate their numbers in slightly different ways. Kirezci et al. (2020) is contrasting the present and potential future (SSP5-8.5 climate projections) coastal flooding hazard from tide, surge and wave setup on coastal populations. The calculation of around 128-171 million population currently impacted is drawn from overlaying just the 1:100 year coastal flood extent on NASA's 2015, ~1km grid, Socioeconomic Data and Applications Centre global population dataset. All Kirezci et al., calculations assume absence of any kind of flood defences. A 1km grid indicates some oversimplification of population distributions.

By contrast, SwissRe (2017) has directly measured the impact of *all* storm surges on the global population, regardless of flood levels or return period. The numbers are calculated based on insurance returns within the topical and mid-latitude zones and will naturally take account of flood defence protection or failure. Because these numbers are based on insurance returns there is a possibility that the currently impacted population is overestimated because the same coastal populations could make multiple claims in a given year.

As such, the authors are happy to update these two sections of the paper to reflect the differences of what is being estimated/measured and highlight why the impacted population is slightly different for episodic extreme flood level vs tropical- and extra-tropical cyclone-induced storm surge inundation.

- **L51-56: Would it be possible to shorten this line? Perhaps, it can go to as a table in the appendix, with corresponding death counts, year, affected country(ies), and reference.**

Agreed. This has been removed to the supplementary materials.

- **L60: Please add 'low confidence' interval defined by IPCC.**

Added as described. 'low confidence' is defined as approximately a 2 in 10 (20%) chance of a correct prediction, of how storm surges may contribute to changes in future sea level extremes.

- **L56: Please avoid inconsistency and mismatching. TCs are feature of Tropical regions. For mid-latitude regions they are extra-tropical cyclones or mid-latitude cyclones, not TCs. So, saying "in particular in tropical and mid-latitude regions" is a bit odd.**

Agreed, this text has been updated to reflect the distinction, with thanks for your feedback.

- **L64-65: By capturing do you mean 'observing'/'recording' ? The next lines tells me so, but "capturing" is a bit of odd choice of word. Do you mean to say that infrequent nature of storm surge is the problem, or having a long-term data with necessary spatial resolution is the problem/challenge?**

Yes in these lines, 'capturing' means recording/observing/acquiring/measuring TCs impact on sea levels within short tide gauge record is a problem. The described sentences have been amended and now reads; '*This knowledge shortfall is in large part because of the essential difficulty of assessing tropical-cyclone induced storm surge datasets associated with events that are, by their very nature, somewhat infrequent (Dullaart et al., 2021; Mori et al., 2019). TCs are not only rare events, but they typically affect comparatively short stretches of coastline (<500km) as they approach land, and so storm surges are under-represented in the data from the sparsely distributed network of global tide gauges (Bloemendaal et al., 2020; Pugh and Woodworth 2014).*' In fact, spatial resolution of the TCs themselves within the atmospheric data record is a secondary problem for this science, this issue is discussed later in the paper.

- **L73: "two variant approaches"? → do you mean to say "two variation of an approach" or "two different approach"? Please clarify.**

We have replaced this text with 'two difference approaches'.

- **L79-81: Please add a reference(s) to this claim. (similar to L85)**

References have been added to Lines 79-81 to justify the statement.

- **L83: ‘gridded domain’ indicates the underlying model is gridded? Is it the case? Although the final distributed output might be ‘gridded’, GTSM model referred just after is a Delft3D FM implementation - meaning unstructured grid.**

Agreed that this sentence is not 100% clear. Hydrodynamic model meshes can be various shapes to suit the bathymetry and shape of the area to be studied. Only some meshes are in square grid format. This sentence has been amended to now read, ‘*The second approach involves the use of hydrodynamic models to generate multi-decadal time-series of surge-driven extreme sea levels across oceanic domains*.’

- **L87: At L69 you refers to > 100 year record. But the datasets you are referencing GTSM they are not 100y long? are they? ERA-Interim data starts from 1979, that means GTSM for current climate is only 40 year long.**

Line 69 refers to the ideal length of tide gauge record if anyone wishes to examine extreme storm surge hazard (up to or exceeding 100 years if the most extreme hazard levels are desired). The Line 87, referred to here, relates to Muis et al. (2022) who utilised ERA-interim data with record of ~40 years of wind and pressure data to represent TC activity in the GTSM, and used tide gauge data with 10 years or longer record to calibrate the model. As such the length of record was acknowledged as a limitation of the methodology and their resulting extreme sea levels were confined to a 1:10 year profile.

- **Table 1. Appears at wrong location. Should be moved around when it is first referred.**

Agreed. This can be changed to a better location in the final paper.

- **L148: ‘interconnected’ is redundant here.**

The authors agree, and have removed this word from this sentence at Line 148.

- **L148-151: Please break these lines.**

Done.

- **L151: What do you mean by ‘reconfigured’ ? And why it was reconfigured?**

We simply updated the land boundary in the model with a dataset that we could source and reference easily. The previous land boundary was no doubt accurate but we were not able to validate or date it. This word has been changed to clarify this point.

- **L172: Last line is redundant as it is standard.**

We would like to keep this sentence in the document to make it plain that currents are (as is standard), not incorporated. This is for the non-specialist reader.

- **L221: Table 1 should contain data range for each station.**

If this comment refers to a date-range, the authors agree and shall incorporate this data into Table 1

- **L295: drop 'smaller'.**

Done.

- **L296: What do you mean by "uncategorized"?**

Thank you for this comment. This sentence has been updated to explain this distinction: '*To create a sub-set of data for the period 1980-2018 for our own past/present hydrodynamic model we extracted all WNP area TCs that reach at least hurricane strength (Category 1 or greater on the Saffir-Simpson Hurricane Wind Scale; Simpson and Saffir, 1974) from the global STORM dataset (Bloemendaal et al. 2020)*'.

- **L324: What is a 'steering' file?**

Thank you for this feedback. We have updated the text to explain what this refers to. In model terms it is the control file that describes the model set up parameters and points to the correct descriptor/forcing files for a model simulation (e.g. in our case bathymetry, wind, tidal boundary data files). Steering files also control the naming and location of output results files. This text now states, '*MATLAB scripts were also used to create associated steering (control) files. The steering file differentiates the parameters of each simulation, by for example pointing to the next cyclone wind/pressure file or by generating a unique output filename. These steering files define the model* *solution technique, to integrate the time and space variables within the shallow water equations via an explicit scheme, utilising a variable time step interval in the calculation*'.

- **L393: What do you mean by 'dampens' surge? May be you wanted to characterize opposite way - like, narrow shelf does not allow surge to amplify as much as . . . ?**

The authors feel that 'dampens' is an appropriate word in this instance, but take on board the comment that it lacks clarity and shall review the document to ensure the meaning is conveyed appropriately, and that it refers to the amplitude of the surge wave.

- **Fig6, Section 4.1: It is interesting to see that basically for all the points, 10th largest surge is generated by tracks crossing the point on the west side, except point 1700 - why would be that?**

We wholly agree that this is a very interesting feature of these images. To explain in an oversimplified manner - firstly by characterising the counter-clockwise turning wind fields around the eye of a TC as a simple circular clock face, and split this face into 4 quadrants (12 to 3 o'clock being the first quadrant, 3 to 6 o'clock being the second, 6 to 9 o'clock being the third and 9 to 12 o'clock being the fourth quadrant). And then secondly declaring that 12 o'clock represents the direction of travel of the TC. In Figure 6, with a TC travelling westwards, the rotating wind fields striking onshore for most of the coastline of Vietnam are from the first quadrant of the cyclone, since the coastline faces east. The largest storm surges, located at the red dot in Figure 6, would therefore be found north-east of this TC for most of Vietnam (east for Chinese) coastlines. But around point 1700 the Vietnam coastline is turning and facing westwards. Hence the TC wind fields in the first quadrant are now coming offshore, and it is the third TC quadrant winds now forcing onshore. This flips the positioning of largest storm surges along the Vietnam coastlines from north-east of the TC, to south-west of the TC.

- **L575: Does this line refer to the current study? Or another study?**

Thank you for this comment. We are referring to the state-of-the-art in this scientific field rather than our own study. But as this is not clear we are happy to amend this sentence.

REVIEWER#2 References

[1] Thomas Guérin, Xavier Bertin, Thibault Coulombier, and Anouk de Bakker. Impacts of wave-induced circulation in the surf zone on wave setup. Ocean Modelling, 123:86–97, 2018.

[2] Xavier Bertin, Kai Li, Aron Roland, and Jean-Raymond Bidlot. The contribution of short-waves in storm surges: Two case studies in the bay of biscay. Continental Shelf Research, 96:1–15, mar 2015.

[3] David Pugh and Philip Woodworth. Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes. Cambridge U. Press, April 2014.

[4] Yann-Treden Tranchant, Laurent Testut, Clémence Chupin, Valérie Ballu, and Pascal Bonnefond. Near-coast tide model validation using GNSS unmanned surface vehicle (USV), a case study in the pertuis charentais (france). Remote Sensing, 13(15):2886, jul 2021.

[5] Y. Krien, L. Testut, A.K.M.S. Islam, X. Bertin, F. Durand, C. Mayet, A.R. Tazkia, M. Becker, S. Calmant, F. Papa, V. Ballu, C.K. Shum, and Z.H. Khan. Towards improved storm surge models in the northern bay of bengal. Continental Shelf Research, 135:58–73, mar 2017.

--

AUTHOR'S ADDITIONAL REFERENCES

Bloemendaal, N., de Moel, H., Martinez, A. B., Muis, S., Haigh, I. D., van der Wiel, K., Haarsma, R. J., Ward, P., Roberts, M. J., Dullaart, J. C. M., & Aerts, J.. A globally consistent local-scale assessment of future tropical cyclone risk. Science advances, 8(17). 2022.

Calafat, F.M., Wahl, T., Tadesse, M.G. and Sparrow, S.N., 2022. Trends in Europe storm surge extremes match the rate of sea-level rise. Nature, 603(7903), pp.841-845.

Cavaleri, L., Abdalla, S., Benetazzo, A., Bertotti, L., Bidlot, J.R., Breivik, Ø., Carniel, S., Jensen, R.E., Portilla-Yandun, J., Rogers, W.E. and Roland, A., 2018. Wave modelling in coastal and inner seas. Progress in oceanography, 167, pp.164-233.

Hinkel, J., Feyen, L., Hemer, M., Le Cozannet, G., Lincke, D., Marcos, M., Mentaschi, L., Merkens, J.L., de Moel, H., Muis, S. and Nicholls, R.J., 2021. Uncertainty and bias in global to regional scale assessments of current and future coastal flood risk. Earth's Future, 9(7), p.e2020EF001882.

Muis, S., Verlaan, M., Winsemius, H.C., Aerts, J.C. and Ward, P.J., A global reanalysis of storm surges and extreme sea levels, Nature communications, 7(1), pp.1-12, 2016.

Muis, S., Apecechea, M.I., Dullaart, J., de Lima Rego, J., Madsen, K.S., Su, J., Yan, K. and Verlaan, M., A High-resolution global dataset of extreme sea levels, tides, and storm surges, including future projections, Frontiers in Marine Science, 7, p.263, 2020.

Saulter, A., Bunney, C., Li, J.G. and Palmer, T., 2017, September. Process and resolution impacts on UK coastal wave predictions from operational global-regional wave models. In Proceedings of the 15th International Workshop on Wave Hindcasting and Forecasting and 6th Coastal Hazard Symposium, Liverpool, UK (pp. 10-15).

Wahl, T., Haigh, I.D., Nicholls, R.J., Arns, A., Dangendorf, S., Hinkel, J. and Slangen, A.B., Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis, Nature communications, 8(1), pp.1-12, 2017.

Amended Figures and Tables:

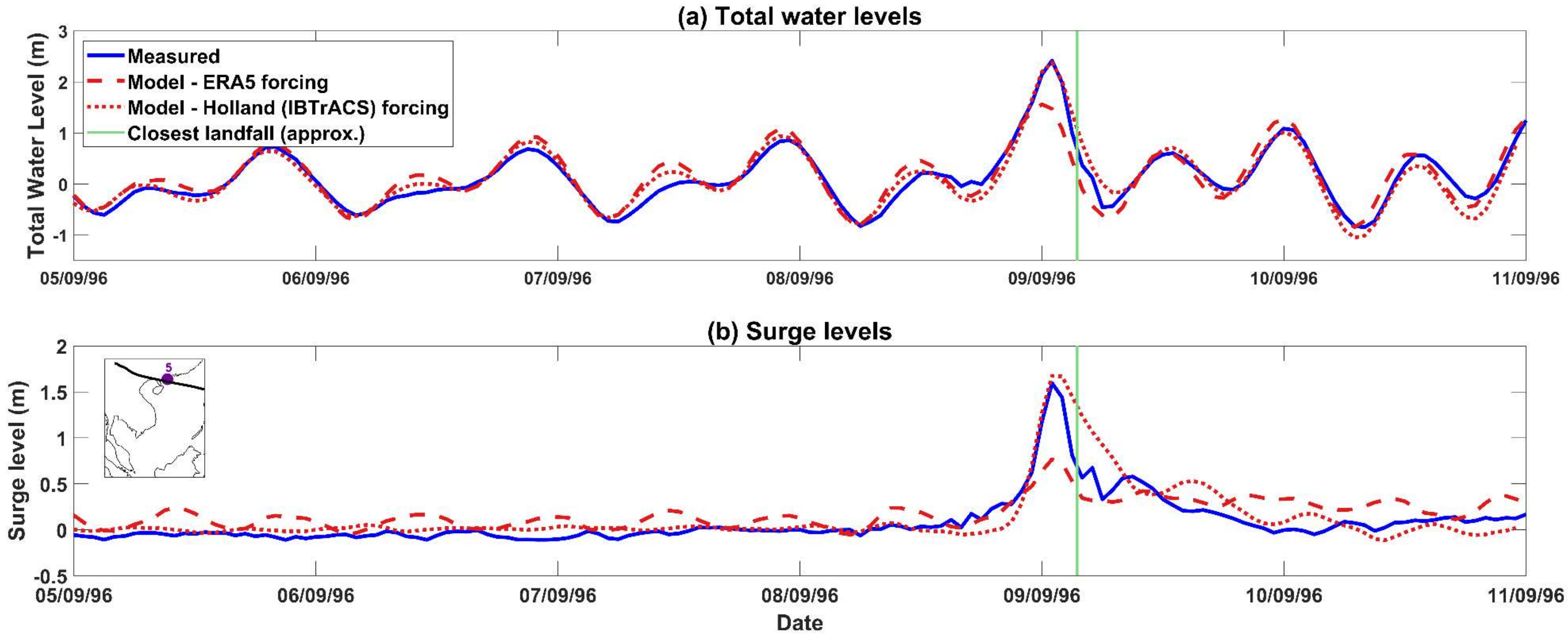


Figure S1 - Validating modelled surges using ERA5 (red dashed) and Holland Model using IBTrACS (red dotted) wind and pressure fields against measured data (blue): Typhoon Sally surge at tide gauge 5: Zhapo, China (inset or see Figure 1 for location), located closest to Zhapo station in the early hours of 9th September 1996 (green vertical line). Firstly (a) comparing total sea levels, and then (b) comparing surge-only water levels.

Table S1 - The Root Mean Square Error (m) between (a) measured tide gauge) and modelled total water levels and (b) tide-removed measured data and modelled surge-only water levels from all the validation hindcast simulations.

Typhoon name (date)	(a) Total Water Level RMSE (cm)		(b) Surge-only Water Level RMSE (cm)	
	ERA5 data	IBTrACS data + Holland model	ERA5 data	IBTrACS data + Holland model
Sally (September 1996)	33	30	20	15
Linda (November 1997)	41	28	39	20
Mangkhut (September 2018)	42	20	36	15
Ketsana (September 2009)	40	18	34	10

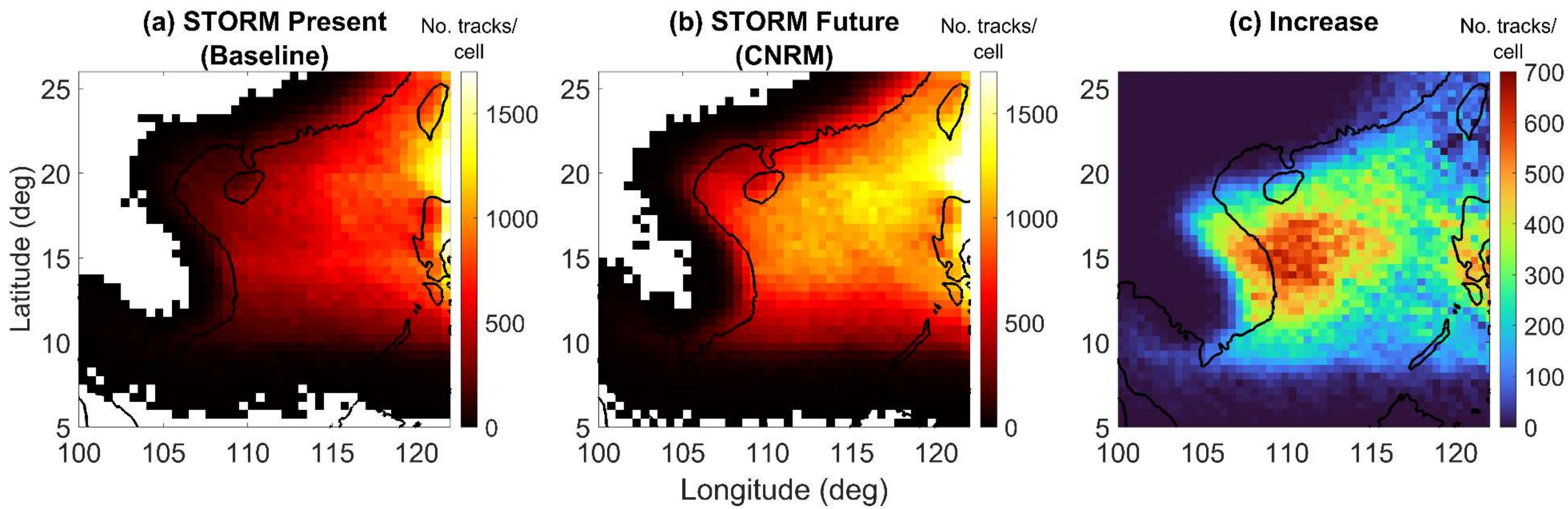
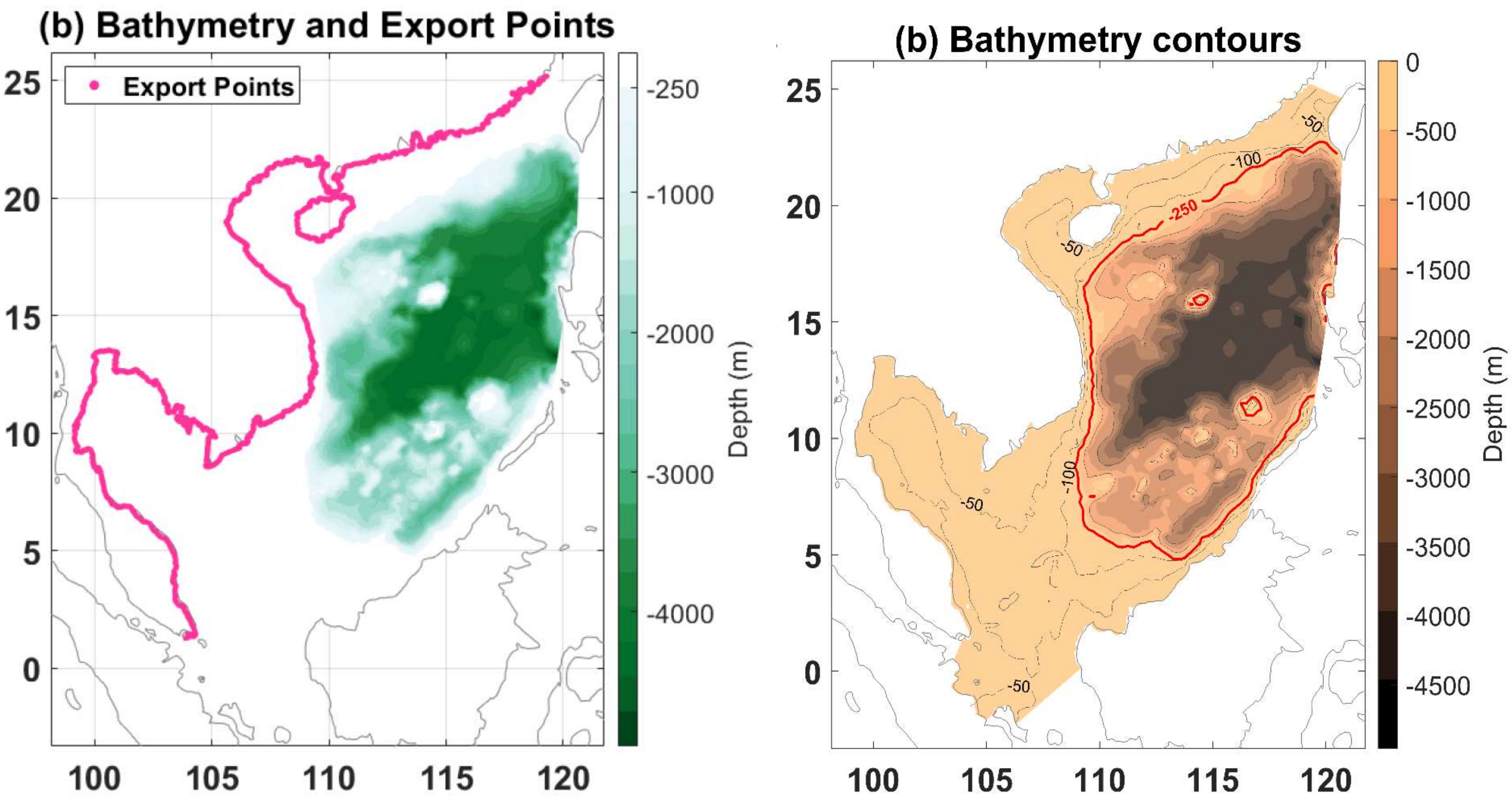


Figure 1 - Track density = the number of Saffir Simpson category 1-5 Tropical Cyclone tracks passing through each 0.5 degree x 0.5 degree grid cell within each ~35 year period. Left: Baseline STORM track density of Saffir Simpson Category 1+ (i.e. excluding Tropical Storms), Middle: CNRM climate model- Future STORM track density, of Saffir Simpson Category 1+, and Right: The cyclone track density difference between them.



Comparison of Figure 2b- Left is original, Right is updated bathymetry with contours and new colorbar. [this is SRTM15+ ocean bathymetry for our South China Sea domain. The 250m isoline is highlighted in red]