We thank the referee for the helpful feedback. Please see individual responses to comments below.

1- Introduction : Although the introduction is interesting and well-written, it primarily presents general information about global cyclone dynamics and lacks specific attention to the regional context and key concerns. Only a few lines in the entire introduction provide an overview of the region of interest. It would be more appropriate to concentrate on the US East Coast or at least the North Atlantic in the introduction.

The Introduction provides background information on the methods used to assess storminess organized by modeling studies followed by empirical studies. While some of these studies are global in nature, we also focus in on the relevant North Atlantic basin and U.S. east coast.

But we agree the Introduction would benefit from more localized context. To that end, we added this paragraph to the Introduction (at line 98) that transitions the Intro from general methodological background to more regional geographic context:

At the spatial scale of the U.S. east coast and centennial temporal scale, natural and potential anthropogenic forcings (e.g., sea-level rise and storms) threaten increasing populations and coastal development and ecosystems, especially given the geographic position of the U.S. coastline relative to extratropical and tropical storm tracks (e.g., Davis and Dolan, 1994; Friedman et al., 2002; Dinan 2007; Little et al., 2015; Doran et al., 2021). While much is known about the rates, spatial distribution, and acceleration of sea-level rise along the U.S. east coast during the twentieth- and twenty first-centuries (e.g., Sallenger et al., 2012; Ezer, 2013; Ezer et al., 2013; Yin and Goddard, 2013; Harvey et al., 2021; Chi et al., 2023; Yin, 2023) and changes to the wave climate over decadal time scales (e.g., Davis et al., 1993; Bromirski and Kossin, 2008; and Komar and Allan, 2007), less is known about changes to the storminess (frequency and changes in strength) over longer coastal reaches and time scales – especially using empirical data. Zhang et al. (2000) investigated water level data from 10 tide gauges from Florida to Maine and found no discernible long-term trend in the number and intensity of moderate and severe coastal storms during the twentieth century.

We also added this sentence at the end of the Introduction (line 116) to provide additional rationale for using SEPI:

A primary advantage of using this method is that sea-level change (i.e., rise) is removed to isolate the impact of storms on beach erosion potential and therefore, a rise in sea-level will exacerbate identified beach erosion potential stemming from storm tides and storm surges.

In addition, considering that numerous storm erosion predictive indices exist, it is important to clarify why SEPI and CSII were chosen, what unique contributions they offer, and what their limitations are.

We added a sentence in the Introduction (line 101) to identify the metric we use for identifying storms:

This study updates Zhang et al. (2000, 2001) Storm Erosion Potential Index (SEPI) assessment of storminess along the U.S. east coast and uses a newly developed index to assess the cumulative impact of storminess (timing and magnitude) on potential beach erosion along the U.S. east coast (Fenster and Dominguez, 2022). Like Zhang et al. (2000, 2001), we use water level data (storm tide and storm surge) to identify storms (rationale provided in 2.1 Storm Identification).

Note this sentence also refers the reader to the justification and rationale for choosing SEPI and its limitations found after line 133 in Section 2.1:

Recent studies have shown that wave runup (swash and setup processes) can contribute to extreme water levels and can induce spatially varying erosion impacts along coastlines due to varying continental shelf widths (Stockdon et al., 2007, 2023; Parker et al., 2023). However, Cohn et al. (2018) used new field datasets and a numerical model to show that anomalously high still water levels (caused by storm surge or spring tides) have a greater potential to produce dune erosion than the largest wave energy. Additionally, the effect of storm surge is purported to be larger (and the wave-driven component smaller) on the U.S. east coast than the west coast because the narrower continental shelves on the west coast limit storm surge (and enhance wave energy) more than the wider east coast shelves (Cohn et al., 2018). Serafin et al. (2017) found that slight increases in wave runup and a doubling of storm surge contribute to increases in extreme total water level events and make the case that the storm surge (high-frequency residuals) can have a 10-fold greater effect on beach erosion on the east coast than the west coast during large storms. While SEPI and water level data do not account for potential wave runup (Stockdon et al., 2007; 2023), Zhang et al. (2001) found a linear relationship between extreme storm surges and storm waves (wave heights > 2 m) indicating that storm surges make excellent surrogates for storm waves in representing the strength of large storms. The use of storm surge data over wave data is further motivated by the reliability and long-term availability of water level, storm tide, and storm surge data.

The importance of the cumulative storm impact index (CSII) was described in Fenster and Dominguez (2022). CSII is a model that can use any storm metric. To clarify in this paper we added text to the Introduction (line 100):

This study updates Zhang et al. (2000, 2001) Storm Erosion Potential Index (SEPI) assessment of storminess along the U.S. east coast and uses a newly developed cumulative storm impact index (CSII) to account for the timing (clustering) and strength of previous storms , ~~to assess the cumulative impact of storminess (timing and magnitude)~~ on potential beach erosion along the U.S. east coast (Fenster and Dominguez, 2022).

and added a sentence after CSII is introduced (line 171):

This index accounts for the timing and strength of previous storms, which make beaches more vulnerable to continued erosion (Fenster and Dominguez, 2022).

2- Method : The SEPI is calculated from S_{2SD}, representing the storm surge above the threshold for detecting storm surges, which is set at two standard deviations, and with a duration of 12 hours. The choice of two standard deviations and a duration of 12 hours is based on previous research. If the threshold were changed to 1.5 or 3 standard deviations or if a different duration were selected, the results would likely be affected. The choice of threshold and duration can influence the identification and quantification of storms, potentially altering the frequency and magnitude trends observed. Therefore, it is crucial to assess the robustness of the results and consider the sensitivity of the findings to different threshold and duration choices.

There was a mistake: there is no minimum duration for a storm. We have corrected the manuscript on line 392 of the discussion (additions are underlined): "We used the Storm Erosion Potential Index (SEPI) to provide thresholds for storm surges and tides that defined a storm by extreme water levels ~~that persisted a minimum of 12 hours~~ (Zhang, 1998; Zhang et al., 2000, 2001)." Following Zhang 2000, there is a criterion of 12 hours to distinguish storms: if the interval between storms is more than 12 hours, they were taken to be distinct storms.

We did not perform a sensitivity analysis of surge threshold (or of other thresholds used to identify storms) and personnel changes have made this task unfeasible. Rather, we used established criteria to identify a storm as the *definition* of a storm, and the results stand on their own using this definition. While a sensitivity analysis of each threshold would make for a very thorough investigation, our results are based on sound rationale, are consistent with previous research (e.g., Zhang et al., 2000, 2001 found a linear relationship between 2s of the storm surge and large waves, $H_s > 2$m), and provide reasonable results (not identifying too many or too few storms relative to named storms, see Figures 10 and 11).

While the methodology for the CSII is presented in the article by Fenster and Dominguez (2022), it would be beneficial for readers if the method were further elaborated in the manuscript. For example, the justification for choosing the exponentially decaying weighting factor and the selection of tc (time constant) as one year for beach systems on the U.S. East Coast should be provided. Additionally, the determination of the delta parameter should be explained, as it plays a role in quantifying the impacts of storm clustering and large magnitude storms on sandy beaches. Justifying these choices would enhance the understanding of the methodology and the interpretation of the CSII results.

We have made the following changes in the manuscript to clarify these decisions:

Line 178: "Assuming that the recovery rate is proportional to the amount of erosion, ~~W~~we use an exponentially decaying weighting factor for W_i (Fenster and Dominguez, 2022) where:..."

Line 188: While an appropriate value of the characteristic time, t_c, is crucial to understanding the meaning of the weighting function, mathematically the two parameters t_c and δ may be combined into one parameter to achieve the appropriate behavior of CSII. See Fenster and Dominguez (2022) for additional details. A reasonable choice of parameters will show accumulation due to storms clustered in time and will show beach recovery (CSII decreasing towards 0) when storms are temporally distant. In practice, there are a range of parameter values that satisfy these conditions and show robust cumulative behavior, though the absolute values of CSII will fluctuate with specific parameter choices. In this comparative study, we choose a value of t_c = 1 year corresponding to the winter-summer beach profile cycle for beach systems on the U.S. east coast, and $\delta=0.3$ for consistency across all tidal gauges studied.

The estimation of $PCTE\ (t)$ is conducted over the period from 1983 to 2001. The specific choice of this time period should be justified to provide a clear rationale for the selection. Additionally, if the analysis did not include the consideration of seasonal and interannual variations of tidal components, it is crucial to explain the reason behind this decision. Providing this clarification will enhance the transparency and facilitate the interpretation of the $PCTE\ (t)$ estimates.

We chose to pull the PCTE(t) values from the NOAA database, rather than calculate them ourselves, because we readily acknowledge that NOAA's collective expertise in this area far exceeds our own. So, we did not produce those estimates, nor did we make the decisions about how they were calculated or add any additional variations to NOAAs calculations. We believe NOAA's data to be appropriate for our calculations.

The time period from 1983 to 2001 is the current National Tidal Data Epoch as determined by NOAA and referred to as the Current Tidal Epoch (CTE) in our manuscript. NOAA centers ALL estimates of PCTE about the MSL of the CTE (rather than the epoch associated with the date of the data). This was not clear to us (the authors) initially, and we appreciate the clarification that we received via personal communication with Todd Ehret of NOAA (cited in the manuscript). NOAA keeps a set of "harmonic constituents" to reconstruct PCTE values at the time that the user requests them by plugging these constituents into a harmonic equation. This equation, though, also requires a parameter to set the average (zero) of the water levels at some chosen value. NOAA chooses this value to be the MSL of the CTE (1983-2001), even if we are querying dates before 1983 or after 2001. Therefore, the PCTE levels are not comparable to the verified water levels outside of the CTE. To calculate the correct value of the storm surge, we had to recenter the predicted water levels (PCTE) on the MSL corresponding to the date of the data. This is the purpose of Eq. 7. This rationale is further explained in the paragraph that begins on Line 241.

To clarify that PCTE(t) data is pulled from NOAA, we have added on Line 213: For each station, we retrieved the following data from NOAA (McManus et. al., 2023a, 2023b; Table 1):

Although the 37 harmonic constituents (listed here: https://tidesandcurrents.noaa.gov/harcon.html?id=9410170) are ostensibly astronomical (and hydrodynamical) in nature, they do incorporate meteorological variations. For example, the harmonic constituents Sa and Ssa7 are determined by seasonal weather changes. Here is a relevant section from NOAA's publication "Tidal Analysis and Prediction" (https://tidesandcurrents.noaa.gov/publications/Tidal_Analysis_and_Predictions.pdf), p.119:

"... the energy at the one cycle per year (Sa) and one cycle per half year (Ssa) found by the analysis is actually meteorological in origin, namely, caused by the seasonal changes in wind, temperature, and atmospheric pressure that affect water level."

https://tidesandcurrents.noaa.gov/publications/Tidal_Analysis_and_Predictions.pdf

It short, NOAA's 37 parameter fit is already VERY good and does incorporate seasonal changes. The method does not require additional corrections.

3- results/discussions :

Figure 5b: Do the results in terms of significance remain the same if a low-pass filter of 3-5 years is applied?

To check this, we performed an analysis of data without years excluded (that is, we use partial data rather than interpolate missing data points as required by a standard low pass filter for discrete datasets). Applying a Butterworth low pass filter of 4 years to these data showed that (1) most slopes are very close to the slopes in Table 2 and (2) the p-value for most stations improved or stayed the same. The analysis showed all but 2 stations have statistically significant results (using $p \le 0.05$). (The two that did not were Newport and Montauk. Newport's low slope was not statistically significant in our original analysis. The Montauk p-value went up slightly but was already just near the p-value cutoff for statistical significance.) Overall, this suggests that our results are conservative for most stations. Results are below:

Station	Slope (m^2h/yr) (from manuscript)	p value (from manuscript)	Slope (m^2h/yr) (with low pass filter)	p value (with low pass filter)
Portland	0.015	0.082	**0.017**	0.003
Boston	0.017	0.055	**0.012**	0.040
Newport	0.004	0.363	0.006	0.853
Montauk	**0.035**	0.049	0.021	0.054
The Battery	**0.045**	0.002	**0.044**	<0.001
Sandy H.	0.028	0.063	**0.028**	0.007
Atlantic C.	**0.033**	0.016	**0.033**	<0.001
Sewell's P.	**0.080**	0.001	**0.084**	<0.001
Wilmington	**0.058**	0.023	**0.058**	<0.001
Charleston	**0.023**	<0.001	**0.023**	<0.001
Fernand. B.	**0.020**	0.019	**0.020**	0.002
Key West	**0.028**	0.001	**0.027**	<0.001

Figure 6: What is the significance of error bars? Are the results presented over the same time period? If not, are the values comparable? It would be helpful to specify this in both the figure caption and the text.

Figure 6 is simply the average and standard deviation of all data in Figure 5. To clarify this, we have made the following change to the caption of Figure 6 (line 318): Average and standard deviation of each data set in Fig. 5. ~~the annual number of storms (a) and the average SEPI per year (b) for all 12 stations, corresponding to the datasets plotted in Fig. 5.~~ Calculations include all years of data plotted in Fig. 5, and similarly ~~Data sets~~ exclude years for which >=10% of data are missing.

We hope this makes it clear that the error bars are simply meant to visually identify the variation of the data in Fig. 5. Similarly, the data included in the calculation is clear from Fig. 5. (The overall ranges of the periods of record are also listed in Table 1, but Figure 5 shows precisely which years have been excluded due to lack of data.) Because we are characterizing each individual station, we chose to use the entire data sets of Fig. 5, rather than restrict the data to a common range.

Line 330 : "the CSII peaks appear to have a periodicity on the order of 3-10 years" Is there any explanation for this observation?

We provided a possible explanation beginning in line 449: "These aperiodic clusters have been thought to correspond with interdecadal to decadal scale variability observed in cyclonic development caused by North Atlantic Oscillation (NAO) and El Niño Southern Oscillation (ENSO) phases (Figs8, 9, 10, and 11; Davis et al., 1993; Zhang et al, 2000; Hirsch et al., 2001; Colle et al., 2015)." However, we do agree that additional work on this topic would make an interesting future study.

Line 448-451 : Is there any variation in the distribution of the **required** recovery time throughout the observation period? Do certain stations **require** more or less time for recovery? As the authors pointed out, the time spans associated with beach recovery range from 3 years to >10 years, depending on the variability of storms in both time and space. It would be interesting to further develop this aspect, particularly in relation to existing studies in geomorphology if available.

Yes, observation of our results suggests there is variation in time, but it's not predictable. Some places are more periodic (e.g., Sewells Point) and some places are less periodic (e.g., Wilmington). Some appear more periodic in more recent times (e.g., Charleston).

With respect to recovery time for certain stations... the answer is the same, some stations would have larger and some would have smaller recovery times.

The informal examination of our results indicate that firm answers would require additional quantitative analyses and comparison to other possible explanatory data which are beyond the scope of our study. We agree with the reviewer that it would be interesting to further develop this aspect of our work and relate it to geomorphology studies (especially NAO and ENSO events) as a standalone project.

Additionally, it seemed to us that the questions indicated a slight misunderstanding of our method. To clarify, we changed the language in the text (line 450) from "require" to "allow" indicating that a time period exists within which recovery can occur and not the actual recovery:

The results from this study show that peaks and troughs tend to vary on time scales of four to 10 years and provide insight into the time scale ~~required~~ allowed for beaches to "heal" after storm clusters and large magnitude storms occur (Figs. 8, 9, 10, and 11).

Line 451-454 : It would be interesting to investigate whether these aperiodic clusters truly correspond to the interdecadal to decadal scale variability observed in cyclonic development attributed to the North Atlantic Oscillation and El Niño Southern Oscillation phases.

We agree, see above.