We thank the reviewer for the careful reading of the manuscript and thoughtful feedback. Please see below for replies to specific comments.

1. I don't agree with some of the jargon used in the title and in the paper's findings related to "storm" and "storm surge." The definition of "storm" comes from the thresholding of extreme water level and non-tidal residual (defined as storm surge here) used in the SEPI, and the "storm magnitude" is the magnitude of the index, NOT of the water level variables.

The referee's distinction between "storm" and "storm surge" is correct, and we agree that some of the manuscript's language is descriptive rather than technical. We have made some clarifying changes and itemized them in responses below.

Similarly, the title suggests the paper is evaluating storm surge frequency, yet the variable that is evaluated is not storm surge. While a high storm surge is necessary for causing the SEPI, the storm tide also has to meet a threshold. This means there is the potential that not all storm surge events on record are analyzed.

For example, there could be events that don't cross the water level boundary, if the duration was low or the tide was low. So the suggestion the paper is evaluating storm surge frequency and magnitude (which is the elevation of the storm surge) is misleading.

We agree that the title is misleading and have changed it from: "Storm surge frequency, magnitude, and cumulative storm beach impact along the U.S. east coast" to "Storm frequency, magnitude, and cumulative storm beach impact along the U.S. east coast."

That said, there are inconsistencies in jargon throughout – using SEPI storm magnitude, average annual SEPI, storm magnitude, SEPI, and SEPI Values, SEPI measures of storm magnitude, all to describe the same value.

Each storm has a single value for SEPI, calculated according to Eq 1. SEPI is defined as the storm magnitude. So, while SEPI refers to the calculation in Eq 1, the terms "storm magnitude" (lines 272, 297, 321, 394, 405, 420, and 434) and "measures of storm magnitude" (line 354) are descriptive and used to remind the reader of the meaning of SEPI. We believe that this adds to the clarity and (in some cases) the grammatical flow of the language.

The "average annual SEPI," however, is NOT the same as the SEPI. While each storm has one calculated value of SEPI (Eq. 1), the average annual SEPI is calculated for each year from the SEPI values of individual storms occurring in that year. It is the average of all SEPI storm values for a given year.

We have clarified this distinction by making the following changes (additions are underlined):

- Line 154: The SEPI ~~storm~~ magnitude <u>for a single storm</u> is defined in terms of these two thresholds:

- Line 291: Added the sentence: Figure 5b shows the SEPI averaged over all storms each year for each tidal station.
- Line 304 in Figure caption 5: Annual ~~R~~results of (a) storm frequency (number of storms/year) and (b) storm magnitude (SEPI averaged over all storms identified in each year) arranged from north (top) to south (bottom).
- Line 318 in Figure Caption 6: 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.

2. The authors justify the lack of inclusion of waves by stating storm tide and duration as primary factors contributing to beach erosion from older studies, but many studies since then (e.g., Stockdon et al., 2007; Stockdon et al., 2023, Cohn et al., 2019, to name a few) show that wave runup (swash and setup processes) are important for spatially varying erosion impacts along coastlines. Other studies have suggested that wave runup/setup can be a large contributor to extreme water levels at the coast (e.g., Parker et al., 2023; Serafin et al., 2017; Stockdon et al., 2023). A brief discussion of the importance of these processes and potential for missing impacts is important.

Zhang et al. (2001) use the results from 11 studies, in part, to justify that storm induced beach erosion is "much more strongly related to storm tide than storm wave height." Additionally, Zhang et al. (2001) made the elaborate case that any beach erosion index should include **storm tide**, **storm wave energy**, and **storm duration**. However, while long-term storm tide data are readily available for the U.S. East Coast, long-term wave records do not exist, and wave records do not coincide with hourly water level data. This lack of empirical data makes it difficult to parse the relative contributions of the various forcings that control the total water level (TWL), i.e., waves, tides, and nontidal residuals (including storm surge) and drive storm-induced beach and/or dune erosion. Therefore, Zhang et al. (2001) used empirical and modeling studies available at the time of their publication to provide rationale and justification for using storm surge to represent storm strength and to be a surrogate for storm wave energy (e.g., Edelman, 1968, Edelman, 1972; Wood, 1982; Balsillie, 1986; Dean, 1991; Hughes and Chui, 1981; Vellinga, 1982, 1986; Steetzel 1991, 1993; Balsillie, 1999). Additionally, empirical data demonstrated a strong linear relationship exists between hindcast significant wave heights and storm surge heights (Zhang et al., 2001). Using storm surges greater than 2$\sigma$ of the annual surge level and wave heights larger than 2m, Zhang et al. (2001) suggested that this linear relationship indicates that storm surges make excellent surrogates for storm waves in representing the strength of large storms.

We acknowledge that spatial variation in beach erosion exists (as described in the suggested references). In fact, the variation exists at many different spatial scales. We also acknowledge that the relative importance of wave runup/setup varies at different spatial scales. To address these concerns, we added a paragraph in the manuscript after line 130 which provides additional rationale for using storm surge as a proxy for waves and incorporates the suggested recent references.

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., 20007; 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.

3. How is storm duration computed? It seems important to the computation of the SEPI. It seems that the SEPI may be the sum of all hourly data over the MHW threshold for the surge "event" but this isn't explicitly stated, beyond interpretation of eqn (1). Line 166 says that there is no minimum time duration for a storm, but Line 392 says a storm needed to persist for a minimum of 12 hours.

This was a mistake: there is no minimum duration for a storm. We have corrected the manuscript on line 392 of the discussion: "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.

The storm duration is the time between the first and last data point of the sum in Eq 1. Data that exceeds both thresholds (and therefore contributes a non-zero term to the sum in Eq 1) are grouped together as a single storm (and comprise the terms of the sum in Eq 1) when they are clustered within 12 hours of each other.

To clarify this, we have added the following to line 166: "Terms in the sum of Eq. 1 will be zero unless both thresholds are met. Data that exceed both thresholds are grouped together as a single storm (and comprise the terms of the sum in Eq 1) when they are clustered within 12 hours of each other. In other words, distinct storms must be separated by 12 or more hours.

~~and t~~The duration of the storm is the time difference between the first and last terms of the sum in Eq. 1, and there is no minimum ~~time~~ duration required for a storm.

4. How is the scaling factor, f chosen for weighting beach recovery, and how much does this choice impact the model result? How sensitive is the periodicity of beach recovery to cumulative storm impacts to the parameters chosen? Is 1 year a good approximation for beach systems along gradient that may experience both ETC and TCs?

Choosing a value for f (which is equivalent to choosing a value for delta) can be done for a single tidal station to characterize the rate at which recovery or return to equilibrium occurs. A good choice of f will show reasonable accumulation (as opposed to "artificial accumulation" described in appendix B of Fenster and Dominguez, 2022) due to storms clustered in time and will show beach recovery (CSII decreases towards 0) when storms are temporally distant. In practice, there are a range of f values that satisfy these conditions, and over this range of reasonable f values, the periodicity of beach recovery (observed in say, Fig. 8) does NOT change. (What does change is the overall range of CSII values: the peaks of CSII values may get higher or lower, but their positions in time do not change. Therefore, periodicity also does not change.) Thus, the overall results presented are robust relative to the choice of f value. More detail on the choice of f (or delta) value, along with sample data for f values chosen too high and too low are given in our previous paper (Fenster and Dominguez, 2022), especially Appendix B.

There is also more detail on the interpretation of t_c in our previous paper (Fenster and Dominguez, 2022), especially Section 3. It makes sense to choose an appropriate t_c to more easily interpret the weighting function; i.e., at time t_c after a storm (that is, t_p=t_c and tau_p=1) the beach has *mostly* recovered to its equilibrium state. Additionally, one could really *calibrate* the value of t_c for an individual tide gauge by validating data for a nearby beach.

But in practice, the index is robust enough to correct for an "incorrect" choice of t_c through the appropriate choice of f. (Note that mathematically, the parameters f and t_c can be redefined as a single parameter.) And in this study, which focuses on comparing 12 different tidal gauges, we did not calibrate any single location, but rather chose one set of parameters (t_c = 1 and delta = 0.3) that were reasonable for the entire set of gauges.

To clarify, we have added the following at 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 $\delta$ may be combined into one parameter to achieve the appropriate behavior of CSII (s~~S~~ee 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.

**Line by line**

Line 45: Typo after intensities ")"

Corrected

Line 97: Seems like Stockdon et al., 2007 would be a good reference to include here too which built off the Sallenger, 2000 publication.

We agree and have added the suggested reference. Note this reference was also used earlier in the paragraph.

Line 103: Nuance here, but I disagree the authors are assessing the frequency of and magnitude of TC and ETCs, as they're evaluating water levels, which aren't necessarily descriptive of JUST the storm climatology.

We agree this sentence is misleading and have clarified it. We are not claiming that we are analyzing TCs and ETCs. Rather we are making an interpretation based on correlation, not causality. We're using our definition of a storm and showing what this definition identified as storms and comparing those results to what's known about and consistent with storm climatology.

To clarify, we made this change in line 101: In particular, we assess the frequency and magnitude of ~~tropical and extratropical cyclones~~storms along the eastern U.S. coast using historical data from 12 tidal gauge stations located from Portland, Maine to Key West, Florida and compare these results to known storm climatology of tropical and extratropical cyclones.

Line 135: Shouldn't it just be SEPI, rather than "SEPI storm index"? Otherwise, you're really saying Storm Erosion Potential Index storm index.

We agree and have deleted "storm index."

Lines 208 – 210: The Wilmington and Battery stations might also be subject to river discharge within the non-tidal residual/storm surge signals.

We agree and have added the following language at line 208 to clarify:

It should be noted that these two stations are most likely subject to tidal wave transformation caused by interactions with complex channel geometries of shallow estuaries and/or fluvial processes not prone to occur at stations located along the open ocean coast (e.g., Aubrey and Speer, 1985; Speer and Aubrey, 1985; van Rijn, 2011; Hein et al., 2021).

Line 226 – 227: While the justification that the selection of storms with MHW vs MHHW is similar is positive, a quantification of how the SEPI or duration of events is affected could be important. I believe MHHW was justified as a threshold in the original paper to infer more wave attack on dunes/back barrier from the storm, and does this relationship hold for MHW?

Yes, Zhang's original justification for using MHHW as a threshold is that waves exceeding the MHHW level will be at an elevation high enough to directly attack dunes. We do not expect this relationship to hold *in general*, especially for mixed semi-diurnal systems such as tidal systems along the US West Coast. We rely only on the similarity of the MHW levels to the MHHW levels for the tidal gauges investigated along the semi-diurnal US east coast as well as the sensitivity analysis mentioned in our paper. To make this more clear, we have made the following additions to the paragraph beginning on line 221:

We note that Zhang et al. (2000, 2001) relied on the condition that water levels exceed the mean higher high water (MHHW) value, rather than the MHW, to identify storms. This is because storm tide above MHHW is high enough to directly and forcefully attack the dunes (Zhang, 2001). Because the U.S. east coast experiences largely semi-diurnal tides and because the difference between MHW and MHHW is small, it was not standard practice for NOAA's National Ocean Service to calculate historical MHHW values at tide gauges located on the U.S. east coast (T. Ehret, NOAA, personal communication). Furthermore, we conducted sensitivity analyses to determine the differences between MHW and MHHW for more recent years when water level data from both datums were available. These analyses revealed no significant differences in storm identification results for the stations considered using MHW compared to MHHW. It should be noted, however, that MHW should not replace the MHHW threshold in general. It is not expected that the MHW level is high enough for waves to do significant work on dunes for mixed semi-diurnal systems such as the the U.S. West Coast.

Line 262: How was the standard deviation over time computed? (if saying they are constant must have looked at time variability in this parameter?)

The data used for this analysis was all *hourly* storm surges available over the entire period of record for each station. We calculated the average and standard deviation of these data for each station over the entire period of record and plotted each distribution to verify that each was Gaussian distributed.

To clarify this in the manuscript, we have changed line 260: "For all 12 stations, the hourly ~~average~~ storm surge data~~s~~ are approximately Gaussian distributed, centered about a value of zero..."

To look for changes in time, we divided the data (over the entire period of record) into 9 time periods for each station. For Wilmington, NC, the period of record is from December 1935 -

December 2022 (87 years). Therefore, each smaller time period consisted of data over about a 10-year period. From these data sets, we calculated the mean and standard deviation of the hourly surges. We found that the standard deviations were approximately constant over the time periods investigated. The only large deviation in surge distributions over time that we identified was for Wilmington, NC (as identified in the manuscript.) The standard deviations quoted in the paper come from this statistical analysis over the 10-year time periods.

We did not think that most of these details were helpful for the paper, but we have changed lines 261-262 to include more detail for clarity: "The standard deviation of the hourly surge values, $\sigma_s$, ~~is approximately constant~~ measured over approximately 10 year time intervals, do not change appreciably over time for almost all stations with values ranging from $\sigma_s$ = 0.07 m (Key West) to $\sigma_s$ = 0.17 m (The Battery and Sandy Hook)."

MHW threshold takes into account sea level rise, why not just remove the MSL trend from the data and use a stationary MHW threshold? How does sea level rise effect results? Or is the inclusion of a time varying MHW/MSL take care of that?

In short, we use a single *surge* threshold because surge distributions do not vary over the time period of record (except Wilmington to some extent, as noted in the manuscript), but we use a time varying *MHW* threshold because the MHW *does* vary in time. Sea-level rise accounts for the long-term upward trend in MHW values, but there are also annual variations. The main reason that we use a moving threshold instead of a single threshold is so that our threshold is more accurate over shorter timescales. Annual calculations of the MHW threshold will more accurately reflect the MHW level for the beach at any particular time compared to a single threshold, and therefore should better predict erosion potential. (That said, we DID test a single MHW threshold and did not find significant differences in the results.)

Yes, the inclusion of a time varying MHW threshold removes the effect of sea-level rise. If we do NOT account for sea level rise (use a single MHW threshold and do NOT adjust the verified water levels), we would get SEPI values that increase dramatically over the period of record. While this DOES reflect increased damage to beaches which are being hit at much higher levels due to sea level rise, it will overestimate the frequency and magnitude of storms during later years and underestimate the frequency and magnitude during earlier years.

The suggestion to remove the MSL trend from the data is certainly an option. We could have subtracted the annual MSL from all verified water level data and used a single threshold. To determine that single threshold, we would also have had to subtract the MSL from the MHW data and find the average of the MHW over the entire period of record. However, our approach also removes MSL rise, uses an appropriately varying threshold (as described above), has the advantage of staying close to the raw data, and follows established protocol of Zhang et. Al, 2000 and 2001.

Line 365 – 370: Hurricane Florence impacted the Carolinas with an incredible amount of precipitation too. Is the signal Wilmington seeing due to river flow rather than coastal driven storm surge? Especially if potentially the duration of the event was impacted, e.g., gauge water levels staying high for much longer due to river flow outletting post storm surge event.

We agree and added language at line 369 to clarify:

~~Note that the~~The tide stations to the south of Wilmington did not record these storms appreciably. Despite the particularly stormy July 2018 along the North Carolina coast, we note that the peak stream stage measured at the stream gauge in closest proximity to (and upstream from) the Wilmington tide gauge (Cape Fear R at Lock 1 NR, Kelly, NC, 02105769) was approximately 11.6% greater than the mean stage (5.4 m vs. 4.9 m, respectively) and 15.5% greater than the median stage (5.4 m vs. 4.7 m, respectively). Consequently, post-storm river flow most certainly impacted the Wilmington water level data by approximately 11-16%.

Line 390: Again, not looking at spatial and temporal trends in storm surge/storm tide, looking at trends in SEPI

We have made the following changes at line 389 for clarity: "This study extracted historical water level data from 12 NOAA tide gauge stations, spanning the early 20th century to 2022 from central Maine to southern Florida, to determine if temporal and spatial trends existed in ~~storm surge and~~ storm ~~tide (as a storm metric)~~ frequency and magnitude along the U.S. Atlantic Ocean coast."

Line 396: What are the typical problems associated with empirical data analyses the authors are referring to?

The typical problems are discussed in the intro lines 68-70, but we have reiterated at at line 396 as well:

Our methods avoid typical problems associated with empirical data analyses such as using heterogeneous historical instrumental data and limited temporal data to detect long-term (decadal to centennial) trends in cyclone frequency or intensity.

In Figure 1, what is considered the duration? This might be a good place to include it

We agree that it would be helpful to illustrate the duration of the storm in a figure. However, Figure 1 only illustrates the first threshold, while Figure 2 illustrates the second. In order to qualify as data that contribute to a storm, BOTH thresholds must be met. Therefore, we cannot point to the beginning and end of the storm on either figure (since it is the product of these two data that identify the storm) and must rely on the mathematical description. As noted above, we have added language at line 166 to clarify what we mean by the duration of the storm.