The authors have made substantial changes to the manuscript in response to my previous concerns. I still believe that this idea is original and worth exploring. Again, I am not commenting on the statistical model, as it is beyond my expertise. The results though seem to make sense at constraining the ensemble members in 20CR dataset.

Despite the undoubtful interest of the work, I still have a number of comments on the manuscript, some minor, some other that would require further exploration. I am listing all of them below, as they appear in the text. Many of my comments arise from wrong statements or parts of the text that are unclear. This makes the manuscript sometimes hard to follow, even in the descriptive parts.

My first comment is about the format of the author tracked changes document. Please, avoid this practice in future submissions. It makes impossible for the reviewer to go through this file and identify the changes and it makes reading very uncomfortable. This is not what I understand should be a version with tracked changes which should contain a comparison between the original and the revised versions. I have therefore gone through the new file, without paying much attention to the tracked version. All the lines below thus refer to the new version.

- The term surge is misleading. Please, use the terminology as defined in the literature to avoid misunderstandings: https://link.springer.com/article/10.1007/s10712-019-09525-z.

- Abstract: the storm surge is not the response to atmospheric pressure. It also includes winds and waves and mean sea level.

- l. 72: .3 mb should be 0.3 mbar (I guess)

- l. 93 "As the latter is driven by a physical phenomenon called the "inverse barometer effect". This statement is incorrect, see my comment and reference above.

- l. 92-97. This is probably a good approximation but it is not strictly correct. First, the anomaly of every ensemble member should be calculated with its own mean. Second, the average should be calculated within the global ocean. According to the responses to my previous revision, the latter has been checked and made negligible differences. You should need to check the first point though.

- please remove figure 2. It does not make much sense to have only one station. This can be mentioned in the text.

- l. 112: storm surge and skew surge are two different metrics

- l. 112: "To access the surge, one first has to remove the tidal part of the signal, and then to remove yearly variations of the mean-sea-level (at interannual and decadal scale), such as sea-level rise”. This is incorrect. The storm surge is the difference between sea level and tides, so it does include mean sea level, so the definition is not exact. Then mean sea level can be removed if the purpose is to understand short term changes.

- figure 3: in the legend and caption, level should be sea level

- l. 127: “These oscillations are either due to tide-surge interactions (Horsburgh and Wilson, 2007) or to measurement errors in the 19th century leading to phase shifts” So you mean that these are 12-h oscillations? it is hard to see in the figure.

- l. 131: “This also implies that these 12-hours-averaged surges will only respond to atmospheric events persisting for more than 12 hours.” I do not think this is correct. It will smooth higher frequency changes but not remove them completely.

- l. 135-142: the inverted barometer acts at periods shorter than 12h, so the statement about longer periods is incorrect. Also, what do you mean by “local time-variations”? In any case, because this record is to be compared to a very low resolution model output, this approach may not be critical. I don’t think it is applicable to high resolution models though.

- l. 151-156: this is not clear to me. I understand that pressure observations have biases. Are you correcting for yearly dependent biases? If so, this is altering the temporal variability of the observations. This means that they cannot be compared as independent from the model. Bias-corrections must apply the same bias during the entire record. If the record is made of two separate set of observations then it makes sense to apply two biases separately, but never every year.

- l. 169-171: “Statistical correlation between pressure variations and wind intensity and direction are responsible for deviations from the inverse barometer approximation of the statistical linear relationship between surges and pressure (Ponte, 1994)” I am unsure about what this means. What statistical correlation?

And after this: “As a consequence of these combined effects of wind and pressure, the statistical relationship between the filtered surges and the pressures from 20CRv3 is expected to be non-linear, and not deterministic. “ Again, unclear. I would say that the effect is linear but there are other processes.

Next: “As showed by Hawkins et al. (2023), using a physical coastal model forced by the values of pressure (and winds) from the 20CR can lead to biases in the estimation of associated surges due to the resolution of the reanalysis, so that a statistical model is needed to correctly represent uncertainties” The need of a statistical model is not really an implication of the inability of the coarse resolution numerical models.

- L. 211-215 on the validation of the LLR model. If I understood correctly, the model is applied to a part of the dataset for the period 1980-2015 (those for which neighbours are 14 days apart). Then, Figure 4 should represent the comparison between the true and the modelled values only for those part of the time series that were not used to fit the model. Is this what it is showing? From the caption and text it seems that it is showing all values, but in this case, it is not a validation because values used to fit the model are also in there. Please, clarify.

- How are the results of the LRR different from purely inverted barometer approximation? Is Figure 4 improved with the LRR with respect to the most simple approach? In principle, if only subseasonal pressure variations are targeted, inverted barometer is likely a good proxy for sea level changes. If another more complex model is to be used, its benefits should be demonstrated. I am afraid that at this point it is still unclear to me why this model is needed.

-l. 358-359: “In 1865 (Fig. 9.b), although the surge-based reconstruction happens to be more consistent with observations than the reanalysis, the reverse is also true.” Please, rephrase, two opposite things cannot be true …

-l. 362: “We attribute these biases to different atmospheric conditions which cannot be estimated from the surges with our simple LLR model, in particular wind directions and intensity.” This seems unlikely given the periods of time of several days and the fact that the data are smoothed. I wonder how these comparisons are with a simple inverted barometer approach, not using the LLR model.