This work uses historical tide gauge records in the French coast as proxy data to improve the atmospheric 20th century reanalysis. Specifically, the sea level observations are used to constrain the number of ensemble members in the reanalysis, selecting and weighting them according to their correspondence to observed storm surges. This is an interesting idea. The approach discards model realisations that, although physically consistent with the scarce available mean sea level pressure observations, are inconsistent with observations of storminess from coastal tide gauges. I am not a statistician and therefore I am not qualified to comment on the details in section 4. My comments are focused on the results and, mostly, on sections 2 and 3 on data description and processing. I have found major issues in these sections, including a basic conceptual misunderstanding of the generation of storm surges. Also, there are missing details in data processing and I have concerns on the simple linear regression model that has been used to relate storm surges and sea level pressure anomalies.

Another major concern is that the authors state in the conclusions that they use the tide gauge time series as a barometer record not assimilated in the reanalysis. However, they model the storm surges from sea level accounting for both pressure and geostrophic winds. I do not understand the interest of including the winds here (which by the way are not modelled correctly, in my opinion). If the purpose is to incorporate only pressure-like measurements, then pressure gradients should not be included at all. Or at least, the pressure-like record should be reconstructed without accounting for wind effects.

All these issues are detailed in the following. My recommendation is that the manuscript should undergo major revisions before being considered for publication.

Section 2: This section needs further details in the description of the data set. Preprocessing of data needs improvement.

- L.62: define the spatial and temporal resolution of the reanalysis.
- L. 70: I assume the variable used is mean sea level pressure (MSLP), correct?
- L. 81-82: please, remove.
- L. 83: anomalies of what? I assume this means that mean SLP anomalies are computed with respect to climatologies for 2 periods. Please, specify.
- Description of tide gauges (section 2.2): please, define the temporal sampling of tide gauges, and their overlapping periods. The statement that one time series can be used to fill in gaps of the another is incorrect. They are highly correlated at low frequencies but not necessarily at high frequencies (i.e. storms). Actually, they are not even in the same location, as stated later in the text, so differences in storm surges may be expected.
- L. 97: actual-> modern, current, present-day?
- The removal of MSL effects in the tide gauge processing is incorrect. The impact of MSL should not be removed in this way, because there are other MSL variations at interannual and decadal time scales that would remain. A better approach is to remove mean (or median) yearly averages. It also avoids the problem of selecting an arbitrary change point to calculate different linear trends.
- L. 108-109 on detiding: which is the period used for detiding?
- L. 113: phenomenons-> phenomena
- L. 110: A 12-h filtering removes part of the storm surge variability, especially that related to storms.

- L. 114-115: smaller scale events are unlikely to be recorded in 20cR at 2deg resolution.

Section 3: This section contains fundamental errors in the understanding of the processes that generate storm surges. Also, the linear regression model is questionable as applied here, and should be better justified. More details:

- L. 123-124: hydrostratic equilibrium is a general law not only affecting this region. Inverted barometer has been widely applied everywhere.
- L. 127-131: this is incorrect. The main mechanism of wind as storm surge generation is not the Ekman transport, but the piling up of waters due to wind blowing perpendicular to the coast. Check your own results, where you can actually see it (e.g. line 163). Please, change this in the abstract too.
- L. 1334-134: I assume that SLP here (and used later in equation 1) is the SLP anomaly with respect to time at the given grid point. If this intend to represent the inverted barometer effect, this is incorrect. The anomaly should be computed with respect to the average mean sea level pressure over the world oceans in the reanalysis. See for example Ponte 1994 (https://doi.org/10.1029/94JC00217).
- L. 134: interpolated how? Linearly?
- Equation 1: please, define all the terms. I do not believe this is a good way to separate the forcings. I think it would be better to use perpendicular and parallel to the coast MSLP gradients (geostrophic winds). In this case, the parallel will be insignificant and the model would be simplified. This separation has a physical meaning and would also allow to interpret the results easily.
- Also, why is the term Cov introduced? How does it change the results?
- Still in the same equation, are all coefficients significant? what happens if a step-wise regression is used instead? I would expect gamma coefficients to be discarded.
- Tables 1 and 2: units are missing.
- L. 170: “1.02 on physical grounds”-> actually not, the inverted barometer effect is a simplification and does not work at the coast.
- L. 171-172: “This justifies…”-> I do not think this statement is correct. I believe a step-wise regression would tell you if the use of these parameters is justified or not. Here it has not been proved, not even with uncertainties of the parameters.
- L. 174: wave setup would be removed with the 12-h averaging, so it does not play any role.
- Section 3.2: To demonstrate that the same coefficients can be used during another period, this is not the correct approach. I think a better approach would be to use the recent period to generate 19th-century-like observations, i.e., downgrading the number and location of observations, and compute the coefficients of the LR. These downgraded, coarse-resolution coefficients would be directly comparable to the ones in the section above with the full resolution. It would be important to calculate uncertainties in all cases in order to compare the coefficients.
- Fig 5: it is unclear what is shown here. Are coefficients of the modern period used here? Please, specify. The description of the results in fig 5 is confusing and does not reflect, in my opinion, the results. For example, in l. 188 the bias referred to one of the cases is visible in all (5a and 5b); l. 193: which interpretation is referred to here?; l. 196 states that there is no clear sign of bias in individual members but I think the bias is similar to that in the ensemble average. The bias is not only due to ensemble mean but also to limitation of coarse resolution data.

- L. 206-220: this discussion on the differences in coefficients should consider uncertainties to ensure that the use of two different periods lead to different values. Values slightly larger/smaller do not provide confidence in the results. Therefore, the interpretations are not reliable (e.g. differences attributed to ensemble averaging).
- L. 225-226: I think the consistency of the coefficients and the explained variance in modern and old periods is a consequence of the dominance of the inverted barometer effect. This is probably not true for extreme values generated by strong winds associated to storms, but it holds for mean storm surge variability driven by pressure changes.
- Summarising, I believe the LR model should be modified to consider winds parallel and normal to the coast in each case, and use step-wise (or similar) approach to remove the terms that do not explain more variance but introduce noise. Also, uncertainties of the parameters should be calculated. The use of the coefficients calculated in the modern period should be tested in downgraded modern data to prove that they are usable in older periods. The entire comparison and discussion of the two periods should be modified accordingly and simplified in case this is proved.

Section 4: I am not an expert in HMM and I am not qualified to comment on the details of the method explained here. It would be good that an expert statistician reviews this part. I nevertheless would recommend defining all variables in the equations, as well as the acronyms (e.g. NWP), as it makes it very difficult to follow as it is now. Other comments follow:

- L. 270: this is a rather large area. Are the results sensitive to this choice?
- L. 336-337: does this mean that the storm surges constrain the data more in winter when they have a stronger signal?
- L. 361: then->than
- Section 4.3 is very illustrative of the potential of the approach. However, more information would be useful to understand why including a single (or two) new records essentially rules out 79 out of 80 model ensemble members. In particular, how many sea level pressure records were included in this period (or shortly before the storms) in the area? How are they distributed? It would be also useful to see a case with more sea level pressure observations (early or mid- 20^{th} century for example).

Conclusions section:

- L. 405-408: I understood from the LR model that winds were incorporated into the model as SLP gradients. Then, why if wind is an important driver of storm surges may limit the use of the tide gauge? In fact, if wind is not accounted for and wants to be removed, then the LR model could include only the inverted barometer effect with the adjusted parameter at each tide gauge location.