Answer to 2nd round of review of manuscript: "Could old tide gauges help estimate past atmospheric variability?" for *Climate of the Past*

Paul Platzer, Pierre Tandeo, Pierre Ailliot, Bertrand Chapron

August 2024

Reviews in black, answers in blue, ***quotes from revised manuscript in bold.***

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.

We thank the reviewer for the new comments and suggestions which again helped make our work much more precise and rigourous.

We have used another tool to create author track-changes, and we hope it allows for a more comfortable reading. In this response document, the underlined line numbers refer to the 2nd revision of the manuscript, but we also include the line numbering of the tracked-changes document.

Comment-by-comment responses are listed below.

1. 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.

1. We thank the reviewer for this clarification. The ambiguous term "surge" was replaced by "surge residual" in most cases, and we have also clarified the additional operations that we make (removing the yearly median, making averages, etc.) where needed. See the response to other comments below.

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

2. Indeed. We have stressed this out by using the word "*including*", writing
"The surge residual is the non-tidal component of coastal sea-level. It responds to the atmospheric circulation, including the direct effect of atmospheric pressure on the sea-surface."

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

3. Corrected.

4. 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.

4. This was changed to, l. 93 (also 93 in tracked-changes document):
"As the latter is driven, in part, by a physical phenomenon called the 'inverse barometer effect' "

5. 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.

5. To check the first point, we have computed :

1. The world-ocean average for each individual member, in year 1870. Then, at each time-step we computed the standard deviation of this average over all 80 members. This first quantity measures intra-member variability in ocean average pressure.

2. The standard deviation over all 80 members of the MSLP at the city of Brest. This second quantity measures the variability of the intra-member variability in local pressure.

We have checked that the first quantity is almost steady, between 22Pa and 35Pa. The second quantity displays more variability, and varies between 200Pa and 800Pa. It thus appears that the potential contribution of the inter-member variability in ocean-averaged pressure to the results of our study is weak.

Furthermore, our estimates of pressure based on the LLR shown in section 5 necessary rely on a unique estimate of the ocean average pressure, for which we have used the average over the 20CRv3 members. It thus makes sense to use this members-average throughout the article instead of one ocean-average per member, although the second solution would be more accurate.

Therefore, we have chosen to keep the computations as is, but to mention this approximation with the following two sentences l. 98 (also l. 98 in tracked-changes document) :

"Note that there is a small variability in ocean-averaged pressure between 20CRv3 members in the 19th century. However, we have checked that this variability is one order of magnitude smaller than the inter-member variability of MSLP at the city of Brest, which justifies our approximation of using simply the members-average of the ocean-averaged pressure as a reference."

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

6. The figure was removed.

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

7. This was removed.

8. 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 anddecadal 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.

8. We thank the reviewer for this clarification. We have modified the definitions and added the reference suggested above, giving l. 114 (l. 116 in tracked-changes document):

"

As mentioned earlier, the part of the sea-level which responds to atmospheric processes is the surge residual (see definition in Gregory et al., 2019). To access the surge residual, one has to remove the tidal part of the signal. Then, as we are interested in sub-seasonal variations, we also remove the yearly variations of the mean-sea-level (at interannual and decadal scale), such as sea-level rise (Cazenave and Llovel, 2010). In this work, we also use moving averages and differences of the surge residual. All these steps are exemplified in Fig. 2.

We first compute the tidal constituents of the raw sea-level (blue curve, Fig. 2.a) using U-Tide (Codiga, 2011), which performs harmonic (Fourier) decomposition with prescribed frequencies corresponding to planetary movements. The tidal constituents are computed over two different periods, one is 1847-1890, and the second is 1981-2015. Removing the tidal part of the signal gives the surge residual (orange dashed line of Fig. 2.a), which has a temporal average value of ∼4m for the Brest tide gauge.

Then, we remove the yearly median value of the sea-level (orange dashed line of Fig. 2.b). We choose to remove the median and not the mean because the mean can in principle be influenced by the number and magnitude of extremes in a given year, which can be linked to the number and magnitude of storms passing in a given year. This second step allows to access the zero-median surge residual which is noted $h(t)$ in the following:

$$h(t) = H(t) - \text{Tide}_H(t) - \text{median}\left[H(t'),\, t' \in \text{year}(t)\right] , \tag{1}$$

where $H(t)$ denotes the raw sea-level, $\text{Tide}_H(t)$ is the tidal part of the signal computed from H, and $\text{year}(t)$ is the year in which time t is found.

"

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

9. This was modified (now in Figure 2).

10. 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.

10. First, when performing 12h-averages we filter *more* than 12h-oscillations, and as pointed by the reviewer we sometimes see oscillations at smaller frequencies that are removed as well. Therefore we have rephrased to give l. 131 (l. 134 in tracked-changes document):

"Note from Fig. 2.b that the surge residual fluctuates at hourly scale, part of which are oscillations which are not due to variations in atmospheric pressure. For instance, these oscillations can be due to tide-surge interactions (Horsburgh and Wilson, 2007) or to measurement errors in the 19th century leading to phase shifts."

Also, we have chosen another example where 12h-oscillations are more obvious. This is also highlighted in the text with the following sentence l. 133 (l. 136 in tracked-changes document):

"Such 12h-oscillations can dominate the surge residual signal in Brest where the tidal amplitude is large (see for instance on the 29th and 30th of January 2014, Fig. 2)."

11. 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.

11. We chose to simply remove this sentence.

12. 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.

12. We have reformulated the rationale for including the variable $\Delta\overline{h}^{3h}(t) := \frac{1}{3}\sum_{t'=-2}^{t'=+1}[h(t+t') - h(t-12+t')]$ in our LLR. It is now as follows l. 141 (l. 145 in tracked-changes document):

" Finally, note that if atmospheric pressure variations are faster than the typical time of adjustment of sea-level, one expects deviations from the inverse barometer approximation (Bertin, 2016). Therefore, fast time-variations of the surge residual are also expected, statistically speaking, to be associated with deviations from the inverse barometer approximation. To allow the model described in section 3.1 to capture this effect, we compute the difference between the surge at time t and at time $t - 12h$, choosing the 12h-interval again to filter out oscillations at a period close to 12h. Furthermore, since the reanalysis is run at 3h-resolution, we perform a 3h-moving average of the surge residual before computing the difference. This difference is noted $\Delta\overline{h}^{3h}(t)$ and defined by the following equation: "

13. 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.

13. We have modified our procedure, applying one bias-correction for each dataset, as described l. 156 (l. 163 in tracked-changes document):

" We have found a shift in average pressure between the EMULATE and Météo France datasets. To overcome this issue, and since we are only interested in sub-seasonal atmospheric variability, we added a constant value of $\sim$ 0.22hPa for the period 1860-1880 (EMULATE dataset) to each value of the independent pressure observation datasets, so that the average pressure are equal between the independent observed pressure and the 20CRv3 mean pressure linearly interpolated at the city of Brest. We did the same operation for the period covered by the Météo France dataset that we are using (1855-1859 and 1881-1894), adding a value of $\sim$ 7.18hPa. "

We have reproduced the figures using this new bias correction (originally numbered Fig. 9, 11, 12, 13, 14, now numbered 8, 10, 11, 12, 13) and found no noticeable difference. Therefore, we kept the figures as is.

14. 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.

14. We have reformulated this paragraph, hoping that it is now more accurate and clear l. 177 (l. 187 in tracked-changes document):

" Since wind is not included in our model, the relationship between the filtered surge residuals and the atmospheric pressures from 20CRv3 should not be deterministic. Also, it is likely that typical wind conditions depend on the amplitude of MSLP anomaly, so that the average value of MSLP anomaly for a given value of surge residual

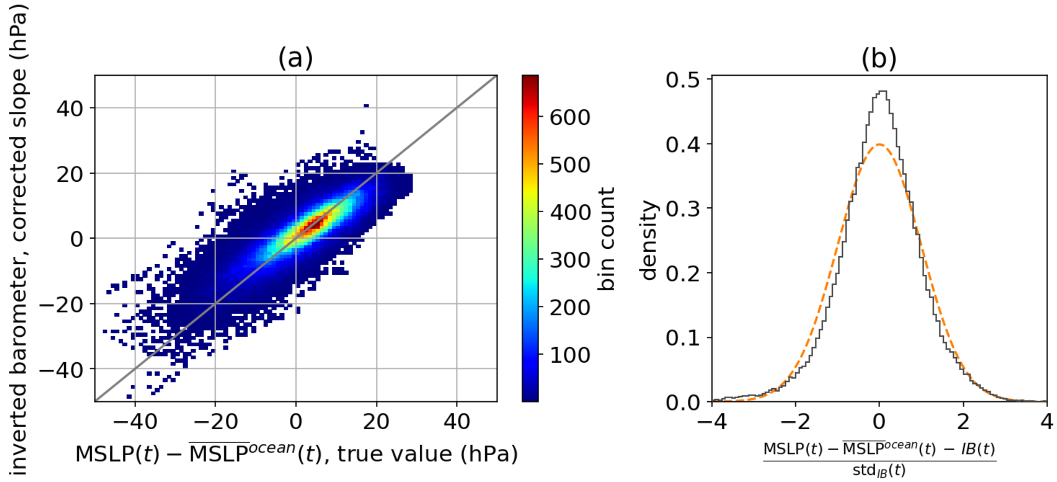

Figure 1: Reproducing a figure of the manuscript but with a simple inverted barometer instead of the more complex LLR.

in Brest may be a non-linear function. 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. A statistical model can thus be used as a tool to correct such biases and represent uncertainties. In our case, since we want to estimate pressure based on the surge residuals only, the effect of unknown wind or other processes must also be taken into account through uncertainty quantification. ”

15. 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.

15. In our procedure, for each time $t \in [1980 - 2015]$, we use 200 samples in $[1980 - 2015]$ but excluding $[t - 14 \text{ days}\ ,\ t + 14 \text{ days}]$ to fit a linear regression. Therefore, the data that is used to fit the model does not include the true values. The 200 samples are chosen at each time t using a similarity criterion (smallest Euclidean distance) on the vector $\left[\overline{h}^{12h}(t), \Delta\overline{h}^{3h}(t)\right]$.

When using the LLR on the 19th century data, since we search for neighbours in the period $[1980 - 2015]$, we do not need to use the twice-14-days window exclusion.

We have reformulated, l. 220 (l. 231 in tracked-changes document):

“ To test the accuracy of this model on the 1980-2015 period, we apply it for all times $t \in [1980 - 2015]$, searching for neighbours' times t_i in the same period but with the condition that there is a minimum of two weeks between t and t_i (i.e., excluding the interval $[t - 14 \text{ days}\ ,\ t + 14 \text{ days}]$). This is called the 'leave-one-out' procedure, ensuring that the data that is used to fit the model does not include the true values. ”

16. 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.

16. For comparison, we show here in Fig. 1 the same plots as in Figure 4 of the 1st-revision of the manuscript but for a simple inverted barometer, understood here as a linear regression between 1. the difference between the local MSLP and ocean-averaged MSLP and 2. the sea-level residual as defined in the text of the 1st-revision of the manuscript.

The left panel of this figure (Fig. 1.a) shows no clear sign of improvement in average prediction with the LLR model compared to a simple IB. However, the right panel (Fig. 1.b) shows that the rescaled IB differs more from the standard normal distribution than the LLR (higher peak probability around zero), indicating that the LLR provides better uncertainty quantification. This is because the variance “var(t)” as defined in the revised manuscript is situation-dependent. A simple linear regression assumes *homoscedasticity* (i.e., that the distribution of the difference between the linear prediction and the true value does not depend on the value itself). On the contrary, the LLR assumes *heteroscedasticity*. This allows for better uncertainty quantification and is a key advantage of the LLR. As we are interested in uncertainty quantification, this justifies the use of a more complex algorithm such as the LLR.

More details on the interest of the LLR are given in the answer to question **18.** in this response document.

17. 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 ...

17. This was poorly formulated and brought confusion. We have reformulated to give l. 368 (l. 381 in tracked-changes document):

" In 1865 (Fig. 8.b), although the surge residual-based reconstruction is sometimes more consistent with observations than the reanalysis, there are as many occasions where it is the reanalysis which is more consistent with the independent observations. "

18. 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.

18. For comparison purposes and at the demand of the reviewer, we reproduce here Fig. 8 of the second revision of the manuscript (numbered Fig. 9 in the first revision of the manuscript), but replacing the results of the LLR with the results of a simple linear regression between surge residuals and atmospheric pressure residuals, simply referred to as "IB" in the figure's legend (Fig. 2 in this response document). A first observation is that the uncertainties associated with the simple IB approach are larger. To allow for a more in-depth comparison, we also plot all estimates in one figure (see Fig. 3 of this response document). We see that the LLR uncertainty range is always comprised in the IB uncertainty range, although smaller. The average from the LLR is close to the IB average. At times, some of which are highlighted in green, the LLR is closer to the independent observations (and the 20CRv3 reanalysis) than the simple IB. Note that this is not always true, and there are a few counter-examples. However, the reduced uncertainties show that the LLR is overall slightly more precise than a simple IB.

References

Xavier Bertin. Storm surges and coastal flooding: status and challenges. *La Houille Blanche*, (2):64–70, 2016.

Anny Cazenave and William Llovel. Contemporary sea level rise. *Annual review of marine science*, 2:145–173, 2010.

Daniel L Codiga. Unified tidal analysis and prediction using the utide matlab functions. 2011.

Jonathan M Gregory, Stephen M Griffies, Chris W Hughes, Jason A Lowe, John A Church, Ichiro Fukimori, Natalya Gomez, Robert E Kopp, Felix Landerer, Gonéri Le Cozannet, et al. Concepts and terminology for sea level: Mean, variability and change, both local and global. *Surveys in Geophysics*, 40:1251–1289, 2019.

Ed Hawkins, Philip Brohan, Samantha N Burgess, Stephen Burt, Gilbert P Compo, Suzanne L Gray, Ivan D Haigh, Hans Hersbach, Kiki Kuijjer, Oscar Martínez-Alvarado, et al. Rescuing historical weather observations improves quantification of severe windstorm risks. *Natural hazards and earth system sciences*, 23(4):1465–1482, 2023.

KJ Horsburgh and C Wilson. Tide-surge interaction and its role in the distribution of surge residuals in the north sea. *Journal of Geophysical Research: Oceans*, 112(C8), 2007.

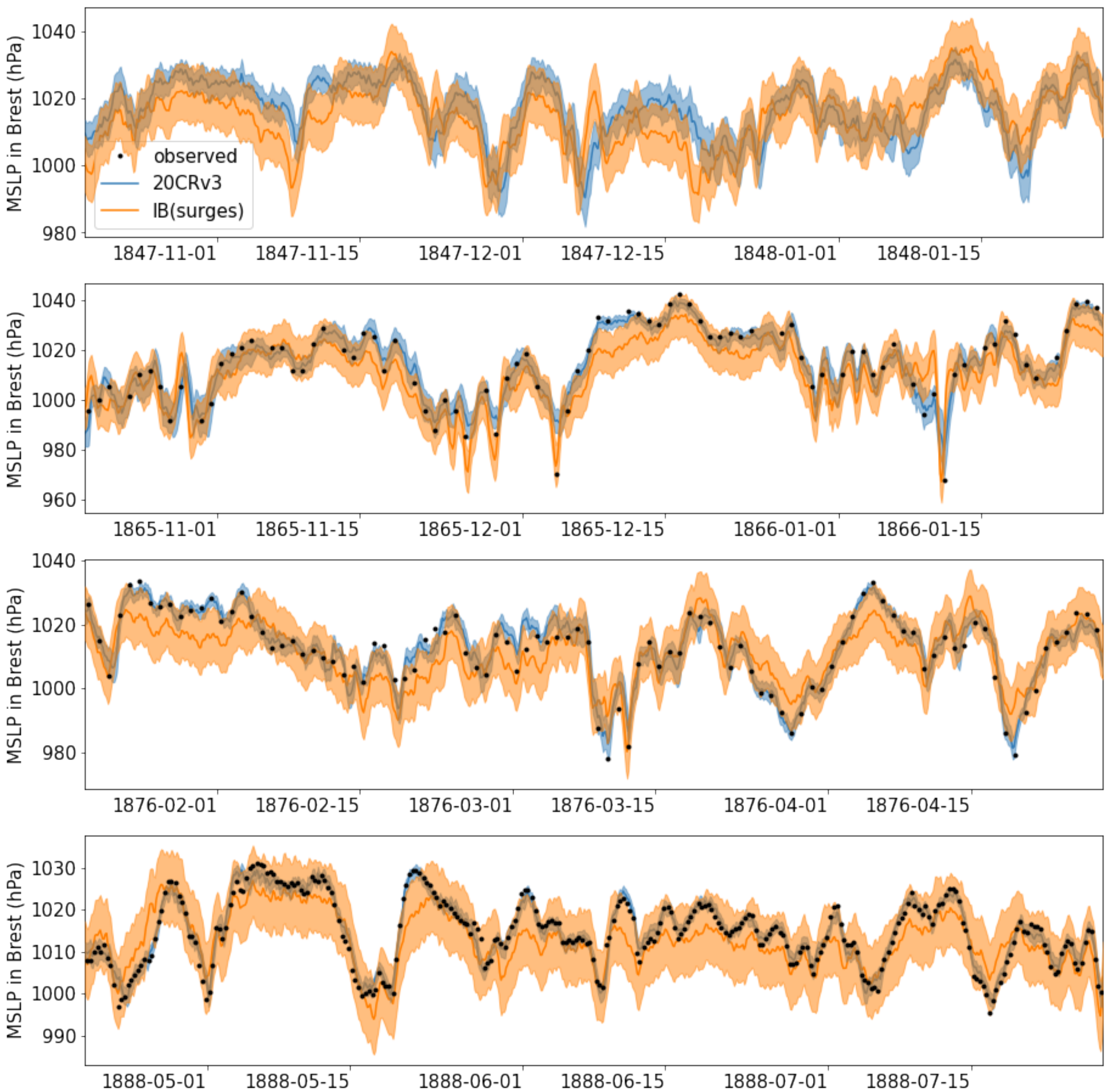

Figure 2: Reproducing a figure from the article but replacing the LLR with a simple corrected inverted barometer (linear regression between surge residuals and pressure anomaly, simply referred to as IB in the legend).

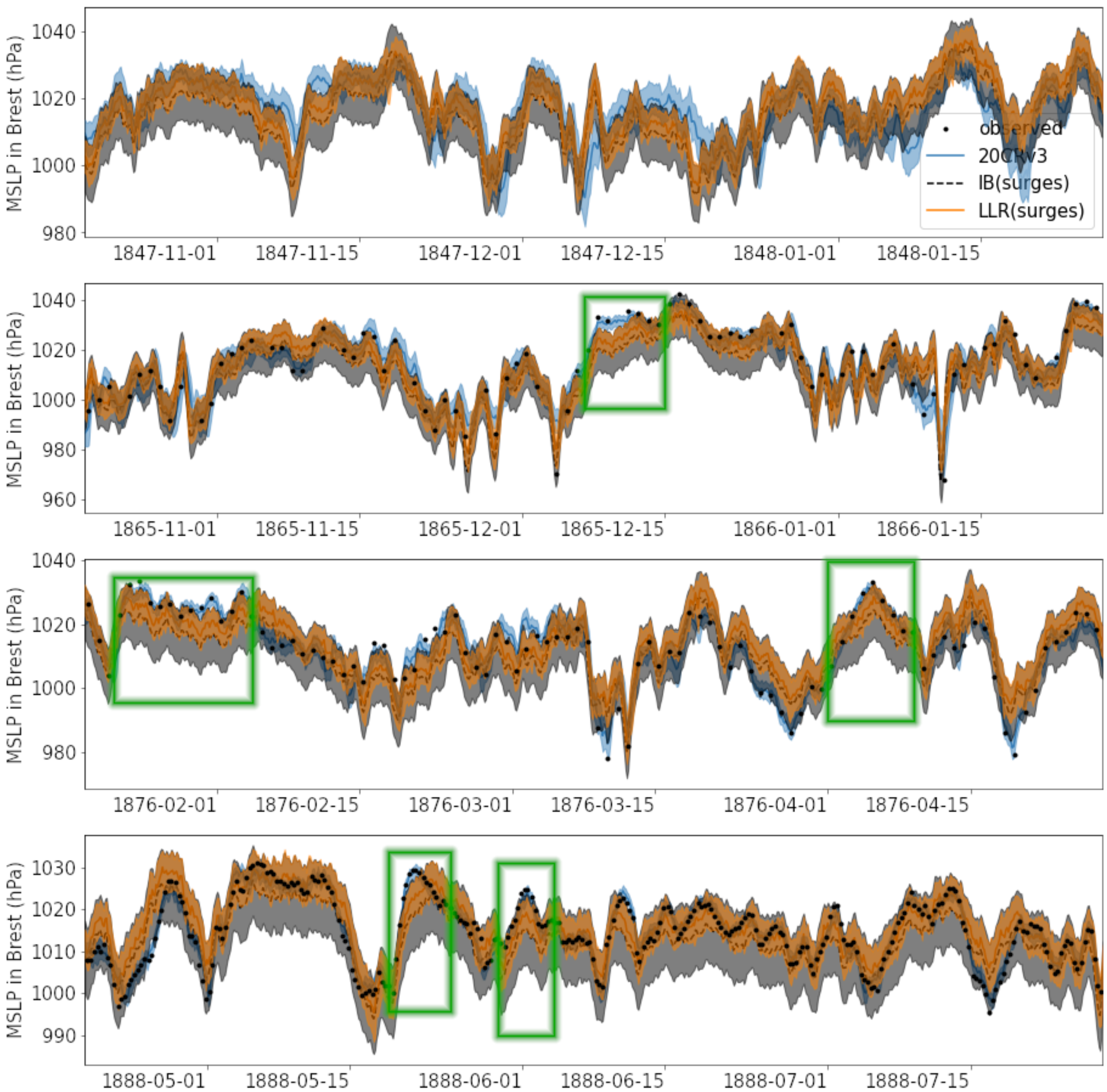

Figure 3: Same as previous figure, but showing all estimations at once. IB uncertainty ranges are in black, and LLR uncertainty ranges are in orange.