Could old tide gauges help estimate past atmospheric variability ?

Paul Platzer[1], Pierre Tandeo[2,3,4], Pierre Ailliot[5], and Bertrand Chapron[1,3]

[1]Laboratoire d'Océanographie Physique et Spatiale, Centre National de la Recherche Scientifique – Ifremer, Plouzané, France
[2]IMT Atlantique, Lab-STICC, UMR CNRS 6285, 29238, Brest, France
[3]Odyssey, Inria/IMT/CNRS, Rennes, France
[4]RIKEN Center for Computational Science, Kobe, 650-0047, Japan
[5]Laboratoire de Mathematiques de Bretagne Atlantique, Univ Brest, UMR CNRS 6205, Brest, France

Correspondence: Paul Platzer (paul.platzer@ifremer.fr)

Abstract. The surge is the non-tidal component of coastal sea-level. It responds to the atmospheric circulation, in particular through the direct effect of atmospheric pressure on the sea-surface. Tide gauges have been used to measure the sea-level in coastal cities for centuries, with many records dating back to the 19th-century or even further, at times when direct pressure observations were scarce. Therefore, these old tide gauge records may be used as indirect observations of sub-seasonal atmospheric variability, complementary to other sensors such as barometers. To investigate this claim, the present work relies on the tide gauge record of Brest, western France, and on the members of NOAA's 20th-century reanalysis (20CRv3) which only assimilates surface pressure observations and uses a numerical weather prediction model. Using simple statistical relationships between surges and local atmospheric pressure, we show that the tide gauge record can help to reveal part of the 19th-century atmospheric variability that was uncaught by the pressure-observations-based reanalysis, advocating for the use of early tide gauge records to study past storms. In particular, weighting the 80 reanalysis members based on tide gauge observations indicates that a large number of members seem unlikely, which induces corrections of several tens of hectopascals in the Bay of Biscay. Comparisons with independent pressure observations shed lights on the strength and limitations of the methodology, in particular for the case of wind-driven surges. This calls for the future use of a mixed methodology between data-driven tools and physics-based modelling. Our methodology could be applied to use other type of independent observations (not only tide gauges) as a means of weighting reanalysis ensemble members.

1 Introduction

Understanding the atmospheric system requires to understand all scales of variation, from daily to centennial. This cannot be done unless long observation records allow to disentangle these scales. The Twentieth Century Reanalysis Project, hereafter "20CR" (Compo et al., 2011), which is now in its third version, hereafter "20CRv3" (Slivinski et al., 2019), is the only atmospheric reanalysis that runs through the 19th century. It relies on the International Surface Pressure Databank Compo et al. (2019), the largest historical global collection of surface pressure observations, and the NCEP Global Forecast System (GFS) coupled atmosphere–land model.

Because it is the longest atmospheric reanalysis available, the 20CR reanalysis is used to study possible long-term trends in atmospheric dynamics (Rodrigues et al., 2018) or for extreme events (Alvarez-Castro et al., 2018). However, although the 20th-century part of 20CR has been compared with other reanalysis (Wohland et al., 2019) and observations (Krueger et al., 2013), comparisons with independent observations in the 19th century (Brönnimann et al., 2011) are scarce. The present work is an effort to compare this reanalysis with tide gauge observations. More generally and to our best knowledge, this paper is the first attempt to use old tide gauges as indirect observations of the atmosphere. However, the opposite direction has been taken by Tadesse and Wahl (2021), who extended surge reconstructions in the past using different atmospheric reanalysis products, in order to estimate past unobserved extreme surges.

Tide gauges are used primarily to measure the tide, which is the largest contributor to sea-level variations in many coastal cities. The astronomical tide is the result of gravitational attraction of the Sun and Moon on the ocean, combined with Earth's rotation. It results in periodic rise and fall of the water level (Melchior, 1983), which have been predicted through harmonic decomposition for centuries. Other physical phenomena impact the water level: a low atmospheric pressure results in a high sea-level, a well-known approximation of which is the "inverse barometer effect" (Roden and Rossby, 1999; Woodworth et al., 2019), and wind stress transport towards (respectively away from) the coast leads to increased (respectively decreased) sea-level. These conditions are usually associated with storms, which is why the associated sea-level variations are called "storm surges". For instance, in Brest (France), the amplitude of tidal variations is close to 4m, and surges can amount to as much as 1.5m.

Tide gauges are numerous, forming a dense global network in recent years, and a sparser one in the last centuries. As an example and from the GESLA-3 sea-level database (Haigh et al., 2023), 10 coastal tide gauge records start before 1907 in the North-American east coast, and 20 start before 1900 in Europe. Old tide gauges have varying observation frequencies, from hourly (Wöppelmann et al., 2006) to daily averages (Marcos et al., 2021). Although the sea-level measured by tide gauges is only an indirect tracer of atmospheric pressure variability, the scarcity of direct sea-level pressure measurements motivates the use of tide gauges to study past atmospheric fluctuations. Indeed, even when pressure measurements exist, they are often not yet digitized and even less available in global repositories (Brönnimann et al., 2019).

It is possible to link sea-level variations with atmospheric phenomena using physical laws and models (Lazure and Dumas, 2008), or using statistical tools (Quintana et al., 2021; Pineau-Guillou et al., 2023). This work adopts the second approach, but the underlying physical phenomena will often be used to motivate and interpret the statistical models. Local-linear regression (LLR) will be used to relate the surge to local mean-sea-level pressure. Hidden Markov Models (HMM) will allow to perform time-smoothing of probabilities given to members of 20CRv3, taking advantage of the time-continuity of each member. The use of a hidden Markov model to smooth the weighting of individual members of a reanalysis based on independent observations (here, tide gauge observations) was not reported elsewhere in the scientific literature. This general methodology could be used for other problems in order to assess and/or enhance available reanalysis products.

Note that a recent study by Hawkins et al. (2023) used tide gauge records to check the ability of the 20CR reanalysis to correctly model storms, in particular with the addition of recently digitized pressure observations. The study used a physics-based coastal model to estimate the surges associated with each member of the reanalysis, and compared to real observations.

One conclusion of the study is that the crude spatio-temporal resolution of the reanalysis is responsible for a systematic underestimation of the observed surges when using a direct physical coastal model forced by 20CR members. This justifies the use of statistical methods to quantify uncertainties in the relationship between reanalyzed pressures and real observed sea-levels. The present study is thus a first step towards using statistical models to assess reanalysis from tide gauge data.

The data and preprocessing are detailed in section 2. Section 3 outlines the local-linear regression and hidden Markov model used in this study. Section 4 shows the global consequences of applying our methodology while section 5 focuses on four specific events and compares with independent pressure observations. Conclusions on the proposed methodology and experiments are drawn in section 6, along with potential applications of this work.

2 Data

2.1 The Twentieth Century Reanalysis version 3 (20CRv3)

The Twentieth Century Reanalysis Project (Compo et al., 2011) aims at producing a global atmospheric reanalysis ending in 2015 and extending back to the 19th century. The present paper uses the latest version, 20CRv3 Slivinski et al. (2019), which extends up to 1806. It is an atmospheric reanalysis with 80 members, using an Ensemble Kalman Filter data assimilation scheme Evensen (2003). It has a temporal resolution of 3 hours, and uses a spectral triangular model in space with truncation of T254 (approximately 75km at the equator). There are 64 vertical levels, up to .3mb. It assimilates only surface pressure observation, from ships and fixed stations, as well as analysed cyclone-related IBTrACs data. These surface pressure observations are taken from the International Surface Pressure Databank (ISPD) which was created for the 20CR project but also exists as an independent product Compo et al. (2019). In 20CR, the sea-surface temperature and sea-ice cover are prescribed as boundary conditions. Sea-surface temperature and sea-ice cover both benefit from satellite observations from 1981 to 2015 (the end of the reanalysis), allowing more precise boundary conditions.

The surface pressure observation density is considerably lower in the 19th century than in the late 20th century. An online platform (https://psl.noaa.gov/data/20CRv3_ISPD_obscounts_bymonth) allows to consult the monthly observation count per $2^\circ \times 2^\circ$ box. Fig. 1 shows yearly averages of the number of surface pressure observations per day, comparing years 1870 and 2000. The maximum value was set to 24 observations per day although in 2000 this value is mostly exceeded. In year 1870, approximately half of Europe's land surface has no observation at all, and only less than 10 points have more than 10 observations per day. Observations coming from ships allow to raise the number of observations to approximately one per day on dense traffic areas. Conversely, in year 2000, virtually all of western Europe's land has more than 24 observations per day. Taking a spatial average over the whole map from Fig. 1 gives approximately one observation every three days in 1870, versus 44 observations per day in 2000. The number of available observations is also highly variable through time, especially in the 19th century. For instance, in the $2^\circ \times 2^\circ$ box centered on 49°-latitude, -5°-longitude, the number of monthly observations in 1870 ranges from 2 (January, 1870) to 85 (May, 1870), while in 2000 it ranges from 2152 (June, 2000) to 3242 (May, 2000).

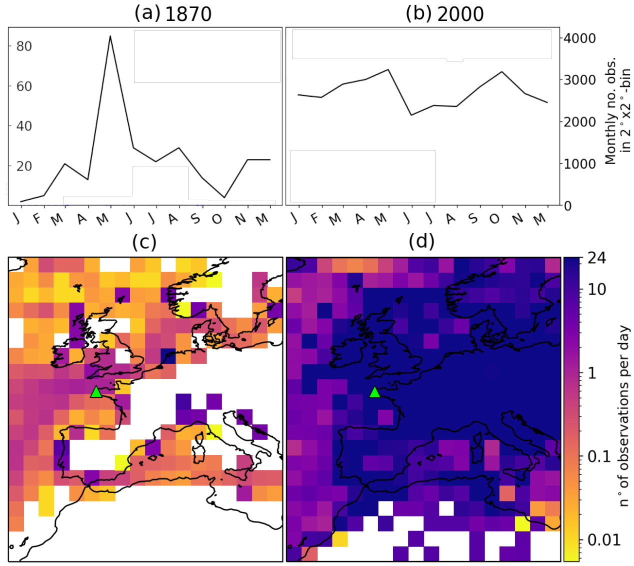

Figure 1. Number of surface pressure observations from the International Surface Pressure Databank (ISPD) assimilated in the Twentieth century Reanalysis version 3 (20CRv3). Top: Monthly in $2^\circ \times 2^\circ$ box centered on Brest, for years 1870 (a) and 2000 (b). Bottom: Yearly average of daily number of observation in 1870 (c) and 2000 (d).

2.2 Preprocessing of mean-sea-level pressure

In this work, we are using only the mean sea-level pressure (MSLP) variable from 20CRv3. We make two different preprocessings of this variable.

A first preprocessing is used for the statistical relationship between the local pressure and the surge. As the latter is driven by a physical phenomenon called the “inverse barometer effect” which will be introduced in the next section, we consider the difference between the MSLP interpolated at the city of Brest (4.49504°W, 48.3829°N), and the MSLP averaged over all members of 20CRv3 and over the North-Atlantic ocean (using the reanalysis’ land mask and averaging from 98°W to 12°E and from 0°N to 69°N), similarly to Ponte (1994). This spatial-averaged pressure is noted $\overline{\text{MSLP}}^{ocean}(t)$ and depends only on time.

A second preprocessing of MSLP is used to compute the probability of transition from one member of the reanalysis to another in the Hidden Markov Model (HMM) presented in section 3.2. For this purpose, we consider seasonal anomalies of MSLP with respect to a climatology computed from the period 1847-1890, because the HMM is run only for those years. The reference MSLP climatology for calendar day d and hour h is given by the average over days between $d-30$ and $d+30$, hours between $h-3$ and $h+3$, and all years 1847-1890. This reference MSLP is noted $\overline{\text{MSLP}}^{clim}$ and depends on latitude and longitude.

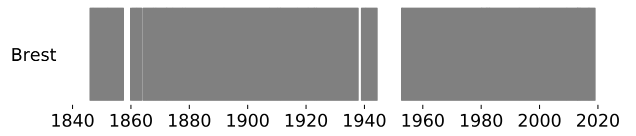

Figure 2. Brest tide gauge record availability trough time, from the GELSA-3 database (Haigh et al., 2023).

2.3 Tide gauge of Brest (France)

In this study, the tide gauge of Brest is used as indirect tracer of atmospheric circulation through surges. The availability of sea-level records in Brest in the GESLA-3 database is shown in Fig. 2. The Brest sea-level record from this database starts in 1846 and has a hourly sampling. Apart from a few large gaps, the record is mostly continuous during periods 1847-1920 and 1953-present. This combination of historical and modern records is at the foundation of the methodology exposed in the next section.

2.4 Preprocessing of sea-level

As mentioned earlier, the part of the sea-level which responds to atmospheric processes is the surge (also called "storm surge" or "skew surge"). 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 (Cazenave and Llovel, 2010). In this work, we are also interested in moving averages and differences of the surge. All these steps are exemplified in Fig. 3.

We first compute the tidal constituents of the raw sea-level (blue curve, Fig. 3.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 orange dashed line of Fig. 3.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, which allows to access the surge (orange dashed line of Fig. 3.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 surge which is noted $h(t)$ in the following:

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

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

Note from Fig. 3.b that the surge fluctuates at hourly scale, part of which are oscillations which are not due to variations in atmospheric pressure. 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. Such oscillations can dominate the surge signal in Brest where the

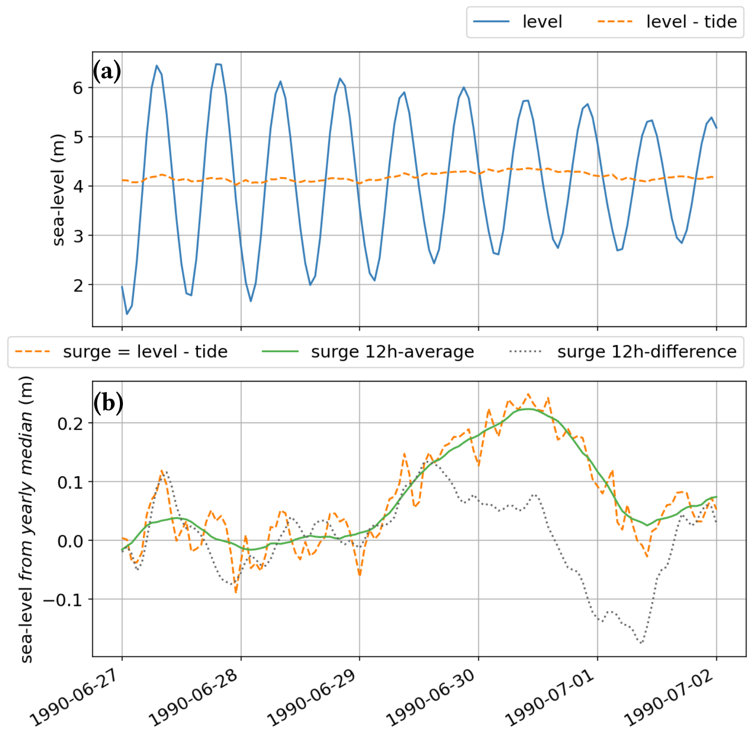

Figure 3. An example output of the different stages of preprocessing of the sea-level signal used in this work. (a) Raw level before (full, blue line) and after (dashed, orange line) removing the tidal part of the signal. (b) Sea-level after removing the yearly median value: the surge $h(t)$ (1h sampling, orange dashed curve), the centered 12h-average of the surge $\overline{h}^{12h}(t)$ (green full curve), and the 12h difference between 3h-averages of the surge $\Delta\overline{h}^{3h}(t)$ (gray dotted curve).

tidal amplitude is large. Furthermore, tide-surge interactions lead to stronger surges in low-tide and weaker surges in high-tide (Horsburgh and Wilson, 2007). As these phenomena are not linked to atmospheric processes, we chose to filter them out with a simple 12h-average (green full curve in Fig. 3). This also implies that these 12-hours-averaged surges will only respond to atmospheric events persisting for more than 12 hours. Given the spatial resolution of 20CRv3, smaller-scale events are likely not to be represented in the MSLP fields used in this study. In the following, we note $\overline{h}^{12h}(t)$ the 12h-average of the surge:

$$\overline{h}^{12h}(t) = \frac{1}{12} \sum_{t'=-6}^{t'=+6} h(t+t') \,. \tag{2}$$

Furthermore, as we are using sea-level observations to estimate atmospheric pressure, we also want to measure the amplitude of local time-variations of the surge. Indeed, as will be further explained in section 3.1, the sea-level response to variations of pressure depends on the time-scale of these variations. More precisely, the "inverse barometer" is an approximation that is only valid for slow variations of pressure. Accordingly, when observing fast variations of the surge, one expects deviations from the inverse barometer approximation. We therefore 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 before computing the difference. This difference is noted $\Delta\overline{h}^{3h}(t)$ and defined by the following equation:

$$\Delta\overline{h}^{3h}(t) := \frac{1}{3}\sum_{t'=-2}^{t'=+1}\left[h(t+t') - h(t-12+t')\right] . \tag{3}$$

2.5 Independent historical pressure observations

In section 5, we use pressure observations for the city of Brest to compare with 20CRv3 and with our estimate of pressure based on tide-gauge observations and our statistical model (local linear regression, LLR). We downloaded these observations from a repository[1] gathering historical pressure observations. These pressure observations come from the EMULATE project (Ansell et al., 2006) for years 1860-1880 and from Météo France archives for years 1858-1860 and from 1880 on. The EMULATE dataset has a daily sampling, while the Météo France archive dataset has a daily to thrice-daily sampling. These observations were not included in 20CRv3 and we did not use them to tune our model, they thus provide an independent validation dataset.

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 for each year to each value of the independent pressure observation datasets, so that the yearly average pressure are equal between the independent observed pressure and the 20CRv3 mean pressure linearly interpolated at the city of Brest. This yearly average correction varies between -2hPa and +2hPa during the period covered by the EMULATE dataset (1860-1880) and between +6hPa and +8hPa in the period covered by the Météo France dataset.

3 Statistical methods

3.1 Local Linear Regression (LLR) between surges and mean-sea-level pressure

To estimate the statistical relationship between surges in Brest and 20CRv3 mean-sea-level pressure, we use the period 1981-2015 during which satellite data is used in 20CRv3 to constrain sea-surface temperature and sea-ice cover, and a large number of pressure observations gives high confidence in 20CRv3 fields of mean-sea-level pressure (MSLP).

The filtered surges described in section 2.4 respond to sub-seasonal variations in atmospheric pressure. First, the sea-level is sensitive to pressure variations. An approximation called the “inverse barometer effect” (Roden and Rossby, 1999) states that an increase (respectively, decrease) of 1hPa in pressure at the mean sea-level leads to a decrease (respectively, increase) in sea-level of approximately 1cm. This approximation is valid only for slow variations of atmospheric pressure compared to the typical time of dynamic adjustment of the sea-level (Bertin, 2016).

[1] https://github.com/ed-hawkins/weather-rescue-data/tree/main

Moreover, the piling up of waters due to wind blowing perpendicular to the coast is responsible for positive (respectively, negative) surge when the wind stress transport is directed towards (respectively, away from) the coast. This effect depends non-linearly on the wind amplitude and direction (Bryant and Akbar, 2016; Pineau-Guillou et al., 2018). 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).

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. 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. In our case, since we want to estimate pressure based on the surges only, the effect of unknown wind or other processes must also be taken into account through uncertainty quantification.

Since our predictor variable is the sea-level measured by the tide gauge, we will use two proxies to estimate the conditional probability distribution of pressure: one is $\overline{h}^{12h}(t)$ and the other is $\Delta\overline{h}^{3h}(t)$. We expect that, for low absolute values of these two predictors, corresponding atmospheric pressure variations should be slow and moderate, and winds should be of low intensity, so that the inverse barometer approximation should hold. For larger absolute values of $\Delta\overline{h}^{3h}(t)$, indicating rapidly changing surges and thus likely rapidly changing atmospheric conditions as well, we expect deviations from the inverse barometer due to the dynamical adjustment of the sea-level. Similarly, the largest absolute values of $\overline{h}^{12h}(t)$ are likely to be caused by the added contribution of wind to the effect of pressure, so that deviations from the inverse barometer are expected as well.

To model all these effects, we use a local-linear regression (LLR in the following, see e.g. Fan, 1993; Hansen, 2022), also called kernel regression (Takeda et al., 2007). More precisely, we borrow our LLR from (Lguensat et al., 2017). In such a model, we will search for similar values (neighbours) of the two predictor variables $\overline{h}^{12h}(t)$ and $\Delta\overline{h}^{3h}(t)$ in the whole dataset, and compute a linear regression on this subset of the dataset. The predicted variable is $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{ocean}(t)$ where $\mathrm{MSLP}(t)$ is the value of the MSLP linearly interpolated at the city of Brest from the reanalysis.

We will assume that, conditionally to the values of $\overline{h}^{12h}(t)$ and $\Delta\overline{h}^{3h}(t)$, the predicted variable $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{ocean}(t)$ follows a Gaussian distribution:

$$\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{ocean}(t) \sim \mathcal{N}\left(m(t), \mathrm{var}(t)\right) \ . \tag{4}$$

We then assume, following Lguensat et al. (2017), that the average $m(t)$ and variance $\mathrm{var}(t)$ of this distribution can be estimated at each time step based on a local linear regression. To perform this local regression, we search for the K nearest neighbours of $\left[\overline{h}^{12h}(t), \Delta\overline{h}^{3h}(t)\right]$ in the satellite-era (1981-2015), where K is an integer set to 200 (other values have been tested and did not yield improvement on the results). The nearest neighbour criterion is the Euclidean distance in the two-dimensional space of values of $\left[\overline{h}^{12h}(t), \Delta\overline{h}^{3h}(t)\right]$. For each time t at which we want to estimate $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{ocean}(t)$, we thus find the set of times $\{t_i\,, i \in I(t)\}$ where $I(t)$ is an ensemble of size K, for which the following distance:

$$\mathrm{dist}(t,t_i) = \left[\left(\overline{h}^{12h}(t) - \overline{h}^{12h}(t_i) \right) + \left(\Delta\overline{h}^{3h}(t) - \Delta\overline{h}^{3h}(t_i) \right)^2 \right]^{1/2}$$

is minimal. To each index $i \in I(t)$, we attach a weight $\omega_i(t)$ according to the following formula:

$$\omega_i(t) = \frac{\exp(-\mathrm{dist}(t,t_i)^2/\lambda(t)^2)}{\sum_{j\in I(t)} \exp(-\mathrm{dist}(t,t_j)^2/\lambda(t)^2)}, \tag{5}$$

where $\lambda(t) := \mathrm{median}\,\{\mathrm{dist}(t,t_i)\,,\, i \in I(t)\}$ is defined as the median of the local values of the distances to the nearest neighbours (Lguensat et al., 2017).

Using this subset of the whole dataset, we compute a weighted linear regression between the subset of regressors $\overline{h}^{12h}(t_i)$, $\Delta\overline{h}^{3h}(t_i)$ and of predicted variable $\mathrm{MSLP}(t_i) - \overline{\mathrm{MSLP}}^{ocean}(t_i)$ using the weights $\omega_i(t)$. This regression has two linear coefficients noted $\alpha(t)$ and $\beta(t)$ and one intercept (constant value) noted $\gamma(t)$. Then, the average is given by applying the local weighted linear model to the actual value of the predictors:

$$m(t) = \alpha(t)\,\overline{h}^{12h}(t) + \beta(t)\,\Delta\overline{h}^{3h}(t) + \gamma(t) \tag{6}$$

while the variance is given by the weighted variance of the prediction error from the weighted linear model over the set of nearest neighbours:

$$\mathrm{var}(t) = \frac{1}{1-\sum_{i\in I(t)}\omega_i(t)^2} \sum_{i\in I(t)} \omega_i(t) \left(\mathrm{MSLP}(t_i) - \overline{\mathrm{MSLP}}^{ocean}(t_i) - \alpha(t)\,\overline{h}^{12h}(t_i) - \beta(t)\,\Delta\overline{h}^{3h}(t_i) - \gamma(t) \right)^2 \tag{7}$$

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 (this is called the "leave-one-out" procedure). Then, we compare the average $m(t)$ with the true value $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{ocean}(t)$ in a scatter-plot (Fig. 4.a). This figure shows that the LLR is able to predict good average values $m(t)$ for moderate absolute values of pressure difference, although it consistently underestimates the most extreme values: this behaviour is expected as the method is limited by the observations it has seen previously. However, as will be seen in section 5, this simple model is still able to capture storms. Then, we test the adequacy of our variability estimate with the parameter $\mathrm{var}(t)$, by checking that the following shifted-rescaled variable:

$$\frac{\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{ocean}(t) - m(t)}{\sqrt{\mathrm{var}(t)}}, \tag{8}$$

follows a standard Gaussian distribution with average 0 and variance 1. To do so, we compare the empirical histogram of this variable with the probability density function of a standard Gaussian distribution, as shown in Fig. 4.b. Although the shape

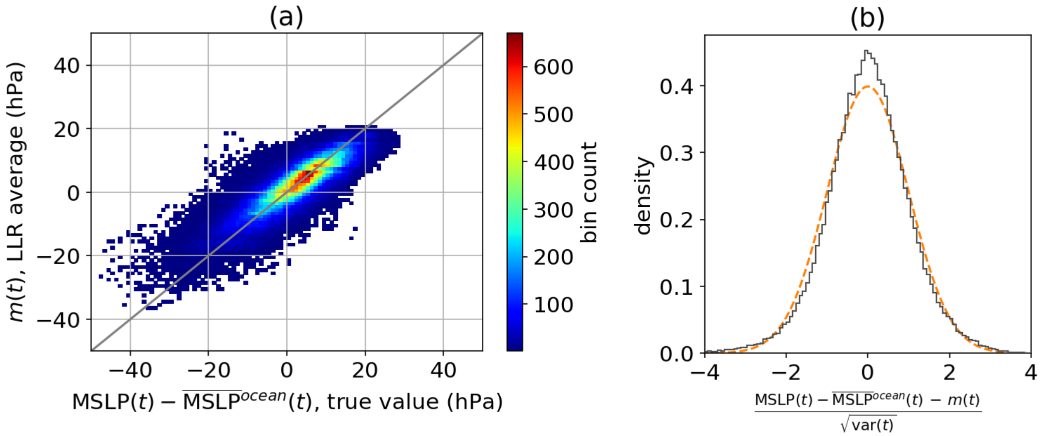

Figure 4. Evaluation of the local-linear regression (LLR) on the period 1981-2015 using a leave-one-out procedure. (a) Histogram scatter-plot of the average estimate from the LLR versus true value of the MSLP difference with the reference ocean-averaged MSLP. (b) Density-histogram of normalized average error of the LLR estimate (black line) compared to theoretical probability density function of standard Gaussian random variable (dashed orange line).

of the histogram slightly differs from a Gaussian probability density function (it is more peaked and has heavier tails), the agreement is satisfying enough for the purpose of this article. This shows, in particular, that the estimate of variance through $\mathrm{var}(t)$ is consistent with the real variability of the estimation process, which is the reason why we advocate for using a statistical method in the first place.

3.2 Hidden Markov Model (HMM)

In the 19th century, the spread between 20CRv3 members is much larger than in the period 1981-2015. One of the aims of this work is to estimate conditional probabilities of each member of the reanalysis, based on surge values in Brest. Note that in the reanalysis, the members are assumed to have uniform probabilities, that is a probability of 1/80, since we have 80 members.

One can estimate conditional probabilities of each member at time t based on the values of $[\overline{h}^{12h}(t)\ ,\ \Delta\overline{h}^{3h}(t)]$. To do that, we use the satellite-era-derived local-linear regression expressed in section 3.1. The average $m(t)$ and variance $\mathrm{var}(t)$ are estimated based on the procedure described in section 3.1 and using the dataset from the period 1981-2015 to search for neighbours of $[\overline{h}^{12h}(t)\ ,\ \Delta\overline{h}^{3h}(t)]$ and compute the LLR.

To differentiate these member probabilities from the ones we will derive later on using a hidden Markov model, we use the notation $p_{\text{HMM}}(i,t)$ for the probability of member i at time t.

$$\begin{cases} p_{\text{HMM}}(i,t) \propto \exp\left\{-\frac{\left(\mathrm{MSLP}(i,t)-\overline{\mathrm{MSLP}}^{ocean}(t)-m(t)\right)^2}{2\,\mathrm{var}(t)}\right\}, \\ \sum_{i=1}^{80} p_{\text{HMM}}(i,t) = 1\,. \end{cases} \tag{9}$$

We also use the convention that, in the absence of surge observations, all members are given equal probabilities $p_{\cancel{\mathrm{HMM}}}(i,t) = 1/80$. Although these probabilities already bear significant information, they have the undesirable property to be time-discontinuous. This is not coherent with the fact that the members of 20CRv3 are time-continuous: they are propagated in time using a NWP model. To remedy this issue, we compute smoothed (or reanalyzed) probabilities using a hidden Markov model (HMM) detailed below, which we write $p_{\mathrm{HMM}}(i,t)$:

$$p_{\mathrm{HMM}}(i,t) := P\left(\mathrm{member}(t) = i \,\middle|\, \begin{bmatrix} H(t=1) \\ \vdots \\ H(t=T) \end{bmatrix}\right), \tag{10}$$

where one uses an observational record of surges from time-index 1 to T and we use the simple notation $H(t) := [\overline{h}^{12h}(t), \Delta\overline{h}^{3h}(t)]$ for vector of surge average and difference. $p_{\mathrm{HMM}}(i,t)$ is a time-smoothed version of $p_{\cancel{\mathrm{HMM}}}(i,t)$ which takes into account past and future values of the surge. For this purpose, a simple Hidden Markov Model (HMM) is used. The first ingredient of the HMM is the transition matrix $\mathcal{T}_{ij}(t)$ from member i at time $t-1$ to member j at time t.

$$\mathcal{T}_{ij}(t) := \; P\left(\mathrm{member}(t) = j \mid \mathrm{member}(t-1) = i\right). \tag{11}$$

To estimate the transition matrix, a strong hypothesis is made:

$$\mathcal{T}_{ij}(t) \propto \; K_\theta\left(MSLP_{\mathrm{map},j}(t), MSLP_{\mathrm{map},i}(t)\right), \tag{12}$$

where $MSLP_{\mathrm{map},i}(t)$ is the i-th member's map of mean-sea-level pressure in a squared box of 18°W $\leq$ lon $\leq$ 18°E, 28°N $\leq$ lat $\leq$ 64°N at time t, and $K_\theta(\cdot,\cdot)$ is a positive real-valued function that measures the similarity between $MSLP_{\mathrm{map},i}(t)$ and $MSLP_{\mathrm{map},j}(t)$ and depends on parameters θ.

Eq. (12) states that transitions from one member to another are more likely if the associated MSLP map at time t are similar. This prevents abrupt transitions to dissimilar atmospheric states. The size and location of the map was chosen to cover an area inside which storms and anticyclones which affect the surges in Brest would lie. Ideally, $K_\theta(\cdot,\cdot)$ should be symmetric, semi-definite. Here, a simple Gaussian kernel of Euclidean distances is used, with normalization factor $\theta > 0$, so that for two fields X and Y:

$$K_\theta(X,Y) = \exp\left\{-\sum_{n\,\in\,\mathrm{lons}} \sum_{l\,\in\,\mathrm{lats}} \frac{(X_{nl} - Y_{nl})^2}{\theta^2}\right\}, \tag{13}$$

where the sum over n and l represents a sum over longitudes and latitudes. We then define Θ through the following equation:

$$\frac{\theta}{\Theta} = \overline{\left(\frac{1}{80}\sum_{i=1}^{80} \frac{1}{T}\sum_{t=0}^{T} MSLP_{\mathrm{map},i}(t)^2\right)^{1/2}}, \tag{14}$$

where the overbar denotes spatial average. This normalization will allow to optimize θ through grid search of Θ for a maximum of likelihood of the surge observations.

One can compute $\mathcal{T}_{ij}(t)$ by setting a value of θ and using the hypothesis of Eq. (12) along with the fact that for all i, t, we have $\sum_j \mathcal{T}_{ij}(t) = 1$. This then allows to estimate $p_{\text{HMM}}(i,t)$ with the forward-backward algorithm (Rabiner, 1989). Let:

$$a_i(t) := P\left(\begin{bmatrix} H(1) \\ \vdots \\ H(t) \end{bmatrix}, \text{member}(t) = i \right), \tag{15}$$

$$b_i(t) := P\left(\begin{bmatrix} H(t+1) \\ \vdots \\ H(T) \end{bmatrix} \middle| \text{member}(t) = i \right). \tag{16}$$

These two quantities can be computed recursively, following the forward procedure:

$$a_i(1) = p_{\text{HMM}}(i,1), \tag{17}$$

$$a_i(t+1) = p_{\text{HMM}}(i,t+1) \sum_{j=1}^{80} a_j(t) \mathcal{T}_{ji}(t), \tag{18}$$

and the backward procedure:

$$b_i(T) = 1, \tag{19}$$

$$b_i(t) = \sum_{j=1}^{80} b_j(t+1) \mathcal{T}_{ij}(t) p_{\text{HMM}}(j,t+1). \tag{20}$$

Finally, this allows to estimate $p_{\text{HMM}}(i,t)$ by noting that:

$$p_{\text{HMM}}(i,t) = \frac{P\left(\text{member}(t) = i, \begin{bmatrix} H_1 \\ \vdots \\ H_T \end{bmatrix} \right)}{P\left(\begin{bmatrix} H_1 \\ \vdots \\ H_T \end{bmatrix} \right)}, \tag{21}$$

which gives, in terms of $a_i(t)$ and $b_i(t)$:

$$p_{\text{HMM}}(i,t) = \frac{a_i(t)b_i(t)}{\sum_{j=1}^{80} a_j(t)b_j(t)} \tag{22}$$

while keeping in mind that Eq (22) implicitly relies on hypothesis (12) and a fixed form of K_θ.

Comparing $p_{\text{HMM}}(i,t)$ with the uniform distribution $p(i,t) = \frac{1}{80}$ allows to see if the surge observations are coherent with the MSLP fields of 20CRv3 (section 4) and to select the most relevant members given surge data (section 5).

To choose the parameter θ, we performed a grid-search of its normalized form, Θ, computed the log-likelihood of the surge observations as an output of the algorithm. Indeed, the log-likelihood $l_\theta(0\dots T)$ is expressed as follows:

$$l_\theta(0\dots T) = \log(\sum_{i=1}^{80} a_i(T)) \; . \tag{23}$$

Figure 5 shows variations of this quantity with Θ, for one year (1885) of surge observations in Brest (i.e. $t = 0$ is 01 January 1885 and T is 01 January 1886). The curve shows a distinct maximum around $\Theta \approx 0.09$, and plateaus for higher values. According to Figure 5, the difference of log-likelihood between the model without HMM ($\theta = +\infty$) and with HMM is close to 1000. The introduction of one extra parameter in the filtering model compared to the static one is thus clearly justified if the two models are compared using standard criteria usch as AIC, BIC or likelihood ratio tests Zucchini et al. (2017).

Note that in the limit $\Theta = +\infty$, we have a constant transition probability $\mathcal{T}_{ij}(t) = \frac{1}{80}$ and $p_{\text{HMM}}(i,t)$ reduces to $p_{\text{\sout{HMM}}}(i,t)$. Figure 5 thus supports the use of the HMM to estimate probabilities of MSLP map conditioned by surge observations.

The choice of restricting the estimation of log-likelihood to one arbitrary year (1885) is supported by the fact that estimation of $\mathcal{T}_{ij}(t)$ is computationally expansive. We assume that the optimal value of θ generalizes well to other years. A better optimization of θ would necessitate further work that is out of the scope of this study. Setting $\Theta = 0.09$ will already enable us to find interesting features of $p_{\text{HMM}}(i,t)$.

4 Modification of 20CRv3 ensemble when accounting for surges

This section is devoted to the study of $\delta\mu_{\text{HMM}}(t)$, the difference between weighted and unweighted ensemble average, defined by:

$$\delta\mu_{\text{HMM}}(t) := \sum_{i=1}^{80} \left(p_{\text{HMM}}(i,t) - \frac{1}{80} \right) MSLP_{\text{map},i}(t) \; , \tag{24}$$

where $MSLP_{\text{map},i}(t)$ is a short notation for the sea-level pressure field of 20CRv3's i-th member. $\delta\mu_{\text{\sout{HMM}}}(t)$ is defined equivalently using $p_{\text{\sout{HMM}}}(i,t)$. This quantity shows how strong is the average deviation when taking into account surge observations. It will also be sometimes normalized by $\sigma_{\text{20CR}}(t)$, the estimated standard deviation of the unweighted ensemble:

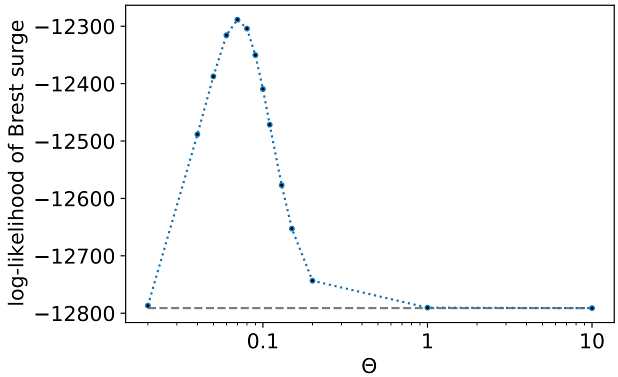

Figure 5. Log-likelihood of Brest surges (dotted blue line) as a function of parameter Θ defined in Eq. (14). For comparison, the log-likelihoood of the simple model without Hidden Markov Model ($\Theta \to +\infty$) is also shown (dashed grey line). The log-likelihood was estimated using data from year 1885 for a first estimation of optimal parameter Θ. Values of Θ used for estimation are 0.02, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.13, 0.15, 0.2, 1, and 10, with 0.07 giving the largest log-likelihood.

$$\sigma_{\text{20CR}}(t) := \left[\frac{1}{79} \sum_{i=1}^{80} \left(MSLP_{\text{map},i}(t) - \frac{1}{80} \sum_{i=1}^{80} MSLP_{\text{map},i}(t) \right)^2 \right]^{1/2} . \tag{25}$$

Note that in this definition, $\sigma_{\text{20CR}}(t)$ depends on time, latitude and longitude. Therefore at each grid point and for each time step the quantity $\delta\mu_{\text{HMM}}(t)$ will be normalized by a different value, indicating the strength of the reanalysis ensemble spread at this location in time and space.

To further interpret the result of our HMM algorithm, we introduce the filtered effective ensemble size $\nu_{\text{HMM}}(t)$ (Liu, 1996):

$$\nu_{\text{HMM}}(t) := \frac{1}{\sum_{i=1}^{80} p_{\text{HMM}}(i,t)^2} , \tag{26}$$

and we define equivalently $\nu_{\not{\text{HMM}}}(t)$. These quantities are estimates of the number of ensemble members that can be retained according to surge observations, assuming one discards very unlikely members.

In Fig. 6, variables $\delta\mu_{\text{HMM}}$, $\delta\mu_{\text{HMM}}/\sigma_{\text{20CR}}$ and ν_{HMM} are shown as a function of time for the period 1846-1890. All these quantities show a strong seasonality. This is due to a much stronger MSLP variability in winter, and a corresponding stronger response of the surges. The figure shows that the amount of correction $\delta\mu_{\text{HMM}}$ and the decrease in ensemble size ν_{HMM} are much stronger using smoothed probabilities with HMM rather than probabilities without HMM. Showing the deviation $\delta\mu_{\text{HMM}}$ in the Bay of Biscay, where the standard deviation of $\delta\mu_{\text{HMM}}/\sigma_{\text{HMM}}$ is strongest (see Fig. 7), substantial absolute values of ~600 Pascals are obtained in early 1850s winters, even after averaging over 3 months. These large deviations correspond to more than one standard deviations of the ensemble size. Using probabilities without HMM, deviations are weaker but still non-negligible ($\sim 500Pa$, $\sim 0.7\sigma$). The slow decrease in $\delta\mu_{\text{HMM}}$ with time is coherent with slowly increasing observations

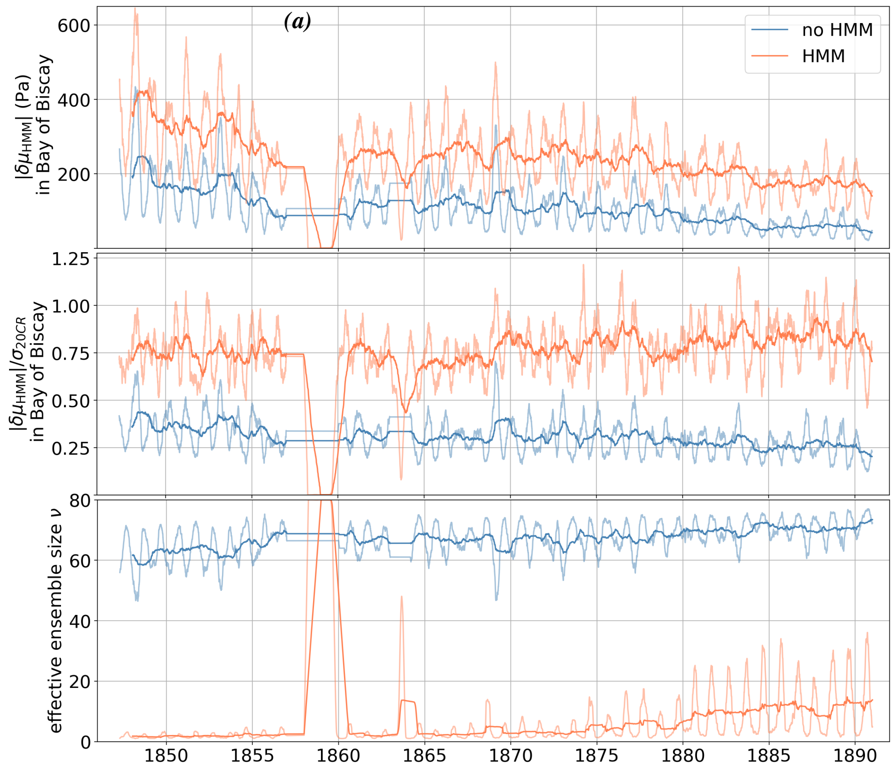

Figure 6. (a) Average MSLP deviation $\delta\mu_{\mathrm{HMM}}$ in Pascals, in the center of the Bay of Biscay. (b) Same as (a) normalized by reanalysis ensemble standard deviation $\delta\mu_{\mathrm{HMM}}/\sigma_{\mathrm{20CR}}$. (c) Effective ensemble size ν_{HMM}. Orange: using smoothed probabilities with HMM $p_{\mathrm{HMM}}(i,t)$. Blue: using probabilities without HMM $p_{\cancel{\mathrm{HMM}}}(i,t)$. Bold: yearly average. Thin: 3-month average. All plots make simultaneous use of data from Brest tide gauge.

used in 20CRv3, although with substantial decadal variations. However, $\delta\mu_{\cancel{\mathrm{HMM}}}/\sigma_{\mathrm{HMM}}$ and $\delta\mu_{\mathrm{HMM}}/\sigma_{\mathrm{HMM}}$ do not show a clear trend, indicating a persisting gain in information from surge observations throughout the 19th century.

In terms of effective size, Fig. 6 shows that the smoothing HMM algorithm imposes a strong member selection, with mostly only 1 member retained at each time step, in winter and before 1880. Probabilities without HMM mostly retain more than half of the members, although peak low values of $\nu_{\cancel{\mathrm{HMM}}}(t)$ show that even without the HMM sometimes more than half of the ensemble members are highly unlikely. Filtered effective ensemble size reaches very low yearly and seasonal average values, indicating that many 20CR members are highly unlikely with respect to surge estimates from tide gauge observations. A strong increase in $\nu_{\mathrm{HMM}}(t)$ is witnessed around year 1880. This can be explained by the availability of a large number of weather station data in Eastern Europe and Russia from 1880-on, and by an intensification of maritime traffic around 1880.

The spatial structure of $\delta\mu$ is examined in Fig. 7. The analysis of time-standard-deviation of $\delta\mu_{\cancel{\mathrm{HMM}}}$ and $\delta\mu_{\mathrm{HMM}}/\sigma_{\mathrm{20CR}}$ shows that the area of greatest influence of the corrections from surge-smoothing from Brest tide gauge is in the Bay of Biscay. This can be explained by the passage of strong storms in the Bay of Biscay, which can cause high surges in Brest, and by

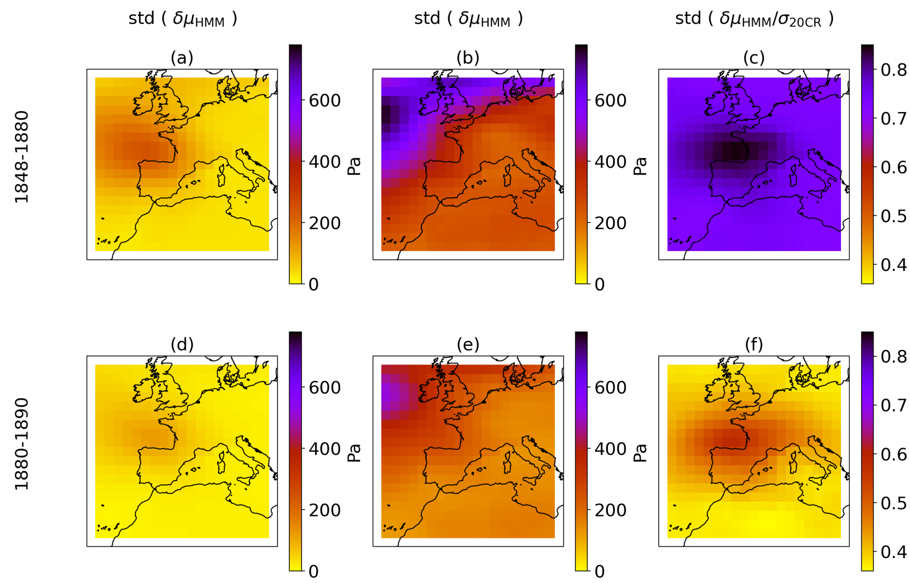

Figure 7. Time-standard deviation of $\delta\mu_{\sout{\mathrm{HMM}}}$ (a,d), $\delta\mu_{\mathrm{HMM}}$ (b,e) and $\delta\mu_{\mathrm{HMM}}/\sigma_{\mathrm{20CR}}$ (c,f) computed from October to March, for years 1848-1880 (a,b,c) and 1880-1890 (d,e,f). Probabilities $p_{\sout{\mathrm{HMM}}}$ and p_{HMM} make use of data from Brest tide gauge.

the sparsity of direct pressure measurements (ship logs) in this area in the 19th century. Standard deviation of $\delta\mu_{\mathrm{HMM}}$ shows largest values to the north-west of the map, which is where strong storms travel. Indeed, the variability of MSLP shows a great north-west gradient, as can be seen from maps of time-standard-deviation of 20CRv3 mean MSLP (not shown). Noticeably, the size of the area of influence of $\delta\mu_{\mathrm{HMM}}$ is smaller in 1880-1890, which can be explained by a greater conditioning of 20CRv3 members by observations, both offshore and in-land. In case of very sparse observations used in 20CRv3, the area of influence of these corrections widens due to continuity of MSLP fields. Note, as well, that the area of influence is greater for $\delta\mu_{\mathrm{HMM}}$ then for $\delta\mu_{\sout{\mathrm{HMM}}}$, because of the time-propagation of corrections thanks to the smoothing HMM algorithm. Finally, this figure confirms the great difference in amplitude of deviations between pre-1880 and post-1880 corrections, already witnessed in Fig. 6. Similar spatial footprints can be witnessed from maps of high and low quantiles of $\delta\mu$, only with different values (not shown). Similarly, computing the time-standard deviations as in Fig. 7 but restricting the times used for computation to April-September rather than October-March shows the same spatial pattern but with much lower values (not shown).

These corrections also have a strong decadal variation, with non-trivial yearly averages persisting for several years, as shown in Fig. 8. The same behaviour can be witnessed for the surge, which is strongly anti-correlated to these deviations (Fig. 8). This can be explained by the fact that 20CRv3 smooths MSLP values in areas of sparse measurements, and that surge-filtering corrections allow to retrieve more realistic intense values (either positive or negative). This interannual variability is related to the variability in storminess (Bärring and Fortuniak, 2009).

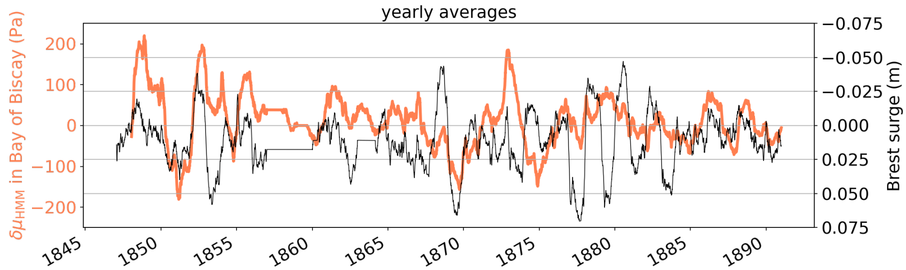

Figure 8. Orange, bold: yearly-average of $\delta\mu_{\mathrm{HMM}}$ at position 47.83°lat, -7.57°lon, in the Bay of Biscay, using data from Brest tide gauge. Black, thin: Brest surge (inverted sign).

5 Focus on four 19th-century events

One of the aim of this study is to show that old tide gauge data can be used to better understand past severe storms. In this section, two storms and one mild situation are studied for illustrative purposes.

To better understand the more general context of the three events studied in this section, we first look at longer time periods (100 days) surrounding the events, and compare the results of the simple LLR based on surges with 20CRv3 and independent observations (when available). These are plotted in Fig. 9. One can thus see that the uncertainties associated with the surge-based LLR does not vary from year to year, while that of 20CRv3 decreases. More precisely, the ratio of the standard deviation of the LLR divided by that of 20CRv3 has an average value of 1.22 in 1847, 1.54 in 1865, 1.95 in 1876, and 2.45 in 1888. That same ratio has a minimum value of 0.47 in 1847 and in 1865, while its minimum value is of 0.69 in 1876 and 0.94 in 1888. This shows that on average the reanalysis has lower uncertainty than the surge-based pressure estimate although with fluctuations, and in 1888 the uncertainty of the reanalysis is always smaller. Comparison with observations also confirms the better precision of the reanalysis.

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. In 1876 (Fig. 9.c), biases of $\sim$5hPa between the LLR and 20CRv3 are found most of the time. For all four periods shown, the reanalysis and the LLR pressure estimates show consistent variations in time, although with persistent biases (either positive or negative) that last from a few days to $\sim$ 15 days. 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. These examples show that the results of our algorithm must be interpreted with care, and that a more in-depth analysis is needed to understand the specifics of an individual event.

Our claim that the wind variations are responsible for the persistent biases between the LLR pressure estimation and the reanalysis is supported by Fig. 13.f, where we also show the direction and amplitude of the 10m-wind intensity as given by the average over all reanalysis members and interpolated at the city of Brest. In March 1876, two low-pressure systems passed to the North of Brest's tide gauge, one around March 10th and a second around March 12th, as indicated by the reanalysis members and the independent pressure observations (Fig. 13.e). However, the first low-pressure system did not induce a surge as strong as the second one. One key difference between the two events is the wind amplitude, which reached 15m/s during the first event and then decreased to 5-10m/s during the second event, with almost steady wind direction. Although wind intensity and direction estimated from the reanalysis must be taken with care, the value of 15m/s is rarely exceeded (only 7 in 1000 times in the period 1981-2015, not shown), indicating exceptional wind intensity during the event, and justifying the inaccuracy of the LLR which is based on already observed events and therefore biased towards typical wind conditions. Our interpretation relies on the fact that the effect of wind on extreme surges acts at small time scales (daily or sub-daily), which is backed by recent work (Pineau-Guillou et al., 2023).

To aid the interpretation of Figures 11, 12, 13 and 14, we also show in Fig. 10 the number of observations assimilated in 20CRv3 in the months of the studied events. In November 1847, observations mostly come from ground stations, indicated by green-blue squares (more 1 observation per day). In November 1865, some more stations are available, and observation density from maritime traffic also grows. In March 1876 and August 1888, the number of observations surrounding Brest increases with respect to 1865 mostly due to an intensification of maritime traffic, although some new stations also constrain the reanalysis but not in the direct vicinity of Brest.

One common feature of Fig. 11, 12 and 13, is that the HMM algorithm tends to be very selective compared to the weighing without HMM. This is the consequence of our optimisation of the parameter θ with the objective of maximizing the likelihood of the surge observations on the 20CRv3 ensemble. Having a low theta allows to give a high weight to the ensemble member which has the highest probability according to the surge-based LLR model. However, as is obvious from comparing Fig. 11.a with Fig. 11.b, Fig. 12.a with Fig. 12.b, and Fig. 13.a with Fig. 13.b, this does not always have a strong influence on the average MSLP field. In the case of Fig. 14, the variability between members of the reanalysis is smaller, and therefore the selection of ensemble members is less acute, with more reasonable effective ensemble sizes. However, again, these figures show that small effective ensemble sizes should not be interpreted as a justification for discarding members with low probability according to the HMM algorithm, but rather as a means to quantify the relative agreement of individual members with the surge observations according to the LLR statistical relationship.

The fact that one member is often much more coherent with the series of surges is the result of 1. the high variability of the ensemble and of the LLR pressure estimation 2. the high dimensionality (or complexity) of the problem, 3. the low size of the ensemble (80 members) and finally 4. the systematic biases between 20CRv3 and the LLR caused by unmodeled atmospheric conditions (winds). Indeed, in case of data scarcity, the variability of the reanalysis is large (point 1.), and a fixed-size ensemble (point 2.) may struggle to correctly span the whole space of possible atmospheric circulations (point 3.), so that a few members are actually much closer to the true atmospheric circulation than all other members. Such a problem is called filter degeneracy (Snyder et al., 2008) and is a common issue in ensemble-based data assimilation schemes. Secondly, since our LLR estimation of pressure experiences time-correlated biases with respect to the 20CRv3 because of unmodeled other variables (winds), this causes the HMM to select the one member which is closest to the biased pressure estimate from the LLR applied to the surge signal. All these issues may remain for other climate-science applications if one uses a similar approach of merging independent observations with a HMM algorithm to weight ensemble members.

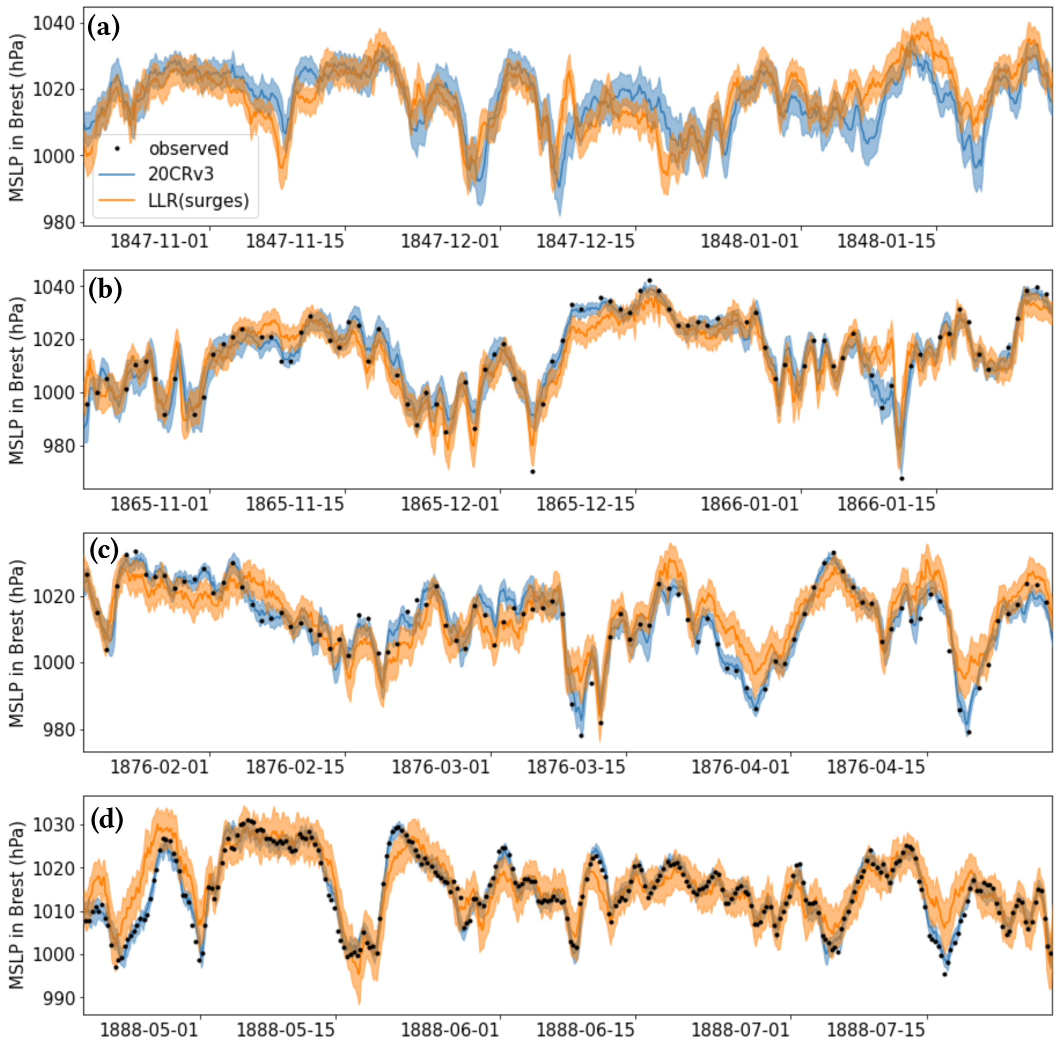

Figure 9. Comparison of MSLP estimation in Brest from 20CRv3 (blue), LLR based on surges (orange), and independent observations (black dots) that were not used to build the orange and blue curves, for three periods surrounding the events studied in this section. Full lines correspond to average values while shaded areas correspond to ± one standard-deviation around the average.

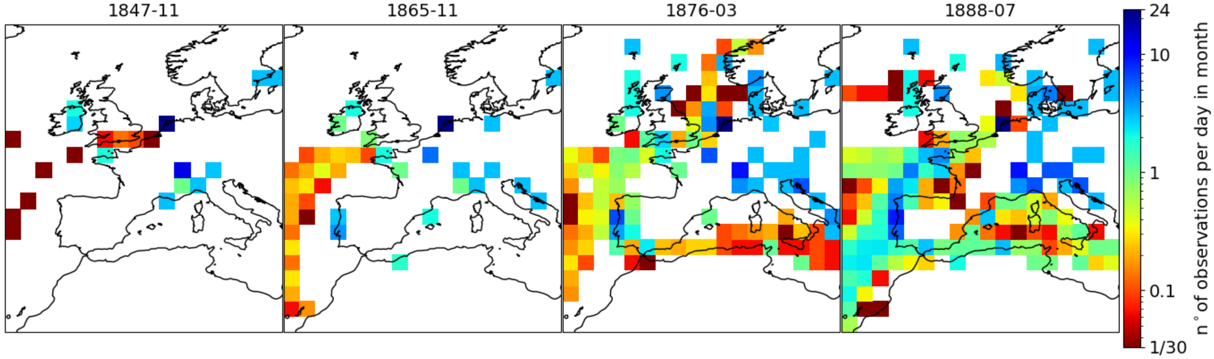

Figure 10. Number of pressure observations in month, divided by 30 (or 31) to give an average number of observations per day, for months 1847-11, 1865-11 and 1888-07.

6 Conclusion and perspectives

This study is a proof of concept for the use of century-old tide gauge data as a means of understanding past atmospheric subseasonal variability. Surges of Brest allow to assess part of the atmospheric variability that was uncaught in global 20CR reanalyses based on pressure observations. Weighing 20CR members according to surge observations reduces the effective ensemble size, and implies significant deviations in members-averaged sea-level pressure in the Bay of Biscay. Through the second-half of the 19th century, these deviations diminish and the effective ensemble size rises, however they remain non-negligible. Independent pressure observations in the city of Brest are coherent with pressure estimations from the reanalysis and the surge-based local-linear relationship. Such comparisons also show that the reconstruction of pressure based on surges is ambiguous due to the influence of winds, so that biases between the surge-base and the reanalysis-based pressure estimates can last for several days.

This work has several potential applications. First, replicating this work with other tide gauges could help to validate reanalyses like 20CRv3 against independent data, and to potentially identify anomalous trends or wrong estimation of specific events. Combining our statistical approach with the physics-based approach of Hawkins et al. (2023) could allow to have both a precise estimate from a high-fidelity coastal model and a good quantification of uncertainties. Second, tide gauges could be used to constrain regional scale atmospheric simulations in order to better estimate the magnitude and spatial extent of known past severe storms. Third, tide gauge records could be combined with direct observations of atmospheric pressure to give statistical estimates of atmospheric fluctuations in the 19th century without the use of a Numerical Weather Prediction model, such as the optimal interpolation of Ansell et al. (2006) based on direct pressure observations only, or the analogue upscaling of Yiou et al. (2014) for the short period 1781-1785 of dense observations in western Europe. Finally, this work could be replicated in a more general context, using other types of variables and observations, learning the relationship between observations and large-scale features using recent observations and precise reanalyses, and applying these statistical relationship in the past to

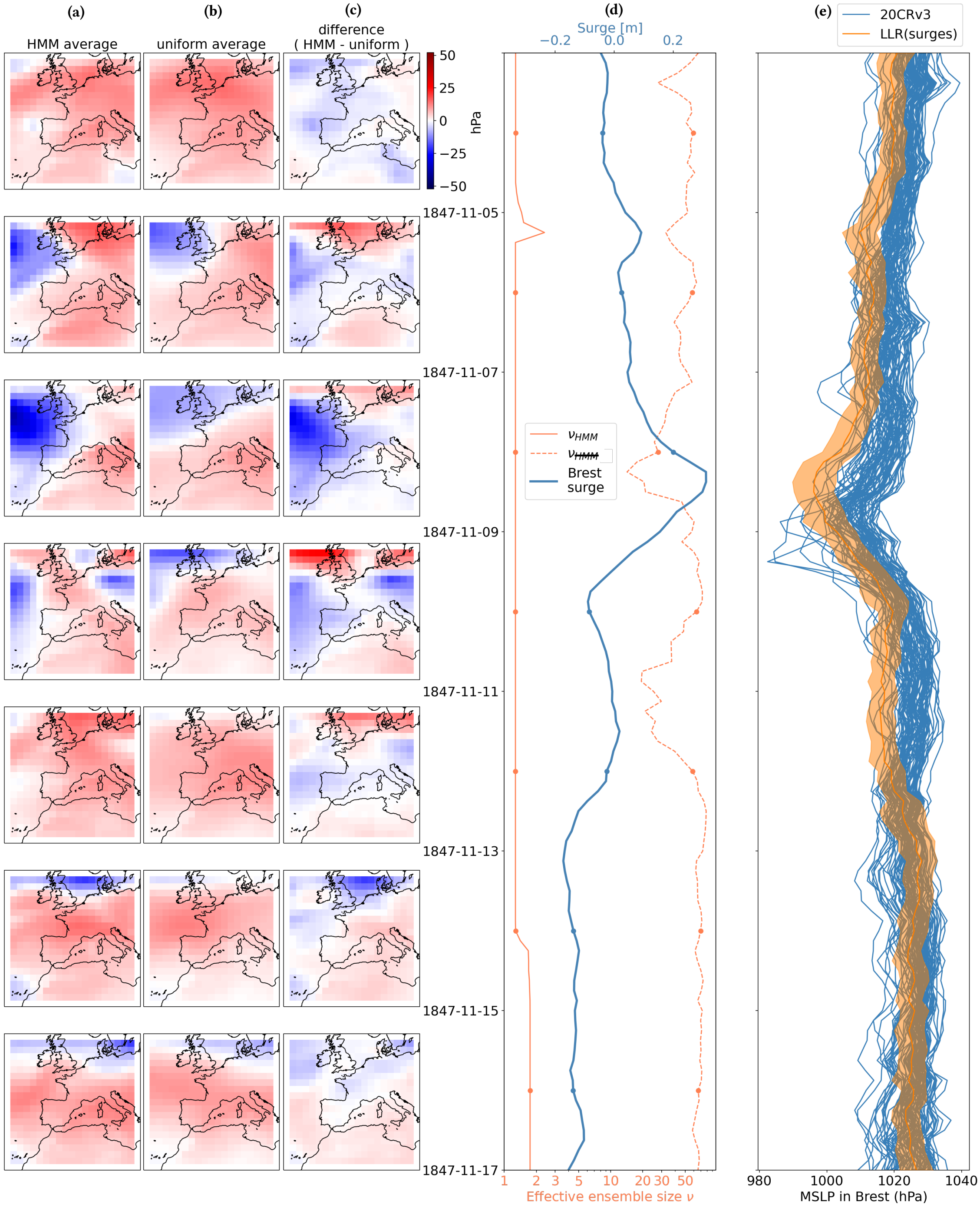

Figure 11. Mean result of the HMM algorithm in november 1847. (a) average according to surge observations and HMM smoothing algorithm (probabilites $p_{\mathrm{HMM}}(i,t)$). (b) average using constant uniform weights on 20CRv3 members. (c) difference HMM - uniform. (d) surge observations and effective ensemble sizes. (e) MSLP in Brest from ensemble members (blue lines) and local-linear regression from surges (average $\pm$ 1 standard deviation). The dates of the left-hand-side plots are indicated with dots in (d) and (e).

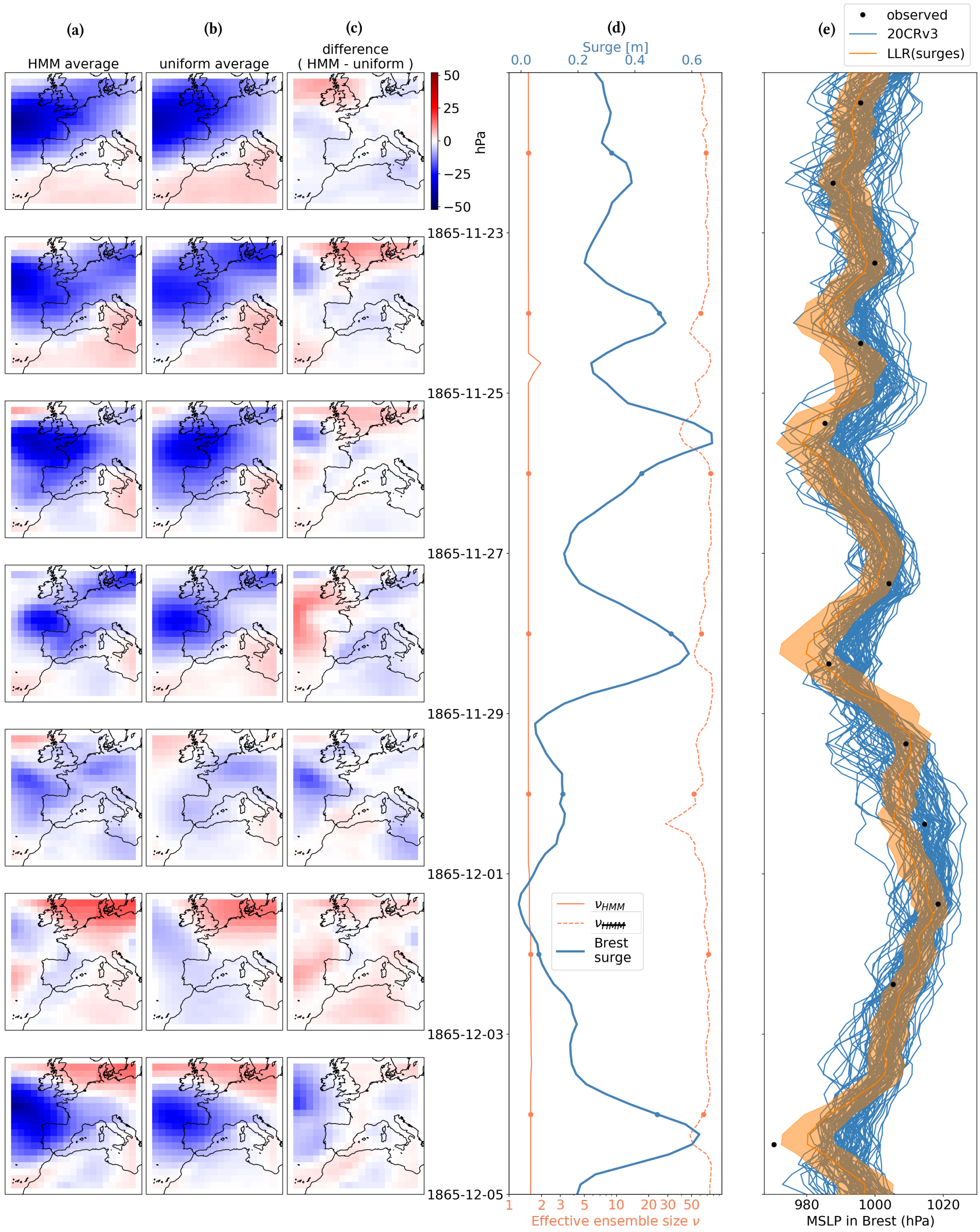

Figure 12. Same as Fig. 11 but for another storm. In this figure independent pressure observations in Brest are also available and shown as black dots in (e).

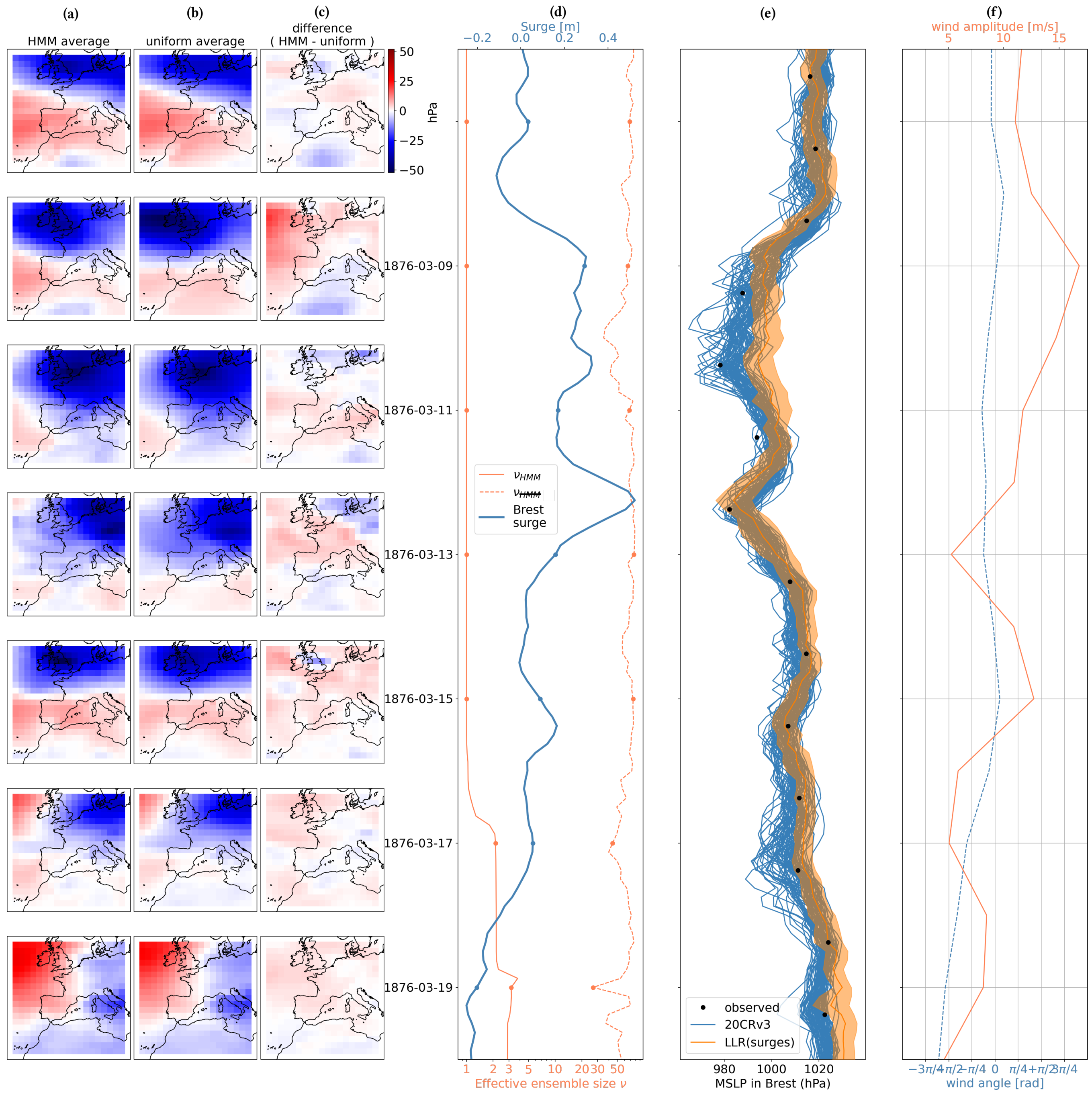

Figure 13. Same as Fig. 12 but for another storm. In (f), we also show the amplitude and direction of daily 10m-winds from 20CRv3 member-mean, linearly interpolated at the city of Brest.

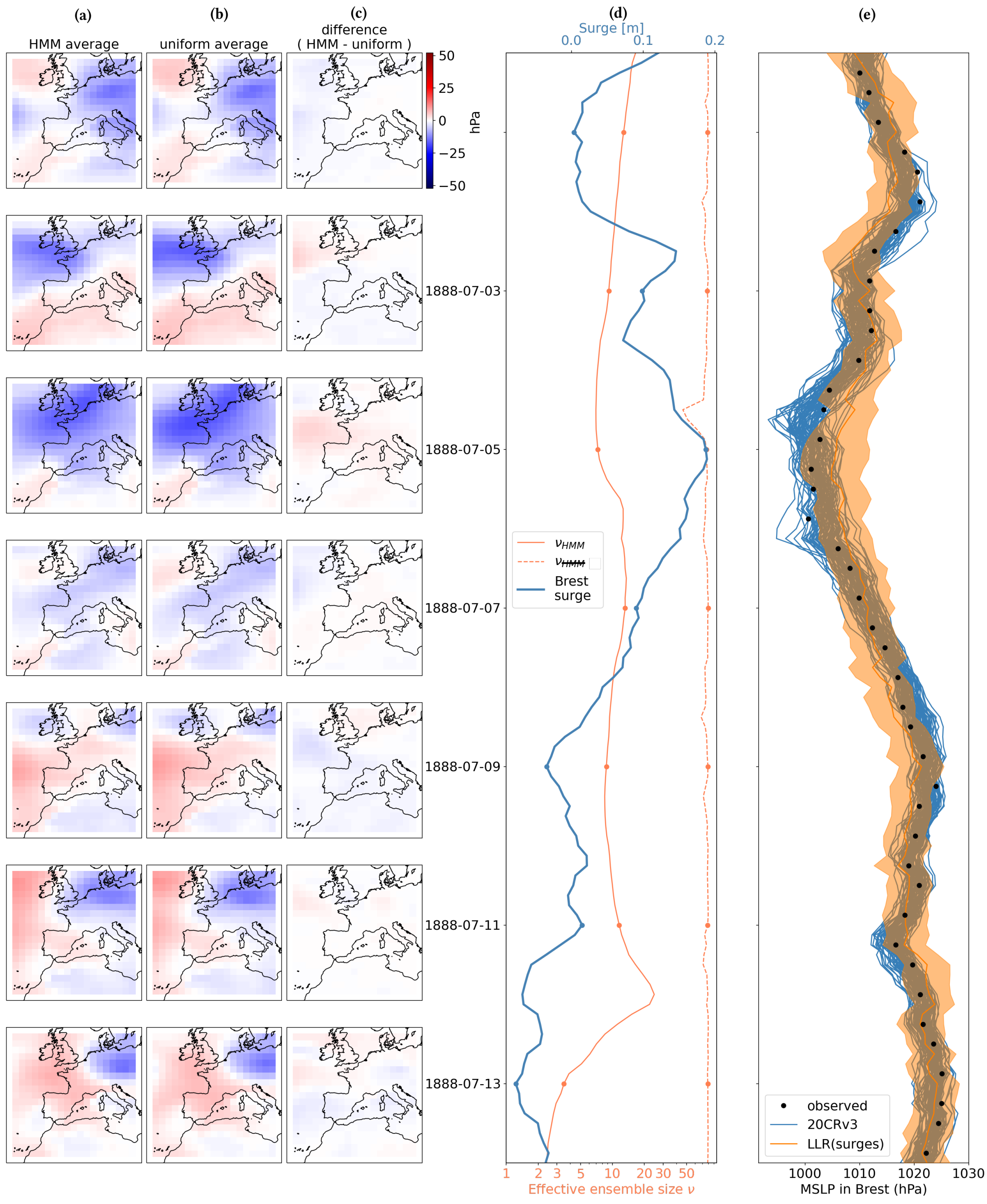

Figure 14. Same as Fig. 12 but for another event.

uncover past large-scale events. In particular, the hidden-Markov model algorithm outlined here could be replicated to weigh ensemble members according to independent observations.

Author contributions. Conceptualization: PP, BC. Methodology,software: PP, PA. Investigation: PP, BC, PA, PT. Writing – original draft preparation: PP. Writing – review and editing: PP, PA, PT, BC. Funding acquisition: BC.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. This work has largely benefited from the suggestions of two anonymous referees, who we thank here. We thank Lucia Pineau-Guillou for sharing her expertise on physical mechanisms driving sea-level fluctuations and on tide gauge data treatment. We thank Simon Barbot for discussions on surges. We thank Jean-Marc Delouis for fruitful discussions and suggestions on our work. Finally, many thanks are owed to Ed Hawkins for contacting us during the open discussion phase, providing information on newly available pressure data, and giving us feedbacks on the idea of using tide gauges as external validation data for atmospheric reanalysis. This work was supported by European Research Council (ERC) Synergy grant STUOD – DLV-856408. Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory.

References

Alvarez-Castro, M. C., Faranda, D., and Yiou, P.: Atmospheric dynamics leading to West European summer hot temperatures since 1851, Complexity, 2018, 1–10, 2018.

Ansell, T. J., Jones, P. D., Allan, R. J., Lister, D., Parker, D. E., Brunet, M., Moberg, A., Jacobeit, J., Brohan, P., Rayner, N., et al.: Daily mean sea level pressure reconstructions for the European–North Atlantic region for the period 1850–2003, Journal of Climate, 19, 2717–2742, 2006.

Bärring, L. and Fortuniak, K.: Multi-indices analysis of southern Scandinavian storminess 1780–2005 and links to interdecadal variations in the NW Europe–North Sea region, International Journal of Climatology: A Journal of the Royal Meteorological Society, 29, 373–384, 2009.

Bertin, X.: Storm surges and coastal flooding: status and challenges, La Houille Blanche, pp. 64–70, 2016.

Brönnimann, S., Compo, G. P., Spadin, R., Allan, R., and Adam, W.: Early ship-based upper-air data and comparison with the Twentieth Century Reanalysis, Climate of the Past, 7, 265–276, 2011.

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., et al.: Unlocking pre-1850 instrumental meteorological records: A global inventory, Bulletin of the American Meteorological Society, 100, ES389–ES413, 2019.

Bryant, K. M. and Akbar, M.: An exploration of wind stress calculation techniques in hurricane storm surge modeling, Journal of Marine Science and Engineering, 4, 58, 2016.

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

Codiga, D. L.: Unified tidal analysis and prediction using the UTide Matlab functions, 2011.

Compo, G., Slivinski, L., Whitaker, J., Sardeshmukh, P., McColl, C., Brohan, P., Allan, R., Yin, X., Vose, R., Spencer, L., et al.: The international surface pressure databank version 4, 2019.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al.: The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, 2011.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean dynamics, 53, 343–367, 2003.

Fan, J.: Local linear regression smoothers and their minimax efficiencies, The annals of Statistics, pp. 196–216, 1993.

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., Arns, A., Bradshaw, E., and Thompson, P.: GESLA version 3: A major update to the global higher-frequency sea-level dataset, Geoscience Data Journal, 10, 293–314, 2023.

Hansen, B.: Econometrics, Princeton University Press, 2022.

Hawkins, E., Brohan, P., Burgess, S. N., Burt, S., Compo, G. P., Gray, S. L., Haigh, I. D., Hersbach, H., Kuijjer, K., Martínez-Alvarado, O., et al.: Rescuing historical weather observations improves quantification of severe windstorm risks, Natural hazards and earth system sciences, 23, 1465–1482, 2023.

Horsburgh, K. and Wilson, C.: Tide-surge interaction and its role in the distribution of surge residuals in the North Sea, Journal of Geophysical Research: Oceans, 112, 2007.

Krueger, O., Schenk, F., Feser, F., and Weisse, R.: Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations, Journal of Climate, 26, 868–874, 2013.

Lazure, P. and Dumas, F.: An external–internal mode coupling for a 3D hydrodynamical model for applications at regional scale (MARS), Advances in water resources, 31, 233–250, 2008.

Lguensat, R., Tandeo, P., Ailliot, P., Pulido, M., and Fablet, R.: The analog data assimilation, Monthly Weather Review, 145, 4093–4107, 2017.

Liu, J. S.: Nonparametric hierarchical Bayes via sequential imputations, The Annals of Statistics, 24, 911–930, 1996.

Marcos, M., Puyol, B., Amores, A., Pérez Gómez, B., Fraile, M. Á., and Talke, S. A.: Historical tide gauge sea-level observations in Alicante and Santander (Spain) since the 19th century, Geoscience Data Journal, 8, 144–153, 2021.

Melchior, P.: The tides of the planet Earth, Oxford, 1983.

Pineau-Guillou, L., Ardhuin, F., Bouin, M.-N., Redelsperger, J.-L., Chapron, B., Bidlot, J.-R., and Quilfen, Y.: Strong winds in a coupled wave–atmosphere model during a North Atlantic storm event: Evaluation against observations, Quarterly Journal of the Royal Meteorological Society, 144, 317–332, 2018.

Pineau-Guillou, L., Delouis, J.-M., and Chapron, B.: Characteristics of Storm Surge Events Along the North-East Atlantic Coasts, Journal of Geophysical Research: Oceans, 128, e2022JC019 493, 2023.

Ponte, R. M.: Understanding the relation between wind-and pressure-driven sea level variability, Journal of Geophysical Research: Oceans, 99, 8033–8039, 1994.

Quintana, G. I., Tandeo, P., Drumetz, L., Leballeur, L., and Pavec, M.: Statistical forecast of the marine surge, Natural Hazards, 108, 2905–2917, 2021.

Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77, 257–286, https://doi.org/10.1109/5.18626, 1989.

Roden, G. I. and Rossby, H. T.: Early Swedish contribution to oceanography: Nils Gissler (1715–71) and the inverted barometer effect, Bulletin of the American Meteorological Society, 80, 675–682, 1999.

Rodrigues, D., Alvarez-Castro, M. C., Messori, G., Yiou, P., Robin, Y., and Faranda, D.: Dynamical properties of the North Atlantic atmospheric circulation in the past 150 years in CMIP5 models and the 20CRv2c reanalysis, Journal of Climate, 31, 6097–6111, 2018.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., et al.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Quarterly Journal of the Royal Meteorological Society, 145, 2876–2908, 2019.

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to high-dimensional particle filtering, Monthly Weather Review, 136, 4629–4640, 2008.

Tadesse, M. G. and Wahl, T.: A database of global storm surge reconstructions, Scientific data, 8, 125, 2021.

Takeda, H., Farsiu, S., and Milanfar, P.: Kernel regression for image processing and reconstruction, IEEE Transactions on image processing, 16, 349–366, 2007.

Wohland, J., Omrani, N.-E., Witthaut, D., and Keenlyside, N. S.: Inconsistent wind speed trends in current twentieth century reanalyses, Journal of Geophysical Research: Atmospheres, 124, 1931–1940, 2019.

Woodworth, P. L., Melet, A., Marcos, M., Ray, R. D., Wöppelmann, G., Sasaki, Y. N., Cirano, M., Hibbert, A., Huthnance, J. M., Monserrat, S., et al.: Forcing factors affecting sea level changes at the coast, Surveys in Geophysics, 40, 1351–1397, 2019.

Wöppelmann, G., Pouvreau, N., and Simon, B.: Brest sea level record: a time series construction back to the early eighteenth century, Ocean Dynamics, 56, 487–497, 2006.

Yiou, P., Boichu, M., Vautard, R., Vrac, M., Jourdain, S., Garnier, E., Fluteau, F., and Menut, L.: Ensemble meteorological reconstruction using circulation analogues of 1781–1785, Climate of the Past, 10, 797–809, 2014.

Zucchini, W., MacDonald, I. L., and Langrock, R.: Hidden Markov models for time series: an introduction using R, CRC press, 2017.