Climate
of the Past

Could old tide gauges help estimate past atmospheric variability?

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

[1]Laboratoire d'Océanographie Physique et Spatiale, Univ. Brest, Ifremer/CNRS/IRD, 29280 Plouzané, France
[2]Odyssey, Inria/IMT/CNRS, 29280 Plouzané, France
[3]Laboratoire de Mathematiques de Bretagne Atlantique, Univ. Brest, UMR CNRS 6205, 29200 Brest, France
[4]IMT Atlantique, Lab-STICC, UMR CNRS 6285, 29238 Brest, France
[5]RIKEN Cluster for Pioneering Research, Kobe, 650-0047, Japan

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

Abstract. 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. 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 earlier to times when direct pressure observations were scarce. Therefore, these old tide gauge records may be used as indirect observations of sub-seasonal atmospheric variability that are 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 surge residuals 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 light on the strengths and limitations of the methodology, particularly for the case of wind-driven surge residuals. This calls for the future use of a mixed methodology between data-driven tools and physics-based modeling. Our methodology could be applied to use other types of independent observations (not just tide gauges) as a means of weighting reanalysis ensemble members.

1 Introduction

Understanding the atmospheric system requires an understanding of all scales of variation from daily to centennial. This cannot be done unless long observation records allow the disentanglement of these scales. The 20th 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, 20CR 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 reanalyses (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, to the best of our knowl-

edge, 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 storm surge reconstructions into the past using different atmospheric reanalysis products in order to estimate past unobserved extreme storm 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 the periodic rise and fall of the water level (Melchior, 1983) and has 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 (away from) the coast leads to increased (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 4 m, and storm surges can amount to as much as 1.5 m.

Tide gauges are numerous, forming a dense global network in recent years and a sparser one over the past few centuries. As an example and from the GESLA-3 sea level database (Haigh et al., 2023), 10 coastal tide gauge records start before 1907 on the eastern coast of North America, while 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; Harter et al., 2024). 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 residual (see definition in Gregory et al., 2019) to local mean sea level pressure. Hidden Markov models (HMMs) will allow us 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 storm surges associated with each member of the reanalysis compared to real observations. One conclusion of the study is that the crude spatiotemporal resolution of the reanalysis is responsible for a systematic underestimation of the observed storm 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 Sect. 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 Sect. 5 focuses on four specific events and compares with independent pressure observations. Conclusions on the proposed methodology and experiments are drawn in Sect. 6, along with potential applications of this work.

2 Data

2.1 The 20th Century Reanalysis version 3 (20CRv3)

The 20th 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 h and uses a spectral triangular model in space with truncation of T254 (approximately 75 km at the Equator). There are 64 vertical levels up to 0.3 mb. It assimilates surface pressure observations from ships and fixed stations and analyzes 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, last access: July 2024) allows us to consult the monthly observation count per $2° \times 2°$ box.
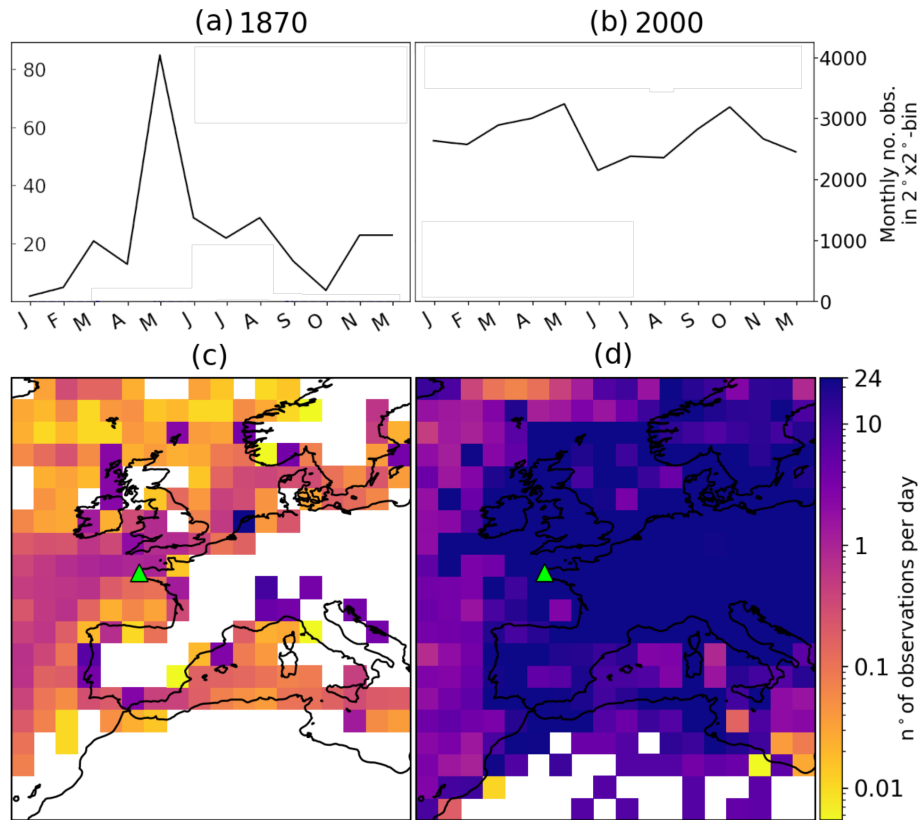
Figure 1. Number of surface pressure observations from the International Surface Pressure Databank (ISPD) assimilated in the 20th century Reanalysis version 3 (20CRv3). **(a, b)** Monthly count over a $2° \times 2°$ box centered on Brest for the years 1870 **(a)** and 2000 **(b)**. **(c, d)** The yearly average of daily number of observation in 1870 **(c)** and 2000 **(d)**.

Figure 1 shows yearly averages of the number of surface pressure observations per day, comparing the years 1870 and 2000. The maximum value was set to 24 observations per day, although in 2000 this value is mostly exceeded. In 1870, approximately half of Europe's land surface has no observations at all, and fewer than 10 points have more than 10 observations per day. Observations coming from ships allow us to raise the number of observations to approximately one per day in densely trafficked areas. Conversely, in 2000 virtually all of western Europe's land area has more than 24 observations per day. Taking a spatial average over the whole map from Fig. 1 gives approximately 1 observation every 3 d 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 $2° \times 2°$ 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).

2.2 Preprocessing of mean sea level pressure

In this work, we only use the mean sea level pressure (MSLP) variable from 20CRv3. We perform two different preprocessing steps on this variable.

The first preprocessing step is used for the statistical relationship between the local pressure and the surge residual. As the latter is driven in part 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 (48.3829° N, 4.49504° W) and the MSLP averaged over all members of 20CRv3 and over the North Atlantic Ocean (using the reanalysis' land mask and averaging from 0 to 69° N and from 98° W to 12° E), in a similar fashion to Ponte (1994). This spatially averaged pressure is denoted as $\overline{\mathrm{MSLP}}^{\mathrm{ocean}}(t)$ and depends only on time. 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 1 order of magnitude smaller than the inter-member variability of MSLP at the city of Brest, which justifies our approximation of using simply the member average of the ocean-averaged pressure as a reference.

A second preprocessing step 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 Sect. 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 denoted as $\overline{\mathrm{MSLP}}^{\mathrm{clim}}$ and depends on latitude and longitude.

2.3 Tide gauge of Brest (France)

In this study, the tide gauge of Brest is used as indirect tracer of atmospheric circulation through surge residuals. The Brest sea level record is taken from the GESLA-3 database starting in 1846 with hourly sampling. Apart from a few large gaps, the record is mostly continuous during periods 1847–1945 and 1953–present. This combination of historical and modern records is at the foundation of the methodology explored in the next section.

2.4 Preprocessing of sea level

As mentioned earlier, the part of the sea level that 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. Following this, as we are interested in sub-seasonal variations, we also remove the yearly variations in the mean sea level (at interannual and decadal scales), such as sea level rise (Cazenave and Llovel, 2010). In this work, we also use moving averages and differences in the surge residual. All of these steps are exemplified in Fig. 2.

We first compute the tidal constituents of the raw sea level (blue curve, Fig. 2a) 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: 1847–1890 and 1981–2015. Removing the tidal part of the signal gives the surge residual (dashed orange line in Fig. 2a), which has a temporal average value of ∼ 4 m for the Brest tide gauge.

Following this, we remove the yearly median value of the sea level (dashed orange line in Fig. 2b). 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 us to access the zero-median surge residual which is denoted as $h(t)$ in the following equation:

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

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

Note from Fig. 2b that the surge residual fluctuates at an hourly scale, part of which is due to oscillations that are not due to variations in atmospheric pressure. For instance, these oscillations can be due to tide–surge interactions (Horsburgh and Wilson, 2007) or measurement errors in the 19th century leading to phase shifts. Such 12 h oscillations can dominate the surge residual signal in Brest, where the tidal amplitude is large (see, for instance, 29 and 30 January 2014, Fig. 2). Furthermore, tide–surge interactions lead to stronger surge residuals at low tide and weaker surge residuals at high tide (Horsburgh and Wilson, 2007). As these phenomena are not linked to atmospheric processes, we chose to filter them out with a simple 12 h average (solid green curve in Fig. 2). Given the spatial resolution of 20CRv3, smaller-scale events are likely to not be represented in the MSLP fields used in this study. In the following, we denote the 12 h average of the zero-median surge residual as $\overline{h}^{12\,\mathrm{h}}(t)$, which is calculated using the following equation:

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

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 in the surge residual are also expected, statistically speaking, to be associated with deviations from the inverse barometer approximation. To allow the model described in Sect. 3.1 to capture this effect, we compute the difference between the surge at time t and at time $t-12\,\mathrm{h}$, choosing the 12 h interval again to filter out oscillations at a period close to 12 h. Furthermore, since the reanalysis is run at 3 h resolution, we perform a 3 h moving average of the surge residual before computing the difference. This difference is denoted $\Delta\overline{h}^{3\,\mathrm{h}}(t)$ and defined by the following equation:

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

2.5 Independent historical pressure observations

In Sect. 5, we use pressure observations for the city of Brest and compare them with 20CRv3 and our estimate of pressure based on tide gauge observations and our statistical model (local linear regression, LLR). We downloaded these observations from a repository (https://github.com/ed-hawkins/weather-rescue-data/tree/main, last access: July 2024) gathering historical pressure observations. These pressure observations come from the EMULATE project (Ansell et al., 2006) for 1860–1880 and from Météo France archives for
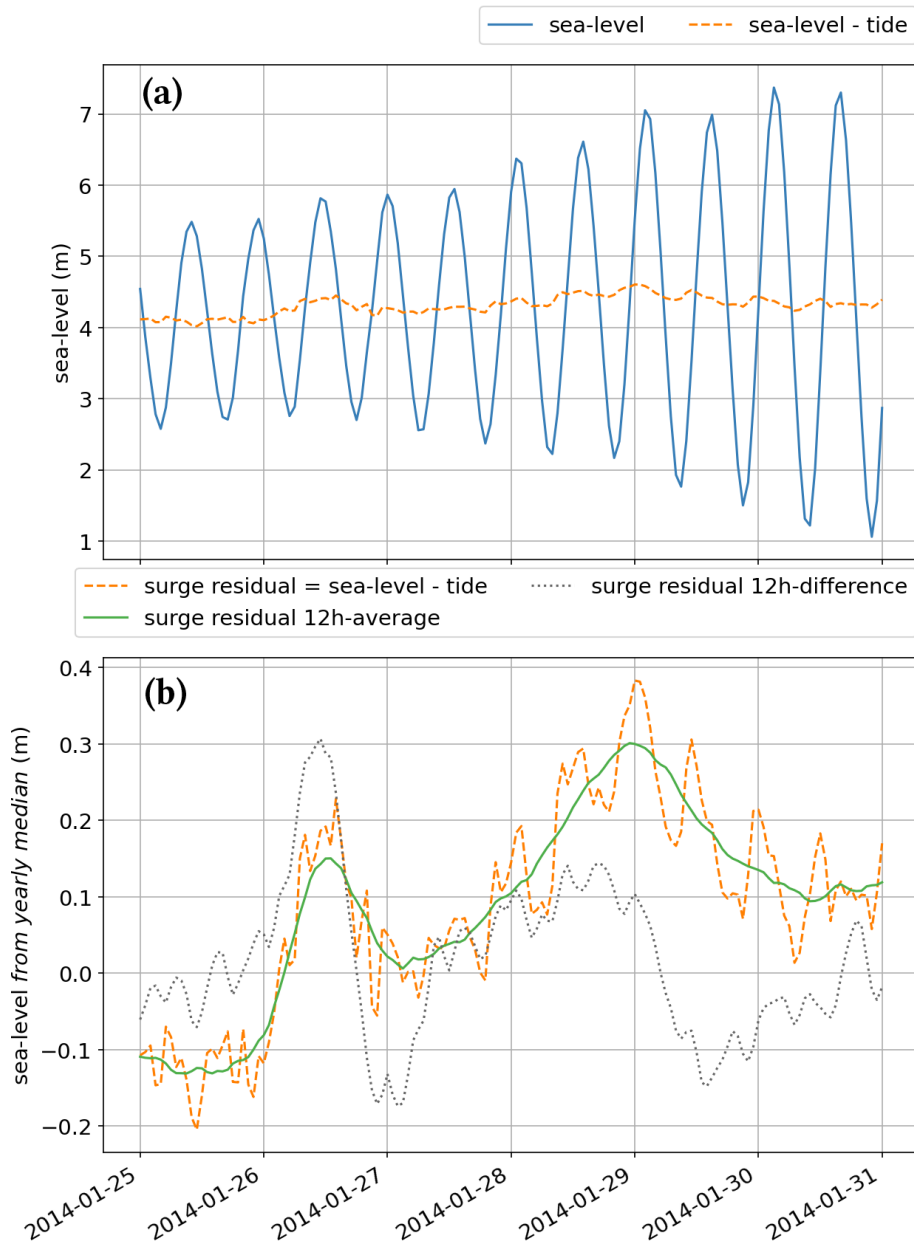
Figure 2. An example output of the different stages of preprocessing of the sea level signal used in this work. **(a)** Sea level before (solid blue line) and after (dashed orange line) removing the tidal part of the signal. **(b)** Zero-median surge residual $h(t)$ (1 h sampling, dashed orange curve), centered 12 h average $\bar{h}^{1\,\mathrm{h}}(t)$ (solid green curve), and the 12 h difference between 3 h averages of the surge residual $\Delta\bar{h}^{3\,\mathrm{h}}(t)$ (dotted gray curve).

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 of ~ 0.22 hPa

for the period 1860–1880 (EMULATE dataset) to each value of the independent pressure observation datasets so that the average pressure is 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 ~ 7.18 hPa.

3   Statistical methods

3.1   Local linear regression (LLR) between surge residuals and mean sea level pressure

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

The filtered surge residuals described in Sect. 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 (decrease) of 1 hPa in pressure at the mean sea level leads to a decrease (increase) in sea level of approximately 1 cm. This approximation is valid only for slow variations in atmospheric pressure compared to the typical time of dynamic adjustment of the sea level (Bertin, 2016).

Moreover, the piling up of water due to wind blowing perpendicular to the coast is responsible for positive (negative) surge residuals when the wind stress transport is directed towards (away from) the coast. This effect depends nonlinearly on the wind amplitude and direction (Bryant and Akbar, 2016; Pineau-Guillou et al., 2018). Statistical anticorrelation observed in most regions between wind-driven and pressure-driven sea level fluctuations causes regressions of surge residual versus atmospheric pressure to deviate from the inverse barometer approximation (Ponte, 1994).

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. It is also likely that typical wind conditions depend on the amplitude of the MSLP anomaly, meaning that the average value of MSLP anomaly for a given value of surge residual in Brest may be a nonlinear function. As shown by Hawkins et al. (2023), using a physical coastal model forced by the values of pressure (and wind) 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.

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: $\overline{h}^{12\,\mathrm{h}}(t)$ and $\Delta\overline{h}^{3\,\mathrm{h}}(t)$. We expect that corresponding atmospheric pressure variations should be slow and moderate for low absolute values of these two predictors and that winds should be of low intensity, meaning that the inverse barometer approx-

imation should hold. For larger absolute values of $\Delta\overline{h}^{3\,\mathrm{h}}(t)$, indicating rapidly changing surge residuals and thus likely also rapidly changing atmospheric conditions, we expect deviations from the inverse barometer due to the dynamical adjustment of the sea level. Similarly, the largest absolute values of $\overline{h}^{12\,\mathrm{h}}(t)$ are likely to be caused by the added contribution of wind to the effect of pressure, meaning that deviations from the inverse barometer are expected as well.

To model all of 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 (neighbors) of the two predictor variables $\overline{h}^{12\,\mathrm{h}}(t)$ and $\Delta\overline{h}^{3\,\mathrm{h}}(t)$ in the whole dataset and compute a linear regression on this subset of the dataset. The predicted variable is $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{\mathrm{ocean}}(t)$, 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 $\overline{h}^{12\,\mathrm{h}}(t)$ and $\Delta\overline{h}^{3\,\mathrm{h}}(t)$, the predicted variable $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{\mathrm{ocean}}(t)$ follows a Gaussian distribution:

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

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 neighbors of $[\overline{h}^{12\,\mathrm{h}}(t), \Delta\overline{h}^{3\,\mathrm{h}}(t)]$ 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-neighbor criterion is the Euclidean distance in the two-dimensional space of values of $[\overline{h}^{12\,\mathrm{h}}(t), \Delta\overline{h}^{3\,\mathrm{h}}(t)]$. For each time t at which we want to estimate $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{\mathrm{ocean}}(t)$, 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 is minimal:

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

We attach a weight $\omega_i(t)$ to each index $i \in I(t)$ according to the following formula:

$$\omega_i(t) = \frac{\exp\left(-\mathrm{dist}(t, t_i)^2 / \lambda(t)^2 \right)}{\sum\limits_{j \in I(t)} \exp\left(-\mathrm{dist}(t, t_j)^2 / \lambda(t)^2 \right)}, \qquad (5)$$

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 neighbors (Lguensat et al., 2017).

Using this subset of the whole dataset, we compute a weighted linear regression between the subset of regressors $\overline{h}^{12\,\mathrm{h}}(t_i)$, $\Delta\overline{h}^{3\,\mathrm{h}}(t_i)$ and of the predicted variable

$\mathrm{MSLP}(t_i) - \overline{\mathrm{MSLP}}^{\mathrm{ocean}}(t_i)$ using the weights $\omega_i(t)$. This regression has two linear coefficients denoted as $\alpha(t)$ and $\beta(t)$ and one intercept (constant value) denoted as $\gamma(t)$. Following this, the average is given by applying the local weighted linear model to the actual value of the predictors:

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

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

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

To test the accuracy of this model on the 1980–2015 period, we apply it for all times $t \in [1980\text{–}2015]$, searching for neighboring times t_i in the same period but with the condition that there is a minimum of 2 weeks between t and t_i (i.e., excluding the interval $[t-14\,\mathrm{d}, t+14\,\mathrm{d}]$). This is called the leave-one-out procedure, ensuring that the data that are used to fit the model do not include the true values. Following this, we compare the average $m(t)$ with the true value $\mathrm{MSLP}(t) - \overline{\mathrm{MSLP}}^{\mathrm{ocean}}(t)$ in a scatterplot (Fig. 3a). 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 behavior is expected as the method is limited by the observations it has seen previously. However, as will be seen in Sect. 5, this simple model is still able to capture storms. Following this, we test the adequacy of our variability estimate with the parameter $\mathrm{var}(t)$ by checking that the following shifted and rescaled variable follows a standard Gaussian distribution with an average of 0 and variance of 1:

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

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. 3b. Although the shape 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 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 residuals in Brest. Note that in the reanalysis the members are assumed to have uniform probabilities, i.e., a probability of $1/80$, as we have 80 members.

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

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

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

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

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

$$\mathbf{T}_{ij}(t) := P(\mathrm{member}(t) = j | \mathrm{member}(t-1) = i) \qquad (11)$$

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

$$\mathbf{T}_{ij}(t) \propto K_\theta\left(\mathrm{MSLP}_{\mathrm{map},j}(t), \mathrm{MSLP}_{\mathrm{map},i}(t)\right), \qquad (12)$$

where $\mathrm{MSLP}_{\mathrm{map},i}(t)$ is the ith member's map of mean sea level pressure in a squared box ($28°\,\mathrm{N} \leq \mathrm{lat} \leq 64°\,\mathrm{N}$,
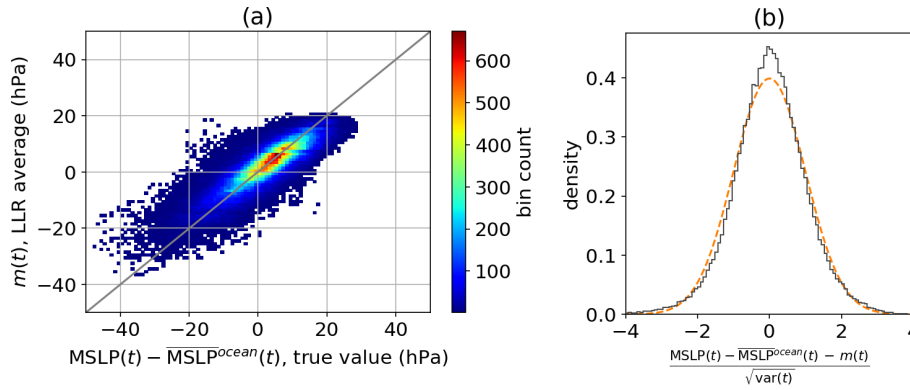
Figure 3. Evaluation of the local linear regression (LLR) on the period 1981–2015 using a leave-one-out procedure. **(a)** Histogram scatterplot of the average estimate from the LLR versus the true value of the MSLP difference with the reference ocean-averaged MSLP. **(b)** Density histogram of normalized average error of the LLR estimate (black line) compared to theoretical probability density function of the standard Gaussian random variable (dashed orange line).

$18°\,\mathrm{W} \leq \mathrm{long} \leq 18°\,\mathrm{E}$) at time t and $K_\theta(\cdot, \cdot)$ is a positive real-valued function that measures the similarity between $\mathrm{MSLP}_{\mathrm{map},i}(t)$ and $\mathrm{MSLP}_{\mathrm{map},j}(t)$ and depends on the parameter θ.

Equation (12) states that transitions from one member to another are more likely if the associated MSLP maps 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 that affect the surge residuals in Brest would lie. Ideally, $K_\theta(\cdot, \cdot)$ should be symmetric and semi-definite. Here, a simple Gaussian kernel of Euclidean distances is used, with normalization factor $\theta > 0$, meaning that for two fields X and Y:

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

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

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

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

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

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

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

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

$$a_i(1) = p_{\mathrm{HMM}}(i, 1), \quad (17)$$

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

and the backward procedure:

$$b_i(T) = 1, \quad (19)$$

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

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

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

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

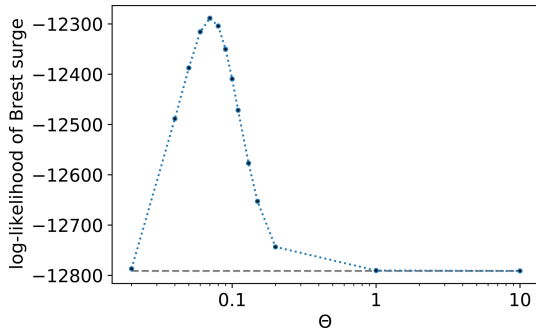$$p_{\mathrm{HMM}}(i, t) = \frac{a_i(t) b_i(t)}{\sum_{j=1}^{80} a_j(t) b_j(t)}, \quad (22)$$

Figure 4. Log likelihood of Brest surge residuals (dotted blue line) as a function of parameter Θ defined in Eq. (14). For comparison, the log likelihood of the simple model without the hidden Markov model ($\Theta \to +\infty$) is also shown (dashed gray 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.

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

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

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

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

Figure 4 shows variations in this quantity with Θ for 1 year (1885) of surge residual observations in Brest (i.e., $t = 0$ is 1 January 1885, while T is 1 January 1886). The curve shows a distinct maximum around $\Theta \approx 0.09$ and plateaus for higher values. According to Fig. 4, the difference in 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 such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), or likelihood ratio tests (Zucchini et al., 2017).

Note that in the limit $\Theta = +\infty$, we have a constant transition probability $\mathbf{T}_{ij}(t) = \frac{1}{80}$, and $p_{HMM}(i,t)$ reduces to $p_{HMM}(i,t)$. Figure 4 thus supports the use of the HMM to estimate probabilities of the MSLP map conditioned by surge residual observations.

The choice of restricting the estimation of log likelihood to one arbitrary year (1885) is supported by the fact that estimation of $\mathbf{T}_{ij}(t)$ 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 $p_{HMM}(i,t)$.

4 Modification of 20CRv3 ensemble when accounting for surge residuals

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

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

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

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

Note that in this definition $\sigma_{20CR}(t)$ depends on time, latitude, and longitude. Therefore, at each grid point and for each time step the quantity $\delta\mu_{HMM}(t)$ 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 $\nu_{HMM}(t)$ (Liu, 1996):

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

and we equivalently define $\nu_{HMM}(t)$. These quantities are estimates of the number of ensemble members that can be retained according to surge residual observations, assuming one discards very unlikely members.

In Fig. 5, the variables $\delta\mu_{HMM}$, $\delta\mu_{HMM}/\sigma_{20CR}$, 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 surge residuals. The figure shows that the amount of correction $\delta\mu_{HMM}$ and the decrease in ensemble size ν_{HMM} are much stronger using smoothed probabilities with HMM rather than probabilities without HMM. Showing the deviation $\delta\mu_{HMM}$ in the Bay of Biscay, where the standard deviation of $\delta\mu_{HMM}/\sigma_{HMM}$ is
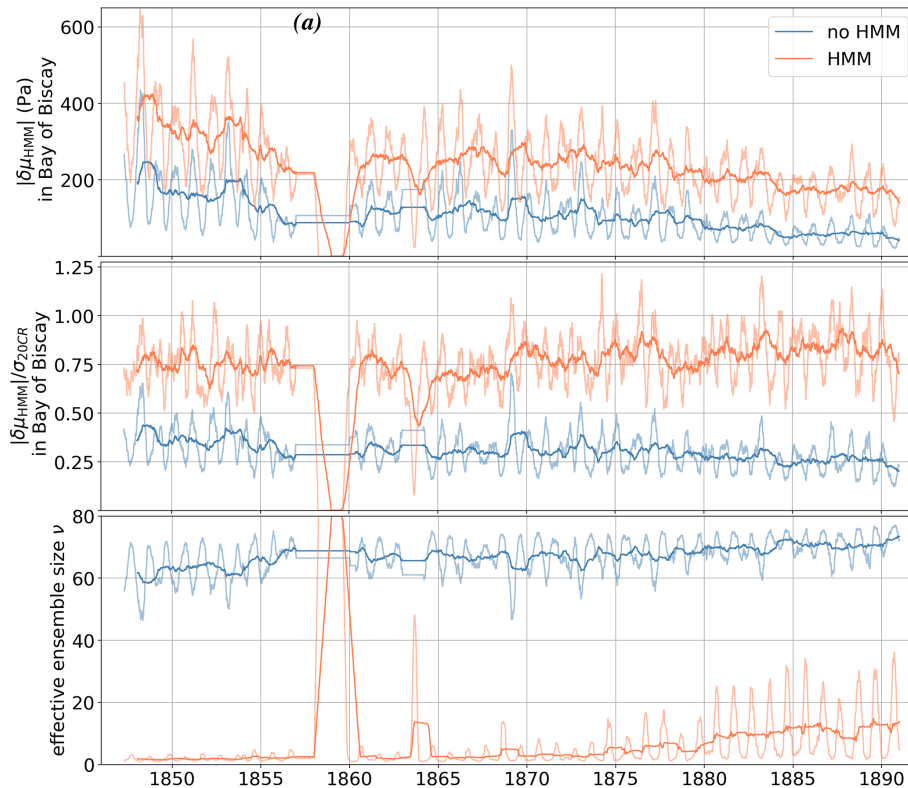
Figure 5. (a) Average MSLP deviation $\delta\mu_{HMM}$ in pascals in the center of the Bay of Biscay. Panel **(b)** is the same as panel **(a)** but normalized by reanalysis ensemble standard deviation $\delta\mu_{HMM}/\sigma_{20CR}$. **(c)** Effective ensemble size ν_{HMM}. Orange is used for data using smoothed probabilities with HMM $p_{HMM}(i,t)$. Blue is used for data using probabilities without HMM $p_{\overline{HMM}}(i,t)$. Bold lines indicate yearly average. Thin lines indicate the 3-month average. All plots make simultaneous use of data from the Brest tide gauge.

strongest (see Fig. 6), substantial absolute values of $\sim 600\,\mathrm{Pa}$ are obtained in early 1850s winters, even after averaging over 3 months. These large deviations correspond to more than 1 standard deviations of the ensemble size. Using probabilities without HMM, deviations are weaker but still non-negligible ($\sim 500\,\mathrm{Pa}$, $\sim 0.7\sigma$). The slow decrease in $\delta\mu_{HMM}$ with time is coherent with slowly increasing observations used in 20CRv3, albeit with substantial decadal variations. However, $\delta\mu_{HMM}/\sigma_{HMM}$ and $\delta\mu_{\overline{HMM}}/\sigma_{HMM}$ do not show a clear trend, indicating a persisting gain in information from surge residual observations throughout the 19th century.

In terms of effective size, Fig. 5 shows that the smoothing HMM algorithm imposes a strong member selection, with mostly only one 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 $\nu_{\overline{HMM}}(t)$ 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 residual estimates from tide gauge observations. A strong increase in $\nu_{HMM}(t)$ 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. 6. The analysis of time standard deviation of $\delta\mu_{HMM}$ and $\delta\mu_{HMM}/\sigma_{20CR}$ shows that the area of greatest influence of the corrections from surge residual 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 surge residuals in Brest, and by the sparsity of direct pressure measurements (ship logs) in this area in the 19th century. The standard deviation of $\delta\mu_{HMM}$ shows the largest values to the northwest of the map, which is where strong storms travel. Indeed, the variability of MSLP shows a great northwest 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 $\delta\mu_{HMM}$ is smaller in 1880–1890, which can be explained by a greater conditioning of 20CRv3 members by observations both offshore and inland. In cases of very sparse observations used in 20CRv3, the area of influence of these corrections widens due to continuity of MSLP fields. Note also that the area of influence is greater for $\delta\mu_{HMM}$ then for $\delta\mu_{\overline{HMM}}$ because of the time propagation of corrections thanks to the smoothing HMM algorithm. Finally, Fig. 6 confirms the large difference in am-
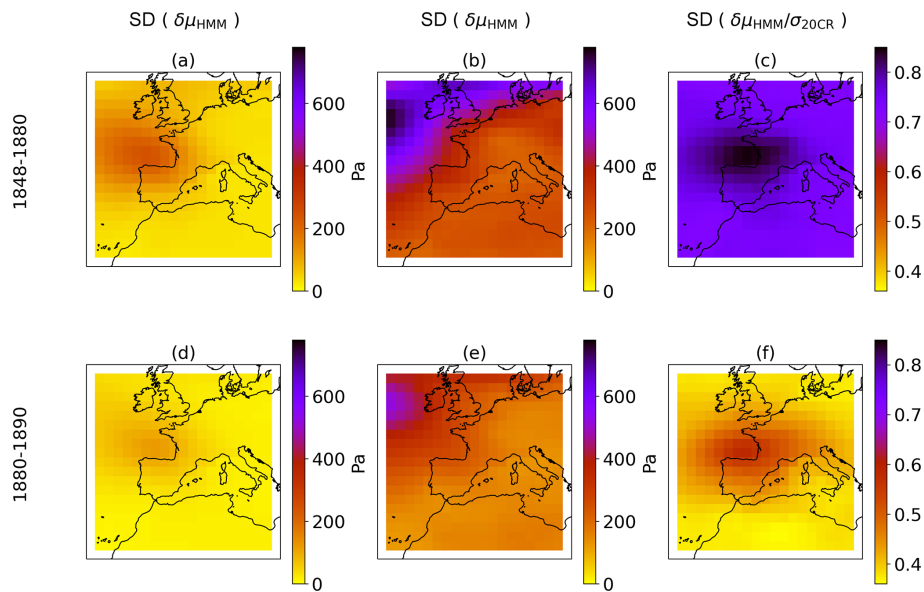
Figure 6. Time standard deviation of $\delta\mu_{\mathrm{HMM}}$ (**a, d**), $\delta\mu_{\mathrm{HMM}}$ (**b, e**), and $\delta\mu_{\mathrm{HMM}}/\sigma_{\mathrm{20CR}}$ (**c, f**) computed from October to March for the years 1848–1880 (**a–c**) and 1880–1890 (**d–f**). Probabilities p_{HMM} and p_{HMM} make use of data from the Brest tide gauge.
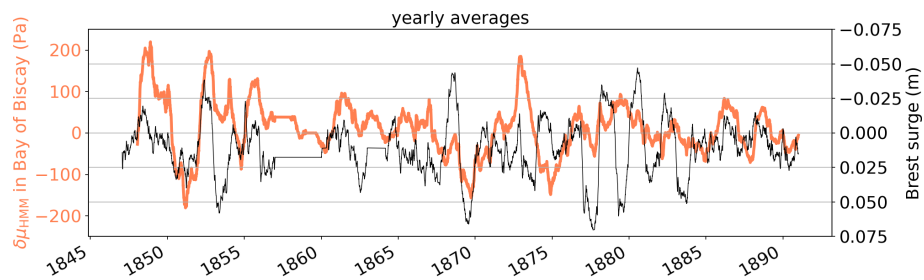
Figure 7. The bold orange line shows the yearly average of $\delta\mu_{\mathrm{HMM}}$ at 47.83° latitude and −7.57° longitude in the Bay of Biscay using data from the Brest tide gauge. The thin black line shows the zero-median Brest surge residual (inverted sign).

plitude of deviations between pre-1880 and post-1880 corrections already witnessed in Fig. 5. Similar spatial footprints can be witnessed from maps of high and low quantiles of $\delta\mu$ but with different values (not shown). Similarly, computing the time standard deviations as in Fig. 6 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. 7. The same behavior can be witnessed for the surge residual, which is strongly anti-correlated to these deviations (Fig. 7). This can be explained by the fact that 20CRv3 smooths MSLP values in areas of sparse measurements and that surge residual filtering corrections allow us to retrieve more realistic intense values (either positive or negative). This interannual variability is related to the variability in storminess (Bärring and Fortuniak, 2009).

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 d) surrounding the events and compare the results of the simple LLR based on surge residuals with 20CRv3 and independent observations (when available). These are plotted in Fig. 8. One can thus see that the uncertainties associated with the surge-residual-based LLR do not vary from year to year, while those of 20CRv3 decrease. 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 both in 1847 and in 1865, while its minimum value is of 0.69 in 1876 and 0.94 in 1888. This
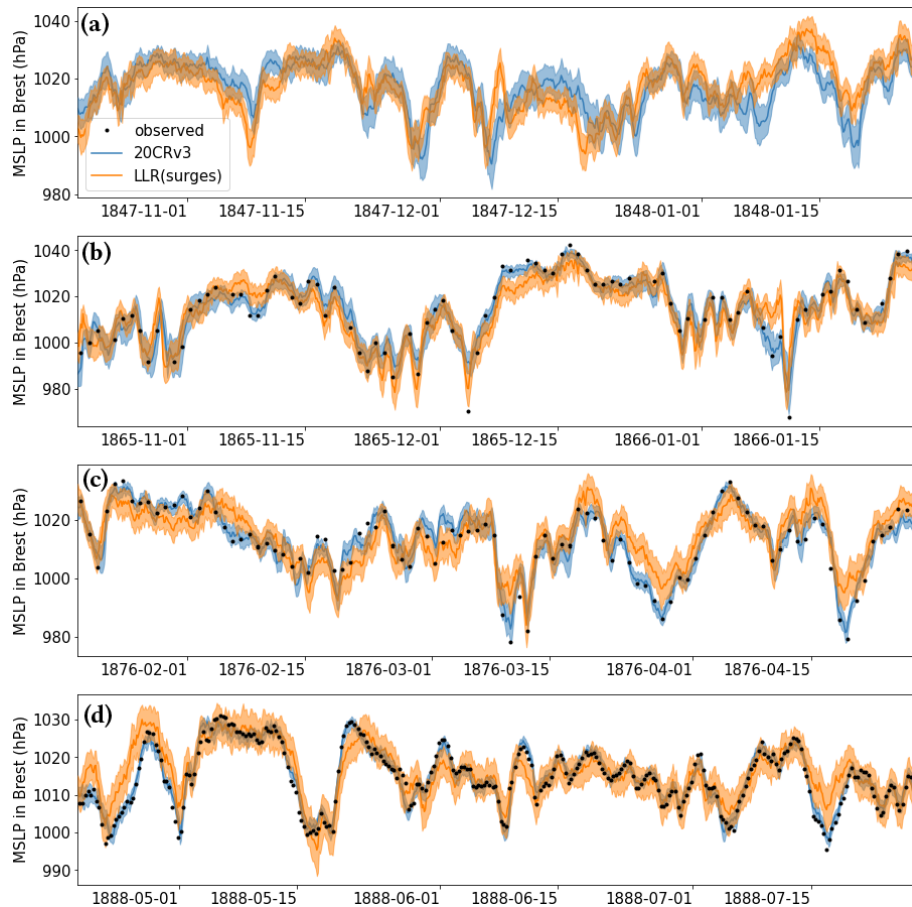
Figure 8. Comparison of MSLP estimation in Brest from 20CRv3 (blue), LLR based on surge residuals (orange), and independent observations (black dots) that were not used to build the orange and blue curves for three periods surrounding the events studied in this section. Full lines correspond to average values, while shaded areas correspond to ±1 SD (1 standard deviation) around the average.

shows that on average the reanalysis has lower uncertainty than the surge-residual-based pressure estimate (albeit 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. 8b), 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 that is more consistent with the independent observations. In 1876 (Fig. 8c), biases of ∼ 5 hPa 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, albeit with persistent biases (either positive or negative) that last from a few days to ∼ 15 d. We attribute these biases to different atmospheric conditions that cannot be estimated from the surge residuals 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. 12f, where we also show the direction and amplitude of the 10 m 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, the first around 10 March and second around 12 March, as indicated by the reanalysis members and the independent pressure observations (Fig. 12e). However, the first low-pressure system did not induce a surge residual as strong as the second one. One key difference between the two events is the wind amplitude, which reached 15 m s^{-1} during the first event and then decreased to 5–10 m s^{-1} 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 15 m s^{-1} 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.
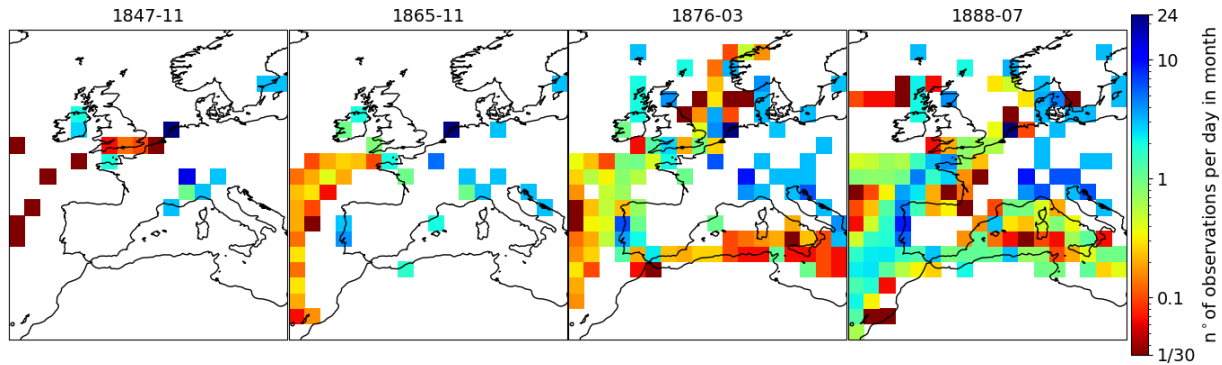
Figure 9. Number of pressure observations per month divided by 30 (or 31) to give an average number of observations per day for the months of November 1847, November 1865, March 1876, and July 1888.

Our interpretation relies on the fact that the effect of wind on extreme surge residuals acts at small timescales (daily or sub-daily), which is backed by recent work (Pineau-Guillou et al., 2023).

To aid the interpretation of Figs. 10–13, in Fig. 9 we also show 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 than 1 observation per day). In November 1865, some more stations are available, and at this point the 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 these are not in the direct vicinity of Brest.

One common feature of Fig. 10–12 is that the HMM algorithm tends to be very selective compared to the weighing without HMM. This is the consequence of our optimization of the parameter θ with the objective of maximizing the likelihood of the surge residual observations on the 20CRv3 ensemble. Having a low theta value allows us to give a high weight to the ensemble member that has the highest probability according to the surge-residual-based LLR model. However, as is obvious from comparing Fig. 10a with b, Fig. 11a with b, and Fig. 12a with b, this does not always have a strong influence on the average MSLP field. In the case of Fig. 13, 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, these figures again 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 residual observations according to the LLR statistical relationship.

The fact that one member is often much more coherent with the series of surge residuals is the result of (1) the high variability of the ensemble and the LLR pressure es-

timation, (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 the 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), meaning 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 that is closest to the biased pressure estimate from the LLR applied to the surge residual 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.

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. Surge residuals of Brest allow us to assess part of the atmospheric variability that was uncaught in global 20CR reanalyses based on pressure observations. Weighing 20CR members according to surge residual observations reduces the effective ensemble size and implies significant deviations in member-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-residual-based local linear relationship. Such comparisons also show that the reconstruction of pressure based on surge residuals is ambiguous due to the influence of winds,
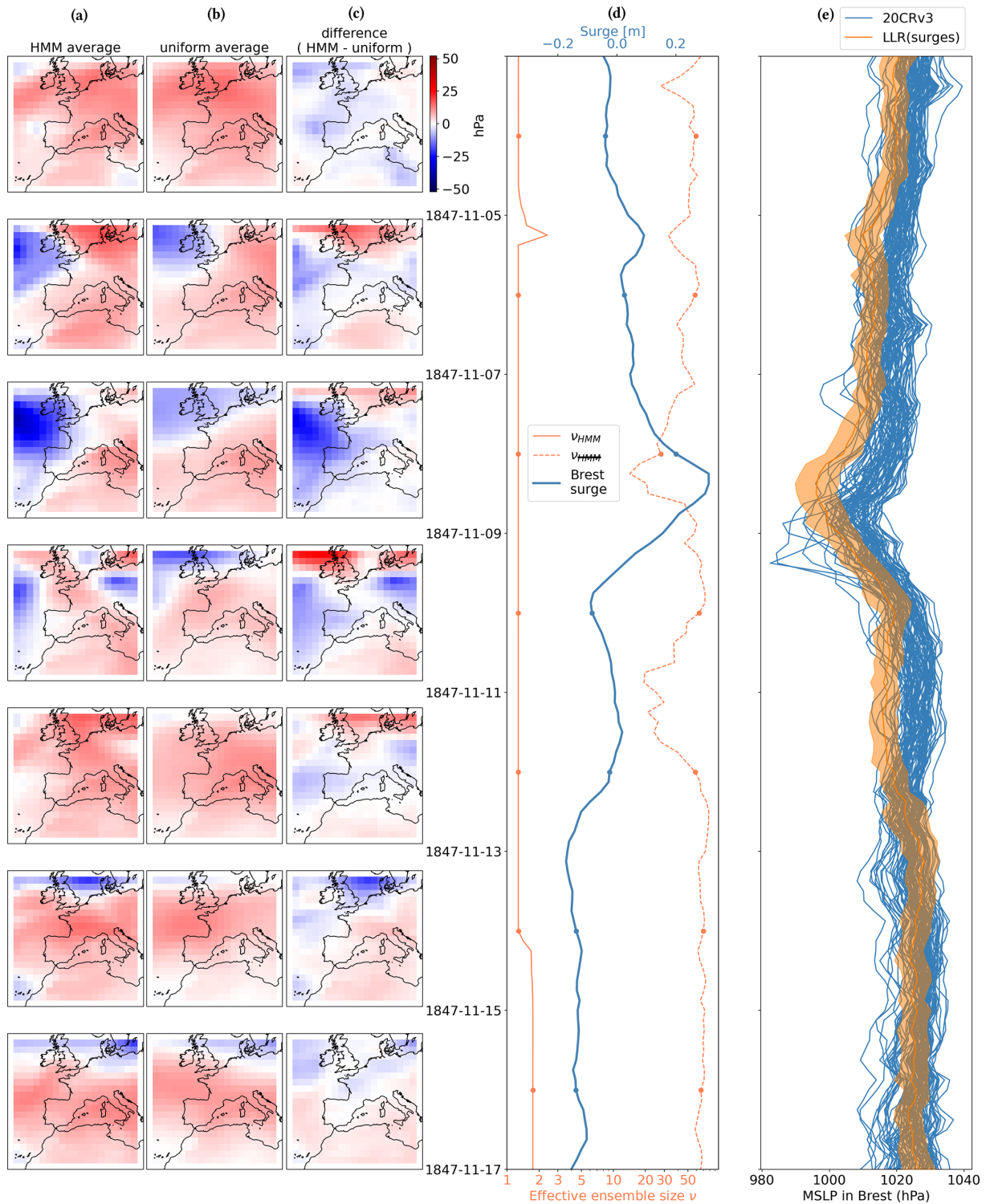
Figure 10. Mean result of the HMM algorithm in November 1847. **(a)** Average according to surge residual observations and HMM smoothing algorithm (probabilities $p_{\mathrm{HMM}}(i, t)$). **(b)** Average using constant uniform weights on 20CRv3 members. **(c)** Difference between HMM and uniform results. **(d)** Surge residual observations and effective ensemble sizes. **(e)** MSLP in Brest from ensemble members (blue lines) and local linear regression from surge residuals as an average ±1 SD (standard deviation). The dates of the left-hand plots are indicated with dots in **(d)** and **(e)**.
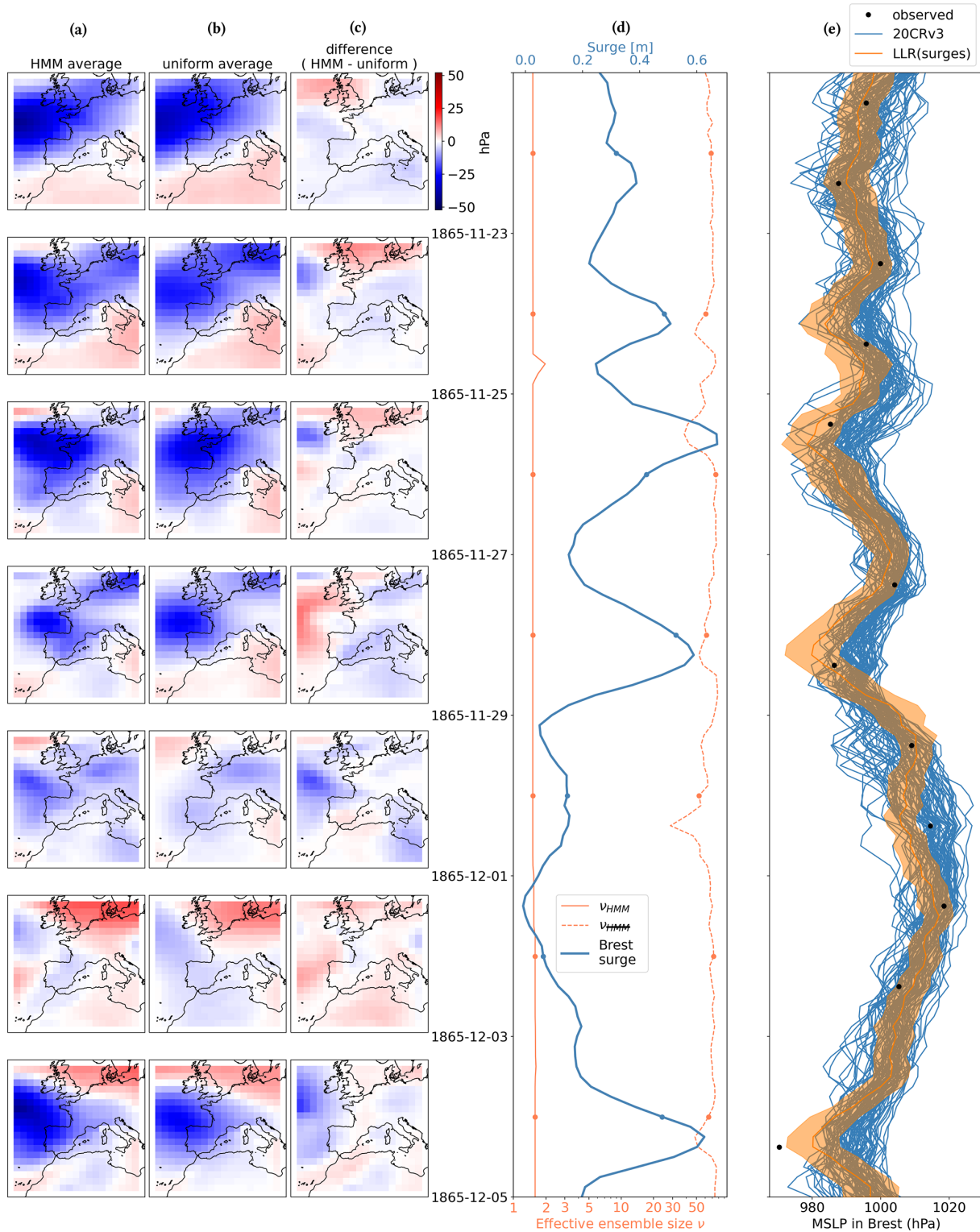
Figure 11. The same as Fig. 10 but for another storm. In this figure independent pressure observations in Brest are also available and shown as black dots in **(e)**.
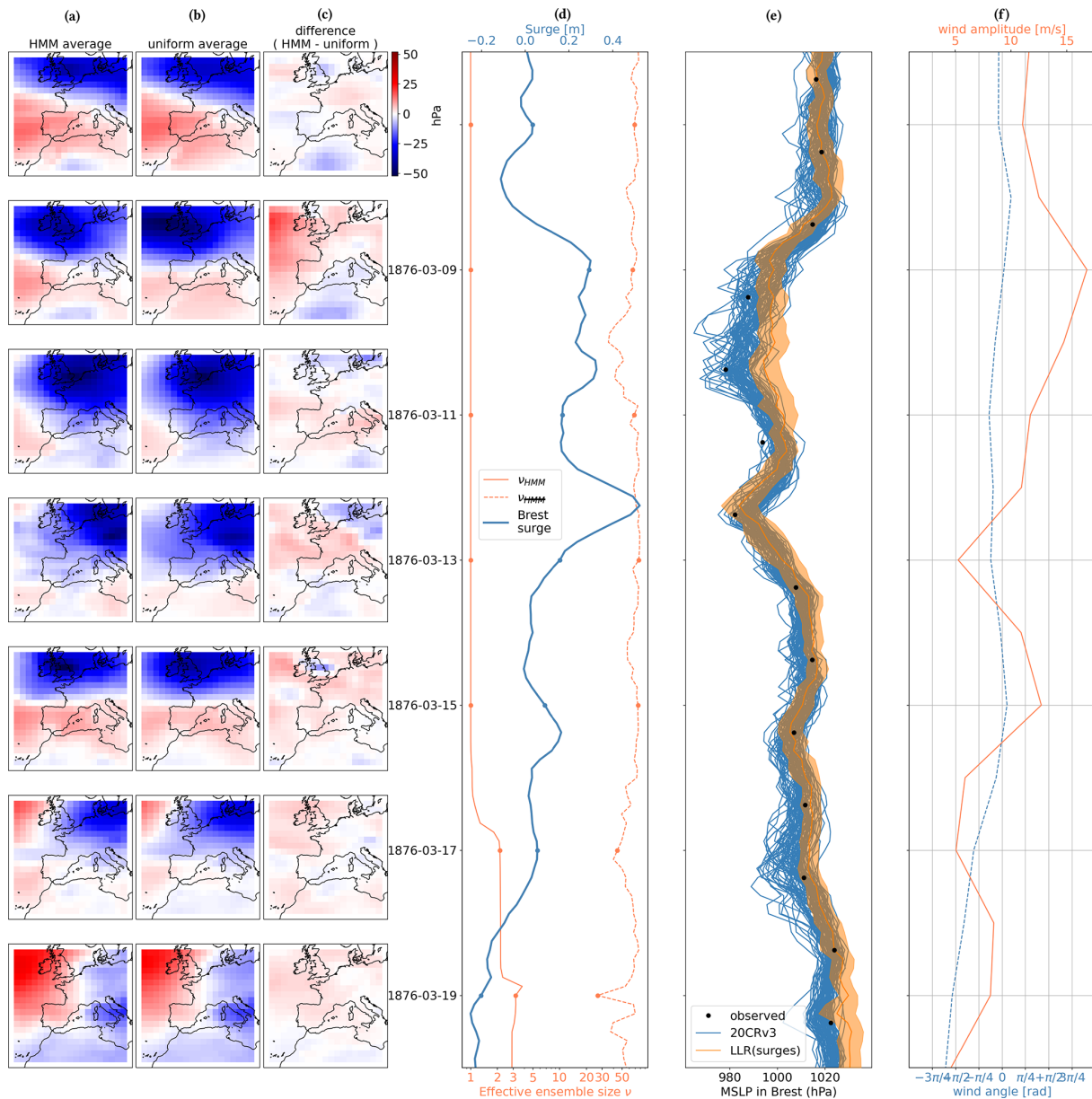
Figure 12. The same as Fig. 11 but for another storm. In (**f**) we also show the amplitude and direction of daily 10 m winds from the 20CRv3 member mean linearly interpolated at the city of Brest.

meaning that biases between the surge-residual-based and 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 us to validate reanalyses like 20CRv3 against independent data and to potentially identify anomalous trends or incorrect estimations of specific events. Combining our statistical approach with the physics-based approach of Hawkins et al. (2023) could allow us 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 es-

timate 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 an NWP 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 pre-
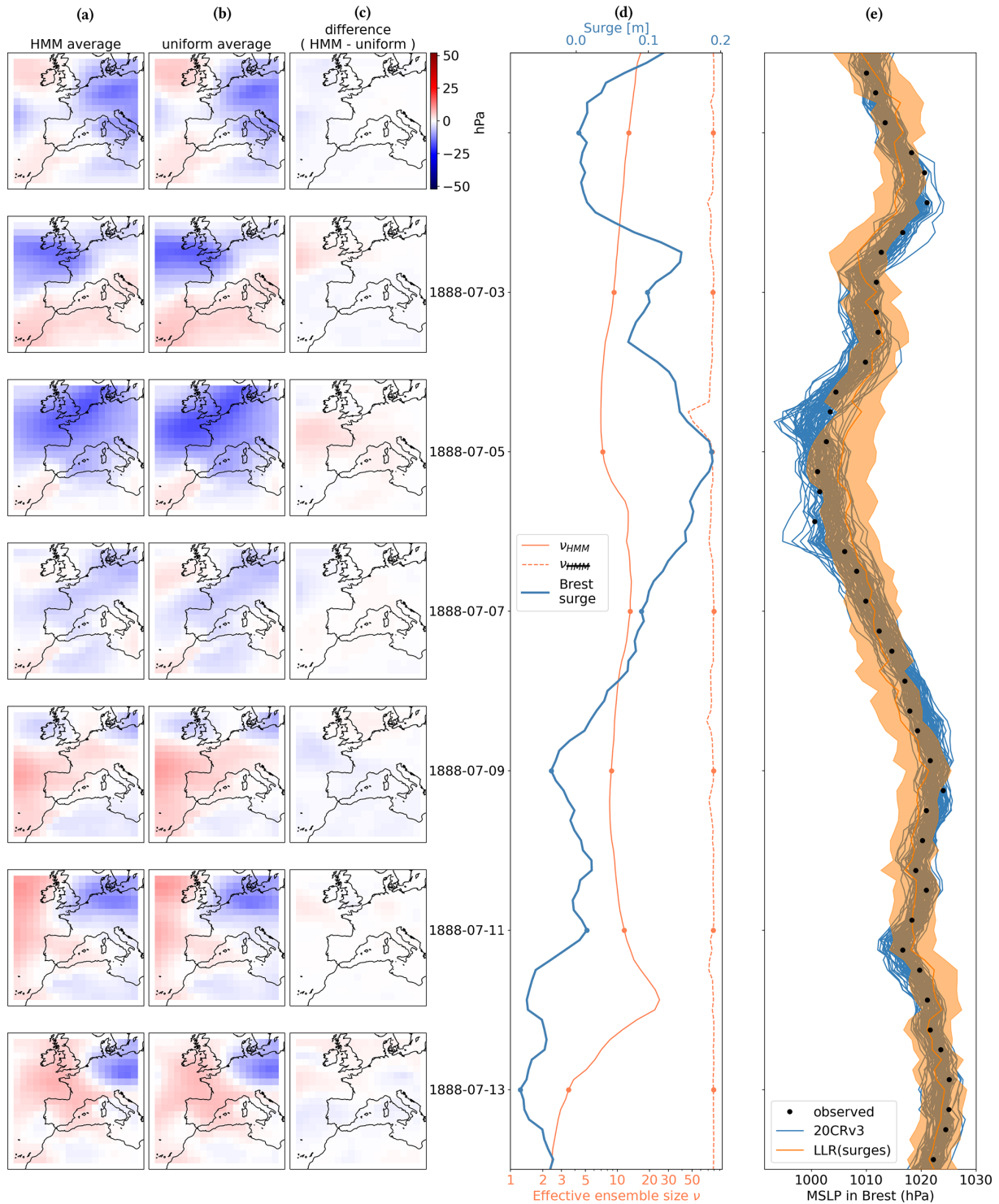
Figure 13. The same as Fig. 11 but for another event.

cise reanalyses, and applying these statistical relationship in the past to 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.

Code availability. The codes used to produce the figures in this article are available upon request from the authors.

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. A., Aguilar, E., Alexandersson, H., Barriendos, M., Brandsma, T., Cox, N. J., Della-Marta, P. M., Drebs, A., Founda, D., Gerstengarbe, F., Hickey, K., Jónsson, T., Luterbacher, J., Nordli, Ø., Oesterle, H., Petrakis, M., Philipp, A., Rodwell, M. J., Saladie, O., Sigro, J., Slonosky, V., Srnec, L., Swail, V., García-Suárez, A. M., Tuomenvirta, H., Wang, X., Wanner, H., Werner, P., Wheeler, D., and Xoplaki, E.: Daily mean sea level pressure reconstructions for the European–North Atlantic region for the period 1850–2003, J. 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, Int. J. Climatol., 29, 373–384, 2009.

Bertin, X.: Storm surges and coastal flooding: status and challenges, Houille Blanche, 2, 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, Clim. Past, 7, 265–276, https://doi.org/10.5194/cp-7-265-2011, 2011.

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., Cornes, R., Domínguez-Castro, F., Filipiak, J., Founda, D., García Herrera, R., Gergis, J., Grab, S., Hannak, L., Huhtamaa, H., Jacobsen, K. S., Jones, P., Jourdain, S., Kiss, A., Lin, K. E., Lorrey, A., Lundstad, E., Luterbacher, J., Mauelshagen, F., Maugeri, M., Maughan, N., Moberg, A., Neukom, R., Nicholson, S., Noone, S., Nordli, Ø., Björg Ólafsdóttir, K., Pearce, P. R., Pfister, L., Pribyl, K., Przybylak, R., Pudmenzky, C., Rasol, D., Reichenbach, D., Řezníčková, L., Rodrigo, F. S., Rohr, C., Skrynyk, O., Slonosky, V., Thorne, P., Valente, M. A., Vaquero, J. M., Westcottt, N. E., Williamson, F., and Wyszyński, P.: Unlocking pre-1850 instrumental meteorological records: A global inventory, B. Am. Meteorol. Soc., 100, ES389–ES413, 2019.

Bryant, K. M. and Akbar, M.: An exploration of wind stress calculation techniques in hurricane storm surge modeling, J. Mar. Sci. Eng., 4, 58, https://doi.org/10.3390/jmse4030058, 2016.

Cazenave, A. and Llovel, W.: Contemporary sea level rise, Annu. Rev. Mar. Sci., 2, 145–173, 2010.

Codiga, D. L.: Unified tidal analysis and prediction using the UTide Matlab functions, GSO Tech. Rep., Graduate School of Oceanography, Univ. of Rhode Island Narragansett, RI, 59 pp., 2011.

Compo, G. P., Slivinski, L. C., Whitaker, J. S., Sardeshmukh, P. D., McColl, C., Brohan, P., Allan, R., Yin, X., Vose, R., Spencer, L. J., Ashcroft, L., Bronnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cra, T. A., Crouthamel, R., Dominguez-Castro, F., Freeman, J. E., Gergis, J., Giese, B. S., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kennedy, J., Kubota, H., Blancq,

F. L., Lee, T., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Tinz, B., Titchner, H., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: The international surface pressure databank version 4, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO, https://doi.org/10.5065/9EYR-TY90, 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., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The twentieth century reanalysis project, Q. J. Roy. Meteorol. Soc., 137, 1–28, 2011.

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

Fan, J.: Local linear regression smoothers and their minimax efficiencies, Ann. Stat., 21, 196–216, 1993.

Gregory, J. M., Griffies, S. M., Hughes, C. W., Lowe, J. A., Church, J. A., Fukimori, I., Gomez, N., Kopp, R. E., Landerer, F., Le Cozannet, G., Ponte, R. M., Stammer, D., Tamisiea, M. E., and van de Wal, R. S. W.: Concepts and terminology for sea level: Mean, variability and change, both local and global, Surv. Geophys., 40, 1251–1289, 2019.

Haigh, I. and Marcos, M.: GESLA (Global Extreme Sea Level Analysis), https://gesla787883612.wordpress.com/downloads/ (last access: 7 October 2024), 2024.

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, Geosci. Data J., 10, 293–314, 2023.

Hansen, B.: Econometrics, Princeton University Press, ISBN 9780691235899, 2022.

Harter, L., Pineau-Guillou, L., and Chapron, B.: Underestimation of extremes in sea level surge reconstruction, Sci. Rep., 14, 14875, https://doi.org/10.1038/s41598-024-65718-6, 2024.

Hawkins, E.: Weather Rescue Data, GitHub [data set], https://github.com/ed-hawkins/weather-rescue-data/tree/main/ (last access: 7 October 2024).

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., McColl, C., Schurer, A. P., Slivinski, L., and Williams, J.: Rescuing historical weather observations improves quantification of severe windstorm risks, Nat. Hazards Earth Syst. Sci., 23, 1465–1482, https://doi.org/10.5194/nhess-23-1465-2023, 2023.

Horsburgh, K. and Wilson, C.: Tide-surge interaction and its role in the distribution of surge residuals in the North Sea, J. Geophys. Res.-Oceans, 112, C08003, https://doi.org/10.1029/2006JC004033, 2007.

Jones, P. D., Folland, C. K., Jacobeit, J., Yiou, P., Brunet M., Luterbacher, J., Moberg, A., Chen, D., and Casale, R.: EMULATE (European and North Atlantic daily to MULtidecadal climATE variability), UEA [data set], https://crudata.uea.ac.uk/projects/emulate/LANDSTATION_MSLP/ (last access: 7 October 2024), 2024.

Krueger, O., Schenk, F., Feser, F., and Weisse, R.: Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations, J. 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), Adv. Water Resour., 31, 233–250, 2008.

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

Liu, J. S.: Nonparametric hierarchical Bayes via sequential imputations, Ann. Stat., 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, Geosci. Data J., 8, 144–153, 2021.

Melchior, P.: The tides of the planet Earth, Oxford, https://ui.adsabs.harvard.edu/abs/1983opp..book.....M (last access: 2 October 2024), 1983.

Météo France: Données climatologiques de base – horaires, https://www.data.gouv.fr/fr/datasets/donnees-climatologiques-de-base-horaires/ (last access: 7 October 2024), 2024.

NOAA-CIRES-DOE: 20th Century Reanalysis, https://psl.noaa.gov/data/20thC_Rean/ (last access: 7 October 2024), 2024.

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, Q. J. Roy. Meteorol. Soc., 144, 317–332, 2018.

Pineau-Guillou, L., Delouis, J.-M., and Chapron, B.: Characteristics of Storm Surge Events Along the North-East Atlantic Coasts, J. Geophys. Res.-Oceans, 128, e2022JC019493, https://doi.org/10.1029/2022JC019493, 2023.

Ponte, R. M.: Understanding the relation between wind-and pressure-driven sea level variability, J. Geophys. Res.-Oceans, 99, 8033–8039, 1994.

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

Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition, Proc. 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, B. Am. Meteorol. Soc., 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, J. 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., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński,

P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. Roy. Meteorol. Soc., 145, 2876–2908, 2019.

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

Tadesse, M. G. and Wahl, T.: A database of global storm surge reconstructions, Sci. Data, 8, 125, https://doi.org/10.1038/s41597-021-00906-x, 2021.

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

UCO/CIRES|DOC/NOAA/OAR/ESRL/PSL: The International Surface Pressure Databank version 4, UCO/CIRES|DOC/NOAA/OAR/ESRL/PSL [data set], https://doi.org/10.5065/9EYR-TY90, 2024a.

UCO/CIRES|DOC/NOAA/OAR/ESRL/PSL: Monthly Maps: Number of Observations per Day for International Surface Pressure Databank Version 4.7, https://web.archive.org/web/20230527064622/https://psl.noaa.gov/data/20CRv3_ISPD_obscounts_bymonth/ (last access: 7 October 2024), 2024b.

Wohland, J., Omrani, N.-E., Witthaut, D., and Keenlyside, N. S.: Inconsistent wind speed trends in current twentieth century reanalyses, J. Geophys. Res.-Atmos., 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., and Merrifield, M. A.: Forcing factors affecting sea level changes at the coast, Surv. Geophys., 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 Dynam., 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, Clim. Past, 10, 797–809, https://doi.org/10.5194/cp-10-797-2014, 2014.

Zucchini, W., MacDonald, I. L., and Langrock, R.: Hidden Markov models for time series: an introduction using R, CRC Press, https://doi.org/10.1201/9781420010893, 2017.