Extension of a high temporal resolution sea level time series at Socoa (Saint Jean-de-Luz, France) back to 1875

Md Jamal Uddin Khan[1], Inge Van Den Beld[2], Guy Wöppelmann[1], Laurent Testut[1], Alexa Latapy[2], Nicolas Pouvreau[2]

[1]LIttoral ENvironnement et Sociétés - UMR 7266, ULR/CNRS, La Rochelle, 17000, France
[2]Service Hydrographique et Océanographique de la Marine (SHOM), Brest, 29200, France

Correspondence to: Md Jamal Uddin Khan, jamal.khan@univ-lr.fr

Abstract. In this data paper, the sea level time series at Socoa (Saint Jean-de-Luz, Southwestern France) has been extended through a data archaeology exercise. We conducted a comprehensive research of national and local archives to catalogue water level records stored in ledgers (handwritten record books) and charts (marigrams from mechanical float gauges), along with other associated documents (metadata). A dedicated effort was undertaken to preserve more than 2000 documents by archiving them in digital formats. Using this large set of rescued documents t~~T~~he Socoa time series has been extended back to 1875, with more than 58 station-years of additional data~~, using this large set of rescued documents~~. The final time series has hourly sampling, while the raw dataset has finer sampling frequency up to 5 minutes. By analysing precise levelling information, we assessed the continuity of the vertical datum. We also compared the new century-long time series to nearby tide gauge data to ensure its datum consistency. While the overall quality of the time series is generally good, siltation of the stilling well has occasionally affected certain parts of the record. We have successfully identified these impacted periods and flagged the corresponding data as doubtful. This extended high resolution sea level time series at Socoa, spanning over more than 100 years, will be valuable for advancing climate research, particularly in studying the decadal scale variations in the North Atlantic, and investigating the storminess and extreme events along the French Basque coast.

1 Introduction

Tide gauge records are among the oldest instrumental datasets. They have a crucial ~~contribution~~ role in~~to~~ our understanding of contemporary sea level variability and climate change (Ekman 1999, Church et al. 2013). Between 1901 and 2010, the Global Mean Sea Level (GMSL) has been rising at a rate of 1.5±0.4 mm/year (Oppenheimer et al. 2019). The assessment of the sea level rise during the 19th and 20th century (which provides a baseline for assessing the changes in the 21st century) relies on long-term tide gauge records (e.g., Dangendorf et al. 2017). A subset of these long time series has become accessible to the scientific community through the process of discovery, digitization and reconstruction, a procedure which is known as sea level data archaeology (Woodworth 1999, UNESCO/IOC 2020).

Over the last few decades, data archaeology has been applied to different parts of the globe to construct and study long-term sea level variability and change (e.g., Woodworth 1999, Hunter et al. 2003, Woodworth et al. 2010a, Talke et al. 2018). For instance, Woodworth (1999) recovered and analysed the mean high-water levels recorded at Liverpool starting from 1768. Similarly, Wöppelmann et al. (2006a) applied data archaeology to reconstruct a sea level series at Brest back to the beginning of the 18th century. Both studies concluded a similar result – a rising trend from the beginning of the 20th century with an acceleration towards the second half (Wöppelmann et al. 2008). In the southern hemisphere, where the data coverage is generally sparse, Testut et al. (2010) recovered water level measurements recorded in 1874 on Saint Paul Island in the southern Indian ocean. Combined with recent measurements they revealed a statistically zero relative sea level trend. Similarly, Hunter et al. (2003) recovered and analysed intermittent sea level records made in Port Aurthur, Tasmania (southern Australia). They ~~reporting~~ reported an average sea level trend of 0.8±0.2 mm/year relative to land and 1.0±0.3 mm/year considering vertical land motion over 1841 to 2002. At local scale, some data archaeology studies combine multiple nearby historical tide gauge records into one long time series for sea level trend analysis (Marcos et al. 2011; 2021; Woodworth, 1999), whereas, regionally, Hogarth et al. (2020) combined data archaeology, numerical modelling, and statistical minimization approaches to further extend the mean sea level record over the British Isles. They estimated a robust regional mean sea level trend of 2.39±0.27 mm/year over 1958 to 2018, with an acceleration of 0.058±0.030 mm/year^2.

Alongside the mean sea level, ~~In long term,~~ tide has also been found to manifests long-term~~has also been through~~ ~~changes which requires~~evolution and high-frequency (typically hourly or less) data are needed to ~~analyse~~ investigate such changes (Woodworth, 2010b, Haigh et al. 2020). During the first half of the 19th century automatic mechanical tide gauges started to appear~~appearing~~, paving the way for systematic continuous measurements of water levels at high frequency (Wöppelmann et al 2006b). Taking advantage of the archaeology of the availability of such high-frequency long-term sea level records, Pouvreau et al. (2006) analysed the secular trend in the evolution of M2 (the main lunar semidiurnal tide) at Brest. They reported no significant trend but long-period oscillations with 141 years period. Recent research on the tidal change shows that the long-term changes are not linear (Ray and Talke 2019). Pan and Lv (2021) reported a quasi 60-year oscillation in the global tide from a global set of long high-resolution sea level time series. These non-linear changes are sometimes with break points around late 19th century (Pineau-Guillou et al. 2021). These ~~contemporary~~ recent results further highlight the necessity for long high frequency sea level time series for studying the evolution of the tide.

High frequency past tide gauge time series are also very useful for the analysis of extreme sea level (ESL), which is a major societal concern due to ongoing sea level rise (Oppenheimer et al. 2019). Dedicated studies have been conducted to understand the dynamics and the drivers of ESL at local (Letetrel et al. 2010, Talke et al. 2014, Talke et al. 2018), regional (Wahl and Chambers 2015, Marcos et al. 2015, Marcos and Woodworth 2017), and global scales (Menéndez and Woodworth 2010). Among various factors, sea level rise is shown to be the first order driver of the observed ESL change in most of the coastline (Menéndez and Woodworth 2010) and projected to be the major factor for future ESL changes globally (Muis et al. 2016, Fox-Kemper et al. 2021). However, ESL variability also varies regionally depending on the local and regional processes (Menéndez and Woodworth 2010). Long high-resolution sea-level time series are particularly interesting to unravel the contribution from mean sea level change (Letetrel et al. 2010), seasonal and decadal variability (Menéndez and Woodworth 2010, Marcos et al. 2015), and local changes (Talke et al. 2014, Talke et al. 2018). Indeed, it is concluded with *high confidence* (meaning high agreement and robust evidence in the available literature) that consideration of localized storm surge processes is essential to monitor the trend in ESL (Oppenheimer et al. 2019). Such monitoring requires reliable long high-resolution observations. A long time series provides an additional benefit by reducing the uncertainty of ESL analysis (Coles 2001), which equates to better flood risk assessment.

Global high-resolution datasets, like GESLA (Global Extreme Sea Level Analysis; Woodworth 2016, Haigh et al. 2022~~0~~) ~~has~~ have been ~~instrumental~~ important for the current global, as well as regional scale studies on ESL. Such ~~a~~ datasets also allowed global and regional analysis of tide (Piccioni et al. 2019), the non-linearity of tide-surge interaction (Arns et al. 2020), as well as data driven modelling of surges (Tadesse et al. 2020). Yet, most of the stations in GESLA have ~~a~~ time series shorter~~less~~ than 50-year long. As demonstrated by previous studies (Wöppelmann et al. 2014, Talke et al. 2014, 2018, Talke and Jay 2017), data archaeology offers a solution to this scarcity of long-term data by tapping into the potential of rescuing numerous instrumental records worldwide (Bradshaw et al. 2015).

As a response to this lack of long-term high temporal resolution records for the assessment of short to long-timescale processes, this article presents a data rescue and archaeology effort to make available a high temporal resolution long-term sea level time series at Socoa. The tide gauge is located in Saint-Jean-de-Luz, France, along the Basque coast in the Bay of Biscay (Figure 1a,b). The region is dominated by strong tides (meso-tidal) and energetic waves (Dodet et al. 2019), making it an important observation location. The tide gauge station at Socoa was established in 1875. However, the earliest available data in the French reference repository before this work (e.g., https://data.shom.fr/donnees/refmar/95, last accessed 10 Apr. 2022) starts from 1942, with continuous recording from 1964 only (Arnoux et al. 2021). The data is available at hourly sampling before 2011, and afterwards both the high-frequency (1min) and the hourly data is available.

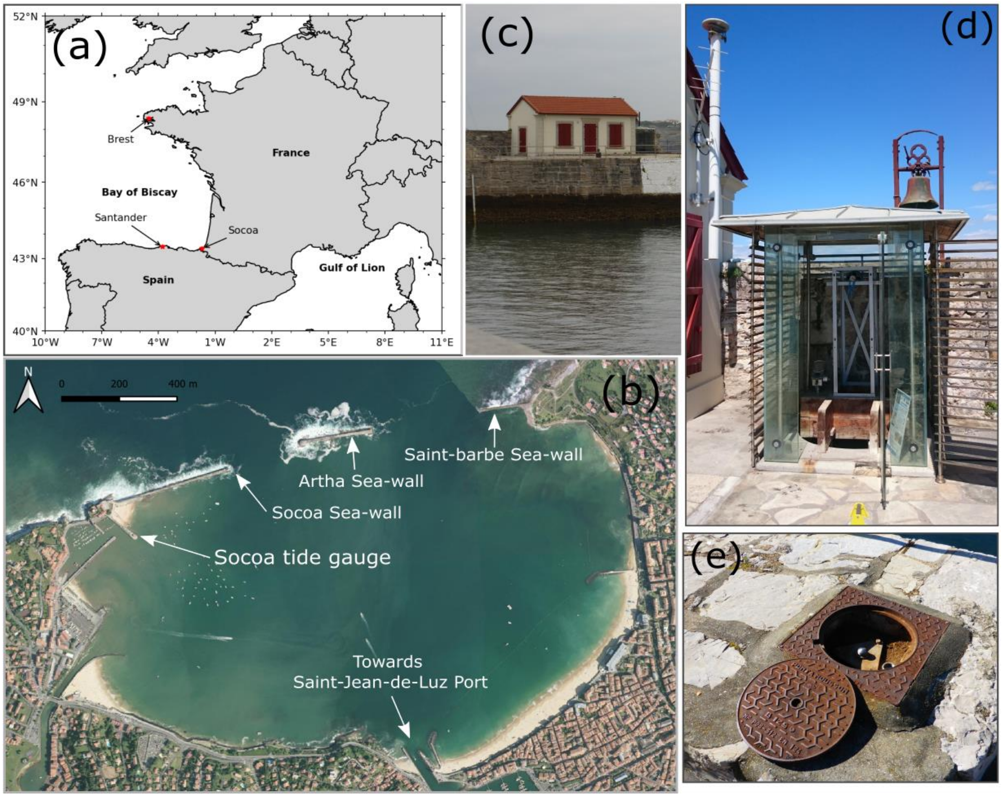

Figure 1: (a) Study area indicating the location of the Socoa tide gauge and other tide gauges used in this study (b) The satellite view of the study area (source: IGN geoservices, https://geoservices.ign.fr/). (c) A view of the tide gauge surroundings and the housing location. (d) The tide gauge house over the top of the stilling well. (e) The nearby tide gauge benchmark IGN O.A.K3L3-5-IV. The photos in panels c, d, e are provided by SONEL (https://www.sonel.org/).

The new sea level time series extended in this work is shown in Figure 2a with existing data in grey and the new data in black. To describe the development of this new time series, the history of the Socoa tide gauge is first presented through various instrumentation periods in Section 2. ~~First, the history of the Socoa tide gauge is presented through various instrumentation periods~~A~~ and a~~ summary of the rescued documents (containing data and metadata) ~~are~~is also presented alongside~~discussed in Section 2~~. The rescue process and the analysis of the time series are described in Section 3, which is followed by the quality control and the data quality assessment in Section 4. In Section 5, we present a trend analysis. The data availability is detailed in Section 6 with concluding remarks in Section 7.

2 History of Socoa tide gauge station and rescued documents

The Socoa tide gauge station was established during the 1873-1875 period. A dedicated housing (Figure 1c) with an adjacent stilling-well system (Figure 1d) was built to host the original tide gauge, for handling the daily tasks of the gauge keeper, as well as to store the paper charts. Several water level instruments were operated during various periods covering the 19th and

20th centuries and followed on into the 21st century with modern technology that is still currently operating (Fig. 1 in Martín Míguez et al. 2008). In the following sub-sections, we provide descriptions of each instrumentation period, along with detailed information about the ~~data and the metadata~~catalogued documents.

2.1 The Chazallon tide gauge period: 1875 to ~~-~~1920

During the 1840s, several float-type tide gauges devised by Antoine M. R. Chazallon (1802-1872) were installed along the French coasts. A schematic of the tide gauge is shown in Figure 2c. Like most of the float gauges, the displacement of the float is reduced through a mechanical system and the resulting sea level variation is recorded on a paper chart controlled by a clock (IOC 1985). One of the Chazallon type tide gauges was installed in La Rochelle (Vieux Port) and operated from 1863 to 1874 (Gouriou et al. 2013). This tide gauge was then transferred to ~~the~~ Socoa in 1875. At Socoa, the float of the tide gauge was installed in a stilling well located near the housing (Figure 1d). The Chazallon tide gauge was operational until 1920 when most tide gauges around the French coast were discontinued. Up u~~U~~ntil then, the tide gauge was operated by the Service Hyrdrographique de la Marine (SHM), the predecessor of the current Service hydrographique et océanographique de la Marine (Shom, https://www.shom.fr/).

Two types of historical records have been found for the Chazallon period – 1) a subset of charts, 2) ledgers. The ledgers are 32x49cm paper documents with water level values obtained by inspection of the charts by an operator. The ledgers (Supplementary Figure S1) and charts are currently stored at the Shom archive located in Brest (France).

2.2 Temporary tide gauge during World War II period: ~~(~~1942 to 1944~~)~~

In the currently available archives, e.g., Shom, or Permanent Service for Mean Sea Level (PSMSL, Holgate et al. 2013), there are data available for~~during~~ the World War II (WWII) period— from November 1942 to May 1944. In fact,~~While~~ 44 charts were found in the Shom archive at Brest covering this period, but the~~,~~ rescued metadata information (discussed in Section 2.5) does not include mention of~~report~~ any tide gauge operating at Socoa. After inspection, we found that these charts had a different paper size ~~of these charts~~ compared to those of the Chazallon or Brillie eras, which indicates that it was a different type of tide gauge. In addition, the paper charts bear German markings. Local historians confirm that there~~it~~ was indeed another tide gauge, installed by the Germans, on the other side of the Socoa bay. The data during WWII appears to be consistent with the rest of the record, as confirmed by a tidal analysis.~~ No record was found between this German tide gauge (1944) and the Brillie tide gauge (1950).~~

2.3 The Brillie tide gauge period: 1950 to 2004

During 1950, a Brillie type float gauge (large model type, Robertou 1955) was installed (Figure 2d) in the stilling well built for the Chazallon gauge. Each chart of this large model type is 72x50 cm in dimension (x-y). The x-axis of the paper is divided into 24 divisions (corresponding to each hour), and each hourly division is further subdivided into 10-minute subdivisions. On the y-axis there is a 1/10 reduction in the water level variation, i.e., the y-axis represents a range of 5m in sea level (50cm x

10). There are further subdivisions on the charts of 25cm and 5cm. Surprisingly, no documentation was found in the catalogued archives regarding the installation and operation of this tide gauge, except knowing the physical existence of the tide gauge itself till 2004.

The recording of water levels by the Service Maritime des Ponts et Chaussées starts in December 1950. In total, 2477 charts spanning the period December 1950 to 2001 were recovered from the local archive - Archives des Pyrénées-Atlantiques at Bayonne. No data were found for 2002-2003.

The water level curves on the charts during the Brillie period were recorded in the legal time of France (see Section 3.2 for further details). Each curve in the charts represents one day of water level record, and each chart was~~is~~ found to contain multiple days of recording (Supp. Figure S1c). Typically, up to 14 days of sea levels were recorded on these paper charts. In most cases, the charts were accompanied by a check sheet, which are an important part of the data rescue. See Figure 2b for the availability of the charts and check sheets.

2.4 The modern instrumentation: 2004 to ongoing

With the advent of the modern RONIM (Réseau d’Observation du Niveau des Mers) sea level measurement network (Martín Míguez et al. 2008), the Brillie tide gauge at Socoa was decommissioned and replaced with a digital radar gauge in 2004 (Figure 2e). This radar gauge is currently co-located with a continuously operating geodetic Global Navigation Satellite System (GNSS) station (https://www.sonel.org/spip.php?page=gps&idStation=835). The antenna of the GNSS station is visible in Figure 1d. This tide gauge is currently maintained by Shom. Its sea level data and metadata are available at the Shom data portal (https://data.shom.fr) in both raw and post-processed quality-controlled form. Raw data is sampled at 1-minute. Data from the tide gauge is accessible through the Global Telecommunication System (GTS) network, which enables a real-time data flow. This data flow enables real-time monitoring of the gauge, for instance via the Intergovernmental Oceanographic Commission (IOC) Sea Level Station Monitoring Facility (see http://www.ioc-sealevelmonitoring.org/station.php?code=scoa2). Note that the data from the IOC Facility should not generally be used for any scientific application, as its main design and procedures have been designed for monitoring the operational status of the gauges (Aarup et al. 2019).

It is worth noting here that at Socoa the position of the tide gauge remained the same over its full period of observation from November 1875 until now through the various instrumentation periods. The modern tide gauge is operating within the same stilling well, hence preserving the spatial and environmental continuity with the past measurements. There is a caveat to this statement concerning modifications made to the stilling well infrastructure in the early recording period, which we will illustrate later in Section 4.3.3.

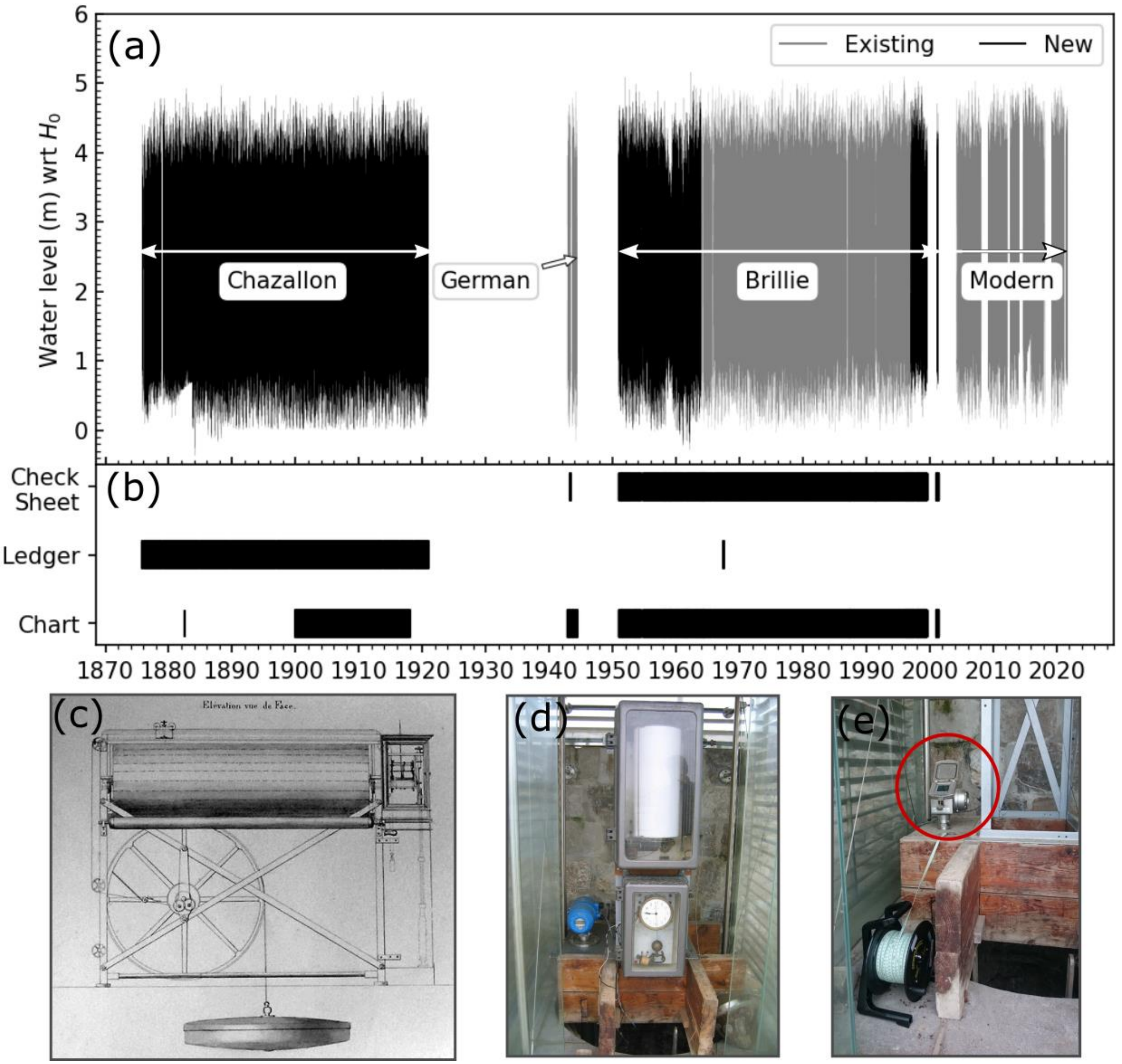

Figure 2: (a) Time series of water level at Socoa with digitized data (in red the new datasets from this study). (b) Coverage of the rescued Registries, Charts, and associated check sheets. (c) A schematic of the Chazallon tide gauge (adopted from Pouvreau~~ et al.~~ 2008). (d) A photograph of the Brillie type tide gauge that was ~~operated ~~installed unti~~l~~l 2004~~1~~ and (e) the modern radar gauge (Photograph taken by authors during a field campaign in 2017).

A chronology of the available measurement periods, instruments, recording mediums, time systems during recording, time sampling before and after digitization and ~~the ~~source archives is summarised in Table 1. The reconstructed time series is in

Universal Coordinated Time (UTC), which is further discussed in Section 3.2. As the modern instrument record starting in 2004 is not part of the data archaeology exercise, it is not further discussed in the following sections.

Table 1: Overview of the instrumentation periods, original data storage mediums, sampling period of the source and digitized data, and time system of the source observations.

Period	Instrument	Medium	Sampling		Source time system	Archive
			Source	**Digitized**		
1875-11-01 ~~10:50:20~~ to 1893-12-31 ~~23:10:37~~	Float (Chazallon)	Ledger	15 min	1 hour	AST	Shom
1894-01-01 ~~00:06:43~~ to 1897-12-31 ~~23:06:43~~	Float (Chazallon)	Ledger	15 min	1 hour	MST	Shom
1898-01-01 ~~00:06:43~~ to 1920-12-14 ~~12:06:43~~	Float (Chazallon)	Ledger	1 hour	1 hour	MST	Shom
1942-11-20 ~~23:00:00~~ to 1944-05-29 ~~22:00:00~~	Float (Unknown)	Chart	Continuous	1 hour	UTC+1/+2	Shom
1950-12-18 ~~11:30:00~~ to 1963-12-23 ~~11:25:00~~	Float (Brillie)	Chart	Continuous	5 min	UTC+1	AD64-Bayonne
1964-01-03 ~~23:00:00~~ to 1997-01-07 ~~08:00:00~~	Float (Brillie)	Chart, Digital	Continuous	1 hour	UTC+1 till 1976, UTC+1/+2 since 1976	AD64-Bayonne, Shom
1997-01-07 ~~10:00:00~~ to	Float (Brillie)	Chart	Continuous	5 min	UTC+1/+2	AD64-Bayonne

1999-08-27 ~~08:45:00~~						
2001-02-20 ~~08:15:00~~ to 2001-05-29 ~~08:20:00~~	Float (Brillie)	Chart	Continuous	5 min	UTC+1/+2	AD64-Bayonne
2004-04-13 ~~14:00:00~~ to 2004-05-31 ~~23:00:00~~	Float (Brillie)	Chart, Digital	Continuous	1 hour	UTC+1/+2	Unknown
2004-06-01 ~~00:00:00~~ to 26/04/2011	Radar (Krone Optiwave 7300C)	Digital	1 hour	1 hour	UTC	Shom
26/04/2011 to date	Radar (Krone Optiwave 7300C)	Digital	1 min	1 hour	UTC	Shom

AST: Apparent Solar Time.
MST: Mean Solar Time.
UTC: Universal Time Coordinated.

Shom : Service hydrographique et océanographique de la Marine
AD64-Bayonne: Archive Departmental 64 - Pyrénées-Atlantiques – Bayonne

2.5 Complementary metadata

During the rescue process, administrative documents were found in the archives in which the Socoa tide gauge was mentioned. These documents include tide gauge journals for the Chazallon period containing logs of tide gauge operations, the correspondence with the ministry (ministry of public works, and ministry of marine and colonies), the engineering and hydrographic survey reports, the quotes for works, drawings etc. The hydrographic survey reports were of particular interest to assess the datum continuity of the tide gauge records (Section 3.3). All the documents form an ancillary part of the available metadata and are provided as supplementary files to the dataset (see Section 6 Data availability).

3 Digitization and reconstruction of the time series

3.1 Scanning and digitization

The water levels recorded by the Chazallon tide gauge during the 1875 to ~~-~~1920 period exist~~s~~ in two different mediums: charts and hand-written ledgers. The large charts were stored for many decades in the archive. They were in an advanced state of deterioration, which prevented them being scanned and rescued. Only the ledgers could be scanned and rescued. The scanned documents are stored as PDF (portable document format) files, each 1-1.2 MB large.

Once the scanning was complete, the process of converting the hand-written text to data (digitization) was done manually from the scanned document to a computer spreadsheet. The paper for the ledgers was designed for transcribing water levels at 15-minute intervals. However, the water levels were transcribed at 15-minute intervals till 1897 only. Afterwards, the transcriptions were done at 1-hour interval. To speed up the manual digitization process, a choice was made to digitize the water level record at hourly interval only.

For the Chazallon period starting from November 1875 to December 1920, 541 ledgers were recovered corresponding to 45 years of sea level records. More than 390,000 values were digitized manually, which corresponds to several weeks of full-time work. During digitization, the time values~~data~~ were digitized directly as in the ledgers. Sea levels from 1875 to 1893 were recorded in Apparent Solar Time, and from 1894 to 1920 they were~~it was~~ recorded in Mean solar time. The conversion of these time records into UTC is described in Section 3.2.

Unlike the early Chazallon-era, the whole recovered archive of the charts covering 1942-2004 period was scanned (with a photo-scanner) and rescued. Most of these charts were accompanied by check sheets. These documents contain relevant information on time and water level at the time of replacing the chart paper. The available check sheets were converted into digital form by a photo camera and later used as metadata for identifying problems, especially related to the slowing down of the clock (See Section 4.2).

Prior to this study, an hourly record of sea level at Socoa from 1964-1996 existed in digital form and was available from the Shom data portal (https://data.shom.fr). Hence, we applied the water level extraction only on the charts that are outside the range of pre-existing digitized data. From 1950 to 2001, the number of available charts amounts to 777. During the scanning phase, the charts were visually sorted into three categories depending on their conditions (good, mildly or badly damaged from mould, and faded) (Supplementary Figure S2). Among the 777 charts, 18.3% (142 charts) were found to be in good condition (Figure S2). Fifty charts were found to have mild mould (mildly damaged), and 32 charts were found badly covered by mould (badly damaged). The majority of the documents, as many as 553 charts (71.2% of the digitized charts), were found to be of the third category with faded water level curve lines.

To extract the water levels from the chart images, a specialized open-access software called Numerisation des Niveaux d'EAU (NUNIEAU) was used (Ullmann et al. 2011). For a given chart image, this software can trace the recorded water curve-line based on a colour-separation technique. Additionally, the software has built-in features to assign time and height scales in the chart (Supplementary Figure S3).

Since the algorithm in NUNIEAU is based on colour separation, the water levels are easier to extract from clean charts. The charts in good condition did not need any further image processing to be applied before passing them~~it~~ through NUNIEAU software. In the second category are the charts which are damaged from mould. These charts were found to be still processable through the processing chain, except for some badly damaged ones which were fully covered by mould. These bad damages essentially translated into a loss of data. Finally, in the third category where the water level curve were very faint, either fully or partially, we applied ~~. Further~~ image processing ~~was applied~~ to enhance the contrast to process them using NUNIEAU (Supplementary Figure S4).

From the charts, the water levels were extracted at a chosen 5-minute interval. ~~We have applied NUNIEAU on these charts to extract the water level at a chosen 5-minute interval.~~ This time interval corresponds to 20 pixels in the x-axis of the scanned charts (scanned at a resolution of 200 dpi). Going below this interval would be pointless as the higher-frequency fluctuations would have been mechanically filtered by the stilling well (IOC 1985).

The overall process of digitizing ledgers and charts was time-consuming which is known for this kind of data archaeology exercise (Latapy et al. 2022). This is obvious for manually digitizing the ledgers from scanned document to a spreadsheet table. But for the digitization of charts, which is a software-based extraction, it is less obvious. In practice, the processing chain could not be automated due to three main reasons. First and foremost, the faded charts needed additional image processing. Second, multiple days of water level were recorded in a single chart, partially overlapping themselves, which required dedicated masks to separate each day. Finally, the zero of the curves was needed to be set manually for each chart within the NUNIEAU software. In consequence of these delicate pre-processing steps, the overall chart digitization process was time-consuming, similar to manual digitization from ledgers, as well as challenging to implement in practice.

3.2 Time systems and conversion

Once the scanning and digitization were performed, the next important step was to reduce the records into a consistent time system, in this case, the Coordinated Universal Time (UTC) in zero-hour time zone UTC±00:00 (henceforth denoted simply as UTC). Over the recording period of the Socoa tide gauge, Apparent Solar Time (AST), Mean Solar Time (MST) and Legal time systems were used as listed in Table 1. The following subsections describe the detail of the conversion from each time system to UTC.

3.2.1 Apparent and Mean Solar Time

From 1875 to 1893, the ledger records are in local AST. As noted by Wöppelmann et al. (2014), AST was used in the earlier days of the Chazallon tide gauge era despite the MST being the legal time in France since the early 18th century. Afterwards, until 1920 the records are in local MST. We have first converted the AST to MST by adding their difference over the year, known as equation of time, E, to AST (Hughes et al. 1989, Müller 1995). Here E is computed using the formulation published by Bureau Des Longitudes (2011):

$$\begin{aligned}E \quad &= 7.362 \times sin(M) - 0.144 \times cos(M) + 8.955 \times sin(2 \times M) + 4.302 \times cos(2 \times M)\\ &+0.288 \times sin(3 \times M) + 0.133 \times cos(3 \times M) + 0.131 \times sin(4 \times M) + 0.167 \times cos(4 \times M)\\ &+0.009 \times sin(5 \times M) + 0.011 \times cos(5 \times M) + 0.001 \times sin(6 \times M) + 0.006 \times cos(6 \times M)\\ &-0.00258 \times t \times sin(2 \times M) + 0.00533 \times t \times cos(2 \times M)\end{aligned}$$

where~~ith,~~ t is ~~is~~ the time difference to 2000-01-01 00:00:00 (in year, negative for earlier years), and $M = 6.240060 + 6.283019552 \times t$ (in radians). Although the equation given by the Bureau Des Longitudes (2011) is ~~fitted~~ specified for 1900-

2100, w~~W~~e have used the same equation for the period late 1800, which induces only minor errors (order of seconds). ~~To convert from~~ MST was then converted to UTC by adding~~,~~ 404 seconds, which equals a correction of ~~a correction of~~ 4 minutes ~~per~~ for each degree of longitude difference between Socoa and Greenwich (zero-longitude)~~ was applied~~. ~~This amounts to 404 seconds to be added to the MST recorded in Socoa to get the time in UTC.~~

3.2.2 Legal time

During the Brillie tide gauge period, the measurements were recorded in the legal time. The history of the legal time in France is long, and we present here a summary to the detailed account of Poulle (1999). Since 1891 the legal time of Metropolitan France was established as the MST in Paris. In a law enacted in 1911, a correction of 9 minutes 21 seconds was applied to Paris MST to define the new legal time as the Greenwich Mean Time (GMT). In 1923, the law related to legal time was amended to introduce ‘summertime’ (Typically last Sunday of March to last Sunday of October), when the clocks are advanced by 1 hour. During 1940-1941, timekeeping was different between German-occupied and -free areas. However, during 1942-1944, the legal time throughout France was essentially GMT+2 during summer and GMT+1 during winter. Post WW2, France switched to using GMT+1 throughout the year on 18 November 1945. The Universal Coordinated Time (UTC), formulated in 1960, gradually replaced GMT. GMT is essentially equivalent to Zero UTC within 1 sec. Hence, in the context of this paper we used GMT+1 and UTC+1 interchangeably. Until 1975, the legal time corresponds to UTC+1. In 1976, daylight saving time was adopted again in Metropolitan France with UTC+2 during summertime (Last Sunday of March to last Sunday of October) and UTC+1 otherwise, which continues to this date.

In theory, to convert a time record from one time zone to another time zone in GMT or UTC is trivial, simply by accounting for the hour difference of the time zone in question. However, the conversion gets complicated due to clock shifts during summer and wintertime. For example, it was found that the charts kept recording at the time system (summer or wintertime) of the paper chart installation. The clock was adjusted to the new shifted time when the chart paper was changed. Thus, the metadata associated with these changes were used to properly apply the time difference between legal time and UTC.

3.3 Vertical datum continuity

Since the installation of the tide gauge at Socoa in 1875, the water level has been recorded relative to the ‘zéro hydrographique’ (ZH), that is the French nautical chart datum. ZH has been in use since the nineteenth century by the French hydrographers (Wöppelmann et al. 2014). A local set of tide gauge benchmarks are usually established around the tide gauge and interconnected by means of levelling to transfer the ZH to each one. Additionally, the practice in France is to include a tide pole and set its zero-measurement mark to the ZH (Wöppelmann et al. 2006b), but this procedure was not adopted for Socoa tide poles. We have found two records of tide poles over the full observation period. For each tide poles, the zero-measurement mark of the poles (ZP) are referenced to~~at~~ different heights from the ZH. Thanks to the rescued documents on the levelling measurements during past hydrographic surveys, it was possible to reconstruct the relationship between the ZH and ZP to the current benchmarks, and subsequently to assess the continuity of the ZH at Socoa (Figure 3). The current primary benchmark of the Socoa tide gauge is identified as O.a.K3L3-4-II, which is also part of the national levelling network under the mapping agency (Institut National de l'Information Géographique et Forestière, IGN) responsibility (SHOM 2020).

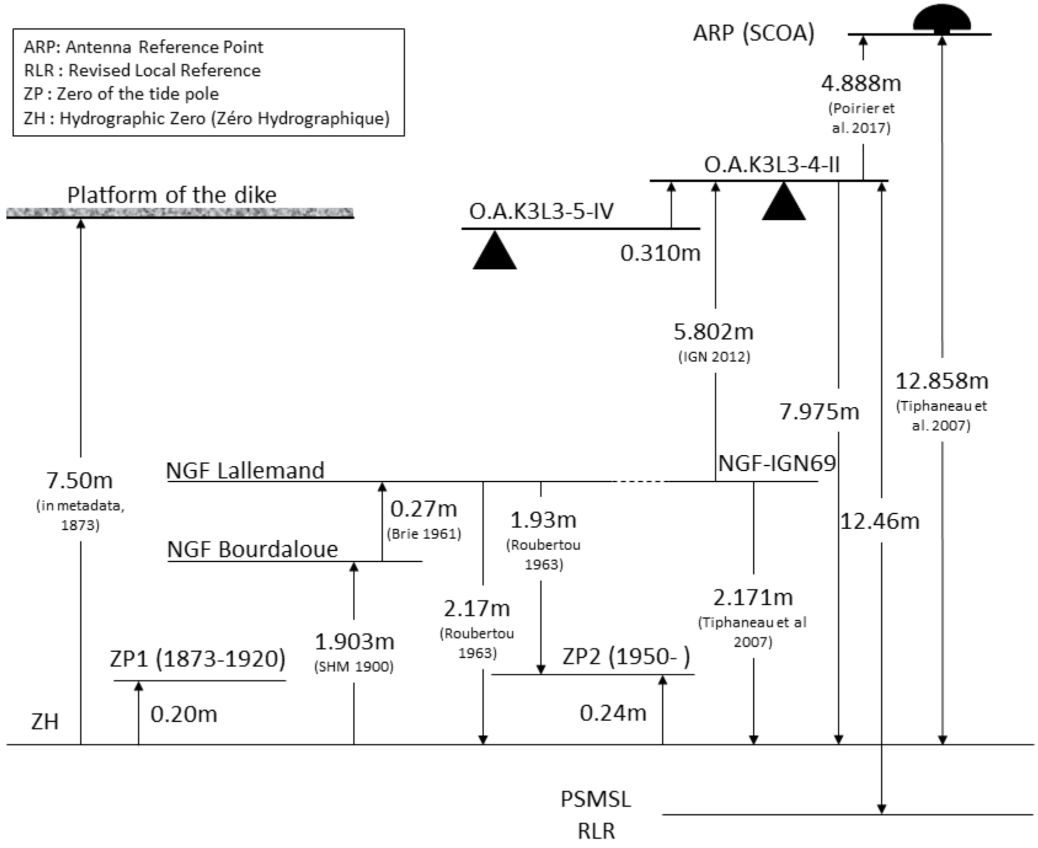

Figure 3: Vertical datum definitions and relationships between benchmarks at Socoa tide gauge. O.A.K3L3-4-II is the primary benchmark, O.A.K3L3-5-IV is the benchmark shown in Figure 1e. The references to the measurements are given inside parentheses (in small fonts).

The first levelling related to the tide gauge was performed in 1873, which established the ZH to be 20cm below ZP, and 7.50m below the dike level. This information was reported in regional department archive AD64- Béarn (Document id: AD64-4S 33) and SHD Vincennes (Document id: DD2-2053). From another published document (Annuaire de marées de 1900, Archive Shom), ZH level was reported to be -1.903m relative to the first national levelling and the associated datum of France established by Bourdaloue (NGF-Bourdaloue) in 1857-1864. However, it is not clear when this datum connection was made. NGF-Bourdaloue has a difference of 27 cm at Socoa to the second national levelling datum later established by Charles Lallemand during 1880-1922 (NGF-Lallemand), locating the hydrographic zero at -2.17m relative to the NGF-Lallemand datum (Brie 1961). No other report of levelling surveys was found during the Chazallon tide gauge period.

In a hydrographic survey done in 1961, the ZH was estimated as being 18 cm above the originally established ZH (Brie 1961). A follow-up investigation in 1963 reveals that the tide gauge was suffering heavy siltation and blockage of the connection ~~with~~

to the sea during the survey of 1961, causing the deviation (Roubertou 1963). Following the investigation in 1963, the ZH was maintained at -2.17m NGF Lallemand, and the ZP was measured to be 24cm above the ZH.

All available documents suggest there was no change in the definition of ZH at Socoa. One false alarm was a letter, dated 9 October 1968 addressed to Shom, where it was mentioned that “the zero of the tide pole” (le zéro de l'échelle~~zero de l’echelle~~) was located -2.178m relative to NGF Lallemand datum, and the primary benchmark was~~is~~ located 5.822m above NGF Lallemand datum. This was identified as a mistake based on the survey done in 2007, which measured the height of OaK3L3-4-II to be 5.805m IGN69 (Tiphaneau et al. 2007). IGN69 refers to ~~NGF-IGN69 is t~~the current national levelling datum (NGF-IGN69), which was ~~-~~established by IGN during 1962-1969. The reported difference between the datum of NGF Lallemand and NGF-IGN69 at Socoa is 0 m (Grid 1245, https://geodesie.ign.fr/contenu/fichiers/grillesorthonormales/GrilleOrthoNormale_Ouest.pdf, last accessed 19-07-2020). Currently, the hydrographic zero (ZH) is reported to be 7.975m below the OaK3L3-4-II benchmark, and 2.171m below NGF-IGN69 datum (Poirier et al. 2017). One of the nearby secondary benchmarks, OaK3L3-5-IV (Figure 1e), sits 0.310m below the OaK3L3-4-II (Figure 3).

4 Data quality assessment

In the previous section, we discussed the method used to reduce the records to a common time system and vertical datum (ZH). These two steps resulted in a merged time series, which was subsequently assessed to detect any potentially erroneous or suspicious water levels (IOC, 2020). Several methods, described in the following subsections, were used to identify potential problems in the data. Based on the rescued metadata, a correction was applied wherever possible, and the corresponding data were~~as~~ flagged.

4.1 Data quality flag

The flag value is defined as a 4-bit number where 1 means the flag is on~~correction is applied~~ and 0 means ~~no correction is applied~~off. Each bit from left to right corresponds to the following:

- Bit 1 – time correction is applied
- Bit 2 – height correction is applied
- Bit 3 – low confidence in the correction in time or height
- Bit 4 – documented siltation period

For example, the 4-bit flag 1010 reads as follows: A time correction is applied (first bit is 1 = True), without height correction (second bit is 0= False), but the data is suspected to be bad (third bit is 1 = True) even if no siltation was reported (fourth bit is 0 = False). This concept is similar to the flag accompanying the PSMSL data (https://www.psmsl.org/data/obtaining/psmsl.hel).

Two files are provided as supplementary material, listing the corrections done on the raw data, for the ledgers (corrections_registry.csv) and the charts (corrections_marigram.csv), respectively. These are concatenated into one file for analysis into a file named ‘correction.csv’ and henceforth identified as ‘Correction file’ (see Section 6 Data availability ~~section~~). In the following sections, different quality control steps are discussed.

4.2 Quality control and corrections

Several basic quality control methods based on visual inspection were~~as~~ applied during the time series construction process. For the ledger data, the digitized tabulated values in spreadsheets were colour-coded, with the colour range between maximum and minimum value to enable visual identification of errors (See Supplementary Figure S5). One of the common errors that this procedure highlights is the wrong transcription of the height by 1m (sometimes 2m). These corrections are flagged as height correction (second bit is 1 in the flag).

Once the obvious height corrections were applied, a tidal harmonic analysis based on validated data was performed, and the recorded water levels were compared with the predicted water levels visually week-by-week (Pugh and Woodworth 2014). This comparison process was useful to identify days with a wrong date (switched with the previous or the following curve in the chart) during transcription, as well as incorrect high and low tides with respect to the tide gauge journal (Section 2.5). The tide gauge journal was checked, and corrections were made if necessary. The high and low tide corrections were typically between 10 and 20 cm.

For the digitized charts, the check sheets were consulted before the time series extraction using NUNIEAU. Anomalies in time were noted for some charts, where the last time of measurement on the chart was different from the one indicated in the check sheet. This type of anomaly is likely due to the faulty placement of the chart on the rotating drum, causing a time difference for the entire measurement period covered by the chart (typically of~~in~~ the order of 5 minutes). The time information of the check sheets was used to apply a time correction. Whenever a constant difference between the time in the check sheet and the tide gauge was noted, a time shift was applied to the final dataset. When~~re~~ the time-shift is different at the beginning and the end, the minimum value of the time-shift was applied to the final data. In some cases, the hourly grid-scale in the charts was relabelled by the observer. The changes induced by grid relabelling were applied directly in the parameterization of NUNIEAU, rather than applying them later.

The above quality controls resulted in a corrected data set with variable time steps depending on the source (ledgers, charts), which was further decimated to hourly values using a linear interpolation. During interpolation, the missing values were computed only if the interpolated timestamp was surrounded by valid data points. This interpolated hourly dataset is the main outcome of this data archaeology exercise and used in the subsequent analysis (See also Section 6 Data availability).

4.3 Unresolved data quality issues

Some additional issues were noticed during the processing of the charts that could not be corrected, which include occasional slowing down of the clocks, mismatch between the height measured by the chart and the tide-pole reading (reported in the check sheets) and delayed rising or falling tidal curve (Supplementary Figure S6). Furthermore, siltation was noted to be a major problem at this tide gauge station. These issues are addressed in here~~the next section~~.

4.3.1 Slowing down of the clock

Thanks to the check sheets, the consistency of the clock at the beginning and the end of a recording period of a chart were able to be cross-checked (Supplementary Figure S6). In some cases, the start time of the clock was found to be correct with a slow-down at the end. The magnitude of the difference varies from 1 to 10 minutes. Only a small portion of the data (less than 2%) are affected by this problem. Given the length of the record in each chart (typically 8-10 days), it is difficult to apply a correction confidently. These values are flagged as values with low confidence (third bit in the flag set to 1).

4.3.2 Possible malfunctioning of the float device

In some instances the tidal curves display a quasi-~~-~~linear rise/fall, instead of the characteristic sinusoidal-like evolution (Supplementary Figure S8). This suggests a malfunction of the mechanical system. In all these cases, the tide gauge regains its normal behaviour in the next tidal cycle. Less than 1% of the recovered data is concerned with this problem. Another issue is linked to siltation (next subsection) which impacted the movement of the float. It concerns about 8% of the total recovered hourly data. The values impacted by this issue are also flagged as values with low confidence (third bit in the flag set to 1).

4.3.3 Siltation

One of the main known issues for the Socoa tide gauge is the siltation of the stilling well (Roubertou 1963, Poirier et al. 2017). The geometry of the stilling well is shown in Figure 4a. The stilling well (Figure 4b) is connected through a pipe of 8cm diameter. The first major siltation problem with the data recording was noticed within the first few years of operation. Notably, the accumulation of silt inside the stilling well restricted the movement of the float (previous subsection) and impacted the recording of the low water levels (less than 70cm). Significant maintenance work was undertaken during 1883-1884 to improve the connectivity of the stilling well to the ocean by creating a duct (Figure 4a,d). The entrance shown in Figure 4c,d was apparently open and accessible through a wooden door. At some (unknown) point, the entrance was partially closed, and the connectivity with the stilling well was severed. After restarting the operation of the tide gauge in 1950, the stilling well exhibited again the siltation and ~~and~~ blockage related problems (Robertou 1963).

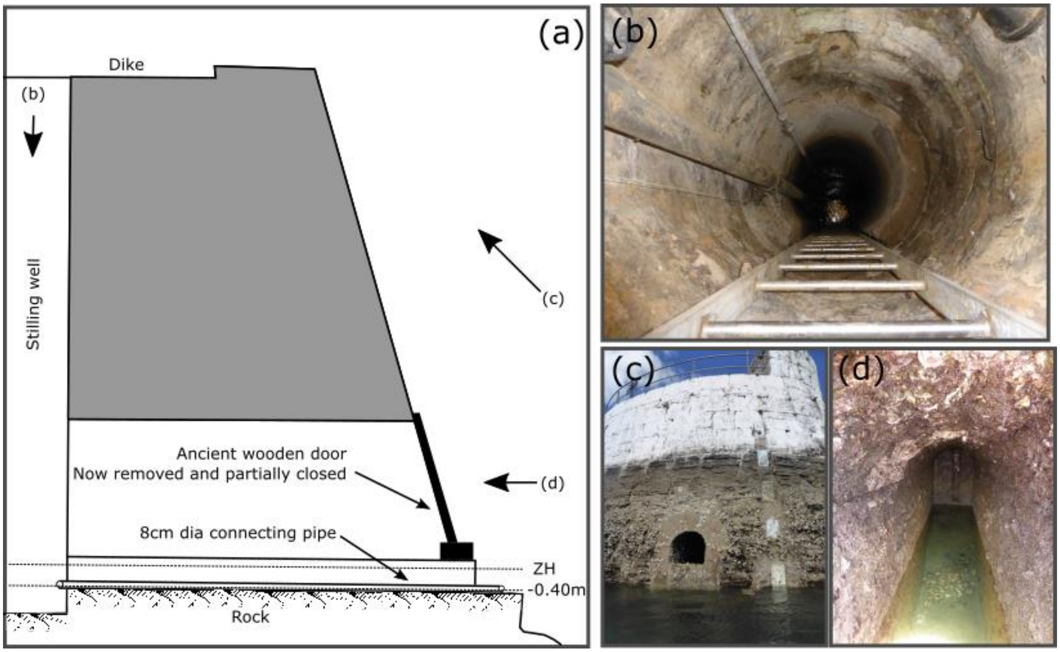

Figure 4: (a) Schematic of the current stilling well. (b) View from above inside the stilling well. (c) Entrance to the stilling well (at about 1m above water level). (d) From the entrance, inside the passage to the stilling well. Images collected during the fieldwork in 2017 (Poirier et al. 2017).

As noted in the literature (e.g., Pugh and Woodworth 2014), malfunctioning of the instrument can be detected by examining the tidal constituents. More specifically, blockage related to siltation ~~inside the stilling well~~ can be detected as a simultaneous amplitude attenuation and phase delay (e.g., Wöppelmann et al. 2014). Here we computed the changes in M2 tide,~~Figure 5 shows the M2 amplitude and phase estimates~~ from a running tidal harmonic analysis with yearly segments from 1875 to 2020 using Utide (python) version 0.2.6 (Codiga 2011, https://github.com/wesleybowman/UTide). Figure 5 shows the M2 amplitude and Greenwich phase lag with ~~T~~the error bars ~~in the figure~~ representing the 95% confidence interval. The most apparent impact of siltation on tide can be observed during 1956-1963 in Figure 5, which is supported by the report of hydrographic surveys carried out at that time (Robertou 1963). Another simultaneous amplitude attenuation and phase delay can also be spotted around the end of the 1990s (1997-2000).

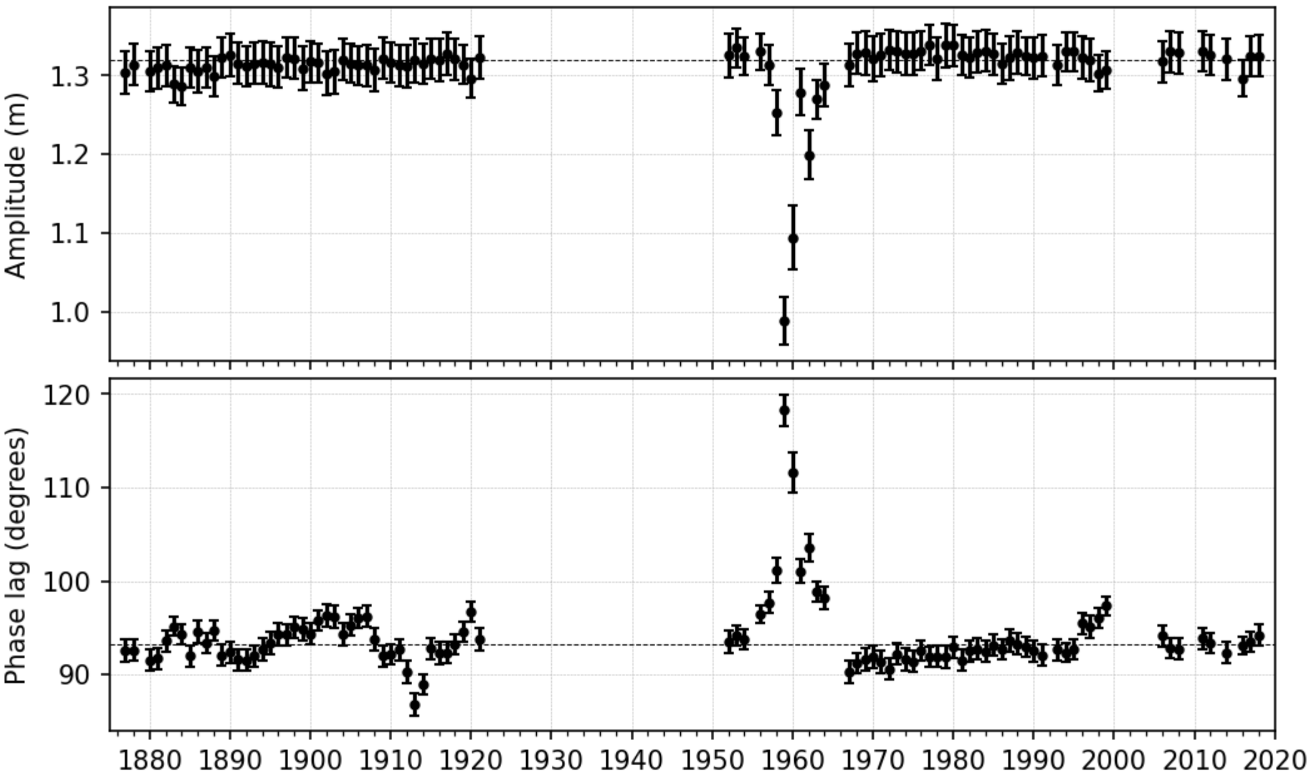

Figure 5: M2 amplitude (top) and Greenwich phase lag (bottom) calculated for one-year segments using tidal harmonic analysis. The dotted lines correspond to the amplitude and phase lag computed for the whole time series. Error bars are 95% confidence intervals.

We have flagged (fourth bit set to 1) the data that are deemed to be impacted by the siltation problem. ~~As noted in the literature (e.g., Pugh and Woodworth 2014), malfunctioning of the instrument can be detected by examining the tidal constituents. More specifically, siltation inside the stilling well can be detected as a simultaneous amplitude attenuation and phase delay (e.g., Wöppelmann et al. 2014). The most apparent siltation problem can be observed during 1956-1963 in Figure 5, which is supported by the report of hydrographic surveys carried out at that time (Robertou 1963). Another simultaneous amplitude attenuation and phase delay can also be spotted around the end of the 1990s (1997-2000).~~

~~Based on the metadata, we apply a siltation flag (fourth bit set to 1) to the data from 1875-11-12 to 1883-08-31.~~ Based on the above-mentioned harmonic analysis, ~~we also flag~~ the data from 1955 to 1963, and 1998-1999 period are flagged. We also added the siltation flag to the data from 1875-11-12 to 1883-08-31 based on the metadata. In total, ~~These amounts~~ ~~to~~ about 29% ~~percent~~ of the total recovered hourly data are flagged as impacted by siltation.

The siltation problem discussed above persists to this date. Currently, the stilling well is cleaned, typically yearly, to maintain an acceptable quality of the data. However, access to the stilling well is challenging, and the cleaning operation is costly. The maintenance is also often perturbed by administrative complications and unforeseen events (e.g., Covid-19 lockdown in 2020-

2021). The stilling well, under current conditions, also does not conform to the recommended 2m depth of water at Lowest Astronomical Tide (IOC 2016). For the Socoa tide gauge, which is currently equipped with a guided wave radar, we recommend a transition from the installation on the stilling well to an installation mounted on the quay of the dike with an unguided open-air radar tide gauge.

4.4 Assessment of the vertical datum continuity

One of the commonly used quality control techniques for sea level is the so-called ‘buddy checking’, which relies on the comparison with mean sea level time series from nearby sites (Pugh and Woodworth, 2014). The difference in monthly mean sea level with nearby tide gauge essentially removes the common part of spatially coherent modes of variability and can reveal malfunctioning at one of the gauges – for instance, step-like features associated with vertical datum discontinuity (Woodworth 2003, Hogarth et al. 2020). Here, we compare our record with the sea level record from the Brest (obtained from https://data.shom.fr), and Santander (obtained from PSMSL, https://www.psmsl.org/) tide gauges. The Brest tide gauge data is one of the well-validated long time series (starting from 17th century) in this region (Wöppelmann et al. 2006a, 2008) covering the whole time series of Socoa. For Santander, Marcos et al. (2021) extended the Santander time series through data archaeology back to 1875 (see Supplementary Figure S9). However, there are multiple apparent datum shifts in the data during the early years (1875-1910), thus we refrained from discussing the extended time series in the buddy checking for Socoa.

We adopted the PSMSL processing scheme for computing the monthly means for Socoa and Brest. First, a Demerliac filter is applied on the hourly data to obtain a detided hourly time series for Socoa and Brest. From the hourly detided water level, the daily mean sea level was obtained using daily averages. A monthly mean is computed only if 50% or more data is available. As the Santander dataset is directly obtained from PSMSL, no further pre-processing was necessary.

The differences of monthly mean sea levels at Socoa with Brest and Santander are shown in Figure 6. For comparison, the mean (computed over 1965-2000) was removed from each dataset before computing the difference. Note that the periods with suspected siltation issues (Section 4.2.3) were removed from the Socoa time series for this analysis.

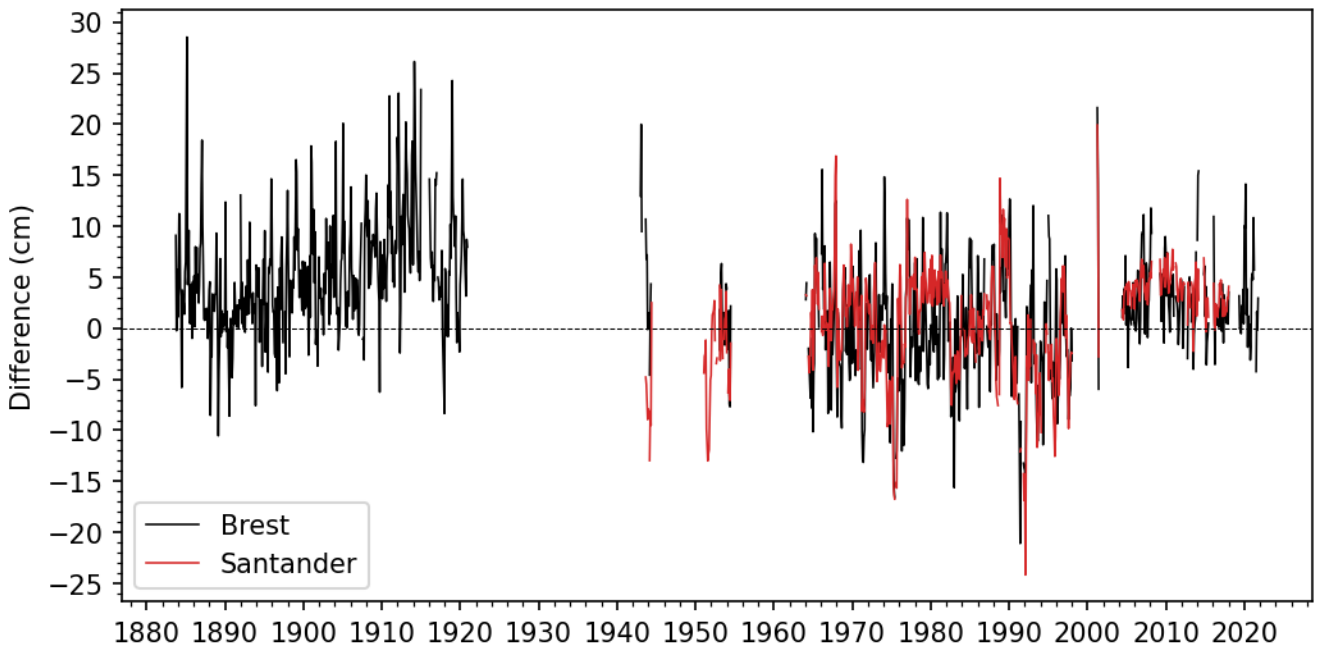

Figure 6: Difference in monthly MSL at Brest (black) and Santander (red) relative to Socoa. The mean over 1965-2000 is removed from each station for comparison.

From Figure 6, no persistent step-like feature is seen in the Brest minus Socoa time series (black), which further strengthen our confidence on the vertical datum continuity established in Section 3.3. In Figure 6, it is interesting to note the gradually increasing difference during the early 20th century. In the literature, this decadal feature is shown to be linked to a large-scale sea-level variability coherent with atmospheric modes of the North Atlantic (Woodworth et al. 2010c, Sturges and Douglas 2011, Calafat et al. 2012, Chafik et al. 2019), and explained by the steric response (Calafat et al. 2012). Both Brest and Socoa tide gauges show this decadal variability (see Supplementary Figure S9), with lower amplitude at Socoa compared to Brest,~~.~~ h~~H~~ence, producing the increasing positive difference from 1900 to 1915 in Figure 6.

The Santander minus Socoa time series also does not indicate any datum shift~~,~~ and is generally consistent with the Brest minus Socoa time series. However, we see a small consistent deviation of 5cm on average during 1976-1980.

5. Trend analysis

From the hourly time series for Brest and Socoa, we have computed yearly means using the yearly PSMSL rules (at least 11 monthly means for a year) and estimated the trends~~ estimate~~ and associated 1-sigma uncertainty (Table 2).

Over the same period (1900-2018), for Brest the trend is 1.50±0.09 mm/year, for Socoa 2.12±0.11 mm/year. The benefit of a long time series is clear here – the longer the time series, the smaller the uncertainty. To compare with the previously published result by Marcos et al. (2021), we also computed the trend for the non-detided time series shown in Supplementary Figure S9, listed in Table 2 under the corresponding~~common~~ period with two~~an~~ asterisks (**). Between Socoa and Santander, the trend estimate is very similar~~close~~ for the common period, 2.08±0.20 mm/year for Socoa, and 2.01±0.12 mm/year for Santander.

Table 2. Estimated linear trends in mm/year at Brest and Socoa over various time periods computed from yearly mean time series.

Period	Brest	Socoa	Santander
Available*	1.30±0.06	1.96±0.08	-
Common (1900-2018)	1.50±0.09	2.12±0.11	-
	1.49±0.09**	2.08±0.11**	2.01±0.12**
Chazallon era (1876-1920)	1.00±0.48	0.82±0.37	-
Brillie era (1963-1997)	1.78±0.52	1.95±0.61	1.44±0.70

* Available period for Brest is 1846-2021, for Socoa is 1875-2021

**Computed from annual mean sea level without using a tide-killer filter (Demerliac)

In the last two rows of Table 2, the estimated trends computed over two periods separated by 40 years - Chazallon era (1876-1920) and Brillie era (1963-1997) – is shown. The sea level trend at Socoa during the Brillie era (1.95±0.61 mm/year) is noticeably increased (i.e., acceleration) compared to Chazallon era (0.82±0.37 mm/year). A s~~S~~imilar magnitude of trend is found at Brest too. During the Chazallon era, the trend at Brest is higher compared to Socoa, which is opposite during the Brillie era. Analysis of the factors that contribute~~s~~ to this observed change in trend is out of the scope of this data paper. However, this leads us to another benefit of a long time series, which allows investigating the non-linear evolution of mean sea levels and associated trends. This benefit is illustrated below through the analysis of inflexion points in the trend at Socoa and Brest.

The analysis is motivated by Wöppelmann et al. (2006a) who noted an inflexion point around 1890 at Brest, which is close to the inflexion point estimated for Liverpool around 1880 by Woodworth (1999). To find the inflexion point of the trend, we have analysed the same yearly time series at Socoa and Brest as above. A linear trend analysis is applied over a running window of 20 years. Windows containing two or more consecutive missing values were removed from the analysis. The trend (mm/year) and 1-sigma uncertainty range are shown in Figure 7. We reproduce an inflexion point at around 1887 in Brest, and an estimated inflexion point between 1895-1900 in Socoa.

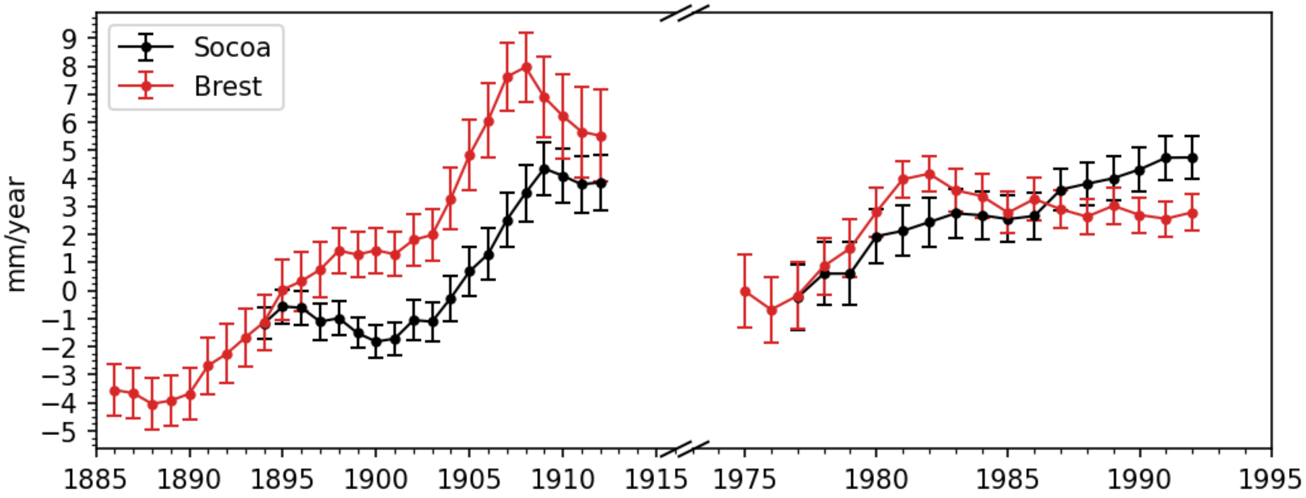

Figure 7: Running trend estimates (20-year windows) for Socoa (black) and Brest (red) during 1875 – 2000. The error bars show the 1-sigma uncertainty range of the trend estimate. Break lines indicates the skipped period when not enough continuous data is available for the analysis.

Multiple drivers may contribute to this inflexion point – decadal variability, long-term climate variability, and climate change induced sea level acceleration. The decadal sea level variability during early 20th century, which is potentially the main contributing factor for inflexion point in trend, is found to be linked with the atmospheric modes of the North Atlantic (Calafat et al. 2012). Jevrejeva et al. (2008) shows that there is a prominent 60-year climatic variation in the trend (acceleration-deceleration), and during the analysis period (1875-1929), the global pattern is a deceleration until around 1910. However, the North-east Atlantic shows a strong deviation from the global pattern with an earlier reversal to acceleration around 1900 (Jevrejeva et al. 2008, Figure 3). Finally, the exact timing of the start of global acceleration of trend due to sea level rise is not accurately answered yet, but studies point towards sometime in the early 19th century (Church et al. 2006, Jevrejeva et al. 2008). Hence, the global trend in sea level rise may have further contributed to the timing of the inflexion point.

6 Data availability

The raw digitized water level, the processed dataset, metadata, and the python notebooks used for processing are available openly at https://doi.org/10.5281/zenodo.7438469 (Khan et al. 2022). The data repository is organized into 4 sub-directories –

1. `data/` : contains the raw and processed dataset for Socoa (`data/socoa`), and other auxiliary dataset (`data/auxiliary`) used in the analysis.
2. `documents/` : contains inventory of the ledgers and charts (`documents/inventory.xlsx`), transcripts of metadata extracted from the regional archives (`documents/archive_records`), and selected transcripts from the tide gauge journal during Chazallon era (`documents/tidegauge_journal`).

3. `figures/` : contains the generated figures used in the manuscript.
4. `notebooks/` : contains the Python notebooks used to process and analyse the data.

The final hourly time series of water level in meters, vertically referenced to local hydrographic zero (ZH), with data quality flags discussed in this paper is distributed as a comma-separated file `data/socoa/socoa_L4.csv` with metadata in the header. The dataset starts at~~from~~ 1st November, 1875 and stops at 4th October, 2021 containing 101 years with data.

In `/documents`, the transcriptions of the relevant part of the tide gauge journals have been provided with the data as yearly documents. The excerpts of metadata documents from the archives of Service Historique de la Défense (SHD) at Brest (SHD-Brest), Rochefort (SHD-Rochefort), Vincennes (SHD-Vincennes), Archives des Pyrénées-Atlantiques - Béarn (AD64- Béarn) and Archives des Pyrénées-Atlantiques - Bayonne (AD64-Bayonne), and SHOM archive are also provided as supplementary files to the dataset.

This dataset is reproducible by applying time and height corrections on the raw uncorrected water level records for Socoa (`data/socoa/socoa_raw.txt`), using the data processing script (`notebooks/01_data_processing.ipynb`) and corrections (`data/socoa/corrections.csv`). Further detail of the files available in the data repository can be found in the include `README.md` text file.

A continuously updated time series of the Socoa sea level can be obtained from the Shom portal (http://dx.doi.org/10.17183/REFMAR#95).

7 Conclusions

We have done a thorough archival research, data-rescue, digitization, and metadata analysis, and increased the coverage of the existing hourly sea level record at Socoa, Saint-Jean-de-Luz (France) back to 1875. Among the total 702 station-months additional data, 693 station-months are with more than 50% data per month. This extension of data amounts to about 58 years' worth of new data.

Data quality flags are assigned to the recovered and distributed final hourly dataset from careful inspection of metadata and dedicated analysis of the dataset. The amount of data when~~re~~ time and height corrections were needed and applied with confidence are small (less than 2% in total). Additional flags were assigned to indicate if time or height corrections were applicable but could not be applied confidently due to insufficient information. A very small proportion of data (<1%) is affected by this issue, without considering siltation related problems.

The largest proportions of the flagged data were related to siltation in the stilling well. A dedicated analysis of the data and metadata was done to identify and document periods with siltation. We have identified three main periods, from 1875 to 1893, 1951-1963, and 1998-1999. A dedicated flag was assigned to these periods, which affects 29% of the recovered hourly data. Considering the gravity and the recurrent nature of the siltation problem in the stilling well, we recommend a transition from the stilling well to an open-air installation for this tide gauge. This transition should be supplemented with a study of the filtering characteristics of the stilling well to track any impact of the installation change on future sea level measurements.

The~~is~~ extended dataset will be communicated and deposited in international sea level databanks (e.g., PSMSL) to further increase the number of long-term sea level records extending back into the 19th century. One of the major features of this sea level record is its location, which has remained the same (buildings and stilling well) since its installation in 1875. The data recovered and rescued in this work would be useful for long-term sea level trend analysis (e.g., Gehrels and Woodworth 2013) and modelling at decadal to interannual scale (e.g., Calafat et al. 2012, Ding et al. 2021). Investigation into the extreme events will specially benefit from the high temporal sampling of the extended time series.

Besides the final hourly sea level dataset from 1875 to 2021, we also provide the raw data, associated corrections that are synthesized above, and computational ~~environment, and~~ notebooks as companion datasets (See Section 6). The objective of this is to promote reproducible research and to increase transparency by allowing validation of our computations.

In this data paper, we have not only extended the sea level time series at Socoa, but also showed that analysing the history of individual tide gauges can reveal important location-specific issues, like siltation, that might not be directly evident from the global dataset at this moment. During the current data archaeology work, we have also found unrecoverable deterioration of historical paper documents, which underlines the urgency of rescuing these invaluable records. Relevant metadata are also in the same danger of being deteriorated beyond rescue.

Finally, there is a vast amount of untapped tide data and associated metadata worldwide (Pouvreau 2008, Bradshaw et al. 2015, Talke and Jay 2017). Talke and Jay (2017) reported identification of more than 6,500 station-years of previously lost or forgotten tide data over United States. Over 60,000 identified documents have been inventoried in France (See Shom inventory, http://refmar.shom.fr/dataRescue), with 70% already rescued (e.g., scanned), but many remain~~s~~ undigitized (Latapy et al., 2022). Given the time critical risk of losing these valuable scientific and historic information, it is crucial to urgently rescue these datasets, digitize them, and make them available for the scientific community.

Author contribution

LT, GW, and AL conceived the idea of the data archaeology for Socoa (Saint Jean-de-Luz) and secured the funding. IB did the data cataloguing, rescue, and digitization under the supervision of AL and NP. JK analysed the data, developed the computational notebooks, and associated figures, curated the data for publishing, and wrote the first draft of the manuscript. GW and LT produced the second draft of the manuscript. All co-authors contributed to editing of the final manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors wish to thank Mr. Alain Roudil and Mr. Christian Ondicola for providing valuable information regarding the history of the tide gauge. We acknowledge Dr. Marta Marcos, University of the Balearic Islands, for providing the monthly mean sea level data for Santander station. We also acknowledge PSMSL for their long-standing effort to maintain a global sea-level databank. The work presented here was carried out within the framework of the FEDER-FSE "Aquitaine" 2014-654, 2020 project EZPONDA (grant No 2018-4619910). Last, but not least, we would like to thank the reviewers for their comments and suggestions that significantly improved the manuscript.

References

Aarup, T., Wöppelmann, G., Woodworth, P.L., Hernandez, F., Vanhoorne, B., Schöne, T., and Thompson, P.R.: Comments on the article "Uncertainty and bias in electronic tide-gauge records: evidence from collocated sensors" by S. Pytharouli, S. Chaikalis, S.C. Stiros in Measurement (Vol. 125, September 2018), 2019.

Arnoux, F., Abadie, S., Bertin, X. and Kojadinovic, I.: Coastal flooding event definition based on damages: Case study of Biarritz Grande Plage on the French Basque coast. Coastal Engineering, 166, p.103873, doi: 10.1016/j.coastaleng.2021.103873, 2021.

Arns, A., Wahl, T., Wolff, C., Vafeidis, A.T., Haigh, I.D., Woodworth, P., Niehüser, S. and Jensen, J.: Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts. Nature communications, 11(1), pp.1-9, doi: 10.1038/s41467-020-15752-5, 2020.

Bradshaw, E., Lesley, R., and Thorkild, A. : Sea Level Data Archaeology and the Global Sea Level Observing System (GLOSS), Geo. Res. J., 6, 9-16, doi :10.1016/j.grj.2015.02.005, 2015.

Brie: Report no. 158. Mission hydrographique de France et d'Algerie (MHCFA), Cherbourg, 1961.

Bureau des longitudes: Guide de données astronomiques pour l'observation du ciel: Annuaire du Bureau des longitudes, IMCEE and Bureau des longitudes, pp. 56-58, ISBN: 9782759805419, 2011. (Retrieved from https://gallica.bnf.fr/ark:/12148/bpt6k9614055r)

Calafat, F. M., Chambers, D. P., and Tsimplis M. N.: Mechanisms of Decadal Sea Level Variability in the Eastern North Atlantic and the Mediterranean Sea, Journal of Geophysical Research: Oceans, 9, n/a–n/a. doi:10.1029/2012jc008285, 2012.

Chafik, L., Nilsen, J.E.Ø., Dangendorf, S., Reverdin, G. and Frederikse, T.:North Atlantic Ocean circulation and decadal sea level change during the altimetry era. Scientific reports, 9(1), pp.1-9, doi: 10.1038/s41598-018-37603-6, 2019.

Church, J.A. and White, N.J.: A 20th century acceleration in global sea-level rise. Geophysical research letters, 33(1), doi: 10.1029/2005GL024826, 2006.

Church, J.A., P.U. Clark, A. Cazenave, J.M. Gregory, S. Jevrejeva, A. Levermann, M.A. Merrifield, G.A. Milne, R.S. Nerem, P.D. Nunn, A.J. Payne, W.T. Pfeffer, D. Stammer and A.S. Unnikrishnan: Sea Level Change. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on

Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Codiga, D.L.: Unified Tidal Analysis and Prediction Using the UTide Matlab Functions. Technical Report 2011-01. Graduate School of Oceanography, University of Rhode Island, Narragansett, RI. 59pp. url: ftp://www.po.gso.uri.edu/pub/downloads/codiga/pubs/2011Codiga-UTide-Report.pdf, 2011.

Coles S.G.: An introduction to statistical modelling of extreme values. Springer-Verlag, New York, 2001.

Dangendorf, S., Marcos, M., Wöppelmann, G., Conrad, C.P., Frederikse, T. and Riva, R.: Reassessment of 20th century global mean sea level rise. Proceedings of the National Academy of Sciences, 114(23), pp.5946-5951, doi: 10.1073/pnas.1616007114, 2017.

Dodet, G., Bertin, X., Bouchette, F., Gravelle, M., Testut, L. and Wöppelmann, G.: Characterization of sea-level variations along the Metropolitan Coasts of France: waves, tides, storm surges and long-term changes. Journal of Coastal Research, 88(SI), pp.10-24, doi: 10.2112/SI88-003.1, 2019.

Ekman, M.: Climate Changes Detected Through the Worlds Longest Sea Level Series, Global and Planetary Change, 9, 215–224, doi:10.1016/s0921-8181(99)00045-4, 1999.

Fox-Kemper, B., H.T. Hewitt, C. Xiao, G. Aðalgeirsdóttir, S.S. Drijfhout, T.L. Edwards, N.R. Golledge, M. Hemer, R.E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I.S. Nurhati, L. Ruiz, J.-B. Sallée, A.B.A. Slangen, and Y. Yu: Ocean, Cryosphere and Sea Level Change. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1211–1362, doi: 10.1017/9781009157896.011, 2021.

Gouriou, T., Míguez, B.M. and Wöppelmann, G.: Reconstruction of a two-century long sea level record for the Pertuis d’Antioche (France), Continental Shelf Research, 61, pp.31-40. doi:10.1016/j.csr.2013.04.028, 2013.

Gehrels, W.R. and Woodworth, P.L.: When did modern rates of sea-level rise start?, Global and Planetary Change, 100, pp.263-277, doi:10.1016/j.gloplacha.2012.10.020 2013.

Haigh, I.D., Pickering, M.D., Green, J.A.M., Arbic, B.K., Arns, A., Dangendorf, S., Hill, D., Horsburgh, K., Howard, T., Idier, D., Jay, D.A., Lee, S.B., Müller, M., Schindelegger, M., Talke, S.A., Wilmes, S-B. and Woodworth, P.L.: The tides they are a-changin': A comprehensive review of past and future nonastronomical changes in tides, their driving mechanisms and future implications. Reviews of Geophysics, 57, e2018RG000636, doi:10.1029/2018RG000636, 2020.

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. Geophysical Data Journal, in press, doi:10.1002/gdj3.174, 2022.

Hogarth, P., Hughes, C.W., Williams, S.D.P., and Wilson, C.: Improved and Extended Tide Gauge Records for the British Isles Leading to More Consistent Estimates of Sea Level Rise and Acceleration Since 1958, Progress in Oceanography, 5, doi:10.1016/j.pocean.2020.102333, 2020.

Holgate, S.J., Matthews, A., Woodworth, P.L., Rickards, L.J., Tamisiea, M.E., Bradshaw, E., Foden, P.R., Gordon, K.M., Jevrejeva, S. and Pugh, J.: New data systems and products at the Permanent Service for Mean Sea Level. Journal of Coastal Research, 29, 493-504, doi:10.2112/JCOASTRES-D-12-00175.1, 2013.

Hughes, D. W., Yallop, B. D., and Hohenkerk, C. Y.: The equation of time, Mon. Not. R. Astron. Soc., 238, 1529–1535, https://doi.org/10.1093/mnras/238.4.1529, 1989

Jevrejeva, S., Moore, J.C., Grinsted, A. and Woodworth, P.L.: Recent global sea level acceleration started over 200 years ago?. Geophysical Research Letters, 35(8), doi: 10.1029/2008GL033611, 2008.

Hunter, J., Coleman, R., and Pugh, D.: The Sea Level at Port Arthur, Tasmania, from 1841 to the Present. Geophysical Research Letters, 4, doi:10.1029/2002gl016813, 2003.

IOC: Quality control of in situ sea level observations: a review and progress towards automated quality control, IOC Manuals and Guides, No. 83, UNESCO, Paris, https://unesdoc.unesco.org/ark:/48223/pf0000373566, 2020.

IOC: Manual on sea level measurement and interpretation, Volume I – Basic Procedures, IOC Manuals and Guides, No. 14, UNESCO, Paris, 1985.

IOC: Manual on sea level measurement and interpretation, Volume V – Radar Gauges, IOC Manuals and Guides, No. 14, UNESCO, Paris, 2016.

Khan, M.J.U., Van Den Beld, I., Wöppelmann, G., Testut, L., Latapy, A., & Pouvreau, N.,: Sea level data archaeology at Socoa (Saint Jean-de-Luz, France) ~~(v1.0)~~ [Data set]. Zenodo, doi: 10.5281/zenodo.7438469, 2022.

Latapy, A., Ferret, Y., Testut, L., Talke, S., Aarup, T., Pons, F., Jan, G., Bradshaw, E., Pouvreau, N.,: Data rescue process in the context of sea level reconstructions: An overview of the methodology, lessons learned, up-to-date best practices and recommendations. Geoscience Data Journal, 00, 1– 30, doi:10.1002/gdj3.179, 2022.

Letetrel, C., Marcos, M., Míguez, B. M., and Wöppelmann, G.: Sea Level Extremes in Marseille (NW Mediterranean) During 1885-2008, Continental Shelf Research, 7,1267–1274, doi:10.1016/j.csr.2010.04.003, 2010.

Marcos, M., Puyol, B., Amores, Gómez, B. P., Fraile, M., and Talke, S. A.: Historical Tide Gauge Sea-Level Observations in Alicante and Santander (Spain) Since the Century, Geoscience Data Journal, 1, doi:10.1002/gdj3.112, 2021.

Marcos, M., Puyol, B., Wöppelmann, G., Herrero, C., and García-Fernández, M. J.: The Long Sea Level Record at Cadiz (Southern Spain) from 1880 to 2009, Journal of Geophysical Research, 12, doi:10.1029/2011jc007558, 2011.

Marcos, M., Calafat, F.M., Berihuete, Á. and Dangendorf, S.: Long-term variations in global sea level extremes. Journal of Geophysical Research: Oceans, 120(12), pp.8115-8134, 2015.

Marcos, M. and Woodworth, P.L: Spatio-temporal changes in extreme sea levels along the coasts of the North Atlantic and the Gulf of Mexico. Journal of Geophysical Research Oceans, 122, 7031-7048, doi:10.1002/2017JC013065, 2017.

Martín Míguez, B., Le Roy, R., and Wöppelmann, G.: The use of radar tide gauges to measure variations in sea level along the French coast, Journal of Coastal Research, 24, 61-68, doi: 10.2112/06-0787.1, 2008.

Menéndez, M. and Woodworth, P.L.: Changes in extreme high water levels based on a quasi-global tide-gauge dataset. Journal of Geophysical Research, 115, C10011, doi:10.1029/2009JC005997, 2010.

Muis, S., Verlaan, M., Winsemius, H.C., Aerts, J.C. and Ward, P.J.: A global reanalysis of storm surges and extreme sea levels. Nature communications, 7(1), pp.1-12, doi: 10.1038/ncomms11969, 2016.

Müller, M.: Equation of time-problem in astronomy. ACTA PHYSICA POLONICA SERIES A, 88, pp.S-49.Vancouver, 1995.

Oppenheimer, M., B.C. Glavovic , J. Hinkel, R. van de Wal, A.K. Magnan, A. Abd-Elgawad, R. Cai, M. CifuentesJara, R.M. DeConto, T. Ghosh, J. Hay, F. Isla, B. Marzeion, B. Meyssignac, and Z. Sebesvari: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)], 2019.

Pan, H. and Lv, X.: Is there a quasi 60-year oscillation in global tides?. Continental Shelf Research, 222, doi: 10.1016/j.csr.2021.104433, 2021.

Piccioni, G., Dettmering, D., Bosch, W. and Seitz, F.: TICON: TIdal CONstants based on GESLA sea-level records from globally located tide gauges. Geoscience Data Journal, 6(2), pp.97-104, doi: 10.1002/gdj3.72, 2019.

Pineau-Guillou, L., Lazure, P. and Wöppelmann, G.: Large-scale changes of the semidiurnal tide along North Atlantic coasts from 1846 to 2018. Ocean Science, 17(1), pp.17-34, 2021.

Poirier, E., Gravelle, M., and Wöppelmann, G.: Contrôles du marégraphe de Socoa (Saint Jean-de-Luz) – Missions du 10-12 mai 2017 et du 23-24 août 2017, SONEL Rapport Nr. 001/17, 2017.

Pouvreau, N.: Trois Cents Ans de Mesures Marégraphiques En France: outils, Méthodes Et Tendances Des Composantes Du Niveau de La Mer Au Port de Brest, PhDThesis, Université de La Rochelle, 2008.

Pouvreau, N., Miguez, B. M., Simon, B., and Wöppelmann, G.: Evolution of the semi-diurnal tidal constituent M2 at Brest from 1846 to 2005, Comptes Rendus Geoscience, 11, 802–808, doi:10.1016/j.crte.2006.07.003, 2006.

Poulle, Y.: La France à l'heure allemande. In: Charter School Library, volume 157, issue 2. pp. 493-502, doi:10.3406/bec.1999.450989, 1999.

Pugh, D., and Woodworth, P.: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes, Cambridge University Press, 2014.

Ray, R. D., & Talke, S. A.: Nineteenth-century tides in the Gulf of Maine and implications for secular trends. Journal of Geophysical Research: Oceans, 124, 7046– 7067, doi:10.1029/2019JC015277, 2019.

Roubertou, A.: The Brillié Tide-Gauge, The International Hydrographic Review, Reproduced from the "Bulletin d 'information du Comite Central d 'Oceanographie et d'Etude des Cotes (C.O.E.C.), 7th year, No. 6, Paris, June 1955", 1955.

Roubertou, A.: Rapport no. 272. Mission Hydrographique de Dragage (MHD). Bordeaux, 1963.

Sturges, W. and Douglas, B.C.: Wind effects on estimates of sea level rise. Journal of Geophysical Research: Oceans, 116(C6), doi: 10.1029/2010JC006492, 2011.

SHOM: Références altimétriques maritimes (RAM), Shom, Brest, France, url: https://diffusion.shom.fr/pro/references-altimetriques-maritimes-ram.html, 2020 (last accessed: 25/03/2022).

Tadesse, M., Wahl, T. and Cid, A.: Data-driven modeling of global storm surges. Frontiers in Marine Science, 7, p.260, doi: 10.3389/fmars.2020.00260, 2020.

Talke, S.A., Orton, P. and Jay, D.A.: Increasing storm tides in New York harbor, 1844–2013. Geophysical Research Letters, 41(9), pp.3149-3155, doi: 10.1002/2014GL059574, 2014.

Talke, S.A., and Jay, D.A.: Archival Water-Level Measurements: Recovering Historical Data to Help Design for the Future, Civil and Environmental Engineering Faculty Publications and Presentations, 412. http://archives.pdx.edu/ds/psu/21294, 2017.

Talke, S.A., Kemp, A.C. and Woodruff, J.: Relative sea level, tides, and extreme water levels in Boston Harbor from 1825 to 2018. Journal of Geophysical Research: Oceans, 123(6), pp.3895-3914, doi: 10.1029/2017JC013645, 2018.

Testut, L., Miguez, B. M., Wöppelmann, G., Tiphaneau, P., Pouvreau, N., and Karpytchev., M.: Sea Level at Saint Paul Island, Southern Indian Ocean, from 1874 to the Present, Journal of Geophysical Research, 12, doi:10.1029/2010jc006404, 2010.

Tiphaneau, P., Breilh, J.-F., & Wöppelmann, G.:Contrôle des performances du marégraphe radar BM70A de Socoa (Saint Jean-de-Luz), Report No. 002/07, May, Centre littoral de Geophysique - Université de la Rochelle, La Rochelle, url: https://www.sonel.org/SoTaBord/ged/Tiphaneau-2007-controle_des_performances_du_m.pdf, 2007.

Ullmann, A., Pons, F., Moron, V.: Tool kit helps digitize tide gauge records, EOS Trans. AGU, 86(38), doi: 10.1029/2005EO380004, 2011.

UNESCO/IOC: Workshop on Sea Level Data Archaeology, Paris, France, 10-12 March 2020. Paris, UNESCO, IOC Workshop Reports, 287, 39 pp. English. (IOC/2020/WR/287) https://unesdoc.unesco.org/ark:/48223/pf0000373327, 2020.

Wahl, T., and Chambers, D. P.: Evidence for multidecadal variability in US extreme sea level records, J. Geophys. Res. Oceans, 120, 1527– 1544, doi:10.1002/2014JC010443, 2015.

Woodworth, P.L.: Some comments on the long sea level records from the northern Mediterranean. Journal of Coastal Research, pp.212-217, 2003.

Woodworth, P. L.: High Waters at Liverpool Since 1768: the UKs Longest Sea Level Record, Geophysical Research Letters, 6, 1589–1592, doi:10.1029/1999gl900323, 1999.

Woodworth, P.L., Pugh, D.T. and Bingley, R.M.: Long term and recent changes in sea level in the Falkland Islands. Journal of Geophysical Research, 115, C09025, doi:10.1029/2010JC006113, 2010a.

Woodworth, P.L.: A survey of recent changes in the main components of the ocean tide. Continental Shelf Research, 30, 1680-1691, doi:10.1016/j.csr.2010.07.002, 2010b.

Woodworth, P. L., Pouvreau, N., and Wöppelmann, G.: The gyre-scale circulation of the North Atlantic and sea level at Brest, Ocean Sci., 6, 185–190, doi:10.5194/os-6-185-2010, 2010c.

Woodworth, P.L., Hunter, J.R., Marcos, M., Caldwell, P., Menéndez, M. and Haigh, I.: Towards a global higher-frequency sea level dataset. Geoscience Data Journal, 3(2), pp.50-59, doi: 10.1002/gdj3.42, 2016.

Wöppelmann, G., Marcos, M., Coulomb, A., Míguez, B. M., Bonnetain, P., Boucher, C., Gravelle, M., Simon, B., and Tiphaneau, P.: Rescue of the Historical Sea Level Record of Marseille (France) from 1885 to 1988 and Its Extension Back to 1849-1851, Journal of Geodesy, 6, 869–885, doi:10.1007/s00190-014-0728-6, 2014.

Wöppelmann, G., Pouvreau, N., and Simon, B.: Brest Sea Level Record: a Time Series Construction Back to the Early Eighteenth Century, Ocean Dynamics, 3, 487–497, doi:10.1007/s10236-005-0044-z, 2006a.

Wöppelmann, G., Zerbini, S., Marcos, and Marcos, M.: Tide gauges and Geodesy: a secular synergy illustrated by three present-day case studies, C. R. Geoscience, 338, 980–991, doi:10.1016/j.crte.2006.07.006, 2006b.

Wöppelmann, G., Pouvreau, N., Coulomb, A., Simon, B. and Woodworth, P.L.: Tide gauge datum continuity at Brest since 1711: France's longest sea-level record. Geophysical Research Letters, 35(22), doi: 10.1029/2008GL035783, 2008.