



Recovering, correcting, and reconstructing precipitation data affected by gaps and irregular readings: The Padua series from 1812 to 1864

Antonio della Valle¹ · Dario Camuffo¹ · Francesca Becherini²  · Valeria Zanini³

Received: 8 September 2022 / Accepted: 10 January 2023 / Published online: 21 January 2023
© The Author(s) 2023, corrected publication 2023

Abstract

The aim of this work is to reconstruct the 1812–1864 period of the Padua precipitation series at the daily level, using a local precipitation Log. Missing readings, cumulative amounts, and gaps often affect early precipitation series, as observers did not follow a precise protocol. Therefore, the daily amount and frequency reported in the register of observations are not homogeneous with other periods, neither comparable with other contemporary series, and need a correction. The correction methodology has been based on the daily weather notes written in the Log in parallel to the readings. Taking advantage of periods in which both weather observations and instrumental readings were regularly taken, the terms used to describe the precipitation type and intensity have been classified, analyzed statistically, calibrated, and transformed into numerical values. The weather notes enable the distribution of precipitation to be determined based on the cumulative amounts collected on consecutive rainy days into the likely precipitation that occurred on every single rainy day. In the case of missing readings, the presence of weather notes enables the missing amounts to be estimated using the relationships found previously. Finally, the recovery of additional contemporary documents made it possible to fill some gaps in this period. Using this approach, 52 years of the long Padua precipitation series have been corrected: precipitation collected for two or more rainy days has been distributed according to the actual rainy days; the rain amount fully recovered and most of the missing values reconstructed; the false extreme events corrected.

Keywords Daily precipitation · Missing values · Early instrumental meteorological series · Descriptive observations · Documentary sources · Climate reconstruction

✉ Francesca Becherini
francesca.becherini@cnr.it

¹ National Research Council-Institute of Atmospheric Sciences and Climate, Corso Stati Uniti 4, 35127 Padua, Italy

² National Research Council-Institute of Polar Sciences, Via Torino 155, 30172 Venice Mestre, Italy

³ INAF-Astronomical Observatory of Padua, Vicolo Dell'Osservatorio 5, 35122 Padua, Italy

1 Introduction

Over the years, several statistical and interpolation techniques have been developed to correct, fill, and validate modern precipitation series datasets with varying levels of complexity (Longman et al. 2020). Identifying an appropriate gap-filling method for broad applications is not possible, as it has to be tailored to the specific case under study. A crucial element is the percentage of missing data and data missingness mechanism (i.e., at random or not) (Aieb et al. 2019). Nevertheless, all these approaches require the availability of predictor station(s). By applying machine learning models, Bellido-Jiménez et al. (2021) assess that the use of neighboring data as a rainfall gap-filling technique is more successful rather than the use of data from the target station from the past and future. The success of such methods depends on the extent of the correlation between the target and predictor stations. Results show that the correlation between the stations is a more important requirement than their proximity (Longman et al. 2020), and has to be verified prior the application of any method. The most performing gap-filling method and its most appropriate parameterization cannot be assessed at all, but has to be studied case by case, as it strongly depends on the relationship between the target and predictor stations. For example, Caldera et al. (2016) compare different techniques and find that some methods provide good predictions in case there is only one neighboring station with a high correlation coefficient, while others are preferable in case of relatively low correlation coefficients with the neighboring stations. Finally, very likely filling a dataset to completion requires the use of multiple approaches. This most important requirement of all the mentioned approaches, i.e., the availability of at least one neighboring station with a certain degree of correlation, is rarely met in early datasets: the available observations are few, especially in the eighteenth century, and were performed in places far from each other.

In addition, the measuring protocols, when known, were not homogenous, as the observational procedure was not standardized. Every observer used their own protocol, i.e., observation time, instrument type, location, and exposure. Some international networks of scientists appeared, each with a different protocol, and these protocols in certain cases ceased soon after, while in other cases they survived for years. This is crucial especially for precipitation, where different rain gauges (operating principle, shape, threshold), exposure (orientation, elevation), and environmental conditions (turbulence, influence of wind drag, wetting, evaporation) could alter the collection efficiency (Brugnara et al. 2020; Camuffo 2022b; Camuffo et al. 2020, 2022b). For example, Camuffo et al. (2022a) show that the ratio between monthly rain amounts in Padua and Venice is not constant, but has changed over time, with the observer and/or site.

Missing data, over short or long periods, constitutes a frequent problem that makes the reconstruction difficult (Camuffo et al. 2022a).

Non-regular reading times used by observers may affect precipitation statistics. Usually, precipitation was recorded one or more times per day, but in some cases, the rain gauge was not read at the scheduled time, or irregularly. Sometimes sub-daily readings were missed, or readings were taken at the end of the rainy period, composed of one or more days. This often happened when meteorological readings were taken by astronomers as a support to their astronomical observations. In the case of cloud cover and long-lasting rain, they missed their observations. When observations are biased or taken irregularly, both daily precipitation amount and frequency are distorted and need a careful correction to undertake climate change studies.

Accurate analyses of documentary sources are extremely important in the early instrumental period, as they can supply useful metadata concerning instruments, measurements, and observation biases. Quantitative data can be obtained by creating indices that are calibrated using single or multiple documentary sources, such as annals, chronicles, memoirs, and weather notes. Finally, the reconstructed data are compared to contemporary instrumental data to estimate weather quantities or trends (e. g. Jones et al. 2009; Domínguez-Castro et al. 2015; Adamson 2015; Harvey-Fishenden and Macdonald 2021; Nash et al. 2021; Camuffo 2022a). These documentary sources allow the climatic analysis of the pre-instrumental period. In most cases, the weather notes reported in sources did not contain information on the duration or intensity of the precipitation, but only general indications about the precipitation type (e.g., rain, snow, hail), so that only the frequency can be estimated (e.g., Raicich 2008). In other more fortunate cases, a short, but more detailed description was given (e.g., light rain, few drops, intense rain, and so forth), this enables an estimation of the daily amount to be made, as done by Camuffo et al. (2022a) to fill the 1764–1767 gap of the Padua series. In this work, an advanced methodology to assess daily precipitation has been developed and used for two aims: (i) to fill missing data attributing an estimated amount; (ii) to distribute day-by-day the cumulative amount collected over some consecutive rainy days. The second goal has been reached for the first time.

This paper is the final part of a long study devoted to the recovery and revision of the Padua precipitation series from 1713 to the present. The recovery and correction of the precipitation series at Padua started in the middle of the 1980s, when Camuffo (1984) recovered the period 1725–1981 from the original Log and analyzed monthly precipitation values. Over time, the data recovery at the daily resolution, its correction and analysis was gradually improved, completely revised and extended to 2018 (Camuffo et al. 2020), and the 1764–1767 gap was filled (Camuffo et al. 2022a). In the early 2000s, the 1725–1998 period of the temperature series at Padua was corrected and homogenized (Camuffo 2002). This work is focused on the critical 1812–1864 period: during these years, different observing protocols and irregular readings strongly affected both rain frequency and amount, generating false extremes at the daily level. The careful correction and reconstruction of this period are crucial to complete the Padua series.

The paper is organized as follows: in Section 2, the original Logs are briefly described, as well as the other nearest contemporary precipitation series; in Sections 3 and 4 the biases that affect the 1812–1864 period and the method used to correct them are presented; Section 5 is devoted to the application of the method and the discussion of the results; conclusions follow in the last section.

To make the text easily readable and understandable, a list of the terms and definitions used in this paper is given in Table ESM1.

2 Data and metadata

2.1 The 1812–1864 Padua precipitation series

The history of the three-century Padua precipitation series has been extensively presented elsewhere (Camuffo 1984, 2002; Camuffo et al. 2020), but the essential items will be recalled when necessary. Some further information can be found in the Electronic Supplementary Material (ESM). The figure representing an overview of the 1812–1864 series,

i.e., observers, exposure, catching level, instrument type, and homogeneity has been reported in Fig. ESM1.

In the 1812–1864 period, the meteorological readings were reported in a Log, structured as a table, two pages per month, and three observations per day: pressure, indoor and outdoor temperatures, humidity, wind direction, precipitation amount (Figs. ESM2 and ESM3). In the last two columns, some weather observations were noted (Camuffo et al. 2020). When Lorenzoni started his observations in January 1865, the structure of the Log was partially modified, with four readings a day. Unfortunately, the original Log from May to December 1838 was lost.

In addition, another useful source of indirect information is a second Log used for the astronomical observations (Fig. ESM4). This Log was conceived for the observations with the telescope; therefore, it reports the astronomical coordinates of the objects, description of their appearance, and sometimes a brief description of the sky (e.g., cloud cover, rainy days). The astronomical Log is a precious source of additional notes, taken by the same or a different observer, in the same location. It has been particularly useful to recognize clear or rainy days not reported in the meteorological Log, or when the meteorological Log has gaps.

2.2 Other contemporary precipitation records

This paper is essentially based on the Logs and documents related to the Padua series, even though there are three near contemporary series in northern Italy, i.e., Venice (30 km east of Padua), Bologna (100 km south of Padua), and Milan (230 km west of Padua).

Venice Observations started with Bernardino Zendrini in 1727. The early period is affected by frequent changes in observers, location, exposure, and reading protocol. The earliest rain gauges consisted of simple cubic funnels. In the nineteenth century, the observations were made at an observatory at S. Anna School; since 1808 at S. Caterina high school; from 1836 to 1951 at the Patriarchal Seminary and from 1958 to present at the Cavanis Institute. In the twentieth century, weather stations belonging to different organizations at national (e.g., Water Magistrate, the Airforce at the Airports of S. Nicolò Lido and Tesserà) or regional level (e.g., the Regional Agency for the Environmental Protection and Prevention (ARPAV)) enabled to continue the measurements. The data quality improved in 1853, when the Patriarchal Seminary adhered to the Wien protocol for meteorology and geo-magnetic observations, and also in 1866 when it adhered to the directives established by the Central Service for the meteorology, Rome. Although the recent period has been considered in some climate analyses (Salon et al. 2008; Brunetti et al. 2012), the early period is still unexploited, except for a very small interval used to fill the 1764–1767 gap of the Padua series (Camuffo et al. 2022a).

Bologna The precipitation series started in Poggi Palace with Jacopo Bartolomeo Baccari in 1723. The precipitation has been recovered from the original Logs and analyzed by Camuffo et al. (2019) for the eighteenth century, while the subsequent period (since 1813) has been presented by Brunetti et al. (2001).

Milan The precipitation series started at the Brera Astronomical Observatory with Louis Lagrange in 1763, but the instrumental data were limited to temperature and pressure, and some weather notes; only the temperature and pressure have been recovered (Maugeri et al.

2002). The rain gauge was a funnel on the top of the *Specola*, connected through a pipe to a collecting vessel in the room below it. The astronomer Angelo Cesaris made regular precipitation readings from 1835 and these data have been recovered and analyzed (Buffoni and Chlistovsky 1992; Todeschini 2012). However, in the eighteenth century, the precipitation was recorded, because Cesaris published the yearly totals from 1764 to 1814 (Cesaris 1815) and Ferrario from 1764 to 1840 (Ferrario 1840). In 1774, Lagrange started the yearly publications named *Effemeridi Astronomiche di Milano* (i.e., Astronomical Ephemerides of Milan) that included only astronomical ephemerides. Since 1804 (observations 1801), the Ephemerides included an Appendix with monthly tables of daily (morning and afternoon) meteorological observations of atmospheric pressure, temperature, sky cover, and related phenomena. At the end of every month, the total precipitation amount was reported in Paris inches and lines.

3 Context and biases that affected the series from 1812 to 1864

Comparing the columns of the rain gauge readings and the weather notes, demonstrates periods of high consistency, indicating that the observer was accurate; however, others have low consistency. Two periods have biased data: (i) from 1812 to 1838, when the observers were prevalently Bertirossi-Busata and Conti (BC period); (ii) from 1839 to 1864 under the Santini direction (SA period). In 1812, after Vincenzo Chiminello was hit by an apoplectic fit, his assistant, the astronomer Francesco Bertirossi-Busata continued the meteorological observations. However, Bertirossi-Busata was in poor health but continued the measurements except for the two last months of his life. This caused a gap between September and October 1825. The new director, Giovanni Santini, had a primary interest in astronomy; as a result, he put the young custodian and technician Giovan Battista Rodella in charge of the meteorological observations and then the astronomer Carlo Conti until January 1865, when the astronomer Giuseppe Lorenzoni started his observations with a rigorous protocol.

Especially in the late BC period, i.e., after 1830, Rodella and Conti took readings after the rain had stopped, and nearly 25% of them were taken as cumulative values after several consecutive rainy days. In 1839, at the beginning of the SA period, there were some improvements with only a few days lost (~5%) and less than 20% of precipitation amounts collected relate to consecutive rainy days.

As an example, the Log of November 1814 is shown in Fig. ESM2: the precipitation amount (column 9, *Pioggia*, i.e., rain) was written only on days 3rd, and 4th, but in the last two columns, the observer classified as rainy days November 2nd, 3rd, 4th, 7th, 8th, and 9th. The following hypotheses/comments are possible: (i) on 4th November, the collected amount was measured regularly. This hypothesis is unlikely (ii) the amount reported on November 3rd is a cumulative value that cannot be referred to November 3rd only, but should be split between the 2nd and 3rd. This is very likely and may be justified because in the weather notes of November 2nd, as it was reported as *piovoso* (i.e., rainy), but the instrumental reading was not taken; (iii) The precipitation amount was not recorded on November 7th, 8th, and 9th (missing readings). Fig. ESM3 shows a particular case of irregular reading: the rain gauge measurement on May 4th, 1842 should be referred to as the rain event of the previous day, when the observer wrote *piovigginoso* (drizzly), but no measurement was taken. Contrariwise, May 4th was defined “cloudy” that justifies the absence of instrumental readings.

The irregular reading had a minor effect on the monthly totals but generated false extreme events because the cumulative amount collected over a number of consecutive days was reported as it was the amount of a single day. This bias increased the intensity of precipitation and decreased the number of rainy days. A secondary effect of the irregular reading was the increase in the risk of evaporation loss during sunny days: the longer the exposure period, the higher the evaporation leakage from the rain gauge, especially in conditions of strong radiation or intense wind.

Except for precipitation, observations and readings of other meteorological variables were taken regularly: there are only two main gaps (September–October 1825 and May–December 1838), and further 18 and 6 spot days for the period BC and SA, respectively. Overall, 329 days are without any information, from 1/1/1812 to 31/12/1864. This can be explained because the barometer was kept indoor and the other instruments were exposed on a terrace that was easily accessible so that all of them could be read without getting wet. Instead, the rain gauge location required the observer to be exposed to the weather to read the instrument. BC used the rain gauge built by the technician Rodella, similar to the one previously used by Toaldo and Chiminello: the funnel was a cubic box, 1 Paris foot side length, i.e., 0.105 m² cross section (Camuffo et al. 2020). Starting from January 1839, SA used the dome of the Meridian Circle Room on the *Specola* tower as a huge funnel, with a catching surface of 27.5 m².

4 Methods

The method is based on the careful analysis of the original Logs of Padua, i.e., the Log for the meteorological observations and the Log for the astronomical observations including notes concerning the state of the sky and precipitation.

The overall method is illustrated in Flowchart 1 (Fig. 1). After the dataset has been recovered from the meteorological Log, all the collected data have been examined. For every day, it is important to verify the consistency between the column with the regular instrumental readings (i.e., quantitative information) and the last column with the weather observations (i.e., qualitative information). The combinations derived from the analysis of this consistency, i.e., full, partial or absent, determine four classes of data as follows:

- (i) a regular reading (green color) is defined when both the instrumental reading (i.e., quantitative information) and the weather note (i.e., qualitative information) were correctly taken and reported according to the protocol. In this case, the observer gave the collected amount and an indication of the precipitation type (i.e., regular reading, green rectangle). This class may also include observations made with some delay after the scheduled time, but that can be referred directly to a precise precipitation event (see for example the particular case of November 4th 1825, Section 5.7). Regular readings constitute 53% of the dataset, do not need any correction and are useful to assess a correlation between precipitated amount and type.
- (ii) a missing reading (orange color) is defined when the instrumental reading was missed, but the weather note was reported (i.e., qualitative information only). They account for 22% of the dataset. From the established correlation between instrumental reading and precipitation type of the regular readings, (i) it has been possible to assign an estimated amount to the weather notes, as explained in Flowchart 2. This method has been successfully used in Camuffo et al. (2022a) to fill the 1764–1767

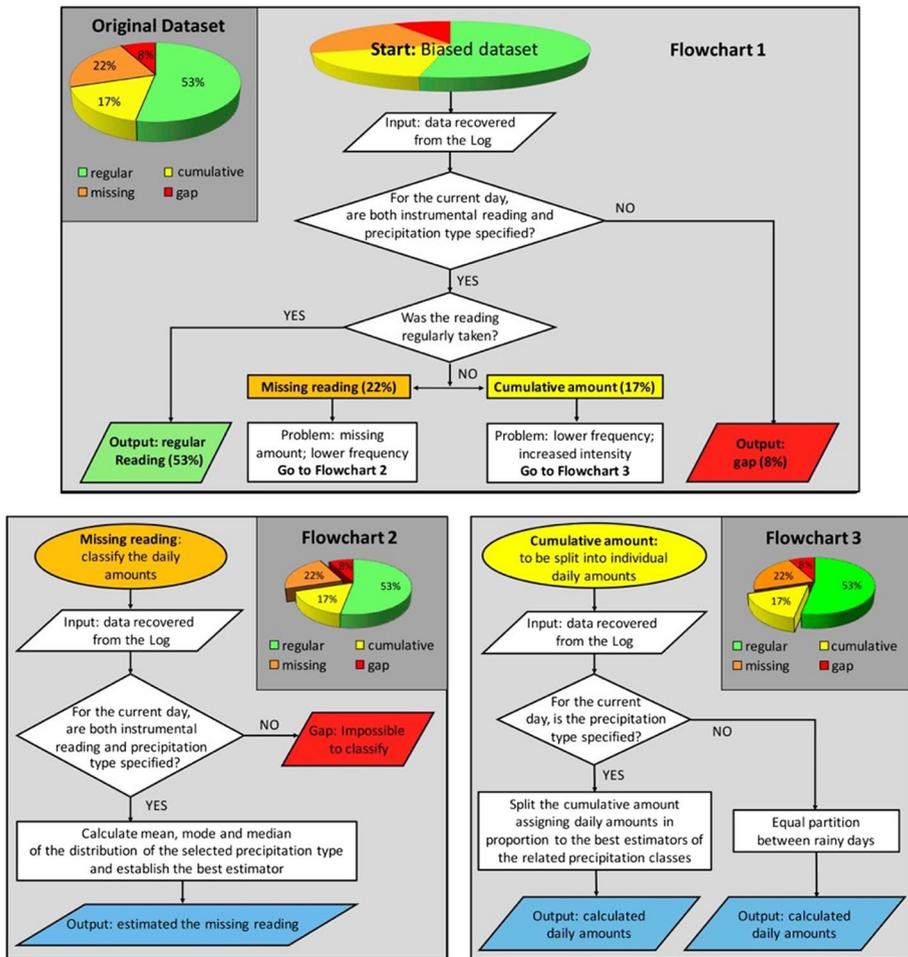


Fig. 1 Flowcharts illustrating the method used to correct and reconstruct the precipitation of the 1812–1864 period of Padua series. Flowchart 1: overview starting from the initial, biased dataset. Flowchart 2: treatment of missing reading, either transformed or to be considered gap. Flowchart 3: treatment of cumulative amount, to be split in relation to the likely contribution of the rainy days

gap, but in this case, the precipitation types and the regular readings used for calibration come from the same source. However, it must be specified that, if the weather note was missing too, this constituted an information gap and it is impossible to classify (see bullet iv).

- (iii) a cumulative amount of the precipitation collected for some days (yellow color) occurs when the instrumental readings report quantitative information that was correct for the total amount, but wrong for individual days. The weather notes provide a qualitative information (i.e., precipitation type). In other words, the observer reported an instrumental reading taken after some consecutive rainy days, that constitutes the total amount of water collected in the period of time between the current (or actual) and the previous reading. The cumulative amounts represent the 17% of the dataset. From the correlation established between instrumental reading and precipitation type

of the regular readings, (i) it has been possible to assign an estimated amount, as illustrated in Flowchart 3. If the Log reports a precise characterization of the previous rainy days it is possible to divide the total amount in relation to the mentioned precipitation types; if the characterization is missing, the total amount is divided in equal parts.

- (iv) a gap (red color) is when both quantitative and qualitative information are missing. The Log is missing or has not been compiled. Fortunately, this affects only 8% of the dataset. To overcome this bias, other kinds of documentation could be consulted but a solution is not certain, at least at the daily level.

In general, the Padua series is characterized by regular observations, except in the BC and SA periods, in which missing readings, cumulative amounts and gaps appeared frequently. The proportion between these classes is shown in Fig. 2.

The data belonging to the above classes have been processed in different ways.

5 Data analysis and discussion

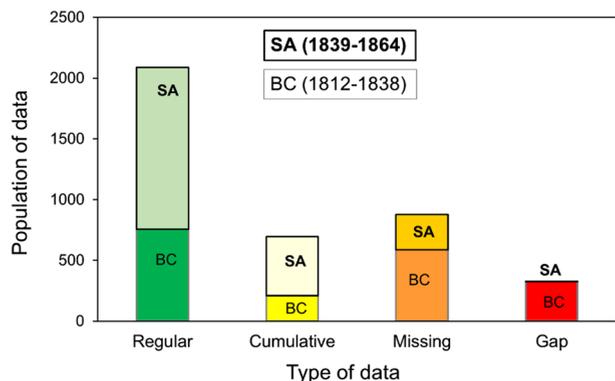
5.1 Generalities of the 1812–1864 biased period

In the 1812–1864 period, raw precipitation data reported in the Log had different features in comparison with the rest of the series, i.e., the light rains disappeared, while the heavy rains increased (Fig. 3a). The frequency of rain dramatically drops during the BC (yellow area) and SA (orange area) periods, and the annual percentiles changed accordingly. It should be underlined that the sub-periods in which the distribution changed can be recognized only if the data are represented with dots: if they are represented with vertical lines (Marani and Zanetti 2015), the tallest lines of the heaviest precipitation mask the low density in correspondence of the small number of the scarce precipitation (Camuffo et al. 2020).

The normalized precipitation frequency (Fig. 3b) shows an overestimation of single-day rains for the BC and SA periods, compared with the previous and subsequent periods, consistent with the 1961–1990 reference period.

The distribution of the daily amounts (Fig. 3c) shows a marked decrease in light rains: the frequency of the daily amounts lower than 2 mm is 15% less than in other periods.

Fig. 2 Type of data, i.e., regular, cumulative, missing or gap, derived from the comparison between the precipitation column and the weather notes. Data refer to the two periods: BC (bottom of the histogram columns) and SA (top of the histogram columns)



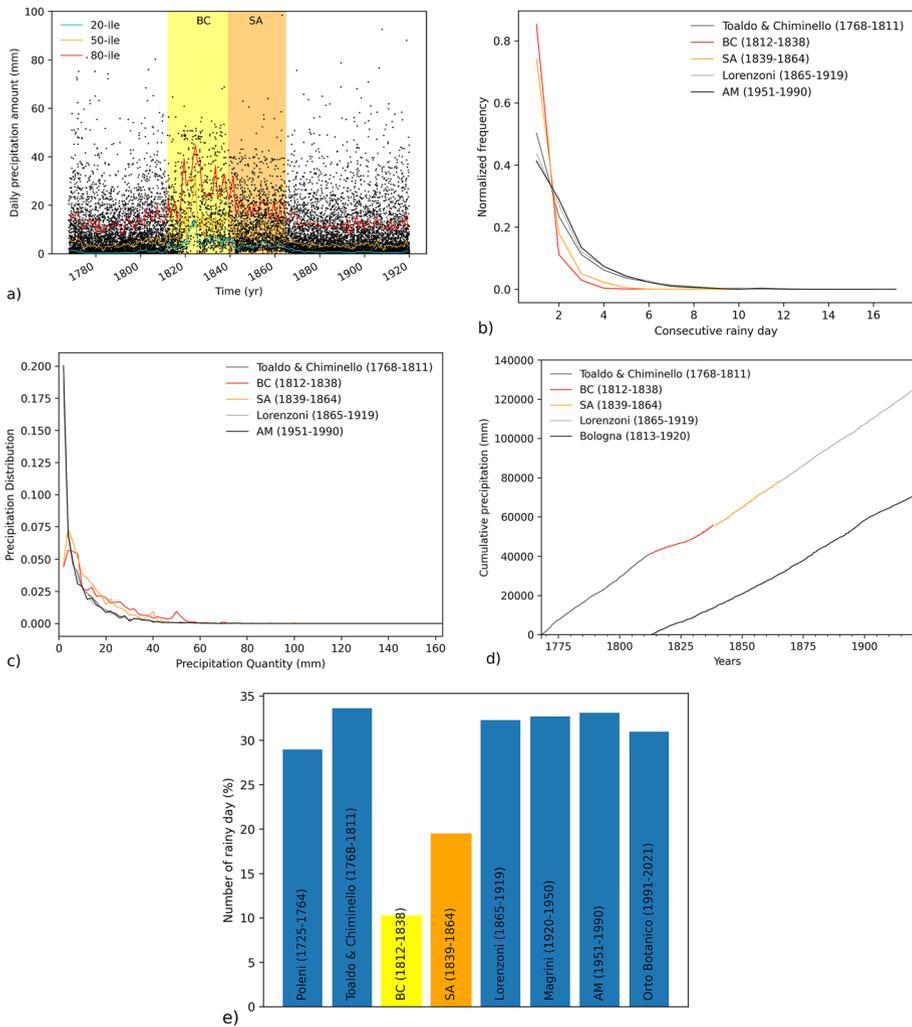


Fig. 3 **A)** Plot of the daily precipitation amount with the indication of the 20, 50 and 80 percentiles over the years in the BC and SA periods. **b)** Normalized frequency of consecutive rainy days over different periods, i.e.: 1768–1811; BC 1812–1838; SA 1839–1864, and the 1961–1990 reference period. **c)** Daily precipitation distribution over the same periods. **d)** Cumulative plot of the precipitation in Padua from 1768 to 1919, compared with Bologna. **e)** Comparison of the percentage of rainy days over the whole Padua series, from 1725 to 2021, with bins related to the main observers (either persons or institutions)

In addition, during the BC period, the slope of the cumulative precipitation amount is lower than in other periods (Fig. 3d): the observations in 1768–1811 and 1839–1919 (gray and orange lines) have nearly the same slope, while in 1812–1838 (red line) slope is 31% lower.

Finally, the percentage of rainy days in the different periods is nearly constant over the whole series, around 30%, except in BC and SA periods (Fig. 3e).

This anomalous behavior might be ascribed to temporary climatic change or to inaccurate observing protocol. The climate hypothesis can be excluded for three reasons: (i) by comparing the precipitation column with the weather notes of the same day, it is evident that readings were taken irregularly, not at the scheduled times; (ii) the slope of the cumulative amounts of the Padua series remains unchanged before and after the 1812–1864 period; and (iii) the slope of the cumulative precipitation amounts of the contemporary series of Bologna (100 km south from Padua) is nearly constant during the considered period (Fig. 3d).

5.2 Classification of the precipitation types

For consistency reasons, two homogeneous subsets of data have been created, one for the BC and one for the SA periods, respectively. For every regular reading of each period, the precipitation amount has been associated with the related term indicating the precipitation type in the weather notes. Since the observers in the BC and SA periods were different, the choice and number of terms used in the two periods were slightly different, i.e., 37 in the BC period, 34 in the SA period, composed as follows: rain BC=30 and SA=26; hail BC=3 and SA=3; snow BC=2 and SA=3; hail or snow mixed to rain BC=2 and SA=2. The identified types are listed in Tables ESM2 and ESM3.

When some rare precipitation types occurred, e.g., hail and snow, the classes related to these kinds of events were unified to increase the population and thus a higher statistical significance.

The amount of precipitation collected by a rain gauge depends on both the intensity and the duration of the precipitation. A challenging problem when using documentary sources is that in general, the descriptions give a general characterization that might be misleading. It would be desirable to identify and separate the precipitation types referred to the duration from those referred to intensity. Camuffo et al. (2022a) considered the problem, with the advantage that the observer Morgagni was very rich in adjectives, adverbs, and their combination, and gave extremely accurate descriptions. During the BC and SA periods, the observers in general described only intensity, with the exception of a few terms, such as “*continua*” (continuous), “*a tratti*” (at intervals), “*di tratto in tratto*” (at times), and “*ininterrotta*” (uninterrupted) that were always used as the only specification. These terms have been considered as additional classes without distinction between duration and intensity.

Daily precipitation has been distributed considering both term meaning, and their consecutive repetition: for example, if the Log reports “rain,” “rain,” one day, “drizzle” and the total amount the next day, the amount has been proportionally distributed in relation to the classes “rain” and “drizzle” and their occurrences.

In the BC period, the most populated class of regular readings is “rain” (without further specification) with 58% of occurrences, followed by “drizzle” (9.2%). In SA, “rain” 26%, “rainy” 26%, and “drizzly” 14%. Once every precipitation event has been attributed to a specific class, a quantitative value can be associated to it, as its most typical value. Unfortunately, the classification does not give a precise value but a certain skew distribution that may be represented in terms of mean, median or mode. The most numerous class, i.e., “rain” is shown in Fig. 4 for the two subsets BC and SA. The column widths of the histograms have been calculated using the Doane method (Doane 1976), an estimator that takes into account non-Gaussian distributions, as precipitation. The mode has been calculated using the half-sample method described by Bickel (2002) and performed using the R

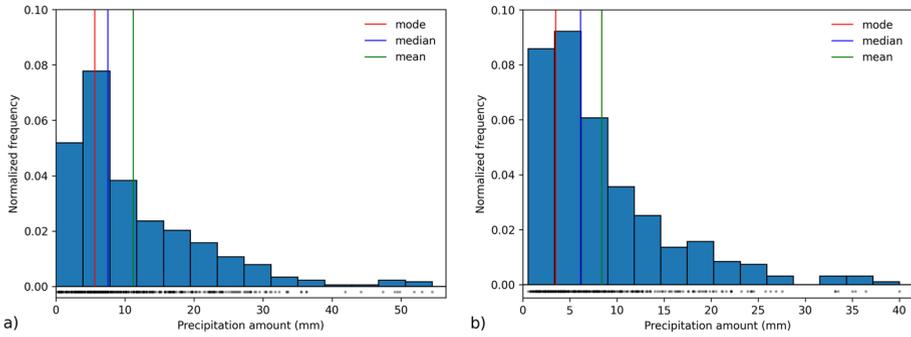


Fig. 4 Normalized frequency of precipitation amount (mm) of the “rain” class for (a) BC, (b) SA

package Modeest. The two distributions are different, depending on the subjective perception of the observer.

5.3 The best estimator to characterize a class

As explained in Section 4, the approach is based on the transformation of weather notes into quantitative values, thanks to the calibration made possible by the contemporary presence in the same Log of instrumental readings and weather observations (Fig. 1, Flowchart 2). Since the observers are the same for both, there is no bias for subjective interpretation, different instruments, observation protocols, and locations that can be problematic when different stations are compared.

The missing daily amounts have been reconstructed by assigning the value correspondent to the precipitation type reported in the weather notes. In this operation, a critical point is the choice of the most representative estimator of every class, i.e., mode, median, and (arithmetic) mean. The results are shown in Fig. 5a for BC and Fig. 5b for SA.

- The mode represents the most frequent and therefore the most probable value. It is not affected by the tail of extreme events. The mode is the lowest of the three estimators (Fig. 4), so the rainfall calculated using the mode can be penalized in the defect. In

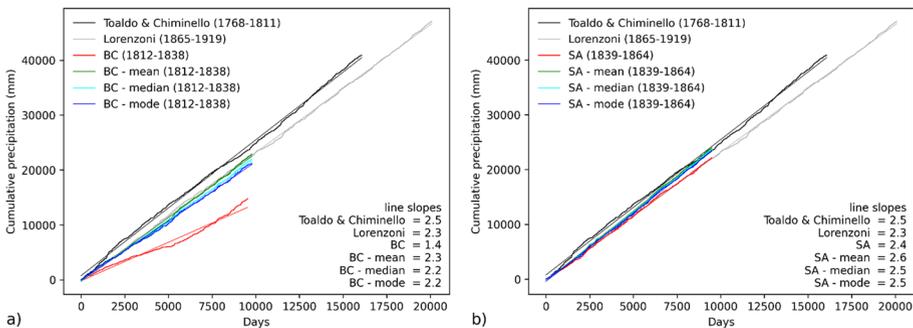


Fig. 5 Comparison of the cumulative precipitation amount of BC (a) and SA (b) periods, and the previous (Toaldo and Chiminello) and subsequent (Lorenzoni) ones. For BC and SA, raw data are compared with data calculated using mean, mode, and median as estimators

fact, the BC and SA periods both appear underestimated in comparison with the previous period (i.e., Toaldo and Chiminello, 1768–1811) and the subsequent one (i.e., Lorenzoni 1865–1919). This fact becomes evident by comparing the slopes of the observed and the calculated cumulative amounts in the various periods. The slope of the observed raw values in the Log in the BC period strongly departs from the other slopes; in SA only slightly. The calculated cumulative amount is much more consistent. In BC, it is still slightly smaller than the observed values in the previous and subsequent periods. In SA, the slopes show better agreement.

- The median represents the central position of the distribution and is not skewed by the extremely large values, therefore provides a good representation of a typical value of that class. The median is the intermediate estimator. The rainfall calculated with the median is more consistent with the values observed in the previous and subsequent periods, although slightly underestimated.
- The mean gives an average representation of the whole population and keeps memory of the small proportion of extremely large values. The mean is the highest of the three estimators (Fig. 4). With the mean, in both BC and SA periods the calculated precipitation seems to be consistent with the other two periods, slightly better than the median. In BC, the result is satisfactory; in SA slightly overestimated. After correction, the SA period becomes the period with the largest amounts in the Padua series. This is realistic, considering that the enormous catching area of the dome had a much higher collecting efficiency than the small cubic funnels used by Toaldo and Chiminello affected by the high turbulence generated by the sharp edges and small borders, and also in comparison with the small circular funnels used by Lorenzoni (Camuffo et al. 2020, 2022b).

When the statistical errors of the linear regression (RMSE and MAE, see Table ESM4) are considered, the mode appears to be the best choice, as the linear regression performed using the mode provides lower values of RMSE and MAE, respect to the mean and the median, i.e., a regression model with higher accuracy. On the other hand, the slope of the precipitation cumulative amount is better aligned with the previous and the subsequent periods if the mean value is used, and less with the mode. With the mode, the total precipitation amount for the period May–December 1838 is underestimated (see Sect. 5.6). In conclusion, the mode gives the best fit, but an underestimation of the total amount; conversely, the regression using the mean is less accurate, but more convenient for the reconstruction of the total amount. Therefore, the mean value of each distribution has been assumed as the best estimator of the related precipitation type.

5.4 Splitting cumulative amounts into daily amounts

Following the descriptions reported in the weather notes, in the BC and SA periods, the 17% of the original data are cumulative values that need to be split into different days (Fig. 1, Flowchart 3). This has been done following two steps: (i) firstly every rainy day has been assigned to a specific class, given by the precipitation type reported by the observer; (ii) then, the cumulative value has been distributed in the previous rainy days proportionally to the best estimator that characterizes the class. For example, the rain recorded on 3rd November 1814 should be split between 2nd and 3rd November (Fig. ESM2). On the 2nd the last part of the day was noted as “rainy,” and the next day as “drizzly.” The amount reported on 3rd November of 20 French points (i.e., 3.76 mm) has been split as follows: 41% on 3rd November (drizzly) and 59% on 2nd November (rain), using the ratio of the

respective estimators reported on Table ESM2. The advantage of this method is that (i) the total amount of precipitation related to some consecutive days is distributed among the contributing rainy days; (ii) the number of rainy days is updated, and consequently, the precipitation frequency is corrected.

It is well known that evaporation loss may affect the collected amount (Goodison et al. 1981; Sevruk and Hamon 1984; Lanza et al. 2006). WMO (2018) estimates the precipitation loss for evaporation up to 4% of the amount of measured water and suggests frequent measurements to minimize this bias. The magnitude of evaporation depends on climatic area, season, and especially rain gauge type (Sevruk 1974, 1982; Leeper and Kochendorfer 2015). However, some precautions were applied to prevent evaporation (e.g., Mariotte 1686; Mordecai 1938): in the BC period, the rain gauge was a cubic box, where the water percolated through a small hole into a cylindrical, vertical vessel. A glass tube, connected to the bottom of the cylinder, is allowed to measure the level of water without opening it (Camuffo et al. 2020). In the SA period, a pipe transported the water collected by the huge funnel into a closed reservoir. In both cases, the evaporative loss was modest. Consequently, it has been decided to avoid corrections not supported by adequate experimental evaluations because the bias they may generate may be larger than the real correction.

5.5 Snow conversion to rain

Snow is rare in Padua, in fact, occasionally snow days were reported in BC and SA periods: a total of 28 cases of snow depth are recorded, with the depth written in the weather note column of the Log. These values have been transformed into water amount following the prescription of WMO (2018), i.e., the snow depth is equivalent to 1/10 of rainfall. These 28 cases accounted for a total amount of 345.4 mm of water, and the individual values have been reported in the recovered series (Table ESM5).

5.6 Filling the gap May–December 1838

The original Log from May to December 1838 was lost, but a copy of the weather notes was found in the *Giornale Astro-Meteorologico* (Journal about Astronomy and Meteorology, GAM) (Pietropoli 1839). In addition, in this volume, both monthly totals and frequency were reported. Therefore, the original weather notes have been recovered from GAM and have been used to fill this gap. The reconstruction of the missing precipitation has been made at the daily level, using the proper estimator for every precipitation type, as explained in Sect. 5.3. In Table ESM6, the original comments concerning the rainy days are reported, together with the estimated amounts.

The resulting monthly amounts are listed in Table 1 and compared with the monthly totals reported in GAM. Even if the reconstructed amounts for May and October are lower than the values reported in GAM, the sum of all the estimated values is nearly 28% higher than the observed value. This might be explained because in the BC period, a number of observations were missed and did not contribute to the observed totals in GAM. The reconstructed daily amount for November and December is much higher than the values reported in GAM. This is due to the poor variety of terms used to describe precipitation in these 2 months, i.e., the term “rain” was used in almost all cases. Note that the estimations made with the median or the mode are lower than the values reported: this confirms the choice of the mean as the best estimator.

Table 1 Comparison between calculated and observed monthly totals for the May–December 1838 gap. The first main column includes the sums of the daily values obtained with the method of the precipitation types and estimators, derived from the weather notes reported in GAM. The second main column, the observed monthly totals, also reported in GAM (Pietropoli 1839)

Month	Calculated from GAM notes				Observed and reported in GAM				Percentage variation
	Amount (mm)			Frequency (days/yr)	Amount			Frequency (days/yr)	
	Mean	Median	Mode		Paris Inches	Paris Lines	mm		
May	102.1	76.4	31.6	9	4	11.1	133.6	9	−31%
June	48.8	32.9	19.3	9	1	8.2	45.7	9	6.4%
July	53.1	36.6	20.4	11	1	2.8	33.4	11	37%
August	44.3	31.9	17.9	6	1	1.6	30.7	6	31%
September	104.8	89.7	77.5	11	2	2.6	60.1	11	42%
October	27.0	19.3	11.3	3	1	4.2	36.6	2	−36%
November	162.1	103.9	69.2	17	2	9.0	74.6	14	54%
December	70.7	45.7	31.2	7	0	11.7	26.4	7	63%
Total	612.9	436.4	278.4	73			441.1	69	28%

The revision has considered also the number of rainy days, but this was affected by bias, except for 4 rainy days left unnoticed.

5.7 Particular cases

This section is devoted to some particular cases, or minor gaps, which required the use of additional documentary sources. The most useful sources are as follows: (i) the Log of Astronomical Observations: when astronomers observed, the sky should be clear, so this is indirect information that it was not a rainy day; (ii) GAM: in some periods, other scientists in other localities took meteorological observations and published them on GAM, which may help to understand the quality of the missing days. In Table ESM7, the missing days are reported; in the “weather” columns, there is the indication of the sky quality deduced by the astronomical Log and GAM. In several cases, astronomers missed both the astronomical and the meteorological observations. It has been possible to recover the sky quality only for 21 days out of 93: 20 were clear, one, October 26th 1825, rainy.

September 1823 The observer wrote that the month had abundant rains that were not recorded, due to a failure of the rain gauge. However, other readings were regularly taken and the weather notes reported. The precipitation amount has been estimated using the method for missing readings.

November 4th 1825 The observer recorded 258 French Points (i.e., 48.5 mm) of rain, but he wrote in the comment that “the night was cloudy with little rain.” This was the first measurement of precipitation for 2 months after Bertirossi-Busata’s death. The value recorded is likely the water found in the gauge at the end of this period. The GAM (1826) gives other useful indications: in the weather comment for the year 1825, Jacopo Penada (1826) wrote that an important thunderstorm with strong winds, heavy rain, and hail happened on October 26th 1825. Probably most of the precipitation reported on November 4th 1825 should be referred to this day. As there are no other indications concerning most of

the days of September–October 1825, the quantity measured on 4th November has been split between 26th October and 4th November, using the procedure described in Sect. 5.3, resulting in 15.3 mm for the 4th November and 33.2 mm for the 26th October.

December 23rd 1842, October 3rd and 10th 1843 The observer wrote that the water on the rain gauge was due to nocturnal condensation, either dew or fog. Therefore, these quantities have been removed by the series and these days have been considered not rainy.

November 15th 1863 The value 17.4499 was written in the column of the precipitation amount, with the annotation “rain gauge full.” So, this quantity should be considered less than the precipitation that has really fallen. Since it has not been possible to estimate the right amount, the original value has been used.

5.8 The corrected dataset

The corrected precipitation subsets, BC and SA, have been compared with the previous and subsequent periods, considering the amounts, the frequency, and the consecutive rainy days.

The density of the dots of the daily amounts (Fig. 6a) indicates the following:

- 1) the frequency of rain has increased in respect to the original raw data (Fig. 3a), with most of the missing readings reconstructed
- 2) the overestimation of (false) extreme events has been corrected since the 80-ile is lower
- 3) the rainy days in the BC period remain with some uncertainty: probably they have not been completely recovered, as shown by 20 and 50 percentiles, especially between 1820 and 1830. However, the cumulative plot (Fig. 5a) shows that the precipitation amount has been fully recovered.

As several rainy days were missed in the Log, after this reconstruction, the percentage of rainy days increases (Fig. 6b): SA period is comparable with the Poleni period; BC remains below the other periods of the series, but some unresolved gaps could change the statistics. After the corrections, the distributions of consecutive rainy days became similar to the other periods (Fig. 6c and d). Finally, the whole rain distributions before and after the correction have been compared (Fig. 6e and f). The original values show small anomalous secondary peaks at high daily amounts (nearly at 50 mm for BC and at 40 mm for SA) that disappear after the correction. The original flat shape of the BC distribution has been corrected to a standard appearance. In Table ESM8, the corrected series are reported, with the indication of how the daily data has been obtained.

The corrections of the BC and SA periods have improved the homogeneity of the yearly anomalies (Fig. 7a and b, red lines). The strong oscillations found first by Camuffo (1984) and then by Marani and Zanetti (2015) disappear, especially in the precipitation frequency. The anomalies over the whole series (1725–2021, Fig. 7c and d) are more regular with an alternation of periods in which precipitation amounts and frequency are greater than, or less than the 1961–1990 reference period. The lower frequency around 1825, when Bertirossi-Busata died, may be explained by the absence of a skilled observer. In fact, the observations were temporarily assigned to the technician Rodella, who was an excellent mechanical builder, but not trained for observations. He may have missed some events, especially

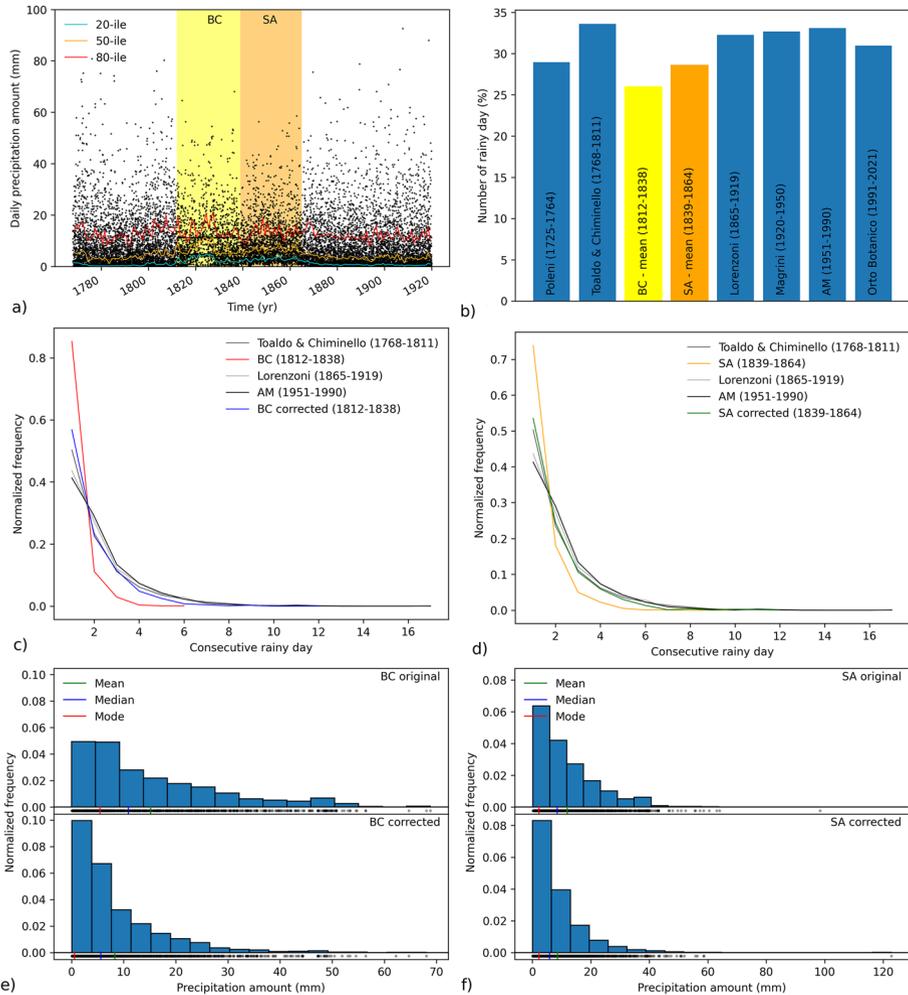


Fig. 6 a) Plot of the daily precipitation amounts with the indication of yearly 20- 50- and 80-iles over BC and SA period after correction. b) Comparison of the percentage of rainy days in the whole Padua series after correction. c, d) Distribution of consecutive rainy days: comparison between the BC and SA periods before and after the correction with the previous and the following periods and with the 1961–1990 reference. e) The distribution of the original raw data. f) The distribution after correction

those near threshold. In addition, the frequency for the year 1825 has been normalized, because there are still 54 missing days of the September–October gap.

6 Conclusions

In this paper, two datasets have been implemented and used to correct and reconstruct precipitation at the daily level, for the period 1812–1864 at Padua. Missing readings, cumulative amounts, and gaps, that often affect early precipitation series, that make

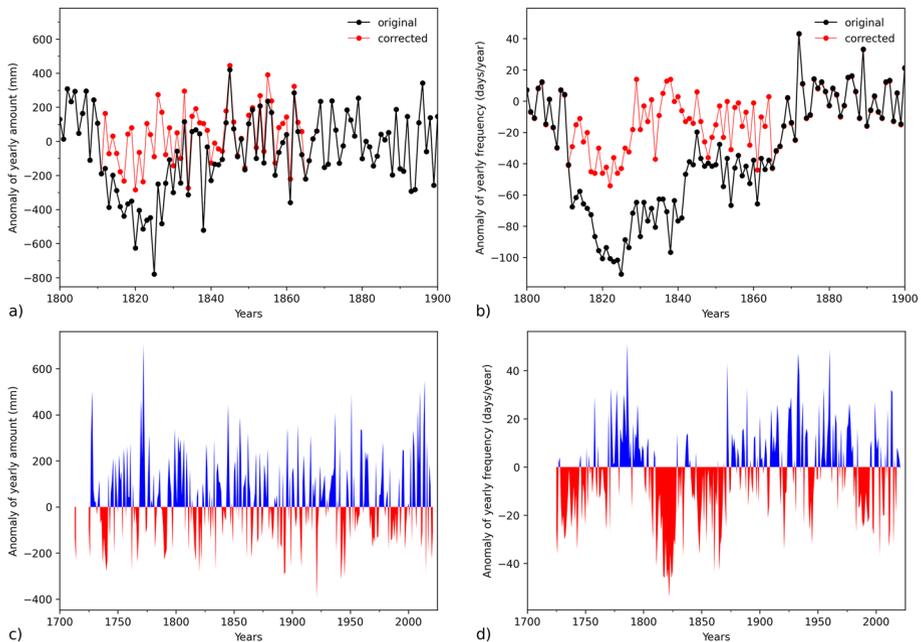


Fig. 7 a, b) Anomalies of yearly precipitation amount and frequency before (black lines) and after (red lines) the corrections. c, d) Anomalies of the whole series using the corrected data and the datasets listed in Camuffo et al. 2020

measurements non-comparable with other contemporary series or other periods are addressed. For modern series, the correction can be made by comparing nearby stations, but in the early period, this is not possible, since scientists operated independently of each other, using different measurement methodologies, that prevent direct comparison. For this reason, a careful analysis of the information extracted from documentary sources has been used.

The methodology proposed in this work has enabled bias to be corrected using the weather notes reported in the same documentary source (Log) of the original data. With this method, the series has been recovered without adding additional bias due to the use of different locations, instrument, or observing protocols. Comparing rain gauge measurements with weather notes, precipitation amounts are classified in three types: regular, cumulative, and missing. The first has formed the datasets that have been used to distribute cumulative values for rainy days and to fill missing measurements.

Using this approach, 52 years of the long Padua precipitation series have been corrected and reconstructed: (i) cumulative values have distributed in the actual rainy days; (ii) most of the missing amounts have been reconstructed; (iii) the series is no longer affected by false extreme events; (iv) the rain amount has been fully recovered; and (v) the strong oscillations in the anomalies found by previous studies disappear.

Starting from the early work of Camuffo (1984), the analysis and the correction of the long precipitation series of Padua required nearly 40 years of careful studies, and three publications (Camuffo et al. 2020, 2022a, this work) to solve, the problems and the bias of this very long series. The end result, a homogenous long precipitation series

for Padua (1725–present), is of considerable value in understanding long-term precipitation variability, patterns, and trends, and is an important addition to the long precipitation series available in Italy and Europe more broadly.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10584-023-03485-5>.

Acknowledgements The authors are grateful to the Historical Archive of the INAF-Astronomical Observatory of Padua for the free consultation and the kind permission to publish documents.

Author contribution All authors have equally contributed.

Funding Open access funding provided by ISP - VENEZIA within the CRUI-CARE Agreement.

Data availability The datasets generated during the current study are available in ESM8.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Historical References

- Cesaris A (1815) *Riflessioni sopra la pioggia che cade a Milan*. In: Carlini F (ed) *Effemeridi Astronomiche di Milano per l'anno bisestile 1816. Con Appendice*. Cesarea Regia Stamperia, Milano, pp 87–100
- Ferrario G (1840) *Statistica medica di Milano*. Guglielmini e Redaelli, Milano
- Mariotte M (1686) *Traité du mouvement des eaux et des autres corps fluides*. Chez Estienne Michallet, Paris
- Penada J (1826) *Relazione meteorologica dell'anno 1825*. In: *Giornale Astro-Meteorologico per l'anno 1827*. Tipografia del Seminario, Padova, pp 72
- Pietropoli G (1839) *Relazione meteorologica dell'anno 1838*. In: *Giornale Astro-Meteorologico per l'anno 1840*. Tipografia del Seminario, Padova, pp 41–54

Modern References

- Adamson GCD (2015) Private diaries as information sources in climate research. *Wires Clim Change* 6:599–611. <https://doi.org/10.1002/wcc.365>
- Aieb A, Madani K, Scarpa M et al (2019) A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria. *Heliyon* 5(2):e01247. <https://doi.org/10.1016/j.heliyon.2019.e01247>
- Bellido-Jiménez JA, Gualda JE, García-Marín AP (2021) Assessing machine learning models for gap filling daily rainfall series in a semiarid region of Spain. *Atmosphere* 12:1158. <https://doi.org/10.3390/atmos12091158>

- Bickel DR (2002) Robust estimators of the mode and skewness of continuous data. *Comput Stat Data Anal* 39:153–163. [https://doi.org/10.1016/S0167-9473\(01\)00057-3](https://doi.org/10.1016/S0167-9473(01)00057-3)
- Brugnara Y, Flückiger J, Brönnimann S (2020) Instruments, procedures, processing, and analyses. Swiss Early Instrumental Meteorological Series, G96(02). *Geographica Bernensia*, Bern, pp 17–32. <https://doi.org/10.4480/GB2020.G96.02>
- Brunetti M, Buffoni L, Lo Vecchio G et al (2001) Tre secoli di meteorologia a Bologna. CUSL, Milano
- Brunetti M, Lentini G, Maugeri et al (2012) Projecting north eastern Italy temperature and precipitation secular records onto a high resolution grid. *Phys Chem Earth* 40–41:9–22. <https://doi.org/10.1016/j.pce.2009.12.005>
- Buffoni L, Chlistovsky F (1992) Precipitazioni giornaliere rilevate all'Osservatorio Astronomico di Brera in Milano dal 1835 al 1990. EdiErmes, Milano
- Caldera HPGM, Piyathissie VRPC, Nandalal KDW (2016) A comparison of methods of estimating missing daily rainfall data. *Engineer XLIX*(4):1–8
- Camuffo D (1984) Analysis of the series of precipitation at Padova, Italy. *Clim Change* 6:57–77. <https://doi.org/10.1007/BF00141668>
- Camuffo D (2002) History of the long series of daily air temperature in Padova (1725–1998). *Clim Change* 53:7–75. <https://doi.org/10.1023/A:1014958506923>
- Camuffo D (2022a) Historical documents as proxy data in Venice and its marine environment. In: Oxford research encyclopedia, climate science. Oxford University Press, pp 1–47. <https://doi.org/10.1093/acrefore/9780190228620.013.875>
- Camuffo D (2022b) Wind-driven rain impinging on monuments and mountain slopes. *J Cult Heritage* 55:149–157. <https://doi.org/10.1016/j.culher.2022.03.007>
- Camuffo D, Becherini F, della Valle A (2019) The Beccari series of precipitation in Bologna, Italy, from 1723 to 1765. *Clim Change* 155:359–376. <https://doi.org/10.1007/s10584-019-02482-x>
- Camuffo D, Becherini F, della Valle A, Zanini V (2020) Three centuries of daily precipitation in Padua, Italy, 1713–2018: history, relocations, gaps, homogeneity and raw data. *Clim Change* 162:923–942. <https://doi.org/10.1007/s10584-020-02717-2>
- Camuffo D, Becherini F, della Valle A, Zanini V (2022a) A comparison between different methods to fill gaps in early precipitation series. *Environ Earth Sci* 81:345. <https://doi.org/10.1007/s12665-022-10467-w>
- Camuffo D, della Valle A, Becherini F (2022b) How the rain-gauge threshold affects the precipitation frequency and amount. *Clim Chang* 170(1):1–18. <https://doi.org/10.1007/s10584-021-03283-x>
- Doane DP (1976) Aesthetic frequency classifications. *Am Stat* 30(4):181–183. <https://doi.org/10.2307/2683757>
- Domínguez-Castro F, García-Herrera R, Vaquero JM (2015) An early weather diary from Iberia (Lisbon, 1631–1632). *Weather* 70:20–24. <https://doi.org/10.1002/wea.2319>
- Goodison BE, Ferguson HL, McKay GA (1981) Comparison of point snowfall measurement techniques. In: Gray DM and Male MD (ed). *Handbook of Snow*. Pergamon, Tarrytown NY, pp 200–210
- Harvey-Fishenden A, Macdonald N (2021) Evaluating the utility of qualitative personal diaries in precipitation reconstruction in the eighteenth and nineteenth centuries. *Clim Past* 17:133–149. <https://doi.org/10.5194/CP-17-133-2021>
- Jones PD, Briffa KR, Osborn TJ et al (2009) High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene* 19:3–49. <https://doi.org/10.1177/0959683608098952>
- Lanza LG, Leroy M, Alexadropoulos C, Stagi, L, Wauben W (2006) Instrument and Observing Methods. Report No. 84. WMO Laboratory Intercomparison of Rainfall Intensity Gauges. WMO/TD No. 1304. World Meteorological Organization, Geneva. https://library.wmo.int/doc_num.php?explnum_id=9313. Accessed 1 Sept 2022
- Leeper RD, Kochendorfer J (2015) Evaporation from weighing precipitation gauges: impacts on automated gauge measurements and quality assurance method. *Atmos Meas Tech* 8(6):2291–2300. <https://doi.org/10.5194/amt-8-2291-2015>
- Longman RJ, Newman AJ, Giambelluca TW, Lucas M (2020) Characterizing the uncertainty and assessing the value of gap-filled daily rainfall data in Hawaii. *J Appl Meteor Climatol* 59(7):1261–1276. <https://doi.org/10.1175/JAMC-D-20-0007.1>
- Marani M, Zanetti S (2015) Long-term oscillations in rainfall extremes in a 268 year daily time series. *Water Resour Res* 51(1):639–647. <https://doi.org/10.1002/2014WR015885>
- Maugeri M, Buffoni L, Delmonte B et al (2002) Daily Milan temperature and pressure series (1763–1998): completing and homogenising the data. *Clim Change* 53:119–149. <https://doi.org/10.1023/A:1014923027396>

- Mordecai A (1938) Report of some meteorological observation made Frankford Arsenal near Philadelphia. *J Franklin Inst* 26(1):30–37. [https://doi.org/10.1016/S0016-0032\(38\)80116-2](https://doi.org/10.1016/S0016-0032(38)80116-2)
- Nash DJ, Adamson GCD, Ashcroft L et al (2021) Climate indices in historical climate reconstructions: a global state of the art. *Clim Past* 17:1273–1314. <https://doi.org/10.5194/cp-17-1273-2021>
- Raichich F (2008) Some features of Trieste climate from an eighteenth century diary (1732–1749). *Clim Change* 86:211–226. <https://doi.org/10.1007/s10584-007-9357-x>
- Salon S, Cossarini G, Libralato S et al (2008) Downscaling experiment for the Venice Lagoon. I. Validation of the Present-Day Precipitation Climatology. *Climate Research* 38(1):31–41. <https://www.jstor.org/stable/10.2307/24870395>
- Sevruc B (1974) Evaporation losses from containers of Hellmann precipitation gauges. *Hydrol Sci J* 19:26. <https://doi.org/10.1080/02626667409493902>
- Sevruc B (1982) Operational Hydrology. Report No. 21. Methods of correction for systematic error in point precipitation measurement for operational use, WMO No 589. World Meteorological Organization, Geneva. https://library.wmo.int/doc_num.php?explnum_id=1238. Accessed 1 Sept 2022
- Sevruc B, Hamon WR (1984) Instrument and observing methods. Report No. 17. International comparison of national precipitation gauges with a reference pit gauge. WMO/TD No. 38. World Meteorological Organization, Geneva. https://library.wmo.int/doc_num.php?explnum_id=9533. Accessed 1 Sept 2022
- Todeschini S (2012) Trends in long daily rainfall series of Lombardia (northern Italy) affecting urban storm-water control. *Int J Climatol* 32(6):900–919. <https://doi.org/10.1002/joc.2313>
- WMO (2018) Guide to Instruments and Methods of Observation, WMO No 8, 2018 edition. World Meteorological Organization, Geneva. https://library.wmo.int/doc_num.php?explnum_id=10616. Accessed 1 Sept 2022

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.