Theoretical and Applied Climatology
https://doi.org/10.1007/s00704-019-02772-5

ORIGINAL PAPER

Quality control process of the daily rainfall series available in Catalonia from 1855 to the present

Alba Llabrés-Brustenga[1,2] · Anna Rius[1] · Raúl Rodríguez-Solà[3] · M. Carmen Casas-Castillo[4] · Angel Redaño[2]

Received: 23 March 2018 / Accepted: 7 January 2019

Abstract

The quality control of weather data is a necessity and a responsibility of meteorological services that store, distribute, and use these data. In the present work, a newly designed quality control procedure for daily rainfall data is presented after it has been adjusted and tested with more than 10^7 data from 1726 daily rainfall measurement sites in Catalonia. It is applicable to data from different origins (e.g., automatic weather stations or manual historical measurements). The procedure is focused on relative comparison of daily data with reference stations that are automatically selected after an initial estimation of their quality and a proximity study regarding location and correlation. The presented procedure has been verified taking advantage of an available network in the study area that has been routinely quality controlled by technicians of the Meteorological Service of Catalonia. The newly designed quality control procedure for daily precipitation yields good results, especially for extreme values: type I error under 10% is found for values up to 150 mm (error decreasing for lower values) and type II error is under 16% when reported values are twice a measure of 50 mm or more (error decreasing for more extreme values). After the application of the quality control procedure, a selection of series with the minimum desired quality is achieved.

1 Introduction

The World Meteorological Organization claims the importance of quality controlling meteorological data and encourages organizations entrusted with data collection and storage to apply such procedures following some guidelines (WMO 2008). Nonetheless, there is not an established quality control procedure for precipitation data (Branisavljević et al. 2009) and a summary of existing methods developed by different authors and organizations is presented below.

Shearman (1975) proposed a quality control routine developed and applied in the Meteorological Office. It was based on an interpolation from surrounding observations that enabled comparison with the value that needs to be checked and could perform a reconstruction in case of accumulations, shifted day, or erroneous time measurements. Romero et al. (1998) presented a study using a 30-year database of daily rainfall data (from 1964 to 1993) for the Mediterranean regions of Spain. The database was created by selecting 410 stations of the study area (with an average distance of 15 km between them) from a number of 3366 available measuring sites. A quality control procedure developed by the authors was performed to the originally available data by means of an iterative estimation of the considered value as an interpolation

✉ Alba Llabrés-Brustenga
alba.llabres@gencat.cat

Anna Rius
anna.rius@gencat.cat

Raúl Rodríguez-Solà
raul.rodriguez@upc.edu

M. Carmen Casas-Castillo
m.carmen.casas@upc.edu

Angel Redaño
angel@meteo.ub.edu

[1] Servei Meteorològic de Catalunya, Berlin, 38-46, 08029 Barcelona, Spain

[2] Department of Applied Physics, Facultat de Física, Universitat de Barcelona, Martí i Franqués 1, 08028 Barcelona, Spain

[3] Department of Physics, ETSEIB, Universitat Politècnica de Catalunya, BarcelonaTech, Diagonal 647, 08028 Barcelona, Spain

[4] Department of Physics, ESEIAAT, Universitat Politècnica de Catalunya, BarcelonaTech, Colom 1, 08222 Terrassa, Spain

from reference stations. Missing values were filled using the same interpolation method. González-Rouco et al. (2001) studied the homogeneity of series of monthly precipitation in the southeast of Europe (including the Iberian Peninsula, the south of France, and the north of Africa) in the period 1899–1989, with a quality control consisting in trimming outliers which overpassed the threshold of three times the interquartilic range above the third quartile, and concluding that the detected outliers were caused by the climatology rather than errors in measurement. Abaurrea et al. (2004), in the development of a daily rainfall database of the Ebro basin, followed a methodology to manually classify rainfall daily series into "optimum," "auxiliary," or "not useful" categories after checking the most common problems in manual observations. Feng et al. (2004) used a series of checks in order to flag erroneous data and remove it before reconstruction of missing values and homogeneity controls. The outcome was the building of a daily dataset for 1951–2000 in China; the quality control checks extreme limits, internal consistency, temporal outliers, and spatial outliers. Viney and Bates (2004) studied the extension and implications of missing values filled with zeros in Australia, in particular, the case of rainfall weekly accumulations reported as daily values. You et al. (2007) explored the performance of three different tests for precipitation based on empirical statistical distributions underlying the observations and decided to adopt the Multiple Intervals Gamma Distribution (MIGD) method for their data. Chen et al. (2008) described an automated and objective protocol established at NOAA's Climate Prediction Center to quality control, on real time, daily information of gauges in the Global Telecommunications System. Their procedure compares historical gauge records, contemporary observations at nearby stations, satellite estimates, and numerical forecasts. Einfalt et al. (2008) presented a procedure to quality control a 5-year database (over 200 stations in central Europe) of subdaily meteorological data. The quality control checked the completeness of the series, errors of physically impossible values, variability of consecutive values (large variations would be suspicious), constant values (not plausible for long time intervals), internal consistency (checking the joint evolution of different meteorological variables measured at the same site), and spatial consistency (comparing values measured at different but spatially nearby sites). Sciuto et al. (2009) developed a procedure to automatically classify daily rainfall data as "validated" or "not validated" before further manual check. This procedure consists in comparing observational data at one station with data at reference stations making the validation by means of a neural network previously trained with historical data which had been already validated. Vicente-Serrano et al. (2010) constructed a database of 828 daily rainfall series for the northeast Spain, with varying coverage, for the period 1901–2002. Data of the initial 3106 observatories were reconstructed, quality controlled (rejecting 0.1% of the values), and homogenized. The quality control applied identified spatial outliers by comparing the rank of each precipitation value with the corresponding rank of the precipitation measured at a nearby observatory. Nie et al. (2012) applied several steps of previously established quality control procedures, namely, outlier filtering, duplication station check, internal consistency check, extreme value check, temporal consistency check, and spatial consistency check. Recently, Serrano-Notivoli et al. (2017) developed an automated quality control for daily precipitation data which compares each daily value with a set of two predictions obtained from reference series (one of them is a binomial prediction showing the probability of dry or wet day; the other is a magnitude prediction that estimates the amount of precipitation for the given day).

Even if there is not an established quality control procedure, most quality checks tend to start by cleaning big errors before facing doubtful or more difficult cases (as mentioned by Einfalt et al. (2008), it is an approach that yields good results) in order to gain increasing quality. Regardless of the methodology, it should be noticed that the dataset can never be completely free from errors and that rainfall variability plays an important role in the trickiness of quality control procedures, especially in the case of daily data.

The Meteorological Service of Catalonia (SMC) is a public company attached to the Department of Territory and Sustainability of the Government of Catalonia, which is responsible for managing meteorological observations and prediction systems in Catalonia. This organization works to recover and store all meteorological data generated by the networks of official weather stations in Catalonia during its instrumental history, since the end of the eighteenth century. These data have been managed by several different organizations over the years. In this way, the set of initially available data is a mixture where several methods of measurement were involved, with different protocols of quality control and metadata coding, if any were applied. New data is continually incorporated into the SMC database, either by new measures (almost in real time) or by historical data recovery; data rescue as well as reconstruction and homogenization of emblematic and representative series of a wide area of the Catalonian territory are tasks that concern the Area of Climatology of the SMC, while the quality controls of the new incorporations are daily operations of the Quality Control Team.

The quality control of weather data is a necessity and a responsibility of the meteorological services given that the assurance of the good quality of stored meteorological data is extremely important for meteorological and climatic studies that could be performed nowadays or to be used in future studies that might require long series of quality controlled data. Besides some automatic control procedures, the operative method used by the SMC since 2009 to perform daily Quality Control is based primarily on closely tracking each episode of precipitation by an experienced technical team for an eventual manual intervention in order to guarantee, for instance, a 100% of completeness in the daily and subdaily series. However, this method cannot be applied to all historical data since it uses, for instance, radar images, not available before the 1990s. Nevertheless, the quality labelling of the historical data is an important goal to achieve. Due to the highly irregular rainfall behavior in Catalonia, the estimation of individual values to eventually complete series will only be possible after performing a detailed study, different for each case, using all the available metadata.

In this way, and after important tasks of data rescue have been accomplished in recent years, the volume of data and metadata that was dispersed as a result of a very convulsive and heterogeneous meteorological past has been significantly increased. Now, the SMC database contains over 10^7 daily rainfall values and a tailored quality control procedure is needed in order to approach a global vision of the quality of the database as well as singling out the best series for specific studies. The set of historically available daily precipitation in Catalonia has recently undergone the quality control which is presented in this paper. It is aimed at trimming gross errors and assessing the quality of the existing series as well as obtaining a quality controlled dataset with uniform procedures that the SMC can subsequently benefit from.

The designed methodology of quality control of daily precipitation presented in this paper has been inspired in the previously reviewed methods but has been tailored to the needs of the SMC data, so an innovative relative control has been incorporated based on the idea of giving maximum credibility to the quality of the data collected between 2009 and 2014 by the automatic meteorological stations of the XEMA network managed by the SMC, which was designed to represent properly the anisotropy that precipitation presents in the territory due to the orography and the meteorological situations originating it. According to the policy of the SMC, the objective was to improve the quality of the database in successive approaches that could be implemented at different phases and produce results at each step. Therefore, the first phase is designed to trim physically impossible data and manually check the few most exceptional cases and correct them if necessary. The second phase consists in increasing the knowledge of the quality of available series in the database. Finally, a third phase consisting in a spatial comparison of daily values is applied in an automated way with adjusted parameters that represent the rainfall climatology of the territory (obtained from a spatially dense network of already validated series of automatic weather stations, although only available in recent years).

The quality controlled dataset is currently being used, for instance, in a project aimed at obtaining maximum expected precipitation and intensity-duration-frequency relationships (Casas et al. 2007; Rodríguez-Solà et al. 2017; Casas-Castillo et al. 2018) at high resolution over Catalonia, which requires a reliable database of daily precipitation observations with a strict quality control.

2 Historical sources and available data

The precipitation database contains daily measures taken at various points located within the territory of Catalonia and, even, in the proximity of its borders (Andorra or Aragon) which are useful in relative quality controls, i.e., those comparing daily rainfall values with measures at nearby stations for the same day. These data come from official weather networks of manual measurements (like the Meteorological Network of Catalonia and the Balearic Islands created in 1894) or automatic weather stations (like the ones managed under the Spanish Automatic Hydrologic Data Collection System (SAIH) which started to operate in Catalonia in 1996). The database contains a total of 1726 daily rainfall measurement sites, each one of them uniquely corresponding to what has been called as a meteorological station. Any change in the location of a measurement instrument has given rise to a different meteorological station, managed under a different and unique code. Table 1 shows the number of stations for every available network used in this work, distinguishing between manual weather stations (MWS) and automatic weather stations (AWS) in each case. The average series length and the first year with an operational station have been included also.

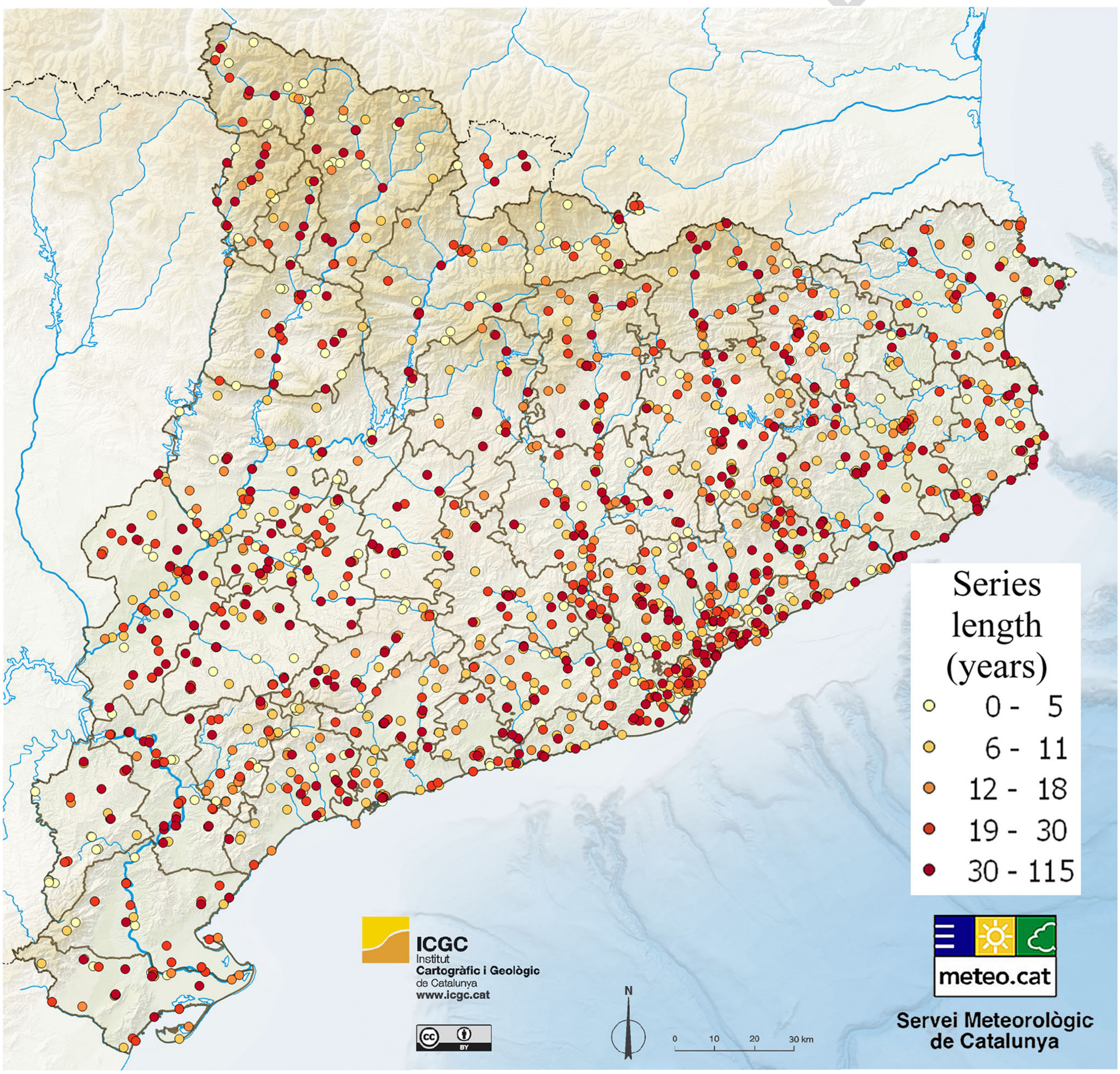
Most available stations were operative before the existence of current meteorological services. In the case of Catalonia, both the Spanish Agency of Meteorology (AEMET) and the SMC are currently managing operative networks and have been rescuing and storing historical meteorological data measured in the study area. SAIH is a network of automatic weather stations that monitor the hydrological basins of Spain and it is organized in a subnetwork for each basin, in the case of Catalonia the oriental basins are managed by a different organization than the occidental basins. XOM is a network of manual observers managed by the SMC and established in 2009 with the objective of expanding the availability of meteorological data and weather surveillance.

The spatial density of measurement points is very high (see Fig. 1), with an observed mean distance between stations of 1.4 km, but most series are short (80% of them are shorter than 30 years), while in a few cases, they are exceptionally long (nearly 3% of the series are longer than 75 years). In any case, it should be noted that given a desired length, it is possible to continue to cover the entire territory in a uniform way, while losing density of stations. For example, by selecting the series of at least 30 years, the observed mean distance between stations is less than 5 km.

Table 1 Number of manual weather stations (MWS) and automatic weather stations (AWS) for every available network, along with the average series length and the first year of operation

Network	MWS	AWS	Total	Average series length (years)	First year with an operational station
AEMET	1112	71	1183	21.8	1855
AEMET–XOM (SMC)	149	–	149	25.5	1901
Andorra	6	1	7	43.6	1926
XEMA (SMC)	–	209	209	15.2	1988
SAIH	–	178	178	12.3	1996
Total	**1267**	**459**	**1726**	**20.4**	

Q4



Q5 **Fig. 1** Location of the meteorological stations with available daily precipitation data in the SMC database, by 2014

The temporal period in which there is availability of daily precipitation data extends from the year 1855 to the present (in this case, the study ends in 2014). Despite the antiquity of the first data, it is not until 1895 that a dozen stations were available simultaneously and, hence, a network was established (Anduaga 2012). The temporal evolution of the number of available series (Fig. 2) presents an overall increasing tendency with a remarkable growth in the decades of 1920s (Prohom 2006) and beginning of 1930s, a drastic decline in the second half of the 1930s (caused by the Spanish Civil War) and a sharp increment in the mid-1990s with the starting of the bulk of automatic stations.

The fact that these meteorological stations were managed by many different organizations has the consequence of initially available data coming from a mixture of different methodologies of measurement (manual and automatic), quality control procedures, and metadata encoding. In this work, the same quality control has been applied to the entire set of data in order to unify criteria.

3 Methodology

As commented in the "Introduction," there is not an established quality control procedure for precipitation data (Branisavljević et al. 2009) and various authors (Shearman 1975; Romero et al. 1998; González-Rouco et al. 2001; Abaurrea et al. 2004; Viney and Bates 2004; Einfalt et al. 2008) use different methods according to the needs and characteristics of their data. The main problem of the quality control of precipitation, especially for daily data, is a great inherent variability, both spatial and temporal, especially in the western Mediterranean region (Romero et al. 1998).

In the present work, the quality control method consists of several steps (Fig. 3) with the object of selecting out the series with the best quality and detecting possible errors, beginning with the more serious ones and progressively enlarging the level of detail. The quality of the database is increased and the selection of series for specific purposes becomes easier once doubtful data is flagged and the quality of each series is estimated by a quality index. Nonetheless, data modification is only done when the correct value from the original source can be checked and removal is only allowed in clear cases of error (e.g., physically impossible data).

In the final step, a relative method is used for comparison of daily values where the issue of variability is dealt with by applying a more lenient threshold at higher distances; in this way, the method reduces the number of erroneously flagged values caused by local episodes. Furthermore, the whole quality control process can be adjusted and verified in our study

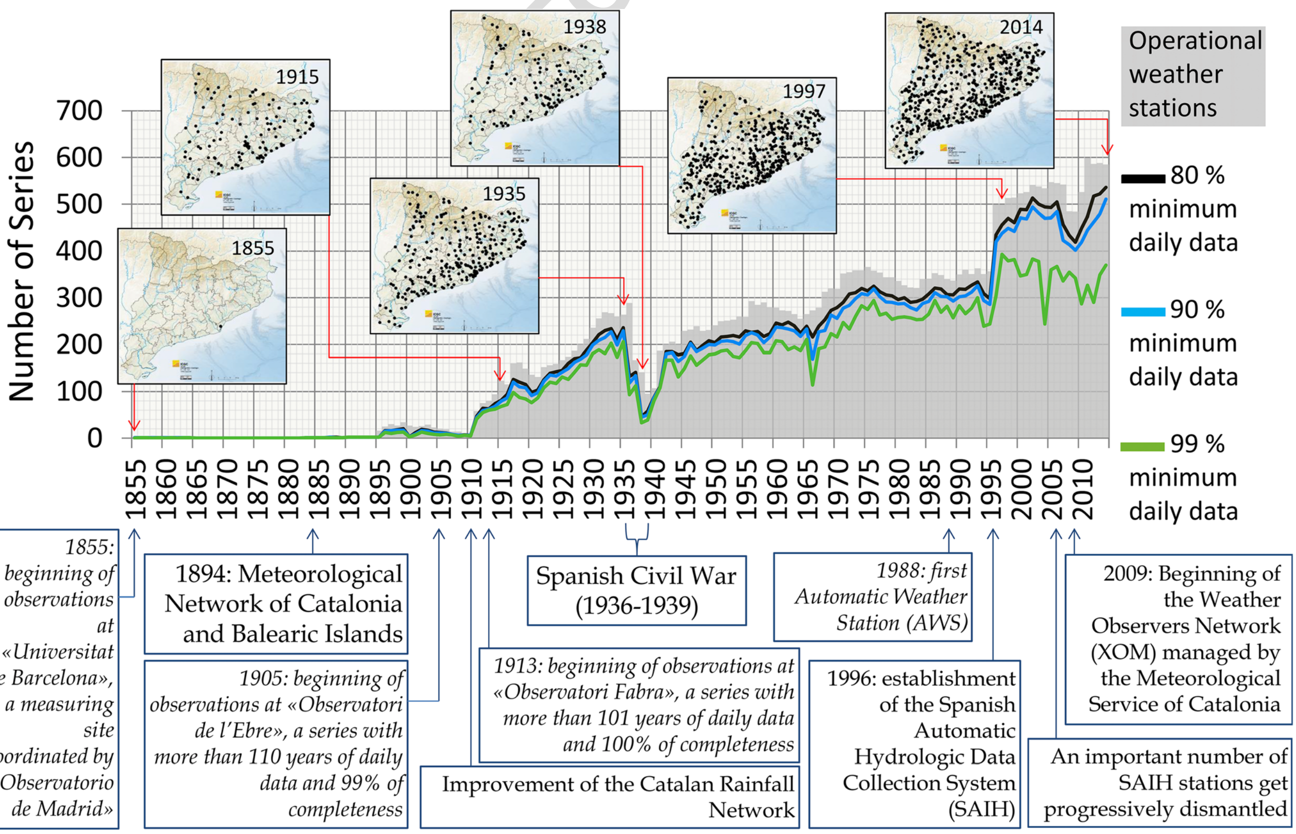


Fig. 2 Temporal evolution of meteorological stations of Catalonia from 1855 to the present. Lines represent number of series with a minimum annual percentage of available data

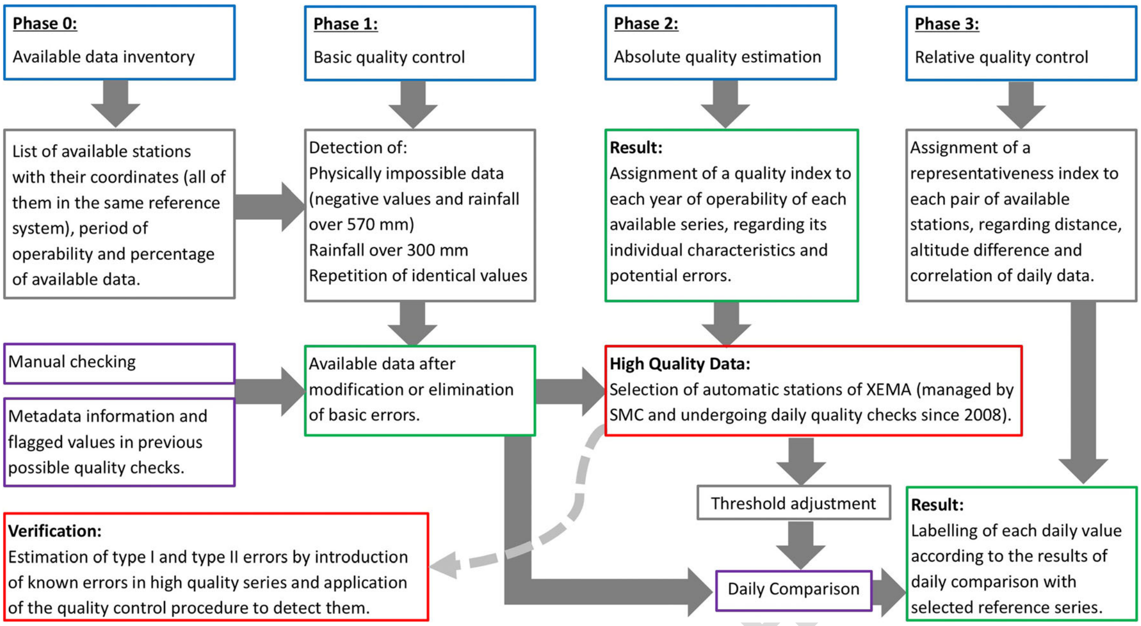


Fig. 3 Quality control protocol scheme

area by taking advantage of a dense network of automatic weather stations whose measurements have been monitored and validated routinely since 2008 by technicians of the SMC.

3.1 Basic quality control

First of all, it is necessary to be sure that the available data are physically possible; in our case, it is necessary to verify that daily precipitation data do not contain negative values or extremely large values. The upper threshold for physically impossible values has been set taking into account the climatology of the study area. The maximum value of the probable maximum precipitation (PMP) in 24 h determined in Catalonia by Casas et al. (2008) was 570 mm, being 430 mm the highest daily rainfall ever recorded in Catalonia, collected on 13 October 1986 in Cadaqués (in the coast, at the northeast) during a flooding episode, comprehensively studied (Vigneau 1987), known as "Aiguat de Sant Eduard." Thus, values above 570 mm incorporated in the database might be errors from the digitization process, values intended to encode metadata and erroneously taken as precipitation values or undetected accumulations.

Within the basic controls, also cases of daily precipitation exceeding 300 mm have been reviewed, analyzing each case with help of graphic representations and consulting, if possible, the original values or the synoptic situation (from reanalysis).

Finally, the last of the basic controls is to detect identical values repeated for two or more consecutive days, using a filter based on the number of consecutive days and the repeated value, in order to manually review the most serious cases (the highest repeated values or the longer spells). Many of these cases are encoding errors, as, for instance, a common old procedure in cases of taking measures after more than 1 day of precipitation was to divide the measured value by the number of days in which the accumulation lasted and to write down the equally distributed value to each one of the days. In these cases, the total precipitation amount can still be used, for instance to calculate monthly precipitation, but it leads to incorrect daily values.

3.2 Absolute quality control

The absolute quality control consists in analyzing the series individually and labelling them, by means of an automatic process, according to their probable quality. With this objective, a quality index has been designed, considering the possible problems that are common in daily precipitation series. This quality index, Q, takes values between 0 and 100 and indicates the absolute quality of the series, so that values above 80 indicate acceptable quality and below 50 very low quality. Q is calculated according to Eq. (1) and its value was assigned to each station for every year of the samples.

$$Q = \frac{1}{5}\left(P + Q_{\text{gaps}} + Q_{\text{zero}}^{m} + Q_{\text{zero}}^{w} + Q_{\text{outliers}}\right) \quad (1)$$

The terms of Eq. (1) are the following:

- P: percentage of annual data, calculated as the number n of available daily data divided by 365 (or 366 for leap years)
- Q_{gaps}: index that takes into account the distribution of gaps, penalizing mostly hollow interspersed with data, more than cases of unique periods without measurement. If n_{gap} is the number of empty days, $L_{\text{gap}}^{\text{max}}$ the length of the

maximum spell with continuous empty days, and n the number of days of the year in which the series was operational, Q_{gaps} is calculated as Eq. (2)

$$Q_{\text{gaps}} = 100 - 100\frac{2n_{\text{gap}} + L_{\text{gap}}^{\text{max}}}{n} \tag{2}$$

- Q_{zero}^{m}: percentage of months with not null accumulated precipitation with respect to the number of complete months during the given year. It aims to penalize series measuring null precipitation for many months, as, in fact, it is very likely to correspond to time periods without measurement that have been filled with zeros (false zeros). If m_0 is the number of complete months with total null precipitation and m the number of complete months of the year for the given series, Q_{zero}^{m} is calculated as Eq. (3):

$$Q_{\text{zero}}^{m} = 100 - 100\frac{m_0}{m} \tag{3}$$

- Q_{zero}^{w}: index that indicates the probability of having systematic cases of accumulations of two or more days which had not been identified. It is based on the fact that the number of precipitation days in a given year should be independent of the day of the week. In the case of detecting a day of the week with rainfall much less often than the rest throughout the year, the value of this index decreases. However, Q_{zero}^{w} is not useful to detect in which moment the accumulation occurred and gives the same treatment to false zeros and to null measures that repeatedly occurred on a specific day of the week. It is calculated from Eq. (4), using the coefficient of variation CV (standard deviation divided by the mean value) of the set n_i (i between 1 and 7, one per every day of the week) that contains the number of days with precipitation equal or higher than 1 mm. It can be calculated as long as the studying year has a minimum of 20 days measuring precipitation equal or higher than 1 mm

$$Q_{\text{zero}}^{w} = 100 - 100\ \text{CV}(n_i) \tag{4}$$

- Q_{outliers}: index related to the proportion of days in which the threshold of outliers is not exceeded with respect to the total number of days of the given year. This threshold is calculated in absolute terms for each month, i.e., it does not depend on the given year, using daily precipitation data equal to or greater than 1 mm and three times the interquartile range above the third quartile

Figure 4 shows the values of the global index Q (averaged over the whole period of availability for each series) calculated for all the series of this study (1726 in total). A distinction is

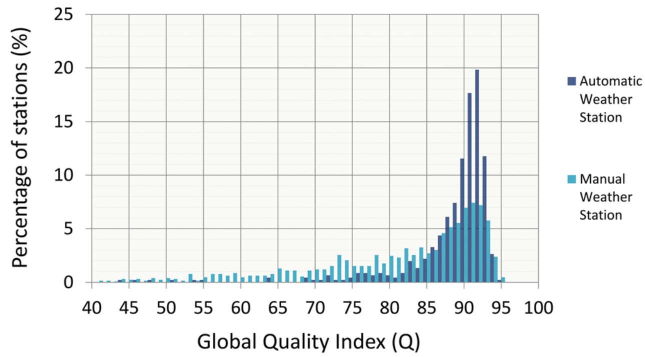


Fig. 4 Quality index values distinguishing between automatic and manual stations

made between manual stations and automatic stations. As expected, the series coming from automatic stations have shown better global quality indexes than manual stations (which are much more numerous and usually longer, so they have potentially more problems). The average value of this global quality index in the whole set of considered stations is 81.7, the median is 87.1, and the first and third quartiles are, respectively, 77.3 and 90.6; we remind the reader that a value of $Q = 80$ was considered, following the definition of the index, as the minimum allowed index for the selection of series according to its absolute quality.

3.3 Relative quality control

This step consists in comparing values measured at a station with those indicated by auxiliary stations available. Apart from the choice of reference series, a source of problems in relative controls at daily level is the comparison between values that are shifted, thus having a correct value but it being assigned to a wrong date (usually shifts of 1 day). The problem of the shifted day is a common source of errors in daily rainfall data basically caused by the need for a convention regarding the date at which rainfall should be assigned when it is measured in the morning of 1 day but has been collected during the previous 24 h (which mainly correspond to the previous day). It is difficult to address this sort of problem: one option is to compare daily values with the corresponding day of the reference series and also with the values shifted 1 day. Another option to address the issue is to obtain the outcome of the quality check as if no shift was present and then go over the cases of sequentially flagged values in order to determine if they correspond to a shifted period. The second course of action is the one we have chosen to avoid an additional complication of the automated procedure, although the results have been difficult to interpret in some cases. For instance, when short periods of shifted values were observed in coincidence with similar values not flagged as errors in the shifted series.

In order to select auxiliary stations suitable for comparison with the candidate station, an index is designed to classify their adequacy or representativeness R according to their spatial coordinates (considering distance and difference of altitude) and the correlation between measured data. This representativeness index is assigned to each couple "candidate station"–"possible auxiliary." Representativeness R is calculated using the expression (5), where d is the distance (in km) between the two stations, h is the difference in altitude (in m), and C_{corr} is Pearson's correlation coefficient between the daily data of the candidate station and those of the auxiliary station. The distance term takes values between 0 and 1 linearly decreasing with distance up to the maximum considered radius (50 km). The altitude term takes also values between 0 and 1 but, in this case, it decreases exponentially with altitude difference increase so that there is no upper limit to permitted altitude difference even if, at the same time, small differences are considered more alike than greater differences which are highly penalized with a term approaching the null value. Regarding the correlation term, in the cases in which the value of C_{corr} is negative, a null value is assigned to this coefficient, as well as in the cases in which the candidate station and the auxiliary do not have a minimum of 25 days in which both have values of nonzero precipitation. Apart from the correlation term which ensures reference series with a similar rainfall pattern, the terms of distance and difference of altitude have also been considered necessary because of the orography of the territory. Indeed, in the study area, great differences of altitude may be present at short distances while this distinction in the terrain characteristics (such as altitude or distance to the coastline) might trigger different drivers for meteorological conditions.

$$R = \frac{100}{3}\left(\frac{50-d}{50} + 0.5^{h/500} + C_{corr}\right) \quad (5)$$

In order to label daily values of the candidate station after performing a daily comparison with reference series, monthly thresholds are established from high quality data and are applied to determine the outcome of the quality control unless the conditions to some special cases are met. At the end of the procedure, daily values will be labelled as valid, "V," doubtful, "D," invalid, "N," or "I" (insufficient information).

Any candidate value with less than three auxiliary values fit for comparison will be assigned label "I." More than one auxiliary is necessary in order to judge if a potential error is present in the measure of the candidate or if it is a fault in the auxiliary. More than two auxiliaries are desirable in order to obtain a solid global outcome based on the average of individual labels assigned after comparison of the candidate with each auxiliary.

Some special cases are labelled directly when the necessary requirements are fulfilled. Only in case of having more than seven auxiliary stations a daily value can be directly labelled according to Table 2 or considered valid with more than 15 auxiliary stations if it falls between first and third quartile of reference values.

In the general case (that is, having at least three auxiliaries and not fulfilling the conditions to be directly labelled as a special case following Table 2), candidate values are given an individual label, L, after the comparison with each auxiliary and afterwards these individual labels are averaged in order to obtain the global label which is the definitive outcome of the quality control for the considered value. The specifics of the final labelling are detailed below, following Eqs. (6), (7), and (8).

Label for each pair candidate–auxiliary can be "V" (valid) or "N" (invalid). It is obtained using a previously set threshold, T, calculated following Eq. (6) which depends on the month of the considered daily measure and the representativeness, R (Eq. (5)), between the candidate and the auxiliary.

$$T = C_m \, ln(101-R) \quad (6)$$

where C_m is a coefficient which depends on the month (see Fig. 4).

T has a lower limit zero ($ln(1)$) when representativeness R is 100 (equivalent to candidate and auxiliary being the same station, i.e., at the same location and with data perfectly correlated) and grows logarithmically with increasing distinction between candidate and auxiliary (i.e., increasing distance, altitude difference and reduction in data correlation).

A difference (Dif) used for comparison (calculated with the measured value at the candidate station, PPT_c, and the measured value at the auxiliary station, PPT_{aux}) is, in fact, a relative difference (divided by the average of both values) and scaled multiplying by the maximum, following Eq. (7).

$$\mathrm{Dif} = \frac{|PPT_c - PPT_{aux}|}{\mathrm{Av}(PPT_c, \; PPT_{aux})} \mathrm{Max}(PPT_c, PPT_{aux}) \quad (7)$$

where PPT is the daily value of the candidate (c) or auxiliary (aux).

In case the difference between the measured values of the pair is under the threshold, the outcome of the individual comparison is label "V," while cases in which the difference exceeds the threshold, the outcome is individual label "N."

In order to achieve a final label "V" after relative comparison, the candidate station must obtain more than 50% in a weighted arithmetic mean, W_m, (see Eq. (8)) of the individual labels of the pairs candidate–auxiliary. In case of obtaining between 20 and 50% of the weighted arithmetic mean, label "D" will be assigned to the candidate value and if less than 20% is obtained, the value will be labelled as "N."

$$W_m = 100 \, \frac{\sum_{Aux}(R-R_{min})^2 \, L}{\sum_{Aux}(R-R_{min})^2} \quad (8)$$

Table 2 Labeling of special cases

Candidate daily value	Reference values of 5 mm or less		Reference values under 1 mm	Label
0 mm	Under 20%	or	Under 1%	N (false 0)
0 mm	Over 90%	Or	Over 75%	V (correct 0)
Over 15 mm	Over 99%	And	Over 90%	N (isolated value)

Where $L = \begin{cases} 1 & \text{if individual label is V} \\ 0 & \text{if individual label is N} \end{cases}$ and $R_{\min}$ is the minimum representativeness allowed, set to 70.

In the general case, where the outcome label is assigned depending on the results of daily comparison between the candidate station and auxiliaries and following Eqs. (6), (7), and (8), the monthly coefficient C_{m} is needed to determine the threshold. Coefficient C_{m} is displayed in Fig. 4 and was established after a study based on daily precipitation values of 150 stations of the Automatic Weather Stations Network of the SMC (XEMA) during the period 2008–2016 (when these stations have undergone quality checks on a daily basis). The threshold has been determined using the representativeness R between stations and organizing the daily relative difference of each pair by months. For each month, the value of 10 times the interquartile range above the 95% percentile of the set of points (i.e., relative differences of each pair) with the same R was found. Subsequently, the obtained values were fitted logarithmically, according to their R, to find the equation later used as threshold. These equations have the form of Eq. (6), where the coefficient C_{m} depends on the considered month. Figure 5 shows the C_{m} coefficients found for each month. Figure 6 shows two examples of observed differences, points obtained at 10 times the interquartile range above the 95% percentile and the adjusted threshold functions for those months.

The empirically obtained coefficients C_{m} show a seasonal cycle with high values in the summer months and low values in winter (Fig. 5). The highest values correspond to months when summer storms are likely to occur (that is from mid-summer to early autumn, i.e., July to September); these storms are characterized by short episodes of heavy rain which yield an extremely irregular pattern on surface measuring stations with potentially high rainfall amounts at some points and, hence, cause high differences in measures taken even at close locations. On the other hand, winter months (December to February) are characterized by episodes of continuous uniform rain which yield homogenous spatial patterns with low differences in measures taken at distant locations. In between, we find months corresponding to spring and late autumn which can produce both kinds of rainfall episodes but tend to present widespread rain and usually coincide with the rainiest seasons in the region of our study.

4 Verification of the quality control

The daily rainfall quality control performed on the set of available stations in the SMC database has been verified by introducing controlled errors and analyzing the obtained results. The quality control method can be verified taking advantage of the availability of data from automatic stations incorporated into the XEMA network managed by the SMC. Data measured at these stations is subject to a daily manual validation by the SMC technicians that has been carried out since 2008 and up to now.

In order to perform a verification of the automatic quality control method, the results obtained in 160 stations across the territory (see Figure 7) have been analyzed after introduction of controlled errors in the data. It is performed by means of a count of false positives (type I error) and false negatives (type II error). Labels produced by the method in the days of the 160

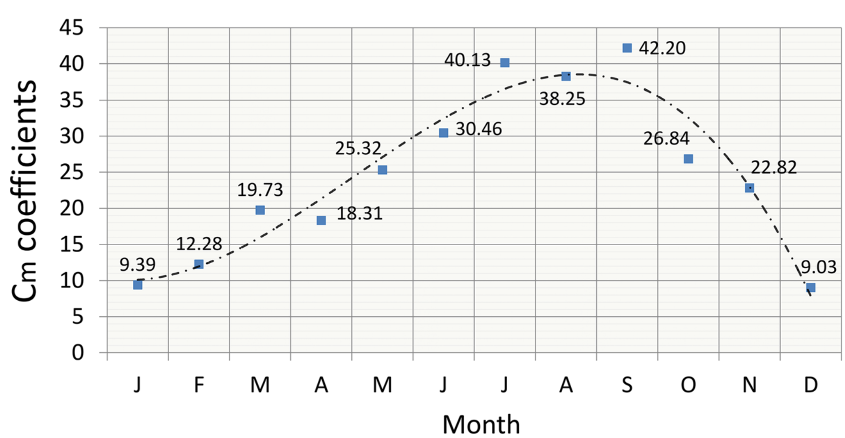


Fig. 5 Coefficients C_m of the threshold's function (depending on the month). Black dashed line is a third-order polynomial tendency line displayed to aid visualization

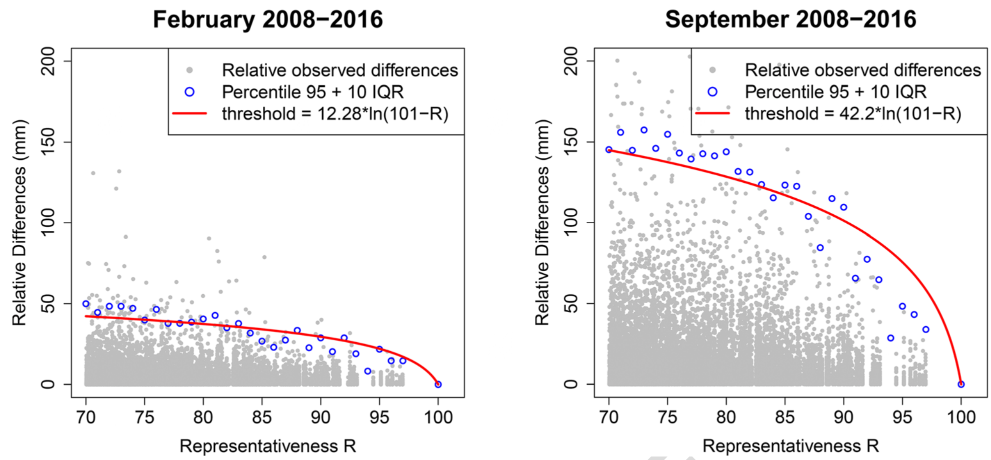


Fig. 6 Threshold functions P95+ 10 IQR fitting for February and September

selected control stations, without any data modification involved, account for type I errors. The introduction of controlled modifications accounts for type II errors. This modifications are introduced by using a multiplicative factor, different at each execution, on each value over 0 mm at one series at a time (i.e., data from the candidate station is modified while auxiliary stations maintain their original values).

In our case, type I errors are cases in which the method labels a correct value as invalid; it can be calculated as a percentage of "N" labelled values over the total number of cases that the method is capable of labelling, considering the control series (which are known to be correct). A total number of 465,551 cases have been considered in order to calculate type I error; it is the number of days that the relative quality control is capable of labelling (considering the 160 control series). The verification result indicates that the type I error committed in the method is on average 0.7%, for values under 50 mm there are a 0.1% of false positives, at most 1.5% for rainfall values under 100 mm, and under 10% for rainfall values under 150 mm; greater amounts have not been considered as these quantities have seldom been measured by the analyzed valid stations (in fact, only 63 cases are measured between 100 and 150 mm).

Type II errors, in our case, are undetected errors in available data, i.e., cases which the method labels as valid but are actually invalid. Given the fact that actual errors in data are uncontrolled and it is precisely what we are trying to flag, the analysis of this type of error produced by the method is performed by introducing controlled errors in valid data, that is, in the subset of control series that are labelled as valid by the method. These controlled errors have been introduced by multiplying the measured value by a factor and the results have been analyzed depending on the rainfall amount indicated by the original value and the magnitude of the introduced error. However, type II error depends on the magnitude of the introduced error (i.e., for errors that yield a value which is very similar to the original measure, failing to reject the error is not a deficiency of the method) and the amount of actual rainfall (i.e., for small amounts, the capacity to distinguish between original measures and values modified by a multiplicative factor is limited by rainfall's inherent variability). In this way, type II error is reduced as the magnitude of the error grows as well as with higher amounts of rain.

Results of type II error obtained on the verification are displayed in Fig. 8, where it can be seen, for instance, that errors of five times the correct value for rainfall under 20 mm are not detected in 16% of the cases whereas a similar type II error is found for errors of twice the correct value in the case of rainfall up to 100 mm.

The method becomes optimal when both types of errors are minimized. Knowledge of the results of this verification ensures that, in the results obtained through the quality control method, data labelling as "invalid" is done with great security (since type I error is very low). On the other hand, data labelled as "valid" is better ensured for high amounts rather than for low values. Therefore, this methodology is suitable for quality controlling extreme values (above 50 mm) yielding type II error (failing to reject invalid data) under 16% when data is twice the correct value.

5 Results and discussion after the quality control

The outcome of the quality control procedure has yield a label for each daily value of the whole set of available series in the database of study. Table 3 summarizes the number of analyzed cases and the obtained labels after the quality control procedure was applied.

After the performed quality control, it is observed that an important part of the days (25.1%) has not been labelled according to their quality because there is not enough information to reach a decision and has been labelled as "I." Regarding the number of series (instead of the total daily values), nearly 5% of the initially available 1726 series have a minimum operating period of 15 years over which it has not

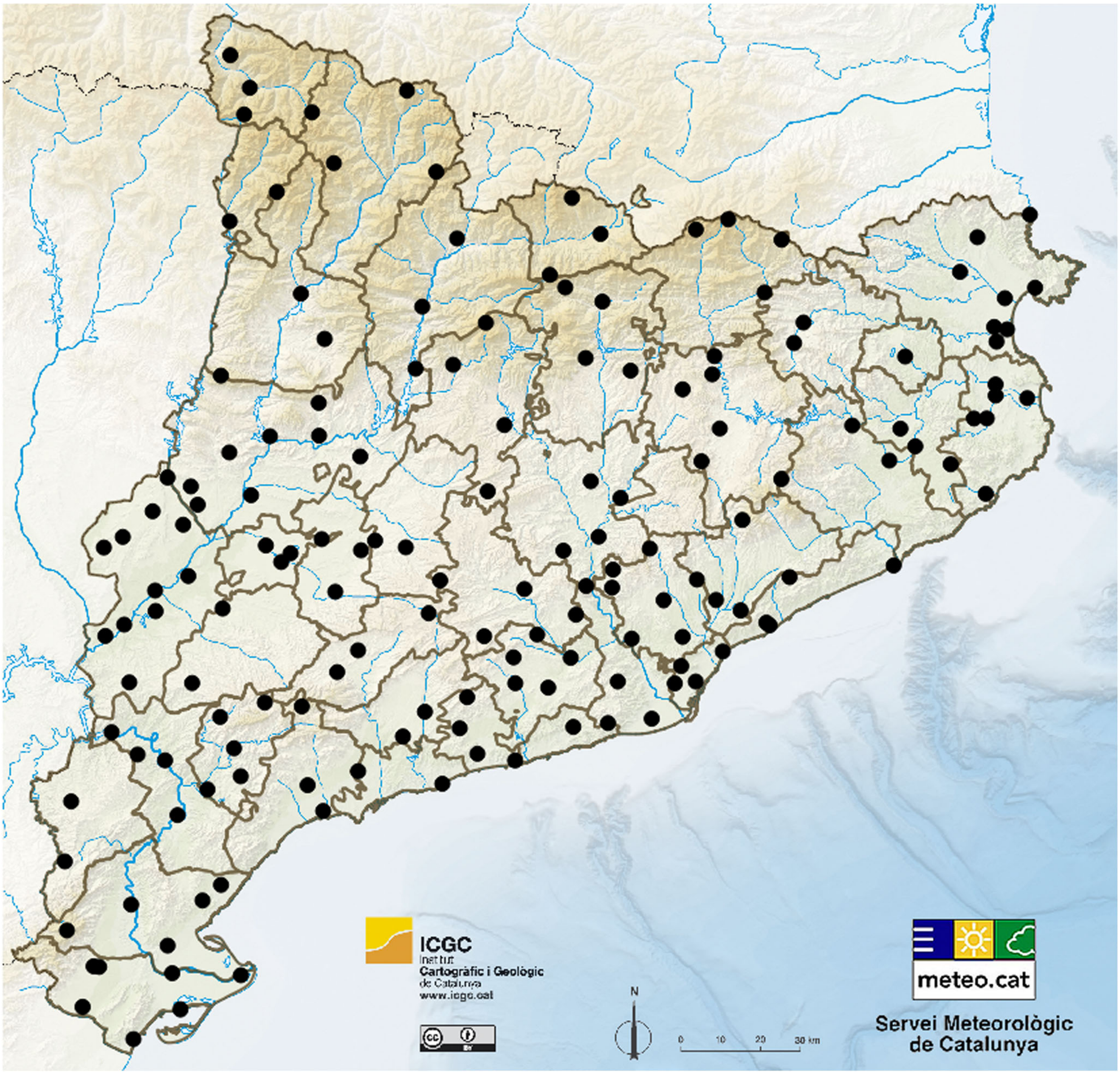


Fig. 7 Distribution of the 160 test automatic stations XEMA used to verify the quality control process

been possible to perform the relative comparison at any time. This is a weakness of the method triggered by low spatial density of measurements because the automatic procedure cannot label values without a minimum number of neighboring stations that comply with some requirements. However, certainty in labelling has been prioritized over forcing to produce an outcome.

The percentage of values labelled as "N" (invalid) that we found (0.2%) agree, for instance, with the rejected values (0.1%) in the study of Vicente-Serrano et al. (2010) in the northeast of Spain or with the flagged and removed values (1.58%) in the study of Serrano-Notivoli et al. (2017) in the island of Majorca (Spain). Although these studies were applied in different areas with a different data volume, a similar percentage of invalid data was expected since the origin of the data is the same.

Regarding the subset of days in which it is possible to perform the comparison, the percentage of data labelled as valid is a 99.6%. These results indicate an outstanding quality of the initial database. Even if a type II error of up to 16% might be committed, the percentage of valid data in the initial database would be between 83.6 and 99.6%.

Apart from the verification process presented in the previous subsection, a manual checking of some cases has been performed in order to further validate the automatic labelling. This analysis has revealed that an important part of correctly Q7

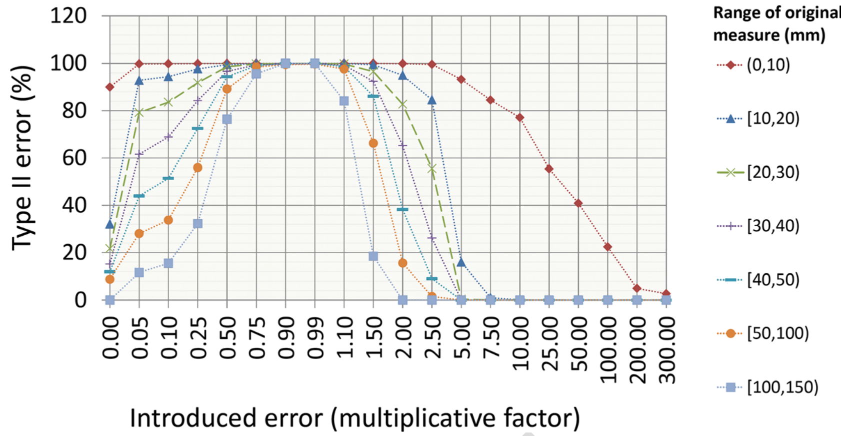


Fig. 8 Type II error obtained on the verification procedure depending on the introduced error and the rain amount of the original measure

detected errors is values with a problem on the day at which the value is assigned, specifically shifts of 1 day. Some of these cases had already been detected at early stages of the planning of the quality control procedure but we chose to overlook them due to the complication associated with approaching the detection of this sort of error.

Another check that has been made is to verify that no spatial pattern of the global index *Q* can be detected, i.e., the results of the quality control carried out seem not to depend on the different precipitation regimes that exist in the studied territory.

Furthermore, an analysis of the obtained results is performed based on the number of available series per year which achieve a minimum quality. It is our proposal to consider as good quality series the ones with a minimum annual index of 85 and a valid percentage of data in the automatic control of at least 95% with respect to data for which the comparative is possible to be made. However, series with acceptable quality would have a minimum annual index of 80 (as defined when this absolute quality index was designed) and detected errors up to 10% could be allowed as it is close to type I error for high amount of collected rainfall (over 150 mm). On the other hand, a stricter quality level has also been considered (90 minimum annual quality index and under 1% of flagged values). Apart from the quality, which applies to measured data that has undergone the quality control procedure, the completeness of data in the series should be considered. As it happens often with daily precipitation, an absolute completeness of daily data might be required, although, depending on the characteristics of the study, a lenient minimum can sometimes be applied. Finally, an important issue to be considered at the selection of series is the length of the series, in this case, the number of years with complete and high quality daily data.

Figure 9 shows the temporal evolution of the number of series that annually comply with the three aforementioned levels of quality (acceptable, good, and excellent quality) as well as the total available series. The temporary evolution of the availability and quality of the series can be analyzed in three broad sections. In the first section (from the beginning and until the mid-1970s), the number of available stations as well as the number of high quality series increase progressively, except for the period corresponding to the Civil War, where both the quality of the series and the number of available measurement points dramatically decrease. In the second section (approximately between 1974 and 1995), the number of available stations is stabilized and so does the number of high quality stations (being the series of excellent quality the more variable with time but being in numbers around 170 stations) so that the percentage of stations that annually exceeds the predetermined excellent quality level is approximately 49%, whereas good quality stations are 77% of total available series. Finally, in the most recent period (as of 1996), there is a sudden increase in the number of available stations caused by the proliferation of automatic stations. It should be mentioned that in the recent period (since mid-1990s), the percentage of detected errors has been reduced progressively even if the number of series that pass the established minimums of quality does not reflect it as it is strongly influenced by the completeness of the series (at the beginning, automatic stations could often present problems that avoided strict completeness of the annual data and later, once the automatic stations were well established, the completeness of the series diminished more often in the case of manual stations).

Table 3 Number of obtained labels for each category

Daily values	Valid cases (label "V")	Doubtful cases (label "D")	Detected errors (label "N")	Cases without enough information (label "I")
11,079,524	8,261,141 (74.6%)	11,823 (0.1%)	24,653 (0.2%)	2,781,907 (25.1%)

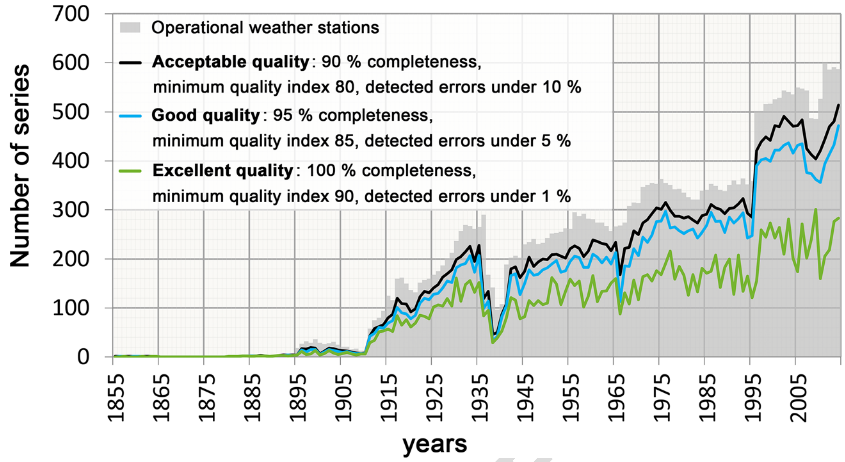


Fig. 9 Temporal evolution of the number of available stations and selected high quality stations

Figure 10 shows the number of series regarding the length of years with high quality data. It can be observed, as expected, that for shorter required length, the number of series is higher. It is interesting to observe the percentages of series that fulfill these requirements against the total number of available series of the same length; in this way, a 47% of the series that had at least 10 years of operability (a total of 1110 series) have excellent quality and 78% good quality; regarding the series with length of at least 30 years (a total of 347 series), 32% of them have excellent quality while 66% pass the level of good quality.

6 Conclusions

In the present study, a quality control procedure for daily precipitation has been presented. The methodology consists in approaching higher degrees of quality by trimming the database of errors and, finally, labelling each daily value as valid, doubtful, or invalid according to the results of a comparison with suitable good quality reference series. The methodology is mostly automatic, although a manual revision is recommended at some point.

The quality control procedure has been designed after adjustment of some newly defined indexes and thresholds from the behavior of high quality data existing in the territory where it has been implemented.

The presented quality control procedure has been applied to the whole set of daily precipitation measures available in the database of the Meteorological Service of Catalonia. This dataset contains more than 10^7 data from 1726 official weather networks that have taken measures in the territory since the beginning of its instrumental history and covers the period from 1855 up to now. In this paper, we have presented and analyzed the obtained results after a thorough description of the method.

A verification of the procedure has been performed taking advantage of a manually quality controlled network present in the territory. The verification results show that on average a 0.7% of the values could be incorrectly labelled as invalid data (type I errors or false positives) and this type of error would be of 10% at the most for daily rainfall values up to 150 mm. On the other hand, the method might commit errors in failing to reject invalid data (type II error or false negative) in, at the most, 16% of the cases of rainfall over 50 mm when the reported value is twice the actual rainfall. Regarding the analysis of type II errors, the

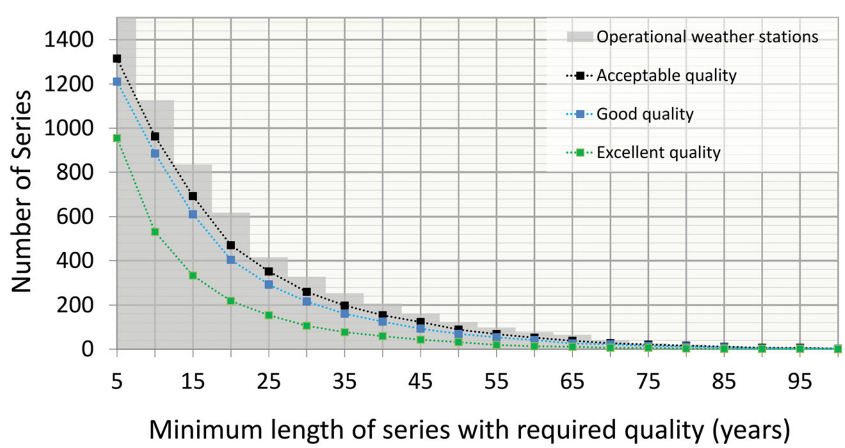


Fig. 10 Number of series with a minimum length of years with acceptable, good, or excellent quality

method gets better at quality controlling extremes of precipitation as the identification of erroneous values performs better at high amounts.

After the application of the designed quality control to the whole database, the number of days labelled as valid, taking into account the verification results, shows that between 83.6 and 99.6% of the values in the database are correct.

It has been verified that no spatial pattern has been detected for the global index *Q*, which means that the results found seem not to depend on the diverse climatic characteristics of the study area.

Furthermore, a selection of the high quality series is achieved. Three levels of quality have been defined, namely, acceptable, good, and excellent quality. The temporal evolution of number of series according to these quality levels is studied in order to achieve a global vision of the quality of the initial database.

The results obtained after applying the quality levels that we propose show that 47% of the series that had at least 10 years of operability (a total of 1110 series) have excellent quality and 78% good quality; regarding the series with length of at least 30 years (a total of 347 series), 32% of them have excellent quality while 66% pass the level of good quality.

A manual checking after the automatic quality control has revealed that an important part of detected errors is values with a problem on the day at which the value is assigned, specifically shifts of 1 day. Some of these cases had already been detected at early stages of the planning of the quality control procedure, but we chose to overlook them due to the complication associated with approaching the detection of this sort of error.

The newly presented quality control procedure was designed according to the needs of the SMC daily rainfall database. It has been tested using high quality series and controlled errors. After it has been applied to the over 10^7 available values, an overall vision of the quality of the database is achieved and daily rainfall series can be selected according to their quality and the requirements of further specific studies.

Acknowledgements The authors thank the Servei Meteorològic de Catalunya (SMC) for providing rainfall data used in this study. The authors would like to thank the editor and the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

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

References

Abaurrea J, Asín J, Cebrián AC, Centelles A (2004) Metodología para el control de calidad y homogeneidad de una base de datos de precipitación diaria. In García JC, Diego C, Fernández de Arróyabe P, Garmendia C and Rasilla D (ed) El Clima entre el Mar y la Montaña. Asociación Española de Climatología y Universidad de Cantabria, Santander, A(4) pp 431-440. ISBN 84-8102-384-1 Q8

Anduaga A (2012) Meteorología, ideología y sociedad en la España contemporánea. Estudios sobre la Ciencia, 61. Consejo Superior de Investigaciones Científicas, Madrid, 450 pp + XVII plates. ISBN: 978-84-00-09421-8

Branisavljević N, Prodanović D, Arsić M, Simić Z, Borota J (2009) Hydro-meteorological data quality assurance and improvement. J Serbian Soc Comput Mech 3(1):228–249

Casas MC, Herrero M, Ninyerola M, Pons X, Rodríguez R, Rius A, Redaño A (2007) Analysis and objective mapping of extreme daily rainfall in Catalonia. Int J Climatol 27(3):399–409. https://doi.org/10.1002/joc.1402

Casas MC, Rodríguez R, Nieto R, Redaño A (2008) The estimation of probable maximum precipitation: the case of Catalonia. Ann N Y Acad Sci 1146:291–302. https://doi.org/10.1196/annals.1446.003

Casas-Castillo MC, Llabrés-Brustenga A, Rius A, Rodríguez-Solà R, Navarro X (2018) A single scaling parameter as a first approximation to describe the rainfall pattern of a place: application on Catalonia. Acta Geophys 66(3):415–424. https://doi.org/10.1007/s11600-018-0122-5.http://rdcu.be/HdU3

Chen M, Xie P, Shi W, Silva V, Kousky V, Higgins W, Janowiak JE (2008) Quality control of daily precipitation reports at NOAA/CPC. AMS 12th conferences on IOAS-AOLS, no 33, 21 January 2008

Einfalt T, Gerlach N, Podlasly C, Demuth N (2008) Rainfall and climate data quality control. 11th International Conference on Urban Drainage, Edinburgh, Scotland, UK, 31/08–5/09, 2008

Feng S, Hu Q, Qian W (2004) Quality control of daily meteorological data in China 1951-2000: a new dataset. Int J Climatol 24:853–870. https://doi.org/10.1002/joc.1047

González-Rouco JF, Jiménez JL, Quesada V, Valero F (2001) Quality control and homogeneity of precipitation data in the southwest of Europe. J Clim 14:964–978. https://doi.org/10.1175/1520-0442(2001)014<0964:QCAHOP>2.0.CO;2

Nie SP, Luo Y, Li WP, Wu TW, Shi XL, Wang ZZ (2012) Quality control and analysis of global gauge-based daily precipitation dataset from 1980 to 2009. Adv Clim Chang Res 3(1):45–53. https://doi.org/10.3724/SP.J.1248.2012.00045

Prohom M (2006) La contribución de la Sociedad Astronómica de Barcelona en la difusión de las observaciones meteorológicas en Catalunya (1910-1923). Investig Geogr 40:141–155. https://doi.org/10.14198/INGEO2006.40.07

Rodríguez-Solà R, Casas-Castillo MC, Navarro X, Redaño A (2017) A study of the scaling properties of rainfall in Spain and its appropriateness to generate intensity-duration-frequency curves from daily records. Int J Climatol 37(2):770–780. https://doi.org/10.1002/joc.4738 http://hdl.handle.net/2117/87312

Romero R, Guijarro JA, Ramis C, Alonso S (1998) A 30-year (1964-1993) daily rainfall data base for the Spanish Mediterranean regions: first exploratory study. Int J Climatol 18:541–560. https://doi.org/10.1002/(SICI)1097-0088(199804)18:5<541::AID-JOC270>3.0.CO;2-N

Sciuto G, Bonaccorso B, Cancelliere A, Rossi G (2009) Quality control of daily rainfall data with neural networks. J Hydrol 364:13–22. https://doi.org/10.1016/j.jhydrol.2008.10.008

Serrano-Notivoli R, de Luis M, Saz MA, Beguería S (2017) Spatially based reconstruction of daily precipitation instrumental data series. Clim Res 73(3):167–186. https://doi.org/10.3354/cr01476

Shearman RJ (1975) Computer quality control of daily and monthly rainfall data. Meteorol Mag 104:102–108

Vicente-Serrano SM, Beguería S, López-Moreno JI, García-Vera MA, Stepanek P (2010) A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. Int J Climatol 30(8):1146–1163. https://doi.org/10.1002/joc.1850

Vigneau JP (1987) 1986 dans les Pyrénées orientales: deux perturbations méditerranéennes aux effets remarquables. Revue géographique des Pyrénées et du Sud-Ouest. Sud-Ouest Européen 58-1:23–54. https://doi.org/10.3406/rgpso.1987.4969

Viney NR, Bates BC (2004) It never rains on Sunday: the prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. Int J Climatol 24:1171–1192. https://doi.org/10.1002/joc.1053

WMO (2008) Guide to meteorological instruments and methods of observation. World Meteorological Organization. Weather Climate Water (WMO-No. 8, the CIMO Guide)

You J, Hubbard KG, Nadarajah S, Kunkel KE (2007) Performance of quality assurance procedures on daily precipitation. J Atmos Ocean Technol 24:821–834. https://doi.org/10.1175/JTECH2002.1