Revision of

“Rescue and quality control of sub-daily meteorological data collected at Montevergine Observatory (Southern Apennines), 1884–1963”

V. Capozzi, Y. Cotroneo, P. Castagno, C. De Vivo, G. Budillon

RC (X) = Referee comment (number X)
AR (X) = Authors’ reply (number X)

Referee #2 - Alba Gilabert Gallart

RC: General impression: For climate analysis (past, present and future) it is essential to rescue past instrumental data. These data rescue projects involve a great effort and a very rigorous work. The paper and dataset presented contribute to recovery a significant amount of sub-daily data and especially data from the 19th century. They have done a great job of digitizing data and recovering some metadata. The paper has an impact on the field. It has a high significance in this scientific field (climatological data rescue) and is within journal scope. I would recommend acceptance of this paper after a minor revision.

AR: Dear Ms. Gilabert Gallart, we are grateful for your positive evaluation of our study. We are glad to clarify the open questions and modify the paper accordingly to your recommendations.

RC (1): Section 2. Materials, data and methods: In an easier way it allows the reader to figure out the characteristics of the dataset and the methodology followed.
Some comments: -Subsection 1: as important as data rescue is metadata rescue, metadata recovered is clear and give an image of the characteristics of observations through the time. It could be great if the authors can add some historical image.

AR (1): We have added some historical pictures of Montevergine Observatory in the Appendix A (Figure A1, see page 12 of this document). We have also included a recent panoramic image of Montevergine Abbey, to highlight that Montevergine Observatory is surrounded by a natural high-altitude environment, whose features have remained unchanged over time.

RC (2): Subsection 2: well explained and correct methodology. Only one comment, about ELP (lines 247-255) is required a reference about methodology applied and if it’s possible add a schematic.

AR (2): In the revised version of the manuscript, we have added three references about snow to equivalent liquid water conversion (Winiger, 2005; Egli, 2008; Egli et al., 2009). Moreover, to better explain the strategy adopted to determine the equivalent liquid precipitation (ELP) parameter, we have produced a new figure (Fig. 4 in the revised version of the manuscript, see page 7 of this document). This figure shows an adapted extract of the rescued MVOBS dataset (available on NOAA’s NCEI repository, Capozzi et al., 2019) focusing on the precipitation measurements collected between 28 and 30 January 1956. In addition, the different scenarios involving an estimation of ELP discussed in the main text are illustrated with specific reference to the sub-daily data recorded in this time segment.

RC (3) Quality control of digitized data - Lines 278-283: improve the way to flag the QC results, specially considering that not all variables pass all the tests. It is necessary to have a clear identificatory to flag (correct, suspicious and wrong values).

AR (3): We are grateful to the referee for this comment, which allow us to better clarify the meaning of the different QC flags considered in our study. The QC labels and the related description are listed below:

- bad data (QC = 8), i.e. data that did not satisfy the gross error test;
- suspicious data (QC = 9), i.e. data that did not passed the manual inspection or that did not satisfy the inter-variable check;
- good data, lower quality level (QC = 1), i.e. data that passed only gross error test
- good data, medium quality level (QC = 2), i.e. data that passed gross error and tolerance tests;
- good data, higher quality level (QC = 3), i.e. data that passed all statistical tests.

According to this classification of QC flags, data that have passed at least one objective statistical check are defined as "good" and are associated to a quality level (ranging from low to high) that is a function of the number of statistical tests satisfied. We have better stressed this concept in the main text. Moreover, we have improved figure 5 of the new manuscript, which offers a schematic diagram of the quality-control procedure (see page 8 of this document), and table 3 that summarizes the results of quality control tests (see page 5). In table 3 caption, we have specified that cloud cover parameter did not undergo tolerance and temporal coherence tests. It should be noticed that cloud amount was estimated by visual observations using a fixed reference scale. Due to the specific nature of this parameter and its strong hour-to-hour variability, it is not possible to define climatological limits for outlier and anomalous jumps detection. Therefore, quality control for cloud cover includes only manual inspection and gross error test and it aims to assess the data plausibility and their consistency with other related meteorological parameters, such as cloud type and, when available, low-level clouds base height and quantity. In addition, it should be noted that, since rainfall and snowfall data are time-integrated values, they were not analyzed in terms of plausible rate of change and therefore for those parameters the highest quality flag is QC = 2.

RC (4) Section 4. Application examples of MVOBS sub-daily dataset: This section tries to give value to the dataset rescued, but I'm not in favour to keep this type of sections. On one hand, I think that for a specialist on this topic is obvious the value of the work done and on the other hand, (especially for the second part) they are trying to do a "climatological analysis" with potential inhomogeneous data (metadata reveals different potential breakpoints). So: -Maybe I'm in agree to keep the first part but, clearly stating that data used is not subjected to any homogenisation procedure and metadata indicates potential breakpoints. -I consider that is better to delete the second part.

AR (4): In our opinion, section 4.2 can be of interest to many readers because it provides concrete evidence on the possible applications of Montevergine sub-daily data in climatological studies. Therefore, we have decided to leave this section. The aim of section 4.2 is to emphasize the potential use of multi-parametric Montevergine time series to analyze "less-studied" atmospheric variables, whose past climate variability is largely unknown especially in Mediterranean region. The hail events frequency of occurrence, showed in section 4.2, is a relevant example: in many regions, such as Italy (Baldi, 2014), the scarcity of historical information does not allow to build a solid long-term climatology for this parameter. This can also be considered valid for the snowfall amount, which is a very important parameter for mountain environment, also from a hydrological perspective.
We acknowledge that a single isolated time-series (although with many distinguish features, such as Montevergine) can only give a partial contribution to the climate reconstruction. However, we hope

that our effort may be an incentive for future initiatives aimed at rescuing historical sub-daily multi-parametric time series.
To conclude, this section has only “illustrative purposes”: it is not our intention to perform a climatologic analysis, which requires homogenized data as rightly highlighted by referee.

RC (5) Table 3: Give information about what means QC =1, QC = 2. . . Due to the way to name the flags, for cloud cover, rainfall and snowfall is not clear if 100 / 99.9 % of values are suspicious.

AR (5): According to referee’s suggestion, in the first row of Table 3 we have added a description for each QC flag:

- bad data (QC = 8);
- suspicious data (QC = 9);
- good data, lower quality level (QC = 1), i.e. data that passed only gross error test
- good data, medium quality level (QC = 2), i.e. data that passed gross error and tolerance tests;
- good data, higher quality level (QC = 3), i.e. data that passed all statistical tests.

As explained in the reply to the RC (3), cloud cover, rainfall and snowfall data did not undergo the entire QC procedure. Cloud cover was checked only with manual inspection and gross error test (i.e, the maximum quality level is QC=1), whereas rainfall and snowfall data quality was evaluated through manual inspection, gross error and tolerance tests (i.e. the maximum quality level is QC=2). The results obtained for these parameters (100% of cloud cover data were flagged as QC=1 and 99.9% of rainfall and snowfall data were labelled as QC=2) should be interpreted as very encouraging signs about their reliability. When other sub-daily time series collected in Southern Italy will become available, an additional quality control assessment will be performed as a mean of spatial consistency check.

RC (6) Figure 1. About the map, please add a reference on the repositories consulted.

AR (6): In the previous manuscript version, caption of Fig. 1 had two references about the consulted repositories (Ashcroft et al., 2018; Compo et al., 2019).
We have added a further reference (see page 6 of this document) to the interactive tool of The International Surface Pressure Databank version 4 (rda.ucar.edu/datasets/ds132.2/index.html?sstn=17606&spart=exact#stationViewer), that allows to search for available atmospheric pressure sub-daily data according to the period and to the geographical area of interest.

RC (7) Figure 4: The diagram is fine a clear, but the way to flag the results need to be improved (see comments above).

AR (7): We have modified the diagram of this figure (Fig. 5 in the new version of the manuscript, see page 8 of this document) according to the referee suggestions in reply to RC (3).

RC (8) Figure 6: review the dots. According the graph and the text there are not blue dots.

AR (8): In Figure 6 (now labelled Fig. 7 in the new version of the manuscript) we show an application of QC procedure to relative humidity sub-daily observations collected in March 1901, where no data are flagged as QC = 1 (i.e. there aren’t outliers). Therefore, we have deleted the phrase “blue dots (outliers, QC=1)” from the caption (see page 8).

RC (9) Figure 7: higher percentages needs to be well explained and maybe some visual information to explain this.

AR (9): We have added a discussion about the results showed by figure 7, which is now numbered as Fig. 8 in the new manuscript version. The higher percentages found in some of the sub-periods (1919-1923, 1924-1928, 1944-1948, and 1949-1953) are related to the very high number of observations flagged as QC=9 (i.e. suspicious data) after visual inspection. In these time segments, data quality was affected by some impairments in thermo-psychrometric measurements, mainly caused by human errors. A brief discussion on such issues was supplied in section 3.1.
We have provided additional details by adding two panels to Fig. 8, labelled (b) and (c), where we show the frequency distribution of dry-bulb temperature measurements in 1919-1925 and 1948-1950 period respectively (see page 9 of this document). In Figure 8b, it can be clearly seen that even temperature values have an absolute frequency much higher than odd ones. Whereas, histogram in Fig. 8c, which only shows temperature records between 10 and 20°C, highlights an anomalous high frequency in the integer temperature values recorded.

RC (10) Figure 8: add a comment about data was only submitted to a QC not to a homogenization procedure.

AR (10): We modified the caption of this figure (now labelled Fig. 9 in the new version of the manuscript) by adding the following sentence (see page 10 of this document): “It should be noticed that such data were subject to a quality control procedure that did not include the homogenization”.

RC (11) Figure 9: I’m not sure that variability is only due to natural evolution. Needs to consider deleting this part.

AR (11): We partially agree with the reviewer. We are aware that artefacts caused by the potential inhomogeneity revealed by metadata probably undermine the climatic signal presented in this figure. Therefore, at this stage, it is not possible to achieve conclusions about climatic variability and trends from Montevergine sub-daily data. This aspect has been highlighted in the new manuscript version.
However, as stated in the reply to RC (4), we feel appropriate leaving this section in our manuscript, because it shows, from a qualitative perspective, some possible future applications of Montevergine sub-daily records in climate fields, with a particular emphasis on some meteorological parameters whose historical variability is largely unknown.

List of cited references

Ashcroft, L., Coll, J.R., Gilabert, A., Domonkos, P., Brunet, M., Aguilar, E., Castella, M., Sigro, J., Harris, I., Unden, P., and Jones, P.: A rescued dataset of sub-daily meteorological observations for Europe and the southern Mediterranean region, 1877-2012, Earth Syst. Sci. Data, 10, 1613-1635, https://doi.org/10.5194/essd-10-1613-2018, 2018.

Baldi, M., Ciardini, V., Dalu, J.D., Filippis, T.D., Maracchi, G., and Dalu, G.: Hail occurrence in Italy: towards a national database and climatology, Atmos. Res., 138, 268–277, https://doi.org/10.1016/j.atmosres.2013.11.012, 2014.

Capozzi, V., Cotroneo, Y., Castagno, P., De Vivo, C., Komar, A., Guariglia, R., Budillon,: Sub-daily meteorological data collected at Montevergine Observatory (Southern Apennines), Italy from 1884-01-01 to 1963-12-31 (NCEI

Accession 0205785). NOAA National Centers for Environmental Information. Dataset. https://doi.org/10.25921/cx3g-rj98, 2019.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis

Egli, L.: Spatial variability of new snow amounts derived from a dense network of Alpine automatic stations, Annals of Glaciology, 49, 51-55, https://doi.org/10.3189/172756408787814843, 2008.

Egli, L., Jonas, T., and Meister, R.: Comparison of different automatic methods for estimating snow water equivalent, Cold Regions Science and Technology, 57, Issues 2–3, 107-115, https://doi.org/10.1016/j.coldregions.2009.02.008, 2009.

The International Surfance Pressure Databank version 4, Interactive Station Viewer: https://rda.ucar.edu/datasets/ds132.2/index.html?sstn=17606&spart=exact#stationViewer, last access: 29 January 2020.

Winiger, M., Gumpert, M., and Yamout, H.: Karakorum–Hindukush–western Himalaya: assessing high-altitude water resources, Hydrol. Process. 19, 2329–2338, https://doi.org/10.1002/hyp.5887, 2005.

List of modified tables according to referee suggestion. We highlighted our changes in yellow.

Table 3. Results of quality control tests applied to MVOBS sub-daily meteorological data. Each column show the percentage of data flagged as QC = 8, QC =9, QC = 1, QC = 2 and QC = 3. It should be noted that cloud cover data underwent only manual inspection and gross error test, whereas rainfall and snowfall measurements quality was evaluated according to manual inspection, gross error and tolerance tests.

Parameter	**% of QC = 8** **bad data**	**% of QC = 9** **suspicious data**	**% of QC = 1** **good data (lower quality level)**	**% of QC = 2** **good data (medium quality level)**	**% of QC = 3** **good data (higher quality level)**
Dry bulb temperature	0.0	13.0	0.2	0.6	86.2
Wet bulb temperature	0.0	13.1	0.4	0.8	85.7
Atmospheric pressure	0.0	0.0	0.4	1.3	98.3
Vapour pressure	0.4	12.8	0.4	1.1	85.3
Relative humidity	0.3	12.8	0.4	1.6	84.8
Cloud cover	0.0	0.0	100.0	Not applied	Not applied
Rainfall	0.0	0.0	0.1	99.9	Not applied
Snowfall	0.0	0.0	0.1	99.9	Not applied

List of modified and added figures according to referee suggestions. We highlighted our changes in yellow.

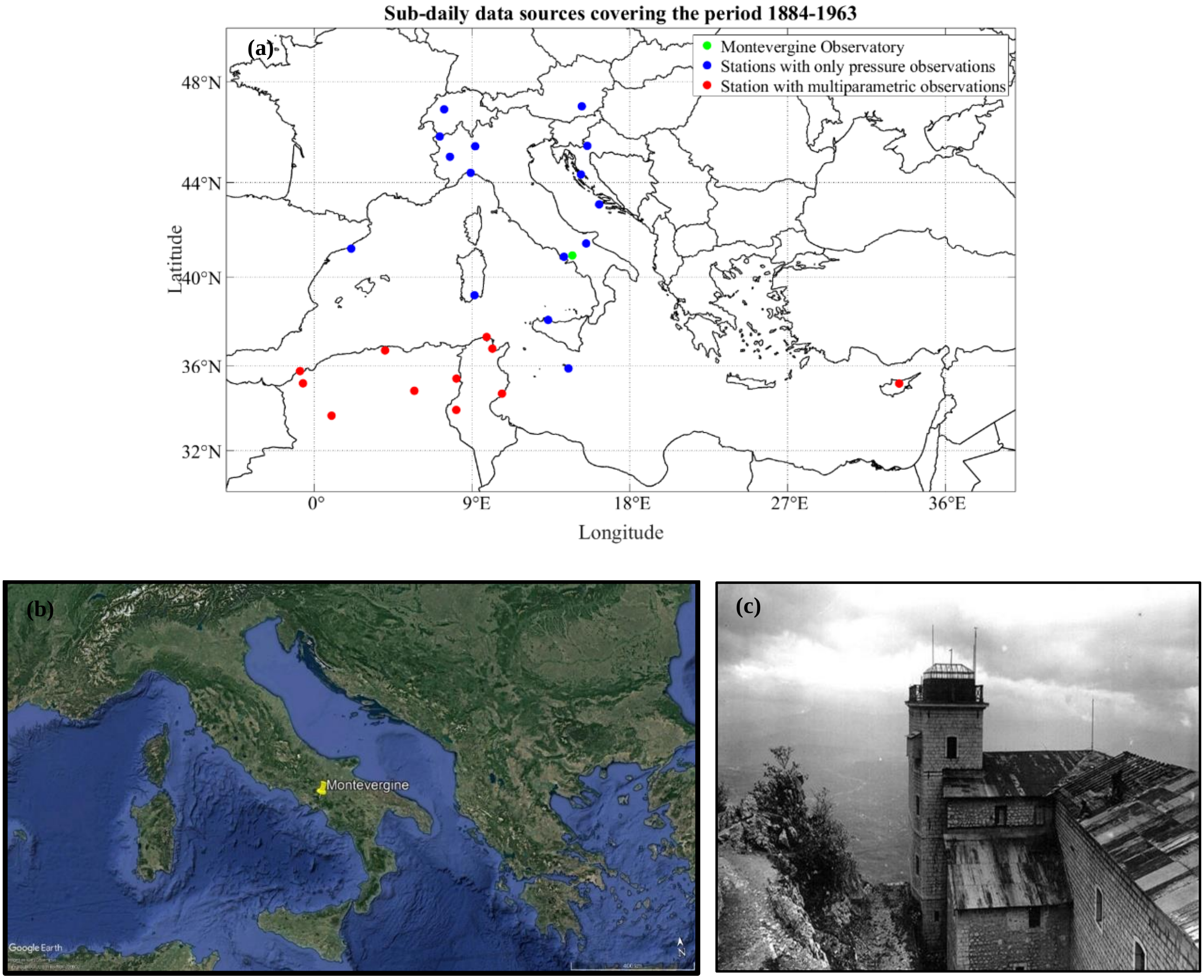
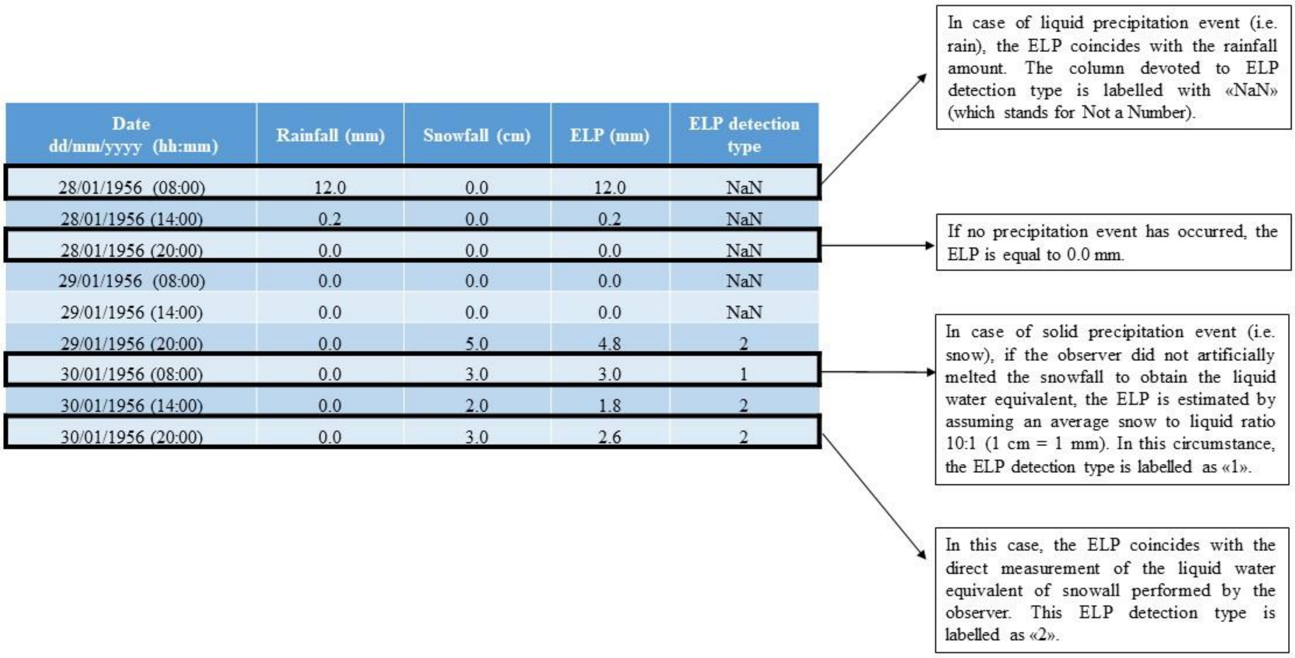


Figure 1: Panel (a): map of Mediterranean region showing the location of sub-daily meteorological data available in digital format for the period 1884-1963. Blue dots represent the stations including only atmospheric pressure measurements, whereas red dots the one for which multiparametric meteorological observations are available. Data sources have been provided by International Surface pressure Databank version 4.7 (Compo et al., 2019; rda.ucar.edu/datasets/ds132.2/index.html?sstn=17606&spart=exact#stationViewer) and by Ashcroft et al. (2018) datasets. Panel (b) shows Central Mediterranean region, including Montevergine location (highlighted as yellow marker). Montevergine position is also marked on panel (as) as green dot. Image credits: © Google Earth, Data Sio, NOAA, U.S. Navy, NGA, GEBCO. Panel (c) presents an old photo of Montevergine Observatory tower, situated near the top of Partenio mountain chain on the north-eastern side of Montevergine Abbey. Image courtesy of Italian Air Force (http://www.meteoam.it/page/montevergine).



Date dd/mm/yyyy (hh:mm)	Rainfall (mm)	Snowfall (cm)	ELP (mm)	ELP detection type
28/01/1956 (08:00)	12.0	0.0	12.0	NaN
28/01/1956 (14:00)	0.2	0.0	0.2	NaN
28/01/1956 (20:00)	0.0	0.0	0.0	NaN
29/01/1956 (08:00)	0.0	0.0	0.0	NaN
29/01/1956 (14:00)	0.0	0.0	0.0	NaN
29/01/1956 (20:00)	0.0	5.0	4.8	2
30/01/1956 (08:00)	0.0	3.0	3.0	1
30/01/1956 (14:00)	0.0	2.0	1.8	2
30/01/1956 (20:00)	0.0	3.0	2.6	2

In case of liquid precipitation event (i.e. rain), the ELP coincides with the rainfall amount. The column devoted to ELP detection type is labelled with «NaN» (which stands for Not a Number).

If no precipitation event has occurred, the ELP is equal to 0.0 mm.

In case of solid precipitation event (i.e. snow), if the observer did not artificially melted the snowfall to obtain the liquid water equivalent, the ELP is estimated by assuming an average snow to liquid ratio 10:1 (1 cm = 1 mm). In this circumstance, the ELP detection type is labelled as «1».

In this case, the ELP coincides with the direct measurement of the liquid water equivalent of snowall performed by the observer. This ELP detection type is labelled as «2».

Figure 4: Summary of the strategy used to assess the Equivalent Liquid Precipitation (ELP) parameter. The left table has been obtained by adapting an extract of the digital version of MVOBS dataset available on NOAA's NCEI repository (Capozzi et al., 2019). It lists the sub-daily precipitation data observed between 28 and 30 January 1956. From left to right, rainfall (mm), snowfall (cm), ELP (mm) and ELP detection type (expressed as numeric or textual label). The rows highlighted in black present different ELP estimation scenarios.

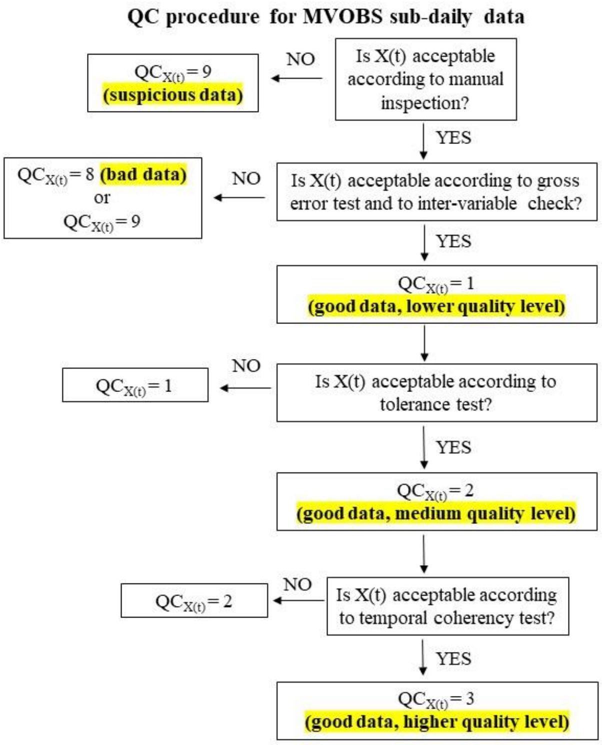


Figure 5: A schematic of the QC strategy developed in this study to check the observation of a determined parameter X collected at the time *t*. It should be highlighted that the cloud cover parameter underwent only gross error test and the temporal coherency test has not been applied to precipitation data (accumulated rainfall and snowfall).

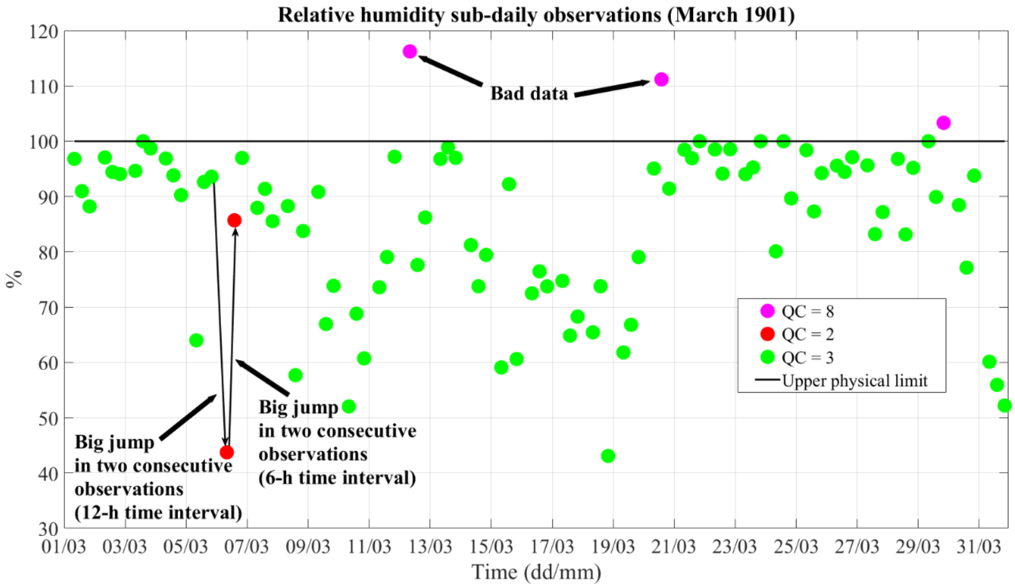


Figure 7: Relative humidity (in %) sub-daily observations collected in March 1901. Each record is color-coded according to its quality flag: bad data (QC=8; magenta), good data with medium quality level (QC=2; red) and good data with higher quality level (QC=3; green). In this example, no good data with lower quality level (QC=1) and suspicious data (QC=9) were detected. Black horizontal line shows upper physical limit.

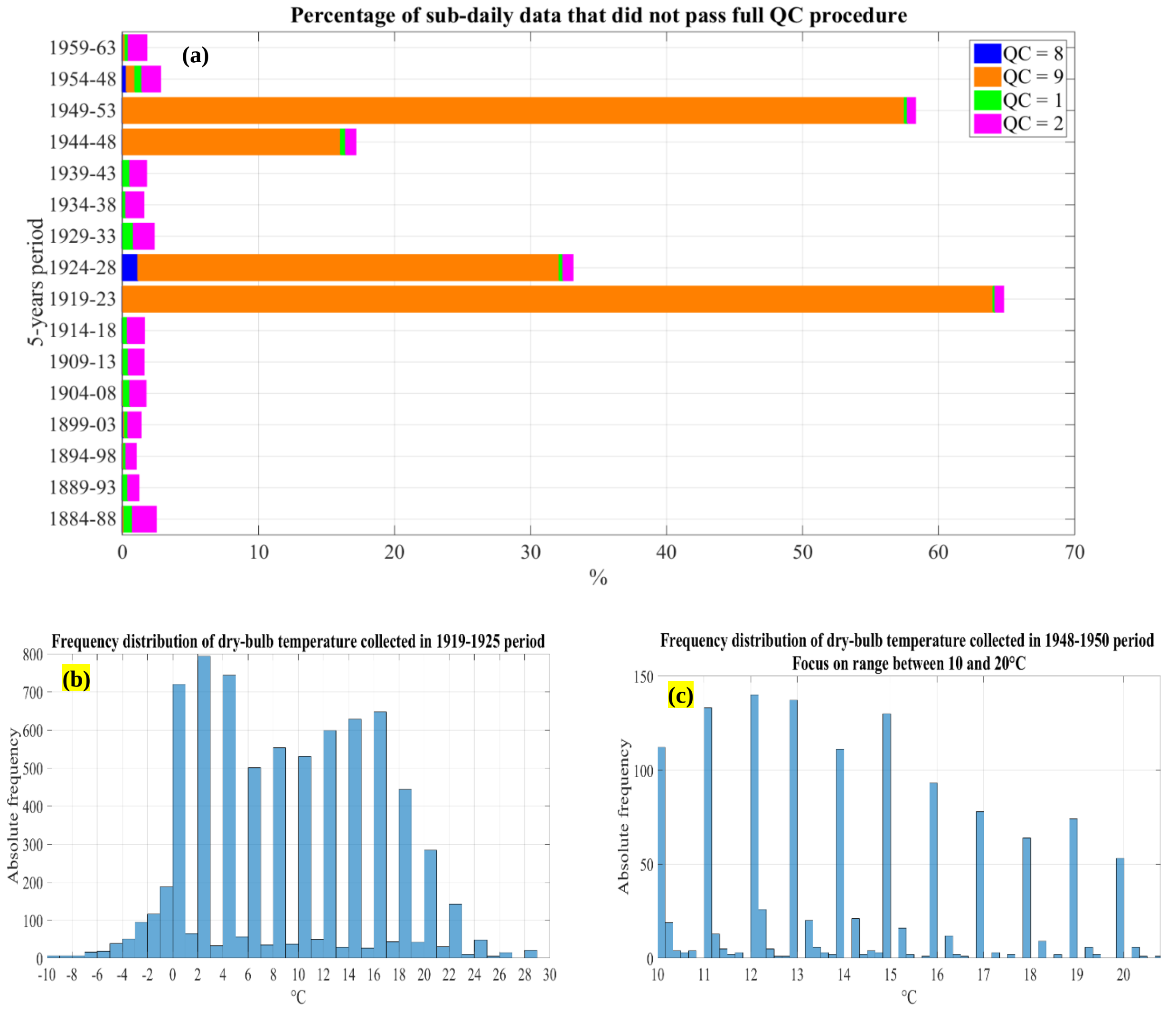


Figure 8: In panel (a), colour coded bars indicate the distribution over time, computed on five-year period, of the percentage of sub-daily data that did not pass the full QC procedure: QC = 8 (blue), QC = 9 (orange), QC = 1 (green) and QC = 2 (magenta). Panel (b) and (c) present the frequency distribution of dry bulb temperature in 1919-1925 and in 1948-1950 period, respectively. It should be noticed that panel (c) only shows temperature values between 10 and 20° C. The bin width is 1.0°C in panel (b), 0.2°C in panel (c).

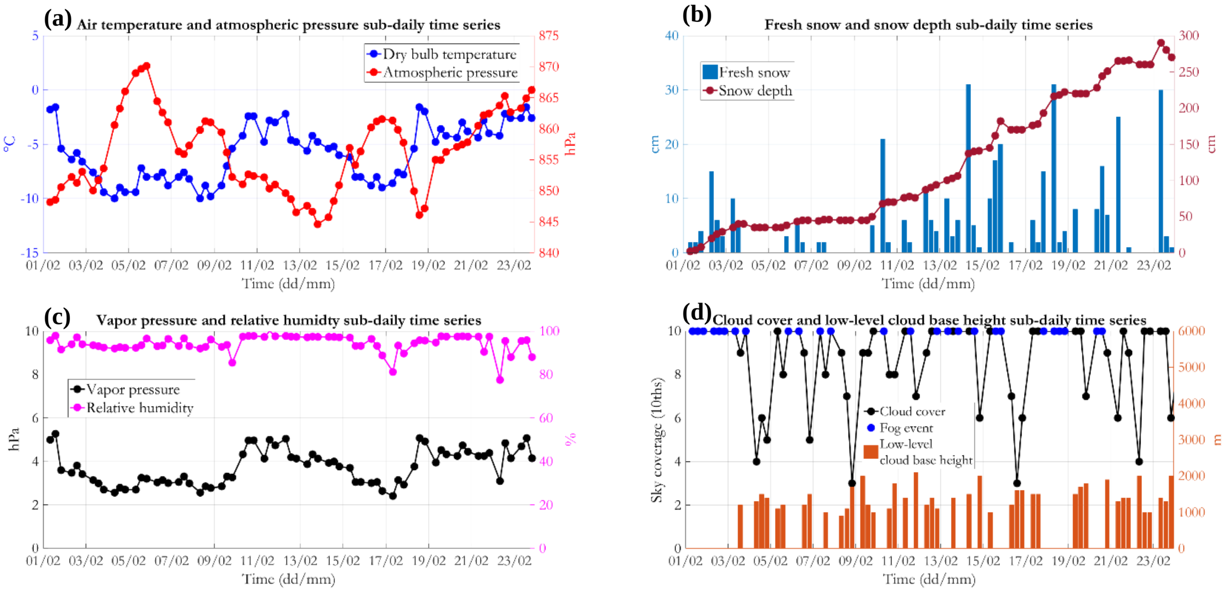


Figure 9: Sub-daily time series of different meteorological parameters observed from February 01 to 23, 1956. Panel (a) shows the dry bulb temperature and atmospheric pressure. Panel (b) presents the fresh snow and snow depth records. Panel (c) shows vapour pressure and relative humidity. Panel (d) plots the cloud cover and low-level cloud base height observations. It should be noticed that such data were subject to a quality control procedure that did not include the homogenization.

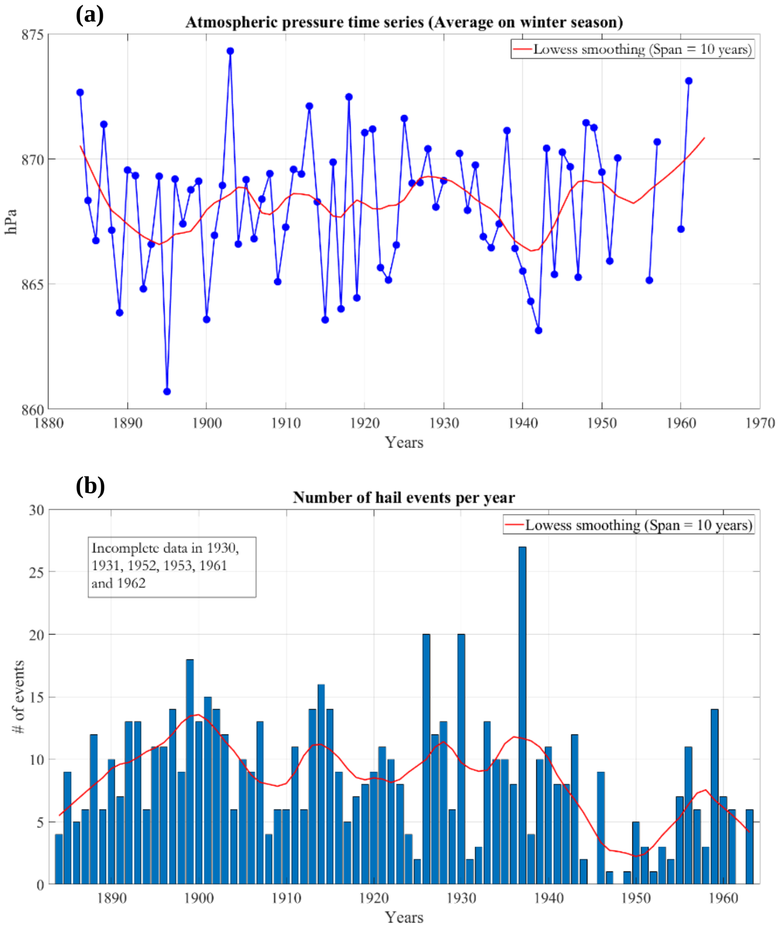


Figure 10: Panel (a): time series (blue line) of winter atmospheric pressure from 1884 to 1963. Each value (blue dot) is the average of the sub-daily observations measured during January, February and March. The red curve is the lowess smoothing filter, computed using a 10 years span. Panel (b): time series (blue vertical bar) of yearly hail events occurred at MVOBS from 1884 to 1963. The hail occurrence has been computed for every year of the investigated period, using the sub-daily observations of precipitation type. The red curve is the lowess smoothing filter, computed using a 10 years span. It should be noticed that such data were subject to a quality control procedure that did not include the homogenization.

Appendix A

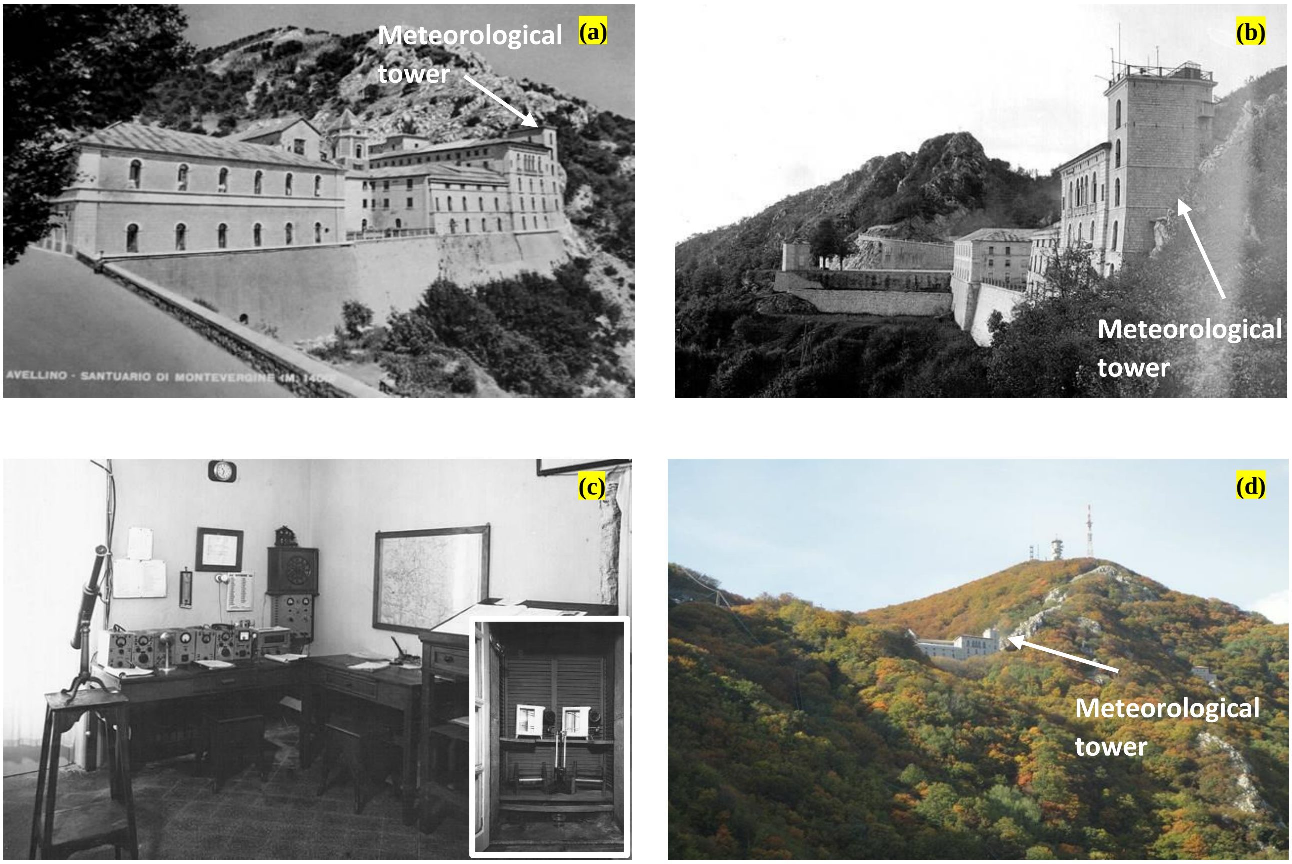


Figure A1: Panel (a) and (b): historical view of MVOBS meteorological tower from a southeastern and a northern direction, respectively. Panel (c): the observatory room in 1950. The small picture in the bottom right corner of the panel shows the inside of Stevenson Screen where thermometric and hygrometric measurements were performed. Panel (d): a recent panoramic view of Montevergine Abbey and MVOBS. Historical and recent images show that MVOBS is surrounded by a natural high-altitude environment, whose features have remained unchanged over time. Photos in panels (a), (b) and (c) are courtesy of the Italian Air Force (www.meteoam.it).