Quality control and gap-filling methods applied to hourly temperature observations over central Italy

Paolina B. Cerlini\(^1\) | Lorenzo Silvestri\(^1,2\) | Miriam Saraceni\(^2\)

\(^1\)CIRIAF-CRC, University of Perugia, Perugia, Italy
\(^2\)Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy

Correspondence
Paolina B. Cerlini, CIRIAF-CRC, University of Perugia, Perugia I-06125, Italy.
Email: paolina.cerlini@unipg.it

Funding information
Programma di sviluppo rurale per l’Umbria 2014/2020, Misura 16 Cooperazone, Sottomisura 16.1 Sostegno per costituzione e gestione Gruppi Operativi dei PEI in materia di produttività/sostenibilità dell’agricoltura.

Abstract
Given the regional surface network of the Umbria region, a mountainous area located in central Italy, the observed hourly temperature time series from 2010 to 2017 were analysed by applying basic and extended quality control procedures following World Meteorological Organization (WMO) standards. The validation procedure consisted of automatic quality control, producing validated data with metadata subsequently recorded in the NetCDF format. After these controls, data were checked manually and an extended procedure was applied to reconstruct the temperature time series for missing data. The spatio-temporal method used to reconstruct the data was a linear interpolation for 1 hr gaps and the empirical orthogonal function (EOF) algorithm for gaps ≥ 2 hr. The introduction of a complete and homogeneous data set of hourly reanalysis ERA5 (from the European Center for Medium-Range Weather Forecasts—ECMWF) allowed for the reconstruction of the longest gaps with statistical and physical consistency. The final product of the study is a continuous station time series of hourly temperatures that will be available to the public by the end of 2020; a daily version of the original time series is already available on the regional website.

KEYWORDS
agrometeorology, empirical orthogonal function, ERA5, missing data, quality control

1 | INTRODUCTION

Meteorological data from ground-based local observation networks are crucial for many applications, ranging from predicting the air quality and weather at a local scale to plant disease prevention and irrigation management, and many other applications requiring maintenance and calibration of the network and the validation of measured values (Begeš et al., 2015).

Usually, these networks are originally built and maintained by local institutions, with very specific goals depending on each institution and network and their history. Each organization has its network, with its processing procedures and its data format, since these networks are generally only partially integrated into the Global Observing System (GOS). In some cases, these networks are not integrated in the GOS whatsoever. Therefore, they are not required to follow the specific quality requirements as imposed by the World Meteorological Organization (WMO 2010b).

In Italy, the importance of these mesoscale meteorological networks and their operational use has been...
recognized and analysed in the literature on both the meteorological and climate scales (e.g. Lanciani and Salvati, 2008; Uboldi et al., 2008; Antolini et al., 2016; Pavan et al., 2018). The local source of meteorological information provided by these networks becomes essential, especially where the GOS network is coarse and the region is characterized by strong orographic forcing.

This is the case of central Italy, and especially of the Umbria region, which is characterized by a complex orography where only one station belongs to the GOS among other local stations that do not comply to a general standard of quality control (QC). Quality procedures are usually applied on daily or monthly processed data by local institutions, but none of the original data fulfils the WMO’s requirements.

In order to gain precious local information, the study submits a first procedure for processing hourly temperature data from the ground-based regional network of the Umbria region.

The applied procedure begins from the raw output of temperature measurements taken from automatic weather stations (AWS) established originally by the Regional Idrographical Service for hydrological purposes. Nowadays, these data are also becoming essential for agrometeorological applications. However, these raw measurements usually contain different types of errors and missing data. Therefore, it is necessary to perform an extended QC (following the WMO’s international standards) together with a gap-filling technique to make the raw data suitable for any application.

The procedure described herein consists of two main steps: validation and reconstruction, including the formatting process of data and metadata, as suggested by the WMO (2010a). In the first part of the study, validation is used to determine whether the data are inaccurate, incomplete, inconsistent or unreasonable (WMO, 2010b, ch. 6.2.2). It contains all the QC checks, such as single-station checks (completeness check or missing data detection, limit checks, temporal and internal consistency) and multiple-station checks (spatial consistency). The validation process was meant to flag data without replacing them.

In the second part, the reconstruction process was used to fix data errors (“data cleaning”) or to fill missing data (“data filling”) identified during the previous step. In this phase, data were modified and reconstructed to fulfill the quality requirements. The original data were kept since the modification introduced by the reconstruction must be reversible. These two steps together with continuous QC monitoring constitute the QC system.

The novelty of this application is the use of the ERA5 data set (ECMWF, 2018) either to fill missing data or to replace wrong data detected through the validation step.

The work was undertaken within the European Union Rural Development project (PSR—Programma di Sviluppo Rurale) for the Umbria region. The final product of the study is continuous station time series of hourly temperatures that will be available to the public by the end of 2020; a daily version of the original time series is already available on the regional website (https://annali.regione.umbria.it/). The main objective of the work is to transform raw and unformatted (text files) meteorological data into a formatted (NetCDF standard format) data set that is spatially and temporally coherent with the observed reality represented by the ERA5, the last generation of the reanalysis described below (ECMWF, 2018).

The paper is structured as follows. After describing the regional network in Section 2, characterizing the stations’ distribution over the orography, a list of the data sets used is added. The QC procedure is then applied to the raw data collected by the network in Section 3.1, which describes the flow chart of the entire procedure and enters the details of validation and reconstruction techniques. The results of all the techniques applied are then presented in Section 4 for validation and reconstruction. A summary of the work with conclusions is presented in Section 5.

2 | DATA

2.1 | Network description

The order of magnitude of the spatial and temporal resolution of a network depends on both the available financial resources and the typical scales of meteorological phenomena to be described. For example, agrometeorological spatial scales, as specified by the WMO (2014), have a characteristic length in the order of 100 m ($L \approx O \{100 \text{ m}\}$), defined as microscale. The typical temporal scales, strictly related to the spatial scales, should be of the order of hours or less ($T \approx O \{\text{hr}\}$), corresponding to the response time of the boundary layer as defined in Stull (1988). These requirements are very difficult to fulfill by a regional network, mainly because reliable automatic weather stations (fulfilling the WMO’s standards for parameters) and their maintenance are very expensive and a micro- or even toposcale (100 m < $L < 3 \text{ km}$) network is too difficult to install and upkeep.

The Umbria region has a mesoscale surface observation system ($3 < L < 100 \text{ km}$). The data analysed herein are temperature time series starting from 2010 to 2017 with an hourly temporal resolution. These data are provided by the Idrographical Regional Service, which is
working with the authors to extend the analysis further. The time range of the temperature time series available is too short for the evaluation of climatic indexes, but it was thought to be long enough to determine the quality of this network by combining it with other climate data, such as the ERA5 reanalysis, which will be described below.

Most of the temperature gauges are located in the proximity of rivers and relevant areas for hydrological purposes. No stations are within forests. Further information is not available on the analysed sensors, although the site inspection and representativeness classification is part of an ongoing project. Nevertheless, the purpose of the study is to give a complete procedure to validate and reconstruct data from stations assumed to be representative.

Umbria is a small region in central Italy characterized by a complex orography distribution (Figure 1c). On the eastern part, the central Apennine mountains (reaching heights of 2,500 masl) make the longitudinal gradient of mean height very high, compared with an inner central region characterized mainly by hills and, as a secondary feature, by the horseshoe-shaped Umbria valley. On the southern part is another important basin in correspondence to the latitudinal mean height minimum, as shown on Figure 1d.

As a whole, Umbria is a hilly region where most of the hills range between 300 and 600 masl. The surface extension is about 8,464.33 km², while the used agricultural area (UAA) is 3,262.39 km², about 40% of the total area, which makes agrometeorological data and applications crucial for this area.

![Figure 1](image-url)
The number of temperature gauges considered in this analysis is 73. Given the short time range available (eight years in total), the station density is constant, and it is about eight temperature gauges and 10 rainfall gauges for 1,000 km². These numbers are comparable with other regional networks such as that for Emilia-Romagna described by Antolini et al. (2016), which is one of the oldest regional agrometeorological networks in Italy following the WMO’s requirements. The average spacing of the stations, together with the main domain parameters, are shown in Table 1.

The station distribution with height is fundamental in order to have a good representation of all the different meteorological conditions, especially in a region with a complex orography (Daly, 2006) as Umbria. From Figure 1a, the number of stations could seem to be very low in the mountainous region compared with the hilly inner part of the region. Looking at the station distribution (Figure 1b), there is good agreement between the network and the digital elevation model, since there are few digital elevation model cells with an altitude > 1,200 masl inside the regional limits. The altitude range that is not well represented by the observations between 800 and 1,200 masl. Indeed, looking at the eastern part of the map in Figure 1c, there is a belt (centred at about 12.9° N) with a mean latitudinal height of about 900 masl with a very low concentration of gauges. The network density and distribution could be an indication for the further installation of new stations, which was one task of the project that has funded the present research.

2.2 | Data sets used

Beside the regional network, the other data sets used in this analysis are as follows:

- ERA5 is the new reanalysis data set of the European Center for Medium-Range Weather Forecasts (ECMWF) in Reading, UK, with hourly time resolution and 30 km grid space resolution (see the ECMWF website; ECMWF, 2018). These data are available for public use at the Copernicus Climate Data Store (CDS) C3S (2017) from 1979 up to three months from the present time. The number of ERA5 grid points contained in the analysed area is 15, then interpolated employing a bi-linear interpolation (interpolation package MIR of the ECMWF; Maciel et al., 2015) to a higher space resolution of 0.125°. This interpolation does not increase the physical information obtained from the model, but it is made to have the analysis points closer to the observations for a better comparison. The distribution of ERA5 grid points with the elevation seen from the model is reported in the top-right corner of Figure 1. No grid points are located below 200 masl, and this is the main reason why ERA5 data cannot be considered accurate enough to represent the local meteorological conditions completely.

- GMTED-2010: the Global Multi-resolution Terrain Elevation Data 2010 (Danielson and Gesch, 2011) is the digital elevation model used in the present analysis. In Italy, it is derived entirely from the shuttle radar topography mission Digital terrain elevation data (Jarvis et al., 2008) and is available at three separate horizontal resolutions. The smallest resolution (7.5″, about 250 m) was chosen for the analysis shown in Figure 1.

3 | METHODS

3.1 | Flow chart

The flow chart of the entire procedure (validation and reconstruction) is shown in Figure 2.

The initial part of the validation was performed on the raw data at the station site (AWS) and will not be treated deeply in the following section, but briefly described. Following the guidelines described by Zahumenský (2004), the raw data (signal samples) were automatically controlled by the control unit present in the station. The products of this step are usually instantaneous data (minute- or hourly-averaged) in common text format. They should be collected and sent to the data-processing centre (DPC) together with their metadata (information describing the sampling methods, instrumental apparatus, site location and geographical location; Aguilar and Llanso, 2003). The importance of metadata is also emphasized by the WMO (2010a). However, agrometeorological or hydrological data are often distributed as text files that rarely contain information about the data themselves. That is why the formatting process has to be considered part of the QC. In the present study, the controlled and processed data were coded in NetCDF format following

### Table 1

| Unit | \(r_{\text{min}}\) | \(r_{\text{max}}\) | \(r_{\text{av}}\) | \(L_{\text{lat}}\) | \(L_{\text{lon}}\) |
|------|----------------|----------------|-------------|-------------|-------------|
| DD   | 0.01           | 1.14           | 0.23        | 1.5         | 1.5         |
| km   | 1.1            | 127            | 25.6        | 167         | 167         |

Note: \(r_{\text{min}}\), \(r_{\text{max}}\) and \(r_{\text{av}}\) are the minimum, maximum and average distances between temperature stations; \(L_{\text{lat}}\) and \(L_{\text{lon}}\) are the latitude and longitude dimensions of the domain size.
the international convention CF 1.7 (Eaton et al., 2003), as suggested by the WMO (2010a, ch. 3).

The other basic aspects of the agrometeorological networks and data that include the fundamental variables to be measured, the layout of a station site, some basic features of the instruments (detailed standard instrumental characteristics such as the resolution, range or measurement units and also data format; the Guide to Meteorological Instruments and Method of Observations (GMIMO) (WMO, 2014) are part of the initial QC, which will be the subject of a subsequent project with the regional authority extending this work to a longer record. Indeed, the
first basic QC also includes regular inspections and maintenance to the site, which represents one of the weak points of the Italian agrometeorological networks (Rete Rurale Nazionale 2004–2020 (RRN), 2017).

The automatic steps of the basic and extended QC with the manual check (the only non-automatic check) constitute the validation procedure applied here, and they will be described below. In contrast to all the other steps, the manual check is crucial in order to understand if the automatic tests have performed properly, and if all the values for the tests have been properly chosen. In other words, this step is fundamental to understand whether the flagged, suspect and wrong data from automatic controls are such, or if the automatic test is too restrictive and some data flagged as correct should be retained. Moreover, it is also important to identify new types of errors and the automatic checks to be implemented and performed.

3.2 Validation

The QC procedures here described are the application of general procedures extensively detailed in various WMO guides and documents, especially for synoptic networks and climatic data. The Guide to the Global Observing System (GGOS) (WMO, 2010b), Guide on the Global Data-Processing System (GDPS) (WMO, 1993), Guide to Agricultural Meteorological Practices (GAMP) (WMO, 2010a), Guide to Meteorological Instruments and Methods of Observation (GMIMO) (WMO, 2014) and Guide to Climatological Practices (GCP) (WMO, 2011), together represent the fundamental WMO literature that defines the international standards to be fulfilled on data QC. However, sometimes these procedures are not complete for all variables that need to be considered or for all the temporal resolutions of the data to be validated. For example, the GGOS contains the standard procedures for processed data with a temporal resolution in the order of minutes, which should be readapted in case of hourly processed data. In this case, it can be assumed that the basic automatic QC on raw data is done automatically by the AWS, while the study proceeds with the basic automatic quality and the extended QC on hourly processed data. Another drawback of these guidelines is that they are limited to single-AWS checks as range, time consistency and internal consistency, while the spatial control is not treated. The temporal consistency test here implemented on hourly data was derived by Estévez et al. (2011) who made a summary of all the QC procedures purposed in the literature from the single-station tests (Meek and Hatfield, 1994; Shafer et al., 2000; Zahumenský, 2004) and also multiple station tests (Shafer et al., 2000; Hubbard et al., 2005).

Chapter 6 of the GDPS (WMO, 1993) is entirely dedicated to the whole validation topic. Although relatively old and focused on climatic aspects, this guide gave the main steps to

| Test | Formulation | Flag | Reference |
|------|-------------|------|-----------|
| Gross error limit | $-30^\circ \text{C} < T_h < 50^\circ \text{C}$ | $2^3 = 8$: Erroneous for the range test | ISPRA and Research (2017) |
| Persistence | $T_h \neq T_{h-1} \neq T_{h-2} \neq T_{h-3}$ | $2^4 = 16$: Suspect for the persistence test | Estévez et al. (2011), Meek and Hatfield (1994) |
| Step | $|T_h - T_{h-1}| < 4^\circ \text{C}$ | $2^5 = 32$: Suspect for the step test | Guide on the Global Data-Processing System (GDPS) (WMO, 1993) |
| Internal consistency | $T_{\text{hmin}} < T_h < T_{\text{hmax}}$ | $2^6 = 64$: Internally inconsistent | GOS, WMO (2010b) |
| | $T_h > T_{\text{dew}} (T_h, \text{RH})$ | | |
| Spatial consistency | $T^* - f\sigma < T_h < T^* + f\sigma$ | $2^7 = 128$: Spatially inconsistent | Hubbard et al. (2005) |
| | $f = 3$, $R = 0.25^\circ$ ($\approx 31 \text{ km}$) | | |
| Climatological limit | $T_{\text{clim hmin}} < T_h < T_{\text{clim hmax}}$ | $2^8 = 256$: Suspect for the climatological limit | WMO (1993) |
| Other flags | Explanation | Flag |
| Not validated data | Original data not checked yet | 0: Not validated data |
| Validated data | Original data already checked | $2^0 = 1$: Validated data |
| Reconstructed data | New data substituting for original data | $2^1 = 2$: Reconstructed data |
| Missing data | Data not transmitted | $2^2 = 4$: Missing data |
follow through the validation process (Figure 2). Most of the other procedures explained in the GDPS are for the synoptic observing system and include hydrostatic checks, upper air data controls and many other controls that will be not applied in the present mesoscale surface network. However, some tests, such as time consistency (Table 2), will be used in the original form described in the guide.

The procedure implemented herein was published by Zahumensky (2004) as a proposal for the Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. This document, now part 6 of the GGOS (WMO, 2010b), contains all the technical information about the QC checks, many of which will be used here. The procedure consists of the basic QC (first level) and the extended QC (second level), with all the corresponding tests (Figure 2) described below. Based on the references cited and the flux chart illustrated in Figure 2, the selected QC checks and the flags used are summarized in Table 2.

3.2.1 Basic QC: Pre-processing checks

Pre-processing checks are meant to detect syntactic errors (i.e. the transcription of station names or the variables' names), to flag missing data, to detect temporal repetition (the same variable at the same station measured twice with

![Reconstruction procedure: Station Polvese 1](image-url)

**FIGURE 3** Example of the application of the empirical orthogonal functions (EOFs) algorithm
completely different values) and to synchronize the data recording time. The last operation is crucial to have a data set homogeneous in time. Some stations could measure the temperature at time intervals 8:12, 9:12 and so on, instead of being centred at the reference sampling time 8:00, 9:00, etc. In this case, the synchronization is made by considering all the data measured within ±15 min from the reference time step. These data are considered as still representative of the correct time step itself. Therefore, all the timestamps of the asynchronous series are simply shifted back to the reference times. Instead, when the asynchronicity is greater than ±15 min, the time lag should be estimated by cross-correlating the test series with a reference one (Antolini et al., 2016).

3.2.2  Basic QC: Gross error limit checks

The gross error limit check (basic QC) is based on the typical instrumental range of temperature gauges of the regional network. This usually differs from the climatological limit check (extended QC) in which the limits are dependent on the season and regional climatic conditions. However, for the present work, these limits will coincide since the short time range of the available hourly data did not allow any climatic calculation.

3.2.3  Basic QC: Temporal consistency

The temporal consistency (basic QC) is evaluated using two different tests: the persistence test, where the minimum variability of temperature is verified by comparing its value with the three previous time steps; and the step test, where the maximum variability is verified by comparing the current temperature with the previous one. Besides the international guidelines of the WMO, another important national guide by the Italian Institute for Environmental Protection (ISPRA and Research, 2017) has been very useful in determining the most appropriate tests to be applied for all the different temporal resolutions; it contains important suggestions for the difficult task of evaluating the temporal consistency of precipitation data, not discussed here.

3.2.4  Basic QC: Internal consistency

The internal consistency test (basic QC), which is the last single station test, should be made by comparing the temperature \( T \) with its minimum and maximum hourly values \( T_{\text{min}}, T_{\text{max}} \) and the dew point temperature \( T_{\text{dew}} \) (estimated through relative humidity (RH) and \( T \) itself) measured at the same stations. This test is reported here for the completeness of the procedure, but it is not applied since humidity data were not available.

3.2.5  Automatic QC: Spatial consistency

At this point the pre-validated hourly data are formatted and the subsequent checks concern the spatial consistency of the data. These checks compare data from multiple stations while the previous checks were implemented on single stations’ data. There are different examples of spatial consistency tests. For instance, an automatic spatial consistency test is built starting from an analysed field provided by an optimal interpolation algorithm as a pre-processing tool of hourly temperature data by Lussana et al. (2010). This is an example of a spatial consistency test (extended QC) based on complex algorithms (complex extended QC).

The fundamental idea used here is that introduced by the Hubbard spatial regression test: the residual between the observed value and that at the same location obtained by an objective analysis has to be confined to a certain interval of confidence.

This test is the first multiple station test and follows the spatial weighted regression method based on Hubbard et al. (2005). First, for each reference station \( (h) \), the neighbour stations \( (n) \) inside a certain search radius \( (R \approx 31 \text{ km}) \) are found. This search radius is approximately equal to the average spacing of the observations (Table 1) and it has been chosen large enough to have at least one neighbour for each station. Once the neighbours have been established, if the number of missing values in each series is < 50%, a linear regression is computed between their temperatures \( T_n(t) \) over all the time steps \( L_T \) (eight years) and the reference station values \( T_h(t) \), in order to find a first estimate \( T_h^* \) of \( T_h \) that should be consistent with \( T_n \). Here \( L_T \) was chosen equal to the total length of the time series. The root mean square error (RMSE) between the reference values \( T_h \) and the estimated values \( T_h^* \) of the neighbour station from the regression line (correspondent to the sample standard deviation of the residuals \( \sigma_n \)) is evaluated to find a measure of the stations’ correlation:

\[
\text{RMSE}_n = \sigma_n = \sqrt{\frac{1}{L_T} \sum_{t=1}^{L_T} \left( \frac{T_h(t) - T_h^*(t, n)}{\text{Residuals}} \right)^2}
\]

This error characterizing each neighbour station is used as a weight in the estimate of the reference value \( T^* \)
(t) and reference standard deviation $\sigma^*$ from the surrounding stations at each instant:

$$ T^*(t) = \frac{\sum_{n=1}^{N} T_n^2 / \sigma_n^2}{\sum_{n=1}^{N} 1 / \sigma_n^2} \quad \sigma^* = \frac{N}{\sum_{n=1}^{N} 1 / \sigma_n^2} \quad (2) $$

Finally, at each time step, a tolerance interval is established by considering a constant factor $f$ and the spatial consistency is verified by ensuring that:

$$ T^*(t) - f\sigma^* < T_h(t) < T^*(t) + f\sigma^* \quad (3) $$

The results of all these procedures are shown in Section 4. Whenever the expression of each test as indicated in Table 2 is not fulfilled by $T_h$, then the data are flagged with their respective flags. Note that during validation data are not modified but only flagged.

### 3.2.6 Automatic QC: Homogeneity

Although the spatial and temporal consistency tests are very effective at evaluating the data homogeneity on small temporal intervals and for large-value discontinuities, they are not enough to identify long-term shifts of the signal measured by a sensor. These shifts are usually because of station relocation, changing of the surrounding environment or other non-climatic factors affecting the sensor. Homogeneity tests such as the Standard Normal Homogeneity Test (SNHT) (Alexandersson and Moberg, 1997), the Craddock test (Craddock, 1979) or the Vincent test (Vincent, 1998) have been developed in order to detect these break-points in the time series by comparing series with a reference one coming either by the same network or by another (also to identify shifts common to the whole network). Anyhow, those tests apply only to climatic daily data with at least 10 segments (Moberg and Alexandersson, 1997), and their results have to be carefully examined not to confuse climatic shifts with artificial break-points. The corresponding WMO guidelines on how to use these tests are explained by Aguilar and Llanso (2003), where the importance of metadata to track any artificial change affecting the measurement is underlined. Given the length of the data set under scrutiny and the need to apply the homogeneity test to a reconstructed data set, this test was not applied at this point of the QC.

After these controls data have to be manually checked.

### 3.3 Reconstruction

Observations coming out from the validation process contain many erroneous, suspect and missing data that need to be reconstructed. Once the suspect and wrong data have been rejected by automatic and manual checks, they are considered as missing values to be replaced by suitable ones. The reconstruction can be made in three main different ways: spatial, temporal and spatiotemporal. Other methods could involve the use of an unbiased external complete data set as reanalysis products, as illustrated by Lompar et al. (2019).

The EOFs reconstruction used here can be classified as a spatiotemporal reconstruction method. It was first purposed by Beckers and Rixen (2003) to reconstruct incomplete oceanographic data sets. However, the same concept of data reconstruction based on orthogonal decomposition was purposed by Everson and Sirovich (1995) and used in many other applications such as the reconstruction of gappy data in particle image velocimetry (PIV) in common fluid dynamic laboratory experiments (Gunes et al., 2006). Spatiotemporal techniques are considering both spatial and temporal dimensions and missing data are filled by taking into account the neighbouring measurements in both time and space. The dominant EOFs are estimated by the singular value decomposition of a matrix composed by the temperature observation vectors ordered in time. At a certain time, the missing data point of the given field is then estimated with the dominant EOFs. This procedure is used to evaluate the EOFs iteratively until convergence.

As explained by Henn et al. (2013), the most suitable reconstruction method is strictly dependent on the length of the missing data period (gap) and the number of available stations. In that paper, a comparison of different reconstruction algorithms was applied to hourly temperature data and the EOF algorithm resulted in the most accurate among the others (Micromet algorithm; Liston and Elder, 2006), especially when a high number of stations was available (> 16) and the gap length was > 2 hr. For shorter gaps, the linear temporal interpolation method was still that with the lowest interpolation error.

### 3.3.1 Extended QC: Reconstruction

Based on the analysis illustrated, the selected reconstruction procedure was divided into two steps which can be seen as a schematic in Figure 2 and its application in Figure 3:

- Linear interpolation: the 1 hr gaps are filled by simple linear interpolation.
- EOF reconstruction: the $\geq 2$ hr gaps are filled following the algorithm described by Beckers and Rixen (2003) and Henn et al. (2013).
As shown in Figure 2, the EOFs algorithm is applied to a matrix $T$ that contains both the station data to be filled $T_{\text{stat}}$ and the ERA5 reanalysis grid points contained in the regional territory $T_{\text{era}}$:

$$
T_{\text{stat}} = \begin{bmatrix}
T_{11} & \cdots & T_{1s} \\
\vdots & \ddots & \vdots \\
T_{t1} & \cdots & T_{ts}
\end{bmatrix}^{\text{stat}},
$$

$$
T_{\text{era}} = \begin{bmatrix}
T_{11} & \cdots & T_{1g} \\
\vdots & \ddots & \vdots \\
T_{t1} & \cdots & T_{tg}
\end{bmatrix}^{\text{era}}
$$

where $t$ is the number of time steps considered for the calculation of EOFs; $s$ is the number of temperature stations available; and $g$ is the number of ERA5-interpolated grid points contained in the considered region. The introduction of the complete and homogeneous reanalysis data in the EOF calculation also allowed the simultaneous filling of the gaps where all the stations were missing.

After assembling the initial matrix $T$, the time average of each station is subtracted by each column to obtain the anomaly matrix $T'_0 = T - \bar{T}$, where:

$$
\bar{T}_{ij} = \frac{1}{t} \sum_{k=1}^{t} T_{kj}
$$

The missing values within $T'$ are then replaced with zeros, which represent an initial guess equal to the means of each time series. The resulting matrix is called $T'_{\text{noise}}$, since those zeros resemble the noise effect on a common signal. At this point the EOFs are calculated by applying a singular value decomposition (SVD) algorithm:

$$
[U, D, V^T] = \text{SVD} \left( T'_{\text{noise}} \right)
$$

where $T'_{N} = T'_{\text{noise}}$ at the initial step; $U$ is the matrix whose columns are the temporal EOFs (or temporal modes); $V$ is the matrix whose columns are the spatial EOFs (or spatial modes); and $D$ is the diagonal matrix containing the singular values $d_k$. A certain number of EOFs ($N_{\text{EOF}}$) are then retained:

$$
U_N = U[:,N_{\text{EOF}}] \quad D_N = D[:,N_{\text{EOF}},:] \quad V_T^N = V^T[:,N_{\text{EOF}},:]
$$

in order to transform the signal back and to obtain a new guess for the missing data:

$$
T_N^{(\text{missing})} = U_N D_N V_T^N
$$

Their new values will be different from zero and dependent on $N_{\text{EOF}}$. The matrix $T_N$ containing the new

\begin{table}[h]
\centering
\caption{Summary statistics for the quality control procedures of the network: Mean, standard deviation (std) and maximum percentage of flagged values over all the temperature stations}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Flags & Miss & Range & Pers & Step & Spatial & Valid \\
\hline
Mean & 4.73 & 0 & 2.42 & 2.46 & 3.09 & 89.02 \\
Std & 12.67 & 0.01 & 9.10 & 2.18 & 6.48 & 17.30 \\
Maximum & 72.65 & 0.07 & 48.12 & 11.91 & 40.14 & 98.10 \\
\hline
\end{tabular}
\end{table}

Note: Miss: missing values, Range for flagged values by range, Pers by persistency, Step by the step test, and Spatial by the spatial test. Valid is indicating the percentage of values to be considered neither suspect, missing nor inconsistent.
estimate of missing data is introduced back into Equation 7 and the EOF calculation restarts again.

The reconstruction procedure starts with $N_{EOF} = 1$ and is repeated waiting for convergence. Convergence is reached once the fraction of variance explained by each mode in one iteration (the squared singular values $d^k$) differs from the previous one $(d^{k-1})$ less a certain tolerance value. The condition to be satisfied is:

$$r_k = \frac{1}{s + \sigma} \sum_{i=1}^{s+\sigma} \frac{(d_i^k - d_i^{k-1})^2}{(d_i^{k-1})^2} \leq rtol$$  \hspace{1cm} (10)

Once this condition is met, the estimated missing data in $T_0$ are used as the initial guess for the new cycle with $N_{EOF} = 2$. This number is increased until a certain maximum number of EOFs to be retained is reached.

This maximum number $N_{EOF}^{max}$ is calculated previously by a cross-validation technique (Beckers and Rixen, 2003): the concept is to replace some valid data with zero values and let the EOF reconstruction technique to estimate them. By comparing the valid data with the estimated data, a measure of the error committed by the reconstruction is derived. This error is a function of $N_{EOF}$; therefore, the optimal number of EOFs to be retained is that obtaining the minimum error.

4 | RESULTS

4.1 | Results of validation procedures

The results of the QC procedures described in Table 2 are summarized here in Table 3. The maximums shown in Table 3 are very high. Some stations contain temporal gaps longer than one year and sometimes those long missing intervals are filled with zeros that are flagged by the persistence test. Moreover, the spatial test shows similar maximums, because it is flagging the persistent values in disagreement with the surrounding stations (Figures 4 and 5).

These extreme values are also reflected by the high standard deviation of the flag percentage. Among all stations, only six have a percentage of valid values that are $< 80\%$ and fortunately just one among them is located at high altitude. It is appropriate to get rid of these stations since their reconstruction would not be reliable. Moreover, the lapse rate calculation and stations’ altitude distribution would not be affected by this modification. Therefore, from this point on, all the procedures and the analysis were conducted on 67 stations, excluding those with a percentage of valid data $< 80\%$.

A more consistent analysis of the validation procedures and their results can be performed by updating the statistics with the new set of stations, as illustrated in Table 4.
the \( \Delta T_{\text{max}} = 4^\circ \text{C} \) limit of the absolute difference between two consecutive measurements is too low. Increasing \( \Delta T_{\text{max}} \) by 1°C halves this percentage (Figure 6). However, the original value has been kept to agree with the WMO’s requirements (WMO, 1993). Most of the time flagged values by the step test are isolated (1 hr) suspect data that will be corrected by linear interpolation. The spatial test parameters \( f \) and \( R \), as explained in Section 3.2.5 and shown in Table 2, have been tuned by considering some real cases where a spatial inconsistency was observed (Figure 7). In this particular example, by choosing \( f = 3 \), the spatial test could detect a repetition in temperature series that indicated an error.

The other tests did not flag those values because they seem reasonable, and this highlights the importance of applying an extended QC with a spatial consistency test.

### 4.2 Results of reconstruction techniques

The EOF reconstruction algorithm (see Section 3.3) is applied for each year, with \( t \) being the number of hours
multiplied by the number of days of the selected year. The number of stations considered is \( s = 67 \) and the number of selected ERA5 grid points is \( g = 221 \), corresponding to a grid with 13 latitude and 17 longitude points. Therefore, for each year, the EOFs are evaluated on a matrix \( T \) whose dimensions are \([ t \times 288] \). The tolerance \( r_{tol} \) for the convergence of each iteration was fixed to 0.01\%. This means that convergence is reached once the fraction of variance explained by each mode in one iteration differs from the previous one by \(< 0.01\% \).

The optimal number of EOF, \( N_{\text{EOF}} \), was selected by the cross-validation technique described in Section 3.3; the results of this iterative procedure are shown in Figure 8.

The final selected number is \( N_{\text{EOF}} = 33 \). After applying the SVD to the anomaly matrix \( T' \), the selected singular values \( d_k \) and their explained variance can be analysed. To connect the singular values to the eigenvalues (just for a matter of dimensionality), the variance–covariance matrix of the anomalies \( T' \) (Wilks, 1995) is defined as:

\[
S = \frac{1}{n-1}T'T' \tag{11}
\]

where \( T'T \) is the anomalies transposed matrix. From this definition, the correlation matrix \( R \), which is equivalent to the covariance matrix applied to the standardized anomalies, can also be calculated. The spatial structure of \( R \) for 2010 can be visualized in Figure 9 which highlights the statistical consistency of the ERA5 data set. Indeed, the top-right quadrant contains correlation co-efficients between 0.96 and 1 almost everywhere, compared with the lower-left quadrant showing the correlation between the temperature stations where this range is much larger (0.7–1). This structure is very similar for all the analysed years.

By applying SVD decomposition as in Equation 7, the following can be obtained:

\[
T' = UDV'^T \tag{12}
\]

By substituting Equation (12) into Equation (11), the relation between the eigenvalues \( \lambda_k \) of the covariance matrix \( S \) and the singular values \( d_k \) of the anomalies matrix \( T' \) is simply:

\[
\lambda_k = \frac{1}{n-1}d_k^2 \tag{13}
\]

The total variance of the system or its total energy (equivalent to the trace of \( S \)) is completely represented by the sum of all eigenvalues. In consequence, the ratio \( f_k = \frac{\lambda_k}{\sum_{k=1}^s \lambda_k} \) represents the fraction of variance explained by the \( k \) mode. Table 5 reports the percentages of \( f_k \) for the first seven modes and the last mode chosen as a truncation number, together with the total amount of the variance.

**TABLE 5** Fraction of the explained variance by each mode expressed as a percentage of the total variance

| Year | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 33    | Total   |
|------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| 2010 | 97.38 | 0.70  | 0.39  | 0.36  | 0.20  | 0.15  | 0.10  | 0.01  | 561,407.72 |
| 2011 | 96.88 | 1.02  | 0.37  | 0.35  | 0.27  | 0.20  | 0.10  | 0.01  | 570,798.28 |
| 2012 | 97.38 | 0.80  | 0.39  | 0.30  | 0.22  | 0.16  | 0.08  | 0.01  | 698,521.81 |
| 2013 | 97.03 | 0.81  | 0.47  | 0.37  | 0.24  | 0.18  | 0.09  | 0.01  | 539,250.55 |
| 2014 | 96.36 | 1.03  | 0.53  | 0.44  | 0.30  | 0.23  | 0.11  | 0.01  | 387,372.69 |
| 2015 | 97.04 | 0.89  | 0.38  | 0.35  | 0.28  | 0.18  | 0.09  | 0.01  | 571,797.29 |
| 2016 | 96.69 | 0.93  | 0.47  | 0.41  | 0.32  | 0.19  | 0.09  | 0.01  | 464,378.88 |
| 2017 | 97.17 | 0.91  | 0.39  | 0.31  | 0.25  | 0.18  | 0.08  | 0.01  | 636,339.42 |

Note: The last column is expressed in \(^{°}\text{C}^2\) and indicates the total energy of the system; all the other columns represent \( f_k \) (%).
As can be seen, the cross-validation technique leads to the retention of a high number of EOFs that are explaining almost all the variance of the system.

5 | SUMMARY AND CONCLUSIONS

A complete procedure for the validation and reconstruction of hourly temperature data from the network of the Umbria region of central Italy was described and applied to observations spanning a temporal range of eight years from 2010 to 2017.

The World Meteorological Organization (WMO) standard basic and extended quality control (QC) procedures were applied to this data set in order to detect observational errors.

A flagging system was designed to take note of all the quality checks and their results, and to recognize missing, suspect, inconsistent and wrong data. Seven stations did not pass the validation procedure since they contained a percentage of valid data < 80%. The remaining 67 stations were considered suitable for the reconstruction procedures.

The reconstruction was made by using the empirical orthogonal function (EOF) spatiotemporal method. This method was slightly modified by the introduction of an external complete and homogeneous data set such as the ERA5 reanalysis. This modification allows one to fill in the missing data and temporal gaps common to all the stations of the network.

The EOFs were calculated over the whole of the domain and updated each year, while their truncation number was calculated by a cross-validation technique.

The final products are hourly temperature series validated and reconstructed, ready to be used for further applications. The methods applied in the present paper will be extended to other variables (precipitation, relative humidity, solar radiation and wind) and to a larger temporal range spanning from the 1990s to the present.

ACKNOWLEDGMENTS

The authors are thankful to Ing. Marco Stelluti and Ing. Piero Nelli for providing the regional data. Part of the study was generated using Copernicus Climate Change Service information (2019).

REFERENCES

Aguilar, E. and Llanso, P. (2003) Guidelines on Climate Metadata and Homogenization. Geneva: World Meteorological Organization. Alexandersson, H. and Moberg, A. (1997) Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. International Journal of Climatology, 17(1), 25–34.

Antolini, G., Auteri, L., Pavan, V., Tomei, F., Tomozeiu, R. and Marletto, V. (2016) A daily high-resolution gridded climatic data set for Emilia-Romagna, Italy, during 1961–2010. International Journal of Climatology, 36(4), 1970–1986.

Barbero, S., Zaccagnino, M., Mariani, S., Lastoria, B., Braca, G., Bussotti, M., Casaiolo, M., Marsico, L., Rotundo, R., Pavan, L., Ricciardi, G., Zenoni, E., Cicogna, A., Micheletti, S., Cazzuli, O., Priolo, S.D., Ranci, M., Rondanini, C., Bianco, G., Eggiati, Montanini, P., Saccardo, I., Campione, E., Pupillo, S., Iocca, F., Lazzeri, M., Tedeschiini, M., Marzano, V., Schena, P., Licciardello, A., Manzella, F., e Sara Ratto, F.B. (2017) Linee guida per il controllo di validità dei dati idro-meteorologici. ISPRA – Manuali e Linee Guida 156/2017. Roma, 2017.

Beckers, J.-M. and Rixen, M. (2003) EOF calculations and data filling from incomplete oceanographic datasets. Journal of Atmospheric and Oceanic Technology, 20(12), 1839–1856.

Bgeš, G., Drnovšek, J., Bojkovski, J., Knez, J., Groselj, D., Černač, B. and Hudoklin, D. (2015) Automatic weather stations and the quality function deployment method. Meteorological Applications, 22(S1), 861–866. https://doi.org/10.1002/met.1508.

C3S. (2017) Copernicus Climate Change Service, ERAS: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), howpublished. Available at: https://cds.climate.copernicus.eu/cdsapp#!/home.

Craddock, J.M. (1979) Methods of comparing annual rainfall records for climatic purposes. Weather, 34(9), 332–346.

Daly, C. (2006) Guidelines for assessing the suitability of spatial climate data sets. International Journal of Climatology: A Journal of the Royal Meteorological Society, 26(6), 707–721.

Danielson, J.J., and Gesch, D.B. (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, 26.

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., Raspaud M. (2003) NetCDF Climate and Forecast (CF) metadata conventions. http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html [Accessed 06 05 2020].

ECMWF. (2018) ERA5. Available at: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 [Accessed 01 04 2020].

Estévez, J., Gavilán, P. and Giráldez, J.V. (2011) Guidelines on validation procedures for meteorological data from automatic weather stations. Journal of Hydrology, 402(1–2), 144–154.

Everson, R. and Sirovich, L. (1995) Karhunen–Loève procedure for gappy data. JOSSA A, 12(8), 1657–1664.

Gunes, H., Sirisup, S. and Karniadakis, G.E. (2006) Gappy data: to Krig or not to Krig? Journal of Computational Physics, 212(1), 358–382.

Henn, B., Raleigh, M.S., Fisher, A. and Lundquist, J.D. (2013) A comparison of methods for filling gaps in hourly near-surface
air temperature data. *Journal of Hydrometeorology*, 14(3), 929–945.

Hubbard, K.G., Goddard, S., Sorensen, W.D., Wells, N. and Osugi, T.T. (2005) Performance of quality assurance procedures for an applied climate information system. *Journal of Atmospheric and Oceanic Technology*, 22(1), 105–112.

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E. (2008) *Hole-filled SRTM for the globe: version 4*. CGIAR-CSI SRTM 90m Database. Available at: http://srtm.csi.cgiar.org.

Lanciani, A. and Salvati, M. (2008) *Spatial Interpolation of Surface Weather Observations in Alpine Meteorological Services*. FORALPS Technical Report, 2. Università degli Studi di Trento, Dipartimento di Ingegneria Civile e Ambientale, Trento, Italy, 40.

Liston, G.E. and Elder, K. (2006) A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). *Journal of Hydrometeorology*, 7(2), 217–234.

Lompar, M., Lalić, B., Dekić, L. and Petrić, M. (2019) Filling gaps in hourly air temperature data using debiased ERA5 data. *Atmosphere*, 10(1), 13.

Lussana, C., Uboldi, F. and Salvati, M.R. (2010) A spatial consistency test for surface observations from mesoscale meteorological networks. *Quarterly Journal of the Royal Meteorological Society*, 136(649), 1075–1088.

Maciel, P., Quintino, T., Raoult, B. and Siemen, S. (2015) MIR – ECMWF’s new interpolation package. Available at: https://www.ecmwf.int/node/13332.

Meek, D.W. and Hatfield, J.L. (1994) Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology*, 69(1–2), 85–109.

Moberg, A. and Alexandersson, H. (1997) Homogenization of Swedish temperature data. Part II: homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology*, 17(1), 35–54.

Pavan, V., Antolini, G., Barbiero, R., Berni, N., Brunier, F., Cacciamani, C., Cagnati, A., Cazzuli, O., Cicogna, A., De Luigi, C., et al. (2018) High resolution climate precipitation analysis for north-central Italy, 1961–2015. *Climate Dynamics*, 52, 3435–3453 (2019). https://doi.org/10.1007/s00382-018-4337-6.

RRN 2014–2020 Rete Rurale Nazionale. (2017) *Lo stato dell’arte sulle reti agrometeorologiche regionali*. Indagine preliminare.

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., Fredrickson, S.E. and Hughes, T.W. (2000) Quality assurance procedures in the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, 17(4), 474–494.

Stull, R.B. (1988). An Introduction to Boundary Layer Meteorology. Kluwer, Dordrecht, 666.

Uboldi, F., Lussana, C. and Salvati, M. (2008) Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks. *Meteorological Applications*, 15(3), 331–345.

Vincent, L.A. (1998) A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, 11(5), 1094–1104.

Wilks, D.S. (1995) *Statistical Methods in Atmospheric Science*, Vol. 59. International Geophysics Series. Academic Press: San Diego, London.

WMO. (1993) *Guide on the Global Data-Processing System* (WMO-No. 305). Geneva: World Meteorological Organization.

WMO. (2010a) *Guide to Agricultural Meteorological Practices* (WMO-No. 134). Geneva: World Meteorological Organization.

WMO. (2010b) *Guide to the Global Observing System* (WMO-No. 488). Geneva: World Meteorological Organization.

WMO. (2011) *Guide to Climatological Practices* (WMO-No. 100). Geneva: World Meteorological Organization.

WMO. (2014) *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8). Geneva: World Meteorological Organization.

Zahumenský, I. (2004) *Guidelines on Quality Control Procedures for Data from Automatic Weather Stations*. Geneva: World Meteorological Organization.

How to cite this article: Cerlini PB, Silvestri L, Saraceni M. Quality control and gap-filling methods applied to hourly temperature observations over central Italy. *Meteorol Appl*. 2020;27:e1913. https://doi.org/10.1002/met.1913