Received: 6 November 2023 | Revised: 30 January 2024 | Accepted: 11 February 2024

DOI: 10.1002/joc.8406

RESEARCH ARTICLE

International Journal of Climatology RMetS

Climatological reference stations: Definitions and requirements

Andrea Merlone[1] | Gaber Beges[2] | Alberto Bottacin[1,3] | Manola Brunet[4] | Alba Gilabert[4] | Drago Groselj[5] | Andrew Harper[6] | Peer Hechler[7] | Mirjana Ivanov[8] | Chiara Musacchio[1,9] | Blair Trewin[10] | William Wright[10]

[1]Istituto Nazionale di Ricerca Metrologica, Turin, Italy
[2]Univerza v Ljubljani Fakulteta za Elektrotehniko, Ljubljana, Slovenia
[3]Politecnico di Torino, Turin, Italy
[4]Universitat Rovira i Virgili, Tarragona, Spain
[5]Agencija Republike Slovenije za okolje, Ljubljana, Slovenia
[6]National Institute of Water and Atmospheric Research, Auckland, New Zealand
[7]World Meteorological Organization, Geneva, Switzerland
[8]Institute of Hydrometeorology and Seismology, Podgorica, Montenegro
[9]Societa Meteorologica Italiana, Bussoleno, Italy
[10]Australian Bureau of Meteorology, Melbourne, Victoria, Australia

Correspondence
Andrea Merlone, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, Turin 10135, Italy.
Email: a.merlone@inrim.it

Funding information
European Metrology Programme for Innovation and Research; European Union's Horizon 2020

Abstract

Ground-based stations are one of many observing systems contributing to the generation of data to evaluate climate trends and variations locally and globally. Networks of stations are made of various numbers of observing sites, equipped with different typologies and varieties of instruments, differently managed and maintained. Among such networks a limited number of stations are required as a reference, to provide top quality traceable measurements and as the top level of a tiered approach; it is these stations which are here designated as Climatological Reference Stations (CRS). At present, there is no agreed definition of the key instrumental and technical features of a CRS, nor are there defined reference measurement procedures. This leads to the situation of a multitude of approaches among different National Meteorological and Hydrological Services (NMHSs), research institutes and other agencies that reduces the comparability of results in space and in time. The lack of CRSs, moreover, is a major contributor to the huge efforts required to harmonize data and detect biases locally, regionally and globally, as in their absence there is a need to fall back on other stations which are more likely to have such biases. This article reports on the outcomes of the work of a group of experts, nominated by the World Meteorological Organization (WMO) Commission for

Climatology (CCI), tasked to promote a standardized and agreed concept about the definitions, specifications and technical characteristics of climate reference data and stations for ground-based networks, with the addition of a practical example of climatological reference station installed in the framework of an European funded project, with the purpose of comparing and confirming the prescribed characteristics and guidance, with respect to practical field use and available instrumentation. The adoption of a unique definition and prescription on the technical setup, measurement procedures and uncertainty evaluation will substantially progress towards a common approach in detecting climate trends. This will improve data comparability in space and time and allow a more robust understanding of climate evolution locally and globally.

KEYWORDS

climate reference data, climate reference stations, climatology, instrumentation, meteorological stations, metrology, uncertainty

1 | INTRODUCTION

In climatology, two main characteristics of data series are required. First, time series should be long, continuous and complete. Second, they should be representative of their surroundings and comparable in space and time. These two aspects are rarely both characteristics of a single station (Thorne et al., 2018).

To try to understand climate evolution, including climate variability and change, it is crucial to ensure the quality of historical observations. In the absence of global coordination, historic meteorological observations were taken with very different procedures, schedules, instruments, shelters, and so forth. Furthermore, the environment of many recognized centennial stations, for example, urban stations, may have changed in the course of the station history, thus affecting the quality and characteristics of their records. Moreover, in many cases changes in instrumentation and methods of observation introduce biases in the series, which may not be fully evaluated when detailed information about the change is missing. The overall effect of all these factors is to compromise the homogeneity, and also to endanger the time series quality in terms of traceability and comparability.

At present, the World Meteorological Organization (WMO) Technical Regulations define that each Member shall establish and maintain at least one climatological reference station (CRS) in order to overcome the limitations of historical observations. Climatological reference stations are described as climatological stations gathering data and associated metadata, intended for the purpose of determining climatic trends. Ideally, the records should be homogeneous and of sufficient length (not less than 30 years) to enable the identification of secular changes of climate (Manual on WIGOS, WMO-No. 1160), where human-induced environmental changes have been and/or are expected to remain at a minimum to avoid the introduction of biases in the series.

The characteristics of a CRS ensure data comparability, which is fundamental in climatology. Indeed, data recorded by the same station should be comparable over time, and can also serve as a reference for other surrounding stations, which may have different measuring systems. In a tiered network concept (Thorne et al., 2016; Thorne et al., 2018; WIGOS, 2019), recommended by the Global Climate Observing System (GCOS) and the WMO Integrated Global Observing System (WIGOS), this high data quality allows reference observations to be placed at the top, allowing meaningful statistical adjustments to lower tier observing data with less rigid observational requirements.

Apart from the recent initiative by GCOS in prescribing the features and promoting the creation of a GCOS Surface Reference Network (GSRN) (GCOS, 2019), and valuable initiatives, such as the United States Climate Reference Network (USCRN) (USCRN, 2020) which now adopt a common layout and unique measuring principles for climatological reference stations, other climatological reference networks are difficult to compare with each other in space and in time. The reason is that they are rarely equipped with the same typologies of instruments and are not always located in sites offering those stable environmental conditions able to avoid introducing undetectable biases. Despite large efforts in infrastructural networking in many regions, such as in Europe, a common approach is missing and previously not planned, in defining how climatological reference stations shall be designed.

The scope of this article is to report on the outcomes of the work of a group of experts, nominated by the World Meteorological Organization's former Commission for Climatology (CCI) (now part of the Commission for Weather, Climate, Hydrological, Marine and Related Environmental Services and Applications – SERCOM), tasked to promote a standardized and agreed concept about the definitions, specifications and technical characteristics of climate reference data and stations for ground-based networks. The adoption of a unique definition and prescription on the technical setup, measurement procedures and uncertainty evaluation will substantially progress towards a common approach in detecting climate trends. This will improve data comparability in space and time and allow a more robust understanding of climate evolution locally and globally. Furthermore, to support the deployment of such definitions and requirements, a prototype of a CRS has been developed and installed by INRiM (Istituto Nazionale di Ricerca Metrologica), whose technical specifications are reported here alongside with first considerations from field tests.

2 | CHALLENGES WITH HISTORICAL DATA SERIES, MOTIVATING A STEP CHANGE FOR PROVIDING BETTER DATA IN FUTURE CLIMATE ANALYSIS

Homogeneous long-term observations are extremely rare globally. Therefore, homogenization efforts are an essential part of climate data-processing. They are needed to adjust for changepoints (abrupt or gradual) and other potential biases artificially introduced into the data through changes in station location or local site environment; the type of screen used to shelter the instruments; instrument type; or observing/reporting methods (WMO, 2020). On the global scale, it is impossible to know exactly when all these changes occurred as they are rarely documented in a way that can be associated alongside the data digitally.

Homogenization algorithms (e.g., Caussinus & Mestre, 2004; Domonkos, 2011) try to detect these signals of changepoints from the background noise and make approximate adjustments. Most currently available algorithms use comparisons with neighbouring stations for detection and adjustment. Large and abrupt changes are relatively easy to detect and correct for, whereas small or time-varying ones (such as the slow land-cover changes or minimal instrument drifts) are harder, although still fundamental to understand climate evolution. Furthermore, some changes will be isolated to a single station, but some will be network wide (e.g., a generalized change in the instrument shelter or in the observing system applied at the same time to all the stations in a network), which makes their detection even more difficult (Begert & Frei, 2018; Trewin et al., 2020). At the daily and sub-daily scale, homogenization is a greater challenge still.

Such mathematical and computational efforts are mainly required due to the lack of data-series originated from stations made for the specific purpose of generating data immune from undocumented changes, continuously traceable to recognized standards and with associated measurement uncertainties.

Many long term and centennial stations are located inside towns, where original historical meteorological observatories were positioned. When cities grow and industrial activities increase, together with traffic, winter heating and summer cooling of buildings, significant effects are introduced in the data quality of such stations, returning a poor representativeness of the surrounding climate, uncertainties and stability. Despite some exceptions (Jones & Lister, 2009; Trewin et al., 2020) such effects are often not stable in time, and often show seasonal and daily cycles and/or weather type dependence in the magnitude of the biases.

Changing the site of centennial stations could avoid the introduction of such biases due to the urban environment, however the impact on the continuity of the time series may outweigh the advantages of moving to a better site. Among the long-term stations, few are located in open fields or other stable environments, with high quality of the surrounding site. They are of very high relevance in climate networks, even when the site is characterized by the presence of minor obstacles if those have been stable over time. Another problem that arises, even for such representative historic stations, is the frequent issue of knowing the exact instrument model, i.e. information related to type of instrument, exposure/orientation and instrument shelter. Maintenance and calibration are also important. Documented maintenance procedures are essential to understand possible breaks in time-series homogeneity and its biases. Few sites, however, have calibrations carried out over their full history.

3 | KEY DIFFERENCES AND COMMON ASPECTS BETWEEN WEATHER AND CLIMATE DATA

Many ground-based stations now associated with climate-relevant networks originated as meteorological stations for weather monitoring and forecasting. In most countries, most of the climate-relevant observing stations are still now managed by NMHSs, even in countries where a different organization has responsibility for climate data management and analysis, such as the US (NWS manages the network, NCDC/NCEI manage the data) and New Zealand (MetService and NIWA manage networks with NIWA managing and curating climate data).

The goal for climate-relevant observations is to describe the long-term aspects of the climate system, while the focus of weather observations is to describe the current state of the atmosphere and its short-term variations. This difference in purpose is reflected in different requirements for observing weather and climate. To evaluate and understand the climate and, particularly, climate variability and change, homogeneous, continuous and good quality observations for a particular climate element are required to derive satisfactory climatological references. Furthermore, a climate observation needs to be associated with an adequate set of metadata that will provide users with information, about how, where, when and by whom the observation has been recorded, and on how it should be interpreted and used. The GCOS climate monitoring principles (WMO, 2019), endorsed by the former CCI, were developed to minimize impacts of inhomogeneities. The minimum time resolution of data is also a point of difference between weather and climatological data. In the modern era, weather forecasting applications have limited use for data with daily time resolution. However, data with only one or two observations per day, as long as they are consistent and reliable, can still be very useful for climatological purposes.

Considering the above requirements, not all weather-related data are suitable for climatological purposes. This can be due to several reasons:

- The length of record is (or is still) too short, although, in some circumstances, a well-equipped station can still contribute to validate other stations in a tiered approach, or be used in gridded analyses.
- The station is equipped with low quality instruments or is located in a site characterized by the presence of nearby obstacles affecting the representativeness and/or stability of the measurement results.
- Metadata are inadequately documented.
- Lack of maintenance is such as to introduce artificial biases in the data series (due to unrecognized drift of sensors not periodically verified or calibrated, or due to instrument change without specific procedures of comparison between old and new systems).
- Biased by changes across time of the observing procedures and schedules.
- Influenced by changes in the local environment and station surroundings and, therefore, too biased to reliably support climate assessments, products or services.
- Other methodological aspects. (An example is that current synoptic observations require measurement of maximum temperatures between 6 AM and 6 PM only and minimum temperatures between 6 PM and 6 AM only, thereby missing 12 h of extreme observations of a 24-h day; such practice is insufficient for climate and extremes analyses).

The differences between climate-relevant and weather observations lead to the need, clearly expressed by WMO and GCOS and in general by the climate community, to make available observing stations and networks designed for the purpose of originating data for evaluating and understanding the climate, although high level meteorological observations can still serve to generate data suitable for the same purpose.

4 | DEFINITIONS AND REQUIREMENTS

Considering the lack of agreed definition and standardized requirements for climate reference data and installation, the first objective of this work has been the discussion and proposal of a clear definition, to be presented to the WMO Editorial Board, for adoption in vocabularies and regulatory material. The structure of the definition is now organized in two parts: a concise and clear definition and a number of notes, to clarify key aspects. The group proposes the two following definitions of climate reference data and climate reference station.

Climate reference data	A series of traceable measurement results able to quantify the variability and change of climate-relevant variables. *Notes*: • The result of the measurement is a single record of the observed parameter, while the overall measurand is the variability and change of the variable. • To be traceable, a measurement result requires that each instrument involved in the measurement process is related to a reference standard of the System of Units (SI) or other standards through a documented unbroken chain of calibrations. • The absolute requirement of a measurement is that it be made in such a way that after accounting for all sources of uncertainty it can be concluded that the true value of the measurand lies within the reported uncertainty interval with specified confidence. The result of a reference-grade measurement is such that it can be used to improve the quality of other (lower-tier) measurements.
Climate reference station	A climatological reference station is an instrumental installation able to generate climatological reference data.

(Continues)

Notes:

- Measured data must be continuous and representative of the local environment.
- The station must be stable in its location and siting characteristics for decades and equipped with top quality instrumentation.
- The instrumentation needs to be well maintained with regular maintenance and calibration to constantly keep documented data traceability.
- Changes in instrumentations must be limited, motivated and documented. Parallel observation periods must be planned prior to any instrumental change.
- All metadata on siting, methods of observation and any changes that may have occurred in these must be documented.

A clear definition of the 'measurand' is also a key aspect in every scientific investigation. Sometimes it is an underrated issue generating difficulties in implementing measurement procedures, including instrumental capabilities. In climatology, normally the measurand is not the single measured record, but is the variation over time of the interested quantity. Such a relative definition of the measurand helps reduce the impact of the uncertainty components on every single record, transferring the overall uncertainty in the measurand (the trend) to statistical methods, under the condition that the measurand is stable and representative of the climate signal under study.

A clear distinction between instrumental stability over the long term and uncertainty in a single measurement has to be considered when developing homogenization and harmonization algorithms. Instrumental uncertainties have typically been neglected in most climate trend analyses. This is acceptable only if there is total trust in the stability of the measuring system and when comparability on trends in different environmental conditions is not required. Now that the existence of a warming climate trend is well established, accuracy in global and local evolution of the observed quantities becomes the key factor under investigation. This is when the accuracy in the measurements becomes fundamental. Considering uncertainty supports reaching traceability, and thus comparability, and the more accurate the measurement the less the time required to identify trends. Completely stable measuring systems and sites, together with traceable measurements, are therefore a fundamental step now required in climate science.

As a further aspect in the definition of measurand, measuring techniques can also introduce problems in interpreting the measurement results. A key example is measuring air temperature, the key variable observed in climate studies for comparing extreme values and anomalies. While it might seem obvious that a thermometer measures the temperature of the air, this is not physically what happens. Indeed, a contact thermometer gives an indication of its heat equilibrium at that time, under those specific conditions of heat exchange. This implies that the temperature value is obtained under a fundamental non equilibrium between the radiative, convective, contact heat transfers and condensation, icing and evaporation phenomena. This is a fundamental thermodynamic principle that turns into large contributions to the uncertainties and even corrections in air temperature measurements. Adopted techniques evolved in time, to reduce errors and the magnitude of the quantities of influence, with reducing the influence solar radiation the first priority. Hence, the result of the measurement is a compromise: a temperature value as representative as possible of the air temperature at that time in those conditions. When reference observations are needed in climatology to compare, harmonize other values, or to represent trends of an area, then a definition of what a 'reference' temperature value is required. This requires a definition of the measurement conditions to be considered as reference. For example: the air temperature value is taken at zero radiation or extrapolated at zero total radiation; under steady conditions, or at air speed minimum of, or even at extrapolated zero wind speed and so on. In the end, it is the definition itself of the measurand, the atmospheric air temperature, that is still missing (Merlone, 2021).

To investigate and solve this problem, a specific task group of the Bureau International des Poids et Measures (BIPM) Consultative Committee of Thermometry (CCT) was formed in 2021, with the objective to propose a definition of air temperature, identify all uncertainty components and prescribe specific guidelines for calibration of thermometers in air.

5 | TECHNICAL FEATURES OF CRSs ACCORDING TO CLIMATE REFERENCE DATA DEFINITION

A climate reference station needs to be equipped with high-quality instrumentation, procedures and technology that will generate the best achievable estimates of values of the observed quantities (see Section 5.1). The station shall be located at a site that is representative of its regional climate and adhering to strict siting requirements (GCOS 226). Locations shall be chosen where no significant changes in the surrounding areas are expected for the next century. Site and equipment maintenance need to be a priority to avoid degradation and uncontrolled instrumental drifts over time due to environmental conditions that may affect the measurements. Instruments shall be calibrated at regular intervals, to correct for drift and maintain traceability (see Section 5.2). Changes in instrumentations, procedures and technology shall be limited, motivated and documented. Parallel observation periods shall be planned prior to any change, and the time series of both new and old stations retained indefinitely (see Section 5.3). Uncertainty on measurement results shall be evaluated including calibration uncertainty, instruments characteristics and field environmental influences (see Section 5.4).

5.1 | Reference-grade measurements and quantities of influence

The core requirement for a climate reference station is a set of instruments and procedures to reliably and consistently measure as many as possible of key surface based atmospheric essential climate variables (ECVs): air temperature, water vapour (relative humidity), surface radiation budget, wind speed and direction, pressure and precipitation (liquid and solid). Originally, climatology mainly relied on evaluating near surface air temperature and precipitation trends from meteorological observations. These two variables are considered as mandatory for CRS, except for precipitation in climatic zones where precipitation is very low.

Other quantities have been subsequently included in the assessment of climate evolution and measurements of such further variables are strongly recommended at CRS: relative humidity, direct total solar radiation (instead of the more complex surface radiation budget), wind speed and direction (at 10 m), air pressure and soil temperature.

Additional variables are also advised such as snow cover, land surface temperature, soil temperature and soil moisture, and surface albedo, especially as they may relate to any terrestrial ECVs that may be measured or to specific effects on instruments. River discharge and ground water shall also be measured if existing in the vicinity. Measurements of permafrost temperature profile, and active layer thickness and other cryospheric measurements, are recommended where permafrost is present. Air quality observations can also be part of reference-grade measurements at a CRS.

Measurements of associated quantities of influence (AQI) are also required: these are needed to produce a reference measurement of another quantity. For example, to have reference air temperature measurements, associated values of total direct and reflected solar radiation, relative humidity, precipitation (and possibly rain temperature) and wind (at the height of thermometers and rain gauges) are also necessary, as detailed in Table 1. These AQIs do not need to be of reference quality (e.g., lower maintenance and recalibration requirements, no overall uncertainty budgets quantified). However, a Quality Check (QC) shall be constantly applied to those instruments used to generate records of AQI at a CRS. The QC shall follow the minimum requirements prescribed for a field verification (WMO 2018/2021).[1]

When an AQI is also one of the reference variables measured at the station, then the same recorded values can be used as values for the associated quantity of influence. In the above example for air temperature, the measurement of precipitation as a mandatory variable will be of reference level, but the remaining AQI do not necessarily need to be.

5.2 | Traceability to establish data comparability in time and space

Data comparability in time and space is the main deliverable of an installation providing measurement results for climatology. The fundamental prerequisite is that recorded data are the result of measurements made by means of traceable instruments. Establishing documented traceability to a measurement process turns the observations into a robust set of values, comparable in time and space and among other different (but traceable) measurements. When climate reference stations are required to validate other systems, such as remote sensing, then the traceability is essential, to guarantee that the response of a system under validation is correctly checked. This also applies when reference stations are used to check other stations in a network.

Instrument traceability shall be established at installation, through calibration prior to deployment, and preserved in time, through procedures involving periodic tests, checks and re-calibration. Regular maintenance and calibration are key aspects of metrological best practice for reference networks and stations. Frequencies of on-site maintenance for a CRS are reported in Table 2.

As a general rule:

- Regular field inspection shall be made every 6 months and/or at need, following, for example, extreme events or evidence of malfunctioning. The inspection can lead to repair / substitution of instruments. In the case

TABLE 1 Schematic table for measured parameters and associated quantities of influence (auxiliary measurements) (from GCOS 226).

	Air temperature	Relative humidity	Solar radiation	Wind speed and direction	Air pressure	Precipitation	Rain temperature	Soil temperature	Soil moisture
Air temperature		**X**	**X**	**X**		**X**	**X**	**X**	
Relative humidity	X		X	X	X	X			X
Solar radiation	X	X				X			
Wind speed and direction	X					X			
Air pressure	X			X					
Precipitation	**X**			**X**					
Soil temperature						X			
Soil moisture						X			

Note: Primary measurements (first column) require other measurements to detect the respective influencing factor. The lines in bold are mandatory variables, for a station to be included in GSRN: the associated auxiliary variables are also mandatory.

of manual observers, refresher training should be provided, especially if there is evidence of systematic errors in the data.

- Field verifications against travelling equipment shall be performed every year, to check instruments' correct working conditions (WMO, 2021). The verification requires a threshold limit for a pass/fail evaluation. Verification failures shall be followed by an immediate recalibration.
- Calibration should be repeated every 24 months. Longer time intervals shall only be considered if warranted by the instruments' quality, their exposure, the environmental conditions of the site, their ageing and the prescriptions from the manufacturers.

Standardized calibration procedures shall be adopted by all network stations to document instrumental traceability. To avoid interruption in the data series during the calibration process, replacement of sensors shall be adopted: (a) in a circular way (calibrated sensors are used at every recalibration to replace sensors to be calibrated) or (b) by temporarily substituting the sensor under calibration with another calibrated one. Field calibration can also be performed in exceptional cases, when difficulties are met in removing the sensors (for example in the case of very old historical systems or in special conditions). Field calibration shall be organized in order to cover the whole range of variability of the sensor, and the likely range of climate conditions, and to establish traceability as for the laboratory calibration. Uncertainties can be larger for field calibration.

5.3 | Managing instruments change

Regardless of whether by choice or necessity, the challenge of an instrument change is to manage all transitions in such a manner that the effects upon long-term series continuity are minimized and the associated uncertainty well understood.

A fundamental metrological principle stipulates that replacing one instrument with another at the same location should pose no problem when the old and the new instrument are both fully traceable to standards. In reality this idealized concept is not fully met since different instruments or sensors may react differently to the same external environmental factors.

Under the conditions that both instruments are calibrated, the change does not affect the accuracy of data, but only the uncertainties associated with the quantities of influence (for example a different influence of radiation on a changed temperature sensor). For compensating such effects, parallel observations of sufficient length and

TABLE 2 Typologies and frequencies of on-site maintenance for a CRS.

Activity	Time interval	Provides	Results in
Inspection	6 months or at any occurrence	Repair/substitution	System OK
Check/verification	12 months	Pass/fail	Tolerance limit
Calibration	24 months	Establish traceability	Uncertainty

encompassing a suitable range of different climate types are required, meeting the whole (or the majority of the) range of variability of each quantity. Such prescription is easily applicable to any station or network but is at present rarely documented in information or metadata.

5.4 | Uncertainty evaluation

'When reporting the result of a measurement of a physical quantity, it is obligatory that some quantitative indication of the quality of the result be given so that those who use it can assess its reliability. Without such an indication, measurement results cannot be compared, either among themselves or with reference values given in a specification or standard. It is therefore necessary that there be a readily implemented, easily understood, and generally accepted procedure for characterizing the quality of a result of a measurement, that is, for evaluating and expressing its uncertainty.' (GUM 0.1).

In line with the International Vocabulary of Metrology (VIM, JCGM 200:2012; JCGM, 2012) and the Guide to the Expression of Uncertainty in Measurement (GUM, JCGM 100:2008; JCGM, 2008), the required measurement uncertainty includes all quantifiable uncertainties. The uncertainty is considered the range within which the true state of the measurand will plausibly reside. Uncertainty should be expressed in units of two standard deviations. The GUM prescribes two main methods of uncertainty evaluation: Type A and Type B, which are referred to statistical methods (A) or to instrumental features (B).

A new way of approaching uncertainties is then required, putting the instrument as focus, and leaving the uncertainty components due to field effects and environmental factors (siting) as separate components, to be determined site by site.

In general, and specifically for meteorological observations, four main sources of uncertainty can be identified:

1. *Instrument components*: including sensor properties (time constant, drift, sensitivity, etc.), calibration, coupling with auxiliary structures such as solar screens, dataloggers. These can be evaluated in laboratory and/or by the manufacturer.
2. *Effects of the environment on the instrument response*: including the effect of precipitation (overcooling), wind, solar radiation, backward radiation, condensation, icing, exposure to extreme limits of use (polar, deserts). These can be evaluated in the specific field and environmental conditions, or by intercomparisons.
3. *Uncertainty on the measurand*: how well the measurand is defined and is representative of the quantity being monitored. This includes siting classification).
4. *Statistical components*: uncertainties in the statistical processes adopted to obtain a single value (for example mean calculation over a certain interval) or in detecting a specific value (for example a maximum or anomalies).

A new way of approaching uncertainties is then required, putting the instrument as focus, and leaving the uncertainty components due to field effects and environmental factors (siting) as separate components, to be determined with further analysis of the site characteristics.

The table in Appendix (Table A1) reports a list of some key sources of uncertainty that contribute to each of these components. Each of the items can generate a component of the overall uncertainty budget, which is calculated through the Gaussian propagation. It is worth to note that the calibration uncertainty is only one among the several components of the overall measurement uncertainty and most frequently not even the largest one. For further guidance on the implementation of uncertainty evaluation see ISO/IEC (2008)/JCGM (2008).

Manufacturers are constantly working to protect the instruments from the effects of the environmental exposure, reducing and minimizing the effects on the sensors' response. Testing and calibration laboratories are requested to control the internal environmental conditions (such as humidity and temperature) in order to reduce, to a negligible contribution, the effects of the environment on the readings of the instruments. Instruments are therefore calibrated in a stable environment to better link their readings to those of reference standards.

Among the various sources of uncertainty in field measurements, the effects of the environment on the instrumental response are among the larger contributions. The variability of the environmental conditions is larger, precipitation (overcooling), wind, solar radiation, backward radiation, condensation, icing, and exposure to extreme limits of use (polar, deserts) in different ways influence the accuracy of the various sensors.

It is impossible to make a field reading immune from these effects. Such influencing factors shall therefore be considered in the overall measurement process, including the evaluation of the uncertainty budget.

Two cases are possible:

a. If the effect can be numerically evaluated, then a correction can be applied to the readings, and a contribution to the uncertainty shall be included, in terms of uncertainty of the correction.
b. If the effect of an influencing factor cannot be corrected, this directly becomes a component of uncertainty which in principle is larger than the uncertainty of the correction.

The effects can be different for the same kind of instrumentation, depending on the environmental conditions and range of variability. Specific experiments should then be made to best evaluate the contributions to the measurement accuracy due to the environmental conditions met at that specific site by that specific instrumentation. Prescriptions do not yet exist, but skilled staff or manufacturers should adopt specific methods to evaluate the effect of the influencing quantities. In most cases such experiments require a couple of identical systems, one exposed to the effect under test and the other protected at best. For example, the overcooling effect of the rain on thermometer readings can be evaluated by deploying two identical instruments (thermometer and shield) in the same site under the rain and protecting one of the two. A detailed protocol to perform such test is published in Musacchio et al. (2019), which assesses the effect of the reflected radiation from a snow-covered surface on the accuracy of thermometers.

6 | PROTOTYPE CRS FOLLOWING THE DEFINITIONS AND REQUIREMENTS

With the purpose of studying, comparing and confirming the prescribed definitions and requirements, a prototype for a CRS has been developed and installed in May 2023 by INRiM.

6.1 | Location and siting

As extensively described, the station shall be located at a site that is representative of its regional climate and adhering to strict siting requirements, according to the GSRN specifications (GCOS, 2019). Locations should be chosen where no significant changes in the surrounding areas are expected, but also should be as uniform as possible, without obstacles in the vicinity of the instrumentation that could induce biases in the measurements. Other characteristics have also been considered, such as security, ease of access for verification and maintenance routine. An example of data product, for the GCOS defined ECV 'Air temperature near surface' is also provided, aligned to the GSRN requirements.

WMO (2018) identifies some of the sources of error due to the site features and proposes a classification scheme. Each site will need to be large enough to house all instrumentation without adjacent instrumentation interfering with one another, with no shading or wind-blocking vegetation or localized topography, and at least 100 m from any artificial heat sources. The 100 m distance is based on a precautionary evaluation to avoid the effect of obstacles, although recent studies (Coppa et al., 2021) demonstrated that at 50 m from the measuring point most of the influences are reduced to negligible effects, within the instrumental uncertainties. This allows including among the climate reference stations of a network also those installations placed in sites with the presence of minor issues at less than 100 m (for example paths or road that are not asphalt or concrete, isolated trees, gentle slope, etc.).

Based on the considerations above, a site has been selected, taking into account specific requirements such as: the proximity to INRiM (ease of access to the site, reduce carbon impact and person time) and the possibility to sign a formal agreement for the use of the land in a public area to avoid possible changes in the use of the terrain with new buildings or similar.

The identified site, shown in Figures 1 and 2, is located in the 'Park of Stupinigi', a public area 3 km from INRiM, in the municipality of Nichelino (Piemonte, Italy). The area is flat, covered by natural grass, and no obstacles are present within 100 m radius, apart from a small dirt path. Thanks to its characteristics and according to the WMO siting classification, the site is in class 1. In the Köppen classification, the climate zone corresponding to the site is Cfa (continental, no dry season, hot summer).

6.2 | Instrumentation

Following the requirements in terms of data quality for primary measurements (at reference level) and need to measure the so-called AQI, the configuration of the

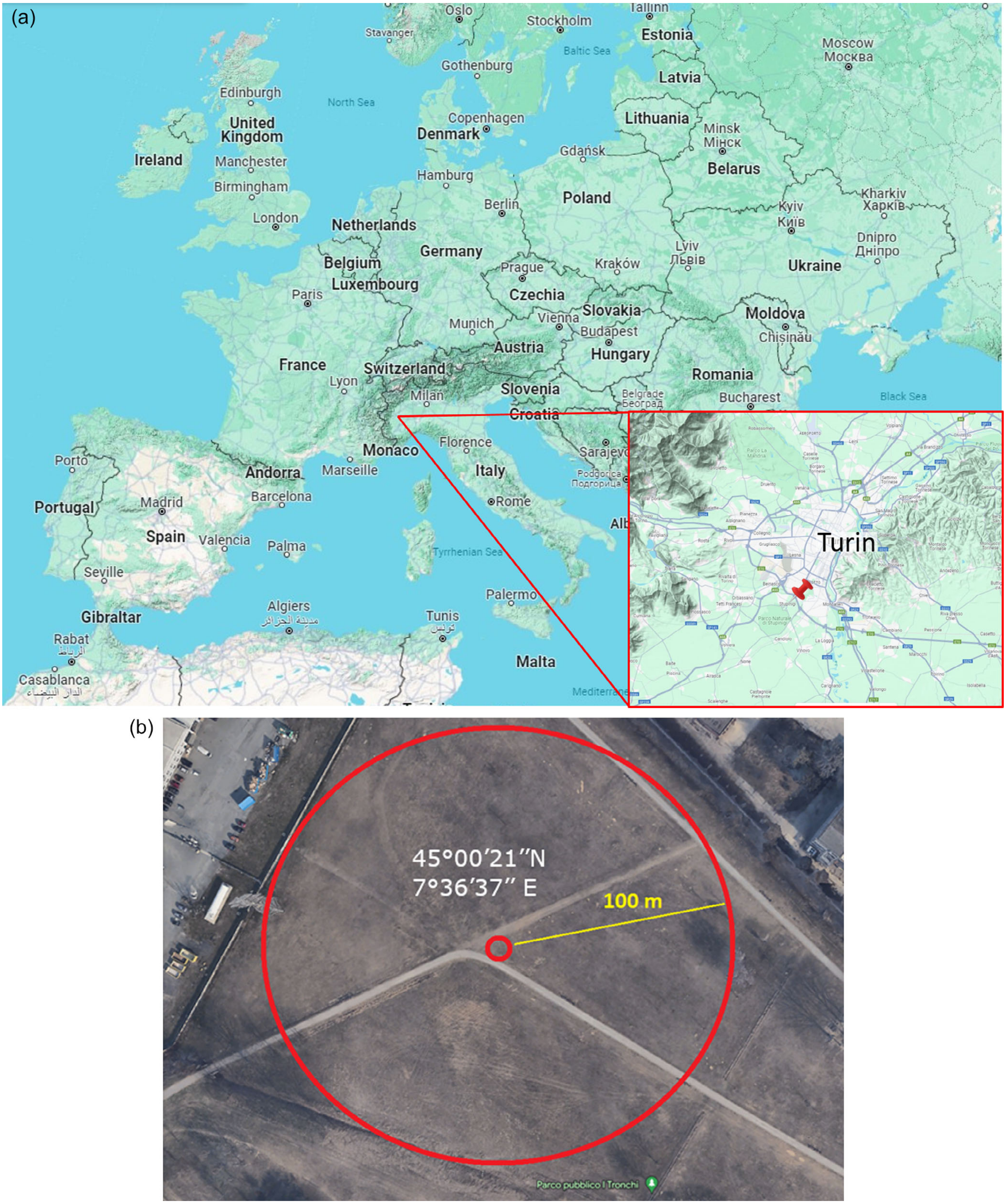

FIGURE 1 (a) Google Maps picture identifying the location of the site selected for the reference station. (b) Google Earth picture of the site, with the 100 m radius free of obstacles around the station, positioned in the centre of the red circle. The coordinates of the station are reported in the picture. It is worth mentioning that all trees have been cut in the area internal to the red circle, according to the WMO prescriptions, in order to reduce uncertainties in the data representativeness and measurement errors. [Colour figure can be viewed at wileyonlinelibrary.com]

INRiM Climate Reference Station, shown in Figure 2, is described in the following Table 3.

The station is based on the instruments, reported in Table 3, and an associated datalogger with transmitter.

FIGURE 2 Picture of the Climate Reference Station. [Colour figure can be viewed at wileyonlinelibrary.com]

The required electric power is generated by solar panels and a battery, in order to test a configuration that can be installed anywhere, including remote locations (polar sites, high mountains, deserts). Measurements are recorded continuously with no missing data and all instruments proved their quality and stability.

6.3 | Measurements data product

In the second half of 2023, the INRiM Climate Reference Station was activated and data started to be recorded and transmitted. The datalogger records values from each sensor at 10 s intervals and transmits data buffers at 1 min frequency. An example data product is reported in Table 4. The logger reads and transmits raw data only: for example, temperature values are recorded as resistance values, since the station uses PT100 thermometers. The same applies for the other quantities. At a post processing level, raw data is transformed in processed values, reported in the associated units. For example, temperature is translated into degrees Celsius by applying the resistance-to-temperature calibration curve's coefficients, while radiation is changed from millivolts to watts applying the pyranometers' sensitivity. All data is also associated to the instrument uncertainty, given by the calibration uncertainty and all other parameters characterizing the response of each instrument.

For each observed quantity, all other associated quantities of influence are recorded and presented at each time. The final values are reported at 30 s intervals, applying a mean over the 10 s values and including the deviation in the overall uncertainty.

Table 4 was presented at the meeting of the Task Team GSRN (September 2023) and will be considered as the basis for the GSRN data product format for air temperature and precipitation.

TABLE 3 Technical configuration of the INRiM's CRS. For each primary quantity and quantity of influence is specified the instrument and the corresponding calibration uncertainty.

Variable	Instrument	Calibration uncertainty
Temperature (reference)	PRT 100 (4 wires connection)	0.012°C
Temperature – relative humidity (AQI)	Vaisala HMP 155	0.05°C 3% RH
Precipitation	SIAP TP200	Total: 2%–10% Intensity: 2%
Air pressure	Paroscientific DIGIQUARTZ	10 Pa
Wind speed and direction (reference)	GILL HS50 3 axis ultrasonic anemometer	Speed: <1 m/s Direction: <1°
Wind speed and direction (AQI)	Gill Windsonic	Speed: 2% @12 m/s Direction: 2° @12 m/s
Solar radiation (direct and reflected)	2 Hukseflux LP02 Second class pyranometers	<1.8%

TABLE 4 An example of data product for near surface air temperature reference data.

Date and time	*T* (°C)	RH (%)	Direct solar rad (W/m^2)	Refl. solar rad (W/m^2)	Wind direction (°)	Wind speed (m/s)	Prec. (mm)	*T* rain (°C)	*U* Instr	*u* RH	*u* Rad	*u* Wind	*u* Prec	*u* Train	*uT*	*UT*
10 October 2023; 1:11 PM	18,114	57	126	13	6	1.98	0	—	**0.012**	0.011	0.063	0.010	TBD	TBD	0.071	**0.14**
10 October 2023; 1:11 PM	18,224	57	126	13	35	3,03	0	—	**0.012**	0.011	0.063	0.015	0	0	0.065	**0.13**
10 October 2023; 1:12 PM	18,235	57	127	13	33	1.99	0	—	**0.012**	0.011	0.064	0.010	0	0	0.071	**0.14**
10 October 2023; 1:12 PM	18,173	58	127	13	46	1.75	0	—	**0.012**	0.012	0.064	0.009	0	0	0.072	**0.14**
10 October 2023; 1:13 PM	18,096	57	127	13	58	1.49	0	—	**0.012**	0.011	0.064	0.007	0	0	0.073	**0.15**
10 October 2023; 1:13 PM	18,011	56	127	13	57	1.53	0	—	**0.012**	0.011	0.064	0.008	0	0	0.073	**0.15**
10 October 2023; 1:14 PM	18,019	57	127	13	61	1.45	0	—	**0.012**	0.011	0.064	0.007	0	0	0.074	**0.15**
10 October 2023; 1:14 PM	18,081	58	127	13	35	1.62	0	—	**0.012**	0.012	0.064	0.008	0	0	0.073	**0.15**
10 October 2023; 1:15 PM	18,045	57	128	13	32	1.74	0	—	**0. 0.012**	0.011	0.064	0.009	0	0	0.073	**0.15**
10 October 2023; 1:15 PM	18,045	57	128	13	30	1.69	0	—	**0.012**	0.011	0.064	0.008	0	0	0.073	**0.15**
10 October 2023; 1:16 PM	18,132	56	129	13	25	1.45	0	—	**0.012**	0.011	0.065	0.007	0	0	0.074	**0.15**
10 October 2023; 1:16 PM	18,183	56	130	13	29	1.03	0	—	**0.012**	0.011	0.065	0.005	0	0	0.077	**0.15**

Note: From the left, after the date and time column, the value of air temperature is reported, already corrected by the instrument calibration. The following seven columns respectively report the values of humidity, direct and reflected total solar radiation, wind direction and speed, precipitation (Prec) and rain temperature (*T*rain). The 10th column reports the instrumental uncertainty (*U* instr) derived from sensor's calibration uncertainty and its characteristics (resolution, drift, sensitivity, etc.) (in bold). The following columns report the uncertainty contributions, *u*, to the air temperature overall measurement uncertainty due to the quantities of influence in columns 3–9, associated to each value. The final two columns report the combined uncertainty (*uT*) and the expanded uncertainty (*UT*) (in $k = 2$) (in bold). Note that the uncertainty values in columns 11–15 are just an estimation since at present most of those contributions of uncertainty are still not possible to evaluate.

7 | CONCLUSIONS AND DISCUSSION

If magic could bring a climatologist back in the past bringing the modern knowledge, the first requirement would be to install and maintain stations able to detect climate signals, aside the existing meteorological observations, and protect them from all those types of changes which disturb the data homogeneity. At the same time a metrological requirement would address instrumentation and sensors would be made traceable to standards through regular calibrations. Metrology keeps track through time of the changes in unit definitions, their realization by means of primary standards, and their dissemination prescribed by calibration procedures. If stations were kept stable and its surroundings unchanged noticeably, traceable data would have achieved full comparability in time and space at the level of the calibration uncertainty. The maintained traceability would have also accounted for instrument change due to renewing technologies.

The main efforts in identifying reference-grade stations among existing historical sites is mainly a matter of evaluating the information which already exists. On the other hand, the creation of 'ad hoc' climate reference stations and sites are a process allowing an active role in managing the implementation and data production. New sites hosting a CRS shall be equipped with top level new instrumentation, which are aligned with the GCOS GSRN prescriptions and with the OSCAR top class requirements. Moreover, climate reference stations shall be organized in a way that upgrading instruments, due to malfunctioning, end of life, obsolescence and renewal does not compromise the continuity of the series together with the value of the long-term data records.

To summarize, three typologies of high-quality stations for climate observations can be identified:

a. Long term stations (including urban environments: these are of fundamental importance in climatology, although they cannot provide reference grade measurement quality);
b. Meteorological or climatological stations already existing at sites with top quality characteristics intending to undergo a process for improving instrument quality and assessing traceability;
c. New stations created at sites identified for the purpose and equipped with the required instrumentation to generate traceable reference grade data; this includes stations in specific environments, such as polar regions, high mountains, deserts.

Certain sites (new or existing) may also act as special centres for research and development on instrument field validation and intercomparison, testing of new and evolving systems, evaluation of drifts and maintenance interval and pros and cons in specific and challenging uses. Establishing such special sites among networks would be strongly encouraged.

All sites shall guarantee recording of data at the prescribed frequency and where possible online data streams or frequent manned download of data, where transmission services (internet, gsm) are not present or not economically sustainable (for example in remote areas).

Ideally, all observations in a reference time series will be traceable to a standard, and all sources of uncertainty will be quantified. In practice, few, if any, existing reference time series meet this standard, but it should be a target that identified reference time series shall meet into the future. For generating climate reference data, accurate measurements of ECVs require two fundamental characteristics: (a) to be traceable measurement results, thus recorded by calibrated and constantly maintained instruments and (b) to be accompanied by the values of all identified associated quantities of influence (the other ECVs influencing the response of the instrument or the measurand itself). The data product must then be a complete set of values, including the uncertainties related to both the primary measurement and the effect of the AQIs on its value and uncertainty.

Nevertheless, the complete evaluation of the uncertainties and the corrections (and associated components of uncertainty), due to the effects of the AQIs, at present is almost impossible for most of the main ECVs of interest in climatology. Reference data products should in any case include, together with the records of the main ECVs, all the values of the AQIs (see Table 4), recorded at the same time. Research is progressing on the evaluation of the uncertainty by the WMO, Metrology Institutes, Universities and Research Groups, also motivated by the recent WMO INFCOM-2 decision 7.4(2) of February 2022, requesting to '*promote organize and coordinate experiments and studies necessary to improve uncertainty evaluation and traceability of measurements*', addressing joint efforts to improve our knowledge on measurement data quality. Together with accurate values of ECVs, making as much metadata as possible available, in terms of instrument characteristics, calibrations, system uncertainty and providing all the values of the identified associated quantities of influence at the time of each record, this work has enabled substantial progress but is still insufficient to allow the complete evaluation of the uncertainty in climatological data. Nevertheless, generating climate reference data series, complete of all the values, will make available the required information to improve homogenization and comparability in the future, but also back in time. This will be a fundamental step in improving our capability and ability to capture climate trends.

AUTHOR CONTRIBUTIONS

Andrea Merlone: Conceptualization; writing – original draft; writing – review and editing; supervision. **Gaber**

Beges: Conceptualization; writing – original draft; writing – review and editing. **Alberto Bottacin:** Writing – review and editing. **Manola Brunet:** Conceptualization; writing – original draft; writing – review and editing. **Alba Gilabert:** Conceptualization; writing – original draft; writing – review and editing. **Drago Groselj:** Conceptualization; writing – original draft; writing – review and editing. **Andrew Harper:** Conceptualization; writing – original draft; writing – review and editing. **Peer Hechler:** Conceptualization; writing – original draft; writing – review and editing; supervision. **Mirjana Ivanov:** Writing – original draft. **Chiara Musacchio:** Conceptualization; writing – original draft; writing – review and editing. **Blair Trewin:** Conceptualization; writing – original draft; writing – review and editing. **William Wright:** Conceptualization; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS

This work is being developed within the framework of the Climate Reference Station – CRS – EMPIR 19SIP03 project, which has received funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Chiara Musacchio https://orcid.org/0000-0001-7473-3678
Blair Trewin https://orcid.org/0000-0001-8186-7885

ENDNOTE

[1] In addition to these, the document 'Field Verification of Meteorological Instruments and Sensors – A Guide to Best Practice' being published by SC-MINT ET QTC has been adopted. It includes minimum estimation of uncertainties in the field verification.

REFERENCES

Begert, M. & Frei, C. (2018) Long-term area-mean temperature series for Switzerland—combining homogenized station data and high resolution grid data. *International Journal of Climatology*, 38, 2792–2807.

Caussinus, H. & Mestre, O. (2004) Detection and correction of artificial shifts in climate series. *Applied Statistics*, 53, 405–425.

Coppa, G., Quarello, A., Steeneveld, G.J., Jandric, N. & Merlone, A. (2021) Metrological evaluation of the effect of the presence of a road on near-surface air temperatures. *International Journal of Climatology*, 41, 3705–3724.

Domonkos, P. (2011) Adaptive Caussinus-Mestre algorithm for networks of temperature series (ACMANT). *International Journal of Geosciences*, 2, 293–309.

GCOS. (2019) GCOS 226–2019. Surface reference network (GSRN): Justification, requirements, siting and instrumentation options. https://gcos.wmo.int/en/news/gcos-surface-reference-network-proposal.

JCGM. (2008) JCGM 100:2008. Evaluation of measurement data – guide to the expression of uncertainty in measurement. https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.

JCGM. (2012) JCGM 200: 2012. International vocabulary of metrology – basic and general concepts and associated terms (VIM 3rd edition). https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf.

Jones, P.D. & Lister, D.H. (2009) The urban heat Island in central London and urban-related warming trends in Central London since 1900. *Weather*, 64, 323–327.

Merlone, A. (2021) Thermal metrology for climate: a review of projects, activities and open issues. *Measurement Science and Technology*, 32, 102001. Available from: https://doi.org/10.1088/1361-6501/abf375

Musacchio, C., Coppa, G. & Merlone, A. (2019) An experimental method for the evaluation of snow albedo effect on near surface air temperature measurements. *Meteorological Applications*, 26, 161–170. Available from: https://doi.org/10.1002/met.1756

Thorne, P.W., Diamond, H.J., Goodison, B., Harrigan, S., Hausfather, Z., Ingleby, N.B. et al. (2018) Towards a global land surface climate fiducial reference measurements network. *International Journal of Climatology*, 38, 2760–2774.

Thorne, P.W., Menne, M.J., Williams, C.N., Rennie, J.J., Lawrimore, J.H., Vose, R.S. et al. (2016) Reassessing changes in diurnal temperature range: a new data set and characterization of data biases. *Journal of Geophysical Research: Atmosphere*, 121, 5115–5137. Available from: https://doi.org/10.1002/2015JD024583

Trewin, B., Braganza, K., Fawcett, R., Grainger, S., Jovanovic, B., Jones, D. et al. (2020) An updated long-term homogenized temperature data set for Australia. *Geoscience Data Journal*, 7, 149–169. Available from: https://doi.org/10.1002/gdj3.95

USCRN. (2020) United States climate reference network, accessed November 2020. https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/us-climate-reference-network-uscrn.

WIGOS. (2019) Manual on the WMO integrated global observing system (WMO-No. 1160). https://library.wmo.int/doc_num.php?explnum_id=10145.

WMO. (2019) Manual on the high-quality global data management framework for climate, WMO-No. 1238. https://library.wmo.int/index.php?lvl=notice_display&id=21686#.X5lZrUeg9PY.

WMO. (2020) WMO No. 1245, Guidelines on homogenization. https://library.wmo.int/index.php?lvl=notice_display&id=21756#.X5mIBUeg9PY.

WMO. (2021) WMO 2018/2021. WMO guide No. 8, Guide to instruments and methods of observation. https://community.wmo.int/en/activity-areas/imop/wmo-no_8.

How to cite this article: Merlone, A., Beges, G., Bottacin, A., Brunet, M., Gilabert, A., Groselj, D., Harper, A., Hechler, P., Ivanov, M., Musacchio, C., Trewin, B., & Wright, W. (2024). Climatological reference stations: Definitions and requirements. *International Journal of Climatology*, 1–15. https://doi.org/10.1002/joc.8406

APPENDIX A

TABLE A1 List of some key sources of uncertainty contributing to the overall budget. The first and the second columns report the general sources of uncertainty and useful references. The corresponding components are listed in more details in the third column.

	Sources of uncertainty	Components
1. *Instrument* References: • Measurement Quality Classification Scheme • Guidelines on calibration • GUM	Measurement System and Calibration	• Construction quality • Resolution • Instrument and logger calibration • Linearity • Hysteresis • Time constant • Drift with temperature • Sampling method • Sampling frequency • Processing algorithm • Digitization and rounding • Response time
	Instrument Coupling	• Radiation screen • Static pressure head • Rain gauge fence screen
	Maintenance and Verification	• Frequency of maintenance • Instrument and system drift with time • Instrument and system ageing • Instrument and system faults (that affect data but do not cause failure) • Cleanliness of instrument and site • Sensor mechanical stress during transport and operation
2. *Environment Effects* References: • Measurement Quality Classification Scheme • Scientific literature • GUM	Effects on instruments, not detectable in laboratory	• Evaporation of precipitation on screen (overcooling), • Wind effects on measurement, • Condensation on temperature instrument, • Solar radiation effects on measurement, including reflected radiation • Icing • Exposure to extreme limits of use
3. *Site* References: • Siting classification • Scientific literature	Effects of obstacles at less than 100 m from the measuring points	• Roads • Trees • Building • Water sources • Slopes
4. *Statistical components* References: • GUM	Type A uncertainties	• Datalogger sampling procedures, mean, standard deviation • Statistics on big data