Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications

2021 edition

WEATHER CLIMATE WATER

WORLD
METEOROLOGICAL
ORGANIZATION

WMO-No. 1269

Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications

2021 edition

WORLD
METEOROLOGICAL
ORGANIZATION

WMO-No. 1269

EDITORIAL NOTE

METEOTERM, the WMO terminology database, may be consulted at https://public.wmo.int/en/meteoterm.

Readers who copy hyperlinks by selecting them in the text should be aware that additional spaces may appear immediately following http://, https://, ftp://, mailto:, and after slashes (/), dashes (-), periods (.) and unbroken sequences of characters (letters and numbers). These spaces should be removed from the pasted URL. The correct URL is displayed when hovering over the link or when clicking on the link and then copying it from the browser.

WMO-No. 1269

Chair, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0) 22 730 84 03
Fax: +41 (0) 22 730 81 17
Email: publications@wmo.int

ISBN 978-92-63-11269-9

PUBLICATION REVISION TRACK RECORD

Date	*Part/ chapter/ section*	*Purpose of amendment*	*Proposed by*	*Approved by*

CONTENTS

ACKNOWLEDGEMENTS

These Guidelines were initially developed by climate data management staff within the Australian Bureau of Meteorology (BoM) and incorporate many of the methodologies used in BoM's effective quality management system software. The efforts of Meaghan Flannery, John Flannery and Lynda Chambers are particularly acknowledged. The initial draft of this publication was reviewed and edited by Reinhard Spengler (German National Meteorological Service (*Deutscher Wetterdienst* (DWD)) and William Wright (formerly of BoM), with the assistance of David Sinclair (BoM), who also added some additional tests, and Peer Hechler (WMO). An open peer review process followed, with comments sought from National Meteorological and Hydrological Services (NMHSs) and members of the WMO Expert Team on Data Development and Stewardship. Their comments considerably improved this document and are gratefully acknowledged.

1. INTRODUCTION

This publication replaces the WMO Technical Document *Guidelines on the Quality Control of Surface Climatological Data* (WMO/TD-No. 111) (WCP-No. 85) of May 1986. It provides a relatively high-level overview of the principles underpinning the effective quality assurance (QA) of climate data as well as considerations for the operational QA and quality control (QC) of meteorological data from surface observing stations at various stages of the data life cycle. It then proposes in the annexes a range of QA and QC tests classified as mandatory, recommended and optional. This publication is intended to be further developed and expanded over time to respond to new developments and to cover other sources of climatological data.

Data homogenization, although part of the climate data QA and QC continuum, is discussed separately in *Guidelines on Homogenization* (WMO-No. 1245) and is therefore not covered in this document. Other value-adding processes, such as the infilling of missing data and the disaggregation of cumulative precipitation or evaporation totals, are also beyond the scope of these Guidelines.

This document focusses on the surface observational data processed by National Meteorological and Hydrological Services (NMHSs); however, the underlying philosophy and principles also apply to other entities with a QC focus. These Guidelines are closely aligned with the publication *Climate Data Management System Specifications* (WMO-No. 1131), which specifies the functionality required to effectively manage climate data across the complete data life cycle from station installation to final archival.

In this document, QA refers to the process for maintaining a satisfactory level of quality in a data set or data collection so that the data available to potential users are sufficiently reliable and complete and can be used with confidence. QA should be understood to be an end-to-end process that extends from the climate archive back through the data transmission channels to the point of observation. The QA process should help diagnose whether any systemic patterns of error are present at any point during the data life cycle; these errors may arise from, for example, non-standard installation, poor observational technique, problems with the measuring technology or its software, maintenance failures, or any upstream processing issues (such as issues within the central processing unit of an automatic weather station (AWS)). Providing feedback to observation providers and rectifying issues identified are also part of the QA process. QC in this document refers to the tools and practices employed to verify whether a reported data value is representative of what was intended to be measured and has not been contaminated by unrelated factors (see *Guide to Instruments and Methods of Observation* (WMO-No. 8)). QC is the process of ensuring that errors in the data, or in the continuity of the data, are detected by checking the data to assess representativeness in time and space and internal consistency, and by flagging any potential errors or inconsistencies.

Although climate services require high-quality time series data, no attempt has been made in this document to tailor the QC process to any specific climate application, such as climate change monitoring, or to specific services which may have additional, value-adding requirements (such as homogenization, infilling of missing data, and so forth). Finally, it should be noted that the QC tests (and broader QA procedures) need to be adapted to the specific climate conditions of the country in question as well as fine-tuned to fit the existing and planned observational and information technology (IT) infrastructure and available human resource capacities.

2. PRINCIPLES AND GENERAL REQUIREMENTS FOR ASSURING HIGH-QUALITY DATA FOR CLIMATE APPLICATIONS

The following sections highlight generic practices regarded as essential to ensure that the surface climate record of an NMHS is of high quality. More detailed guidance is provided in Chapter 3.

2.1 Overarching principles

- A QA process that covers the life cycle of the data, from equipment installation and observation through to final archival, must be established.

- All methods and practices employed to ensure quality and traceability across the entire data life cycle must be documented in a comprehensive manner; all the relevant documents must be maintained and accessible.

- Arrangements with data providers must be established in order to ensure that observations for climate purposes are carried out and maintained and that the ten Global Climate Observing System (GCOS) principles of climate monitoring (see *Guide to Climatological Practices* (WMO-No. 100)) are adhered to.

- Objective QC tests and procedures must be used, and the outcomes of the tests for each data element must be clearly flagged.

- A copy of the originally read or received data, even if considered doubtful, must always be kept.[1]

- Each NMHS should have a data QA manager responsible for monitoring the complete QA process across the data life cycle.

- QA processes require feedback on sources of errors. It is recommended that there be a regular exchange of information between the QA manager and the personnel responsible for the observing systems of NMHSs so that sources of systemic errors can be addressed and the relevant issues rectified.

2.2 Requirements underpinning the adequacy of observations for climate applications

- The observing site location must be representative of the defined environment.

- Each observing site and variable, in particular the siting classification and the sustained performance classification, must be classified according to the *Guide to Instruments and Methods of Observation* (WMO-No. 8).

- AWSs should have data loggers (automatic data logging) that can be accessed and used to restore information to the climate data management system (CDMS) and to allow the automatic re-computation of derived parameters, including dewpoint, mean sea level pressure and so forth.

- Data loggers must meet the technical requirements set by the NMHS and must be tested and accepted for use.

- Pre-deployment testing, calibration and acceptance testing are required to ensure that the instrumentation sensors or AWS software and sensors meet the specified requirements. This includes ensuring that communication channels are sufficiently robust.

---

[1] Sometimes, an original reading will clearly not represent meteorological conditions but will have scientific value in its own right. For example, an elevated temperature reading may be due to the occurrence of a wildfire near a station. In these cases, the original reading will not form part of the official climate record but will still be available for research and other purposes. See also Footnote 4.

- Observation manuals must be an integral part of an NMHS's process documentation and must be readily available.

- Observers must undergo regular training, especially when there any changes to observational practices.

- NMHSs must ensure that stations are regularly inspected, including by means of site visits at appropriate intervals, in order to verify whether or not maintenance is needed and sensors should be changed and in order to maintain adequate observing conditions at the site. Sensor recalibration checks should be performed on a regular basis.

- Changes in sensor hardware (such as type or manufacturer) or changes to an observation site require parallel measurements to be made (from both the current and the new sensor or the current and the new site) over a period of at least one year[2] (two years is preferred). Both the original and parallel data and associated observations metadata must be archived and retained in perpetuity.

- NMHSs should develop minimum quality guidelines, including observational metadata requirements, for external data providers. (The provision of training materials may help external data providers adhere to these standards.)

2.3 Requirements for technical monitoring

- **Data ingest error monitoring:** An alert system is needed to ensure that circumstances in which the data can be corrupted or inadvertently overwritten or deleted either cannot occur or can rapidly be detected and corrected if they do occur. This system must undergo user acceptance testing before it is made operational, and the outcomes of QC processes must be analysed on a regular basis. There should also be automated processes which ensure that expected data are received, that timestamping is accurate, and that ingested data are not unrealistic.

- **Data flow monitoring:** A system for real-time monitoring of technical equipment (AWSs, sensors, data transmission equipment) should be implemented so that possible malfunctions can quickly be identified and addressed (failure to do so has been known to lead to significant losses of data from the climate record that may not be detected until months later).

- **Data recovery processes:** Where alerts reveal that data are missing or somehow corrupted, there must be standard operating procedures to recover the original version of the data in question; for instance, data from an AWS should be buffered, or recoverable from a data logger, with a data retention period of sufficient length. Such data outages are frequently due to communication issues, rather than the failure of the sensor itself. Prompt action is required because AWS data loggers or databases upstream from the CDMS (where the data arriving at an NMHS are often first stored) typically retain data for only a few days before the data are overwritten. Consequently, if failure to adequately track these processes means that data losses are not revealed for some time, the result could be a significant loss to the climate record.

- Any issues with sensors and data dropouts must be addressed as soon as possible and documented, with both the occurrence and rectification of the issue communicated to the data users.

- It is the responsibility of the data provider and/or IT department to investigate and fix any processes that result in lost or corrupt data, and proper internal protocols should be in place to ensure this.

[2] The *Guide to Climatological Practices* (WMO-No. 100) recommends that, where feasible and practical, both the old and the new observing stations and instrumentation should be operated for an overlapping period of at least one year, and preferably for two or more years, to determine the effects of changed instruments or sites on the climatological data.

2.4 Requirements underpinning the adequacy of climate data management practices

- From a data traceability perspective, it is important to be able to trace the provenance of the data from sensor to archive. This can be accomplished by maintaining a record of the original measurements, along with any changes made to the data.
- Quality flags are an essential component of data management and are used to provide a measure of confidence in the veracity of the data[3] (the subject of source and quality flags is dealt with in more detail in Subsection 3.4.5).
- QC procedures should be documented in detail and if required, made available to data users, preferably online.
- Any alteration made to measured or observed values must be flagged and documented as part of the metadata, with both the original and altered value retained.
- Data typically pass through various quality assurance steps (described in Section 3). The outcomes of QC tests at each step should be documented, including by assigning a quality flag. Thus, for instance, quality flags may be assigned for data at the point of observation, on ingest, and following delayed mode QC within a CDMS.
- Doubtful/suspect values that have been changed must not be deleted but stored along with their quality assessment.[4]
- When retrieving measured/observed data, data users should have access to all available quality information (explanations included). A distinction may be drawn between users internal to the NMHS, who may require extremely detailed information, and external users, who will likely only be interested in accessing data that has passed through the appropriate quality checks.
- Observational metadata supports the data, provides provenance information, can provide assistance in investigating suspect data, and is essential for the application of homogenization. Therefore, retention of metadata for at least the life of the data is to be regarded as a mandatory practice. It is important to have such metadata easily accessible to the QC operator, as well as to the data users.

Notes:
1. *Climate Data Management System Specifications* (WMO-No. 1131) provides detailed information on appropriate data management and quality management practices.
2. The *Manual on High-quality Global Data Management Framework for Climate* (WMO-No. 1238), a part of the WMO regulatory material, sets out standards and recommended practices for climate data stewardship.

2.5 General considerations when designing a quality control system

2.5.1 *Automation considerations*

QC procedures may be applied manually, automatically or semi-automatically at each stage along the data life cycle. While the details of such procedures will vary among AWSs, a typical sequence of QC gateways may contain the following:

- QC at the point of observation. For instance, a manual observer about to record an observation electronically may be prompted to check the proposed value if algorithms

---

[3] At the time of writing of this document, there is no standard WMO guidance on the definition of data quality flags; nevertheless, it is expected that such guidance will be developed by WMO in the near future.

[4] Of course, it is incumbent on NMHSs to ensure that the climate record presented to users is of the highest standard possible and capable of reliably supporting climate services. Erroneous values must not appear in this public version of the climate record but must be available if required for public enquiry and/or to assist in detecting and addressing measurements or software problems.

suggest that the value is unlikely. In the case of observations from AWSs, various automated checks may be built in based on details of sampling frequency and results and range checks;

- QC at databases upstream from the CDMS to support real-time product generation;

- QC at the point of ingest into the CDMS, which may automatically eliminate values failing an extended range or other checks (although a copy of the value itself should be retained);

- QC within the CDMS post-ingest (manual, semi-automated, or automated).

The discussion below refers solely to the QC of data within a CDMS. This is sometimes termed *delayed mode* QC because the QC may occur at some time after the observation has been made and the data has been ingested. Delayed mode QC may involve manual, automated or semi-automated checks, or any combination of these. This document makes no specific recommendation about which checks are optimal as this will depend on such considerations as the nature of the data, the number of stations and the QC resources available (noting, however, that QC checks should take place as soon as possible so that, should the need arise to verify a value with an observer, the observer still remembers the event).

The general approach in delayed mode QC is to apply QC testing, assign a quality flag which should be part of the observations metadata, and archive the data and metadata, along with the details of any corrections, as part of an audit trail. As the names suggest, manual QC refers to checks made to the data by QC operators. It is generally labour-intensive, subjective, and heavily dependent on the knowledge and experience of the QC operators. Manual techniques may be suitable where there are few stations and there is little access to QC diagnostic software. On the other hand, automated QC refers to the process of scanning the data and assigning quality flags fully automatically, that is, without manual intervention. It is the fastest way to process incoming data. However, fully automated testing may tend to flag too many potentially suspect values (false positives), or too few, depending on how screening parameters are set, and it may flag some genuine extremes as suspect, which is not desirable given increasing scientific interest in changes in extremes for climate monitoring.

Semi-automated techniques represent a compromise between manual and fully automated techniques. Here, the data are subject to an automated check against a range of test criteria, and values that fail one (or more) tests are flagged as suspect and are subject to a manual follow-up investigation by trained QC staff. To properly diagnose whether a suspect value is erroneous or not, it is recommended that QC operators be equipped with a readily accessible suite of tools which experience suggests are useful in diagnosing errors. Such tools may facilitate the use of complementary satellite and/or radar data (for assessing rain/no rain situations), as well as radiosonde data, and may provide means of accessing the original message (which may turn out to be incorrectly coded). Other means of diagnosing whether a suspect value is erroneous or not include plotting values on a map to test spatial consistency and plotting a time series at one station and comparing it with those at neighbouring stations (in the case of rainfall, this type of check often helps identify cases where rainfall was recorded on the wrong day or represents an accumulated total) and assessing the data against the numerical weather prediction (NWP) model output for the same time/location.

Note that with high frequency data extremes, such as sub-hourly data from AWSs, automated techniques are probably the only realistic option for QC, although even then, some form of follow-up check should investigate, if possible, whether the error was a software glitch or a physical event such as a wildfire or a sprinkler being turned on.

In the future, it is likely that machine learning techniques, including expert system approaches, could lead to improvements in QC automation. NMHSs are advised to stay abreast of such advances in data science and their potential applications.

Another consideration in applying QC tests is whether to apply them in parallel or sequentially. For instance, it may be helpful when doing comparisons against neighbouring stations to first eliminate from the testing any likely erroneous stations.

Finally, it is worth mentioning that in the case of data transcribed into a CDMS from paper forms, key entry mistakes are likely to be a rich source of errors. These can be mitigated by inserting on-entry checks to alert the key entry operator to a possibly mistyped value. If the resources of the NMHS so permit, double keying has been shown to be very effective in mitigating key entry errors. Double or multiple key entry is especially recommended when digitizing historical data as part of a data rescue process.

2.5.2 *Designing and implementing a quality control system*

- In the case of manual or semi-automated systems, it is strongly recommended that when the system is being designed, the QC for a new variable is being introduced, or the existing QC system is being modified, the QC operators who are to use the system be involved in the planning and implementation phases from the beginning. This will ensure that the system conforms to the actual operational environment in which the operators are to work and will give the operators a sense of ownership. In addition, QC operators can recommend the diagnostic aids they find useful. It is further recommended that where a computer-based system is being designed and/or implemented, agile methodologies should be followed – with developers working directly with QC operations staff to discuss, for instance, operational decision-making processes and shortcomings in the system.

- Statistics on errors and false positives (where checking routines flag a value as suspect, only for it later to be verified as reliable) should be compiled and updated. These provide information about systemic errors to be communicated to the data providers, as well as information that may potentially help fine-tune the QC system and its algorithms.

- Following on from the previous point, the effectiveness of the QC system should regularly be assessed to investigate which testing techniques work best on particular types of data and which can assist in fine-tuning the system.

3. ELEMENTS OF DATA LIFE CYCLE QUALITY ASSURANCE

3.1 Overview

QA across the data life cycle can be broken down into the five major steps described below.[1]

Step 1: Observation

This step involves the data quality aspects associated with ensuring that the observation process is sound, including ensuring that the observing station, whether manned or automated, has been set up and operates correctly. These aspects should be reflected in the observational metadata, for example, site characteristics, instrument metadata, and metadata confirming that the site installation specifications have been met. Observational metadata should also reflect the data capture process, whether automatically or manually observed, and the regular maintenance and technical monitoring processes. There may also be software checks to query whether a submitted reading is reasonable or not or whether or not an instrument has malfunctioned. For manual recording on paper records, the QA process should ensure the timely submission of paper forms to the NMHS.

Step 2: Data delivery and ingest

This step comprises the data quality monitoring associated with the flow of data between the observation site and the CDMS. (There may also be upstream databases that perform their own quality checks on the data.) QC checks on the data should occur at the point of electronic ingest into the CDMS (to identify "impossible values"). If the data are key-entered, additional checks should be in place to verify that the data have been entered correctly. This step should also screen for data expected but not received.

Step 3: Climate database management (within the CDMS)

This step typically comprises the extensive area of (central) climate data management[2] following data ingest, at which point delayed mode QC occurs, using either manual, semi-automated or automated quality control tests. It is this step that is primarily addressed in the remainder of this document.

Step 4: Final archival

This step includes the requirements for final archival within the CDMS as well as the requirements relating to archiving hard copy paper records in a suitable repository.

Step 5: Disaster recovery

This step includes protection against loss of data through processes involving, for example, electronic backups and scans of hard copy records. Note that recovery processes should be in place for the intervening steps as well.

---

1 In this document, the "data life cycle" extends only to the point at which the data are archived, whereas in standard usage, the data life cycle extends up to the point at which the data are disposed of.

2 Considerably more detail on data management requirements can be found in *Climate Data Management System Specifications* (WMO-No. 1131).

3.2 Step 1: Observation

This section addresses the data quality aspects associated with the observation of meteorological elements as well as with ensuring that the observing stations, whether manned or automated, have been set up and operate correctly.

3.2.1 *Observation standards*

The search for a site and the decision of whether to install an observing station at a specific location influence the data quality obtainable relative to the observation requirement. For climate purposes, the site has to satisfy data users' requirements[3] over as long a period of time as possible. The observation conditions should therefore remain largely unchanged over a long period of time. The site should also be representative of the meteorological conditions of a larger area.

Standards and guidance on meteorological observations can be found in many WMO publications, including the following:

- *Manual on the WMO Integrated Global Observing System* (WMO-No. 1160)
- *Guide to the WMO Integrated Global Observing System* (WMO-No. 1165)
- *Guide to Instruments and Methods of Observation* (WMO-No. 8)
- *Guide to Climatological Practices* (WMO-No. 100)
- *WIGOS Metadata Standard* (WMO-No. 1192)
- *Challenges in the Transition from Conventional to Automatic Meteorological Observing Networks for Long-term Climate Records* (WMO-No. 1202)

The critical elements of observing standards and practices from a QC and QA perspective regarding climate applications and services are detailed below.

- Sensor pre-deployment processes: The sensors need to meet specified standards covering the following areas:
    1. Manufacturer's test specifications;
    2. The NMHS's review or an external review of the sensor's capacity to deliver to standards that cover the expected climatology (it is recommended that sensor testing replicate the expected field range of the parameter);
    3. WMO standards for sensors;
    4. Calibration.
- AWS pre-deployment processes: AWSs need to meet specified standards covering the following areas:
    1. Documentation and encoding specifications for AWS algorithms (where available[4]), which must be maintained by the NMHS as part of the observations metadata;

[3] These will be many and varied, and data for climate change monitoring will have more stringent quality requirements than data for other purposes. However, in all cases, the aim is to ensure that the climate record is of the highest possible quality.

[4] Unfortunately, many AWS providers do not allow access to the sensor algorithms.

2. Test beds and field trials to ensure that the AWS platform is fit for purpose (that is, robust, reliable and serviceable) and to ensure compliance with NMHS and WMO standards;

3. Data logger capacity (which is dependent on the capacity of the NMHS to recover data);

4. Adequate communications, power supply, and so forth to minimize down-time.

All existing sites must be classified according to the *Guide to Instruments and Methods of Observation* (WMO-No. 8) for each parameter observed at the site; the classification must be included in the station's observations metadata.

3.2.2 *Observations metadata*

Observations metadata must be documented and stored in the NMHS metadata system according to WMO and NMHS standard practices (*Guide to the WMO Integrated Global Observing System* (WMO-No. 1165)) and the *WIGOS Metadata Standard* (WMO-No. 1192)).

For the purposes of climate data QC, it is recommended that observations metadata contain as much of the following information as possible:

- Site and exposure information, including:
  - Latitude, longitude and elevation;
  - Skyline survey, if possible (including description);
  - Site photographs from the North, South, East and West;
  - Local environment details, along with a site survey (polar diagram), including the elevation of instruments and the distance from instruments to trees, buildings, and so forth.
- Instrument/sensor information, including:
  - Serial number;
  - Details regarding calibration checks;
  - Deployment/serviceability history;
  - Reporting method and time of observation.

The absence of adequate, correct and up-to-date observations metadata may compromise the entire quality management process. Note that, for homogeneity assessment practices, these metadata should be regularly updated to reflect any changes, such as growing trees or land use changes, that might affect the measurements.

3.2.3 *Quality control at the station level*

This section applies to both manually performed observations and automatically generated electronic observations from unmanned stations.

It is a great advantage if the first stage of QC is carried out at the station in real time. The raw data of all sensors are available on site, and, in the case of manned stations, observers can perform a consistency check against other parameters. This prevents data that are clearly erroneous from being ingested into the data archive and potentially incorporated into products.

Manual observations

All observers must be supplied with standard instructions for meteorological observations, which they are required to follow. These instructions are an integral part of process documentation and compliance. If there are any changes to these instructions, observers must be informed accordingly. Observers should undergo refresher training at regular intervals. Ideally, replacement sensors should be readily available in the event of sensor damage to minimize missing data.

Records which observers have made manually should be sent to a central office (usually the NMHS) immediately after the end of the month. The station should maintain a copy of these records.

For observations that are either fully or partially manually read, the following checks should be performed (a checklist might help ensure that these are carried out in a systematic manner):

- Consistency checks against other manual observations;
- Range and constraint checks against reasonable climatological norms;
- Consistency checks against the station record (the historical record for that station, including past extremes);
- Calculations on observations (such as the relative humidity computation);
- For manned sites, independent checks, which may be performed by the officer in charge or another observer.

More details of these checks are presented in Annex 1 and Annex 2.

Automatically generated observations

For fully automatically generated observations (generated via AWSs), the following checks should be performed:

- Constraint checks imposed by the database ingest;
- Constraint checks to determine whether the observations are climatologically reasonable;
- Consistency checks against other parameters at the station (internal consistency).

For examples of these checks see Annex 1 and Annex 2.

Ideally, outcomes of the automated quality checks should be stored as a quality flag which is transmitted to the regional or central office together with the corresponding data value.

For automatically generated electronic observations from unmanned stations, it is recommended that the data logger capacity should be sufficient to store data during communications outages and that, for critical sites, a robust data recovery method should be considered (for instance, satellite polling).

In the event of technical problems with a sensor, the data logger or the AWS, the systems should automatically generate technical alarms which can be analysed and acted upon by the NMHS.

3.2.4 ***Maintenance***

Regular visits to all stations are required to check the site conditions and – even more importantly – to carry out maintenance work, including changing sensors. The results of these site visits must be documented.

Where sensor changes are made at an individual station, recalibration is required, and the calibration results must be included in the observations metadata.

In case of sensor changes across an observing network, parallel measurements are required (current versus new sensor) over a period of at least one year, and preferably for two years[5] or more (see *Guide to Climatological Practices* (WMO-No. 100)) at all stations affected, or at least at a selection of reference stations representing the range of climatological conditions across the network, before introducing the new sensor to the entire network.

The results from these parallel measurements must be documented in the observations metadata and thus be accessible to data users for further analysis. It is important to record the exact date of the sensor change for each site.

3.2.5 ***Technical monitoring***

This section applies to both semi-automatic and fully automatic stations. It concerns semi-automated technical monitoring systems which can detect and, to a degree, intervene regarding malfunctions relating to sensors, data loggers and communication systems.

Errors in sensors and data loggers should be identified as expeditiously as possible by real-time monitoring of technical parameters. This reduces the likelihood of missing data, or that faulty data will continue to be processed over a lengthy period of time to the detriment of the climate data time series and derived climate products.

It is highly recommended that when developing monitoring system detection mechanisms, observational network operators consider the following:

- Intelligent sensors can detect whether their power supply, the internal calibration mechanism, or any internal system, such as the heating system, is operating correctly. If the sensors detect any incorrect operation, the corresponding signals are sent to the data logger, and from there to the monitoring software (control centre) during data retrieval. Ideally, if the sensor power supply, the internal calibration mechanism or an internal system is not operating correctly, this information is made visible as a quality flag attached to the measured values during signal processing in the data logger ensuring that detection of the incorrect operation is straightforward.

- Either the data logger or the data processing unit at the station, or both, may malfunction. This can be detected from the fact that the measured values are no longer encoded or the retrieval of data from the station is no longer possible.

- Technical monitoring of data communication provides rapid information about any process errors that might exist. This information provides the basis for investigating gaps in data inputs (that is, data expected versus data actually received), which helps to explain why certain data may be missing, and hence provides a basis for initiating rectification procedures. Note that such faults may occur at various points within the data infrastructure of the NMHS.

---

[5] The main reason for documenting parallel measurements over a period of two years, rather than one year, is to provide additional supporting data should the first year of the comparison be climatically unusual.

Details of all technical errors should be recorded in a database. The database should contain at least the following information:

- Type of error
  - Sensor
  - Data logger/AWS
  - Communication line;
- Start time of the error;
- End time of the error;
- Measures taken to mend the error (Note: Ensure that the measures are clearly described);
- Whether or not the error has caused any data gaps;
- Who dealt with the error.

3.3 Step 2: Data delivery system and ingest

This section considers data quality issues associated with the data delivery system from the observation site to the (central) database, including data ingest.

3.3.1 *Manual key entry*

Many NMHSs receive hard copy station records that are then digitized by NMHS staff. As part of the QA/QC process it is recommended that:

- The record be imaged, and the image stored,[6] as part of the NMHS data rescue strategy;
- Multiple key entry digitization be used to mitigate against digitization errors (double keying is recommended);
- An operator ID be attached as metadata to the digitized record. This enables auditing of the performance of operators and assists with training;
- The digitization date/time be appended to the digitized record. This supports the auditing of any changes;
- Digitization-level QC be used to reduce the incidence of errors. For digitization into a CDMS, ingest-level QC algorithms should include:
  - Range and consistency checks;
  - Checks to ensure that data for the station, time and date do not already exist in the database;
  - Metadata checks to ensure that the station name and location correlate to the assigned station number;
  - Arithmetic checks of the parameter calculation (for instance, calculation of the dewpoint using dry bulb and wet bulb temperature values);

---

[6] Note that in general, the paper record should be retained as well unless the country's national archival authority provides legally binding advice to the contrary.

– In the case of cumulative variables such as rainfall and evaporation, an arithmetic summation of daily values to ensure consistency with the recorded monthly total.

3.3.2 *Electronic data ingest*

For ingest-level QC, it is recommended that tests be confined to identifying data that can be statistically identified as clearly erroneous through domain range checks. If an additional QC analysis is performed at ingest, it is recommended that any such doubtful data be flagged for assessment through the NMHS's delayed mode QC system (see Section 3.4.1).

Recommendations:

- The ingest must reliably decode data and metadata for all parameters;
- Ingest QC software must be independent of the software used to ingest the data (to reduce problems associated with upgrades and updates);
- Disaster recovery – it must be possible to re-ingest recovered data where necessary;
- Software should be readily extensible to accommodate new data types, data formats, and data flows;
- Error logs should be compiled and used to highlight potential problems with the data ingest system.

3.3.3 *Systemic error detection*

Even in a fully automated system, there must exist the appropriate tools to detect and analyse systemic network issues. In addition, any changes to the QC system require user acceptance testing as well as testing in the operational environment, and the tests and results must be fully documented. End-to-end (sensor to archive) testing is also required.

Ideally, there should be a process by which identified data issues are documented and error pattern analysis is performed. If a particular type of error is detected repeatedly, this may indicate that there is a systemic problem – at the point of observation, in the observing system, in the ingest software, or involving key entry operators.

Systemic network issues for a given parameter may be identified by a higher-than-usual frequency of data flagged as "suspect".

3.4 **Step 3: Quality control within the climate database**

This section addresses QC within the CDMS, including rules, procedures and tests.

The screening processes defined to date on observational data arriving at, and being ingested into, the CDMS are generally designed to capture the more obvious errors. However, before the data are archived as part of the official climate record, a final, comprehensive QC assessment is required. QC processes differ depending on the source of the data, the parameter measured and the temporal resolution. Depending on the structure of the NMHS, QC may be performed by a central unit or at the regional level following a consistent, well-defined NMHS-wide process. The important point is that the final QC should be carried out using a uniform procedure, with consistent tools and approaches, prior to final archival.

3.4.1 *Introduction*

The QC tests described here are recommended in order to provide a comprehensive, though not exhaustive, analysis of the data received in the climate database. They should be carried out after ingest,[7] and as such, are often referred to as delayed mode QC. There are five main types of tests:

- **Constraint tests** (sensor-based range tests, which check whether observations fall outside the theoretical limits of a sensor, and domain tests, which determine whether the value is scientifically feasible);

- **Consistency tests** (which test whether the data from two or more sources are consistent (for example: Do sunshine hours exceed the maximum possible for the time of year? Is the dewpoint temperature greater than the wet bulb temperature? Is there an inconsistency between the three-hourly values and the daily maximum/minimum temperature? Is the precipitation improbable considering the humidity, clouds, and so forth?));

- **Heuristic tests** (which rely on experience and knowledge of observation processes and instrumentation to test for inconsistent or unlikely values, for example, tests to check whether the wet bulb wick has dried out);

- **Data provision tests** (for example, tests which compare local clock time with the time when data are received, tests which check whether there are unreasonably large gaps between observations, and so forth);

- **Statistical tests** (which analyse data received against historical data or spatial variability patterns to detect inconsistent or unlikely values, for example, flat line or spike tests; this category also includes spatial consistency tests via linear regression or mapping and temporal consistency tests).

An overview of these tests, including advice on whether the tests should be regarded as mandatory, recommended or optional, is provided in Annex 1. Further details regarding the tests are provided in Annex 2.

Specific quality monitoring parameters (QMPs) are used to determine whether the values being tested have “passed” or “failed” the above tests. A QMP is a numerical reference value that acts as a constraint, often as an upper or lower limit. QMPs may be defined for a given meteorological element/station/month/test type and may be statistically generated (for example, more than x standard deviations from the mean). Values that fall outside of the range of QMPs for a particular test are generally considered to have failed that test and may be considered suspect, at least until follow-up investigations can either verify those values, or confirm that they are suspect or wrong.

The sensitivity of a test is often determined by whether the test is automated or semi-automated. Automated QC tests can result in lower confidence data and may either overestimate the number of errors or underestimate extremes.[8] However, automated QC tests are suited to high-frequency data due to the limited value changes in the data over short time periods and the high volumes of data the tests are required to process.

3.4.2 *Dealing with site differences*

Site differences need to be accounted for when running QC tests. This is particularly important for NHMSs that operate over a wide range of climate zones and elevations.

Coastal stations, for instance, have vastly different ranges in some variables compared to continental stations, and this can have an impact on the expected number of false errors if

[7] It should be noted that some of the tests described may additionally be carried out at other stages of the data life cycle.

[8] This situation could be improved by various means, such as ensuring that there is a large training sample size, multi-variate testing (see below), or specific scenario situations. The application of machine learning techniques may help in this regard.

the same test parameters are used. For NHMSs covering different climate zones, there are two main approaches for dealing with site differences and differing false error rates. One option is to calculate QMPs and normalization constants separately for each zone. Alternatively, a single system could be used in a semi-automated procedure, with trained operators used to identify false errors related to seasonal variations and known variability between sites.

Differences in site elevation can also have an impact on the expected number of suspects if the same test parameters are used. Quality control operators should check whether neighbouring stations are at significantly different elevations or in significantly different environments (for example, inland stations, where the candidate station is coastal); if this is the case, the use of nearest neighbour-based tests, for example, may not be appropriate. Comparing data anomalies from the sites' climate normals, however, can solve the challenge of differences in site elevation.

The length of the climate record can have an impact on the QMPs and hence on the reliability of the tests applied. For example, QMPs generated by a site that has 30 years of record will be significantly more representative than those generated by a site that has only three years of record. When running spatial statistics, if the neighbours are relatively new stations with short records, the test performance will be affected. Observations metadata will assist operators in determining the length of the record and the expected effectiveness of the test. As an example, the climate-based range test described in Annex 1 and Annex 2 compares against the local climate record extremes, and therefore, a site with a short period of record without a wide range of extremes would be expected to trigger false positive suspects more frequently.

The density of the observation network has an impact on quality control tests that use spatial statistics, for example, spatial variability tests and mapping-based spatial tests (see Annex 1 and Annex 2). Limits may need to be placed on spatial algorithms to restrict the distance of a neighbouring station from the target station to a "reasonable" distance, which would be determined by a number of factors, including climate zone, topography, the length of the record and the parameter to be tested. For very remote sites or very sparse networks, the first day of the month may need to be triggered as suspect to ensure manual checking for errors. Alternative methods and/or parameters may be necessary to validate any suspect observations. Using the example of rainfall observations, alternative methods may include the use of satellite images, radar, river heights or streamflow. It may also be possible to contact the observer directly to confirm the recorded quantity, preferably in near to real time.

Multivariate tests can be used to assist the QC operator. For example, if the daily rainfall total for a given day is recorded as high, this total can be compared with hourly rainfall observations, present and past weather reported, cloud types associated with heavy rainfall, humidity, data from other rainfall measuring instruments (for example, pluviometers or rainfall intensity charts), radar images, satellite images, numerical weather forecasts, impact analyses, media reports and so forth. This assumes, of course, that the alternative parameters and images are readily accessible. Note that such multivariate approaches can also potentially assist in the automation of the QC process.

3.4.3 ***Dealing with high-frequency data***

At present, many NMHSs operate AWSs which provide data at high temporal resolutions (1 minute, 10 minutes) that are transmitted at short intervals from the observing station to the headquarters. In addition to any automated QC that takes place directly at the station within the central processing unit of the AWS, the opportunity exists for further QC to be carried out on the data in near real time within the CDMS. This additional QC helps avoid major errors in derived daily or monthly values within the climate record because the high-resolution data from which the daily or monthly values are derived have already been flagged and/or corrected.

For traceability purposes, the original high-frequency data need to be stored, which requires complex storage systems and data models (or at least an audit trail of the changes made). The QC of high-frequency data is highly time-consuming. Therefore, some NMHSs choose to perform near-real-time QC on the basis of the derived hourly values: high-frequency data are only given a closer look if the hourly values appear to be doubtful. Note, however, the desirability of also incorporating spike or flat line tests, as spikes, for instance, may not be adequately captured in an hourly value.

Where possible, it is recommended that an automated QC process be used for all high-frequency data parameters and that QC testing include:

- Consistency tests;
- Spike tests;
- Rapid change tests;
- Flat line tests;
- Domain tests.

The risks associated with fully automated tests, as outlined in Section 2.5, can be ameliorated by multivariate analysis.

Post-event QC in the context of high-frequency data

There may be cases where a significant and unexpected change in the high-frequency data parameter, such as a data spike, warrants further investigation to assess the broader context of the sudden change. The event could, for instance, be due to a real severe weather event or a localized phenomenon such as a microburst, a non-meteorological event such as a nearby wildfire, or a spike generated by an internal voltage surge.

The context of the event is evaluated using a multivariate analysis to confirm whether the event is representative of the meteorological conditions at the time or due to some other cause.

3.4.4 ***Value-added quality control***

Value-added QC can take various forms, including the estimation of replacement values in the case of missing or erroneous values.[9] However, it is beyond the scope of this document to address such methods here.

Homogenization may also be thought of as a form of value adding. It involves the adjustment of climate records (when required) to remove the effects of non-climatic factors (for example, station relocations), so that the resultant data reflect unbiased variations due to real climate processes. It is essential that the QC/QA process provide sufficient reliable metadata to support data homogenization.

Guidelines on Homogenization (WMO-No. 1245) provides guidance on dealing with homogeneity problems. The four key steps commonly used in the homogenization process are:

- Analysis of observations metadata: This step looks for changes in the measurements and looks at which QC procedures have been performed.
- Creation of a reference time series: Generally, a reference time series is created by using a weighted average of neighbouring stations (post-QC), but other techniques, such as principal component analysis, can also prove useful.
- Breakpoint detection: This step involves a search for inhomogeneities by observing divergences between a reference time series and the candidate time series.

---

[9] This type of substitution is especially likely to be preferable to simply stating “not available” in the case of rainfall, where one missing value can invalidate an entire month’s/year’s total.

- Data adjustment: In this step, a decision is made on which breakpoints are accepted as inhomogeneities by comparisons with the available metadata and by using expert judgement. The assessed discontinuities in the data are then corrected to match the conditions of its most recent homogeneous section.

To support the traceability and trustworthiness of the data, it is important to document every adjustment made to the data and to always preserve the original data.

3.4.5 ***Source and quality flags***

Although at the time of publication of this document, no generic WMO guidance on QC flagging was available, for traceability purposes, NMHSs should adopt some form of quality-indicator flagging for tracking data provenance and source. Some suggestions are provided below.

Source flags identify the type of observation source. For example, they distinguish between manual and automated observations or between NMHS-derived observations and those from external sources (which may be subject to differing observational practices and requirements for metadata completeness). Source flags can also be used to compare the same observation received via different methods, for example, daily rainfall sent electronically, as well as on a hard copy manuscript. These values do not always agree, and in such cases, they require further investigation. Source flags can also be used to compare values from co-located instruments, for example, daily rainfall values where both manual rainfall observations and automated rain gauge readings are available, and values from third-party observations, such as flood warning gauges.

Quality flags provide a qualitative indication of the level of confidence in the data. These flags should be allocated from any automated QC carried out at the station (AWS, datalogger, and so forth), during the ingest/key entry process (which may itself involve one or more steps, as described above), and during delayed mode QC. Typical quality flags should indicate whether or not a datum has been quality controlled, and if so, whether it has been assessed as valid, suspect, doubtful or wrong and whether or not it has been estimated or amended.

One example of a situation in which a quality flag may be warranted would be if a manually read maximum thermometer value at a station did not compare favourably with its neighbours and the QC operator suspected a reading error. In this case, it may be appropriate to flag the datum as a gross estimate or, in more extreme circumstances, as a suspect or incorrect value.

QC flags can also refer specifically to the type of QC applied: ingest QC, automated QC and automated flagging of data, semi-automated QC with multivariate operator analysis, and so forth.

Allocated flag values, along with any changes to them, need to be maintained in an audit trail, and these changes need to be available for manual analysis or scrutiny.

The flagging model adopted should not be complex. It is recommended that, rather than a single large listing of flags that encompasses every possible metadata scenario, a series of flag subsets be used to indicate various aspects of the metadata.

The amount of effort involved in planning a quality flag system will dictate how future proof the system will be. The flagging system should be flexible enough to accommodate future changes in technology, networks, and so forth. A flagging system incorporating flag subsets has a number of advantages, several of which are indicated below.

- The flag subset can be extended when required. For instance, an additional flag can be added to an existing flag series to indicate a new source of data, and an extra flag series can be added to describe the quality of instrumental exposure or the specific QC already performed by an external data provider.

- Retrospective QCs can be carried out as the QC system improves.

Guide to Instruments and Methods of Observation (WMO-No. 8) gives examples of types of quality flags. While it does not contain an exhaustive list, it does provide a basic guide. Types of quality flags include:

- Type of data code;
- Acquisition method code;
- Validation stage code.

Quality flags allow for audit trails and data provenance – whether changes to the data were made, by whom, how and why. They also facilitate the monitoring of the performance and accountability of QC operators. For example, if a training issue was identified for a particular operator, it is possible, through the use of quality flags, to recover all the corrections applied by that operator which relate to that specific training issue, to recover the original data, and to reprocess the data. All the decisions made regarding flagging data as suspect need to be scientifically supported. If the operator cannot reasonably and confidently make a decision, it is recommended that the data be retained as is.

The use of quality flags is a mandatory component of QC. It provides a level of quality assurance for data users, as well as a record of the sources of the data and the results of QC processing. It ensures that suspect or erroneous data are not deleted, but rather flagged (as mentioned earlier, although suspect/erroneous data are retained, they should not be incorporated into the user-facing climate record). It provides data users with the capacity to choose the level of data quality appropriate for their needs.

Data from non-NMHS sources

Data held by an NMHS may have been collected by, and belong to, external networks and may be expected to vary in quality due to differing observing practices and/or the amount of observations metadata available. For this reason, data may be considered to belong to a particular class of network, either NHMS-operated or operated by external parties. As data (and metadata) quality may vary according to the data source, it is recommended that appropriate flagging (or other means of distinction) be used to indicate the source and to document what processing, if any, is applied (including, for instance, whether the data were subject to QC by the data provider). More detailed guidance on the management of externally sourced data may be found in the *Manual on the High-quality Global Data Management Framework for Climate* (WMO-No. 1238).

3.4.6 *Future prospects for quality control*

The QC processes described above are mainly based on algorithms which use ground surface data and the manual verification of those data by the operator using satellite and radar data and products and other information and tools.

Future QC processes have the potential to integrate these types of remote sensing data, as well as any derived products, directly into the verification algorithms as part of efforts to increase automation. Based on remote sensing data, it may be possible to fully verify the following parameters automatically, in particular for areas where observing stations are sparsely distributed and spatial comparisons with other surface stations are not possible:

- Solar radiation;
- Cloud cover;
- Snow cover;

- Temperature (daily mean);
- Precipitation (daily sum).

NWP products (for example, first guess fields) can potentially provide valuable information for near-real-time QC of measured/observed data to detect and rapidly correct faulty data. Site-specific weather forecasts can be used to determine the historical difference (anomalies) between forecast and observed values. If the difference between the forecast and the observation is larger than a certain threshold, the observation can be flagged as suspect. The advantage of this approach is that the magnitude of the differences is generally small and hence, the threshold value above which an observation is considered suspect can be determined fairly tightly. The disadvantage of this approach is that if the forecast is wrong, a correct observation may be flagged as suspect. Similarly, reanalysis technologies have the potential to assist with the detection of anomalies due to suspect data.

Advances in information and communication technology (ICT), such as machine learning and expert system approaches, should also be monitored given their potential to assist with the automation of QA/QC.

3.5 Step 4: Final archival

For electronic archives, at the point of final archival, it is important to ensure that:

- The data to be archived are accompanied by a full audit trail of changes, and/or separate versions of the data following QC and any value-adding performed;
- The original electronic message is retained and can be easily retrieved along with the observation value and metadata;
- The provenance, indicating all changes made to the data, can be extracted if needed;
- The data are fit for purpose, that is, suitable for climate applications;
- The data are of known quality, the quality tests applied have been properly documented and quality flags have been assigned (it might be useful for public confidence in the QC process to have a "fact sheet" available that describes what tests have been carried out);
- Suitable governance procedures have been applied such that data cannot be changed without proper authorization. Note: Post-archival changes could arise, for instance, if a user identifies a clear error in the data, or if there is a subsequent change to the QC methodology and the data are retroactively QC'ed using new methods.

Regular technology migration is needed to ensure that final archives remain accessible in perpetuity.

For hard copy storage, the records management system should have a controlled environment:

- That uses storage materials that are acid-free;
- That is both fire- and water-protected;
- That is pest-free;
- That contains a full inventory of the records held;
- That is subject to appropriate handling procedures;

- That preserves all records; and
- That contains adequate security.

For a more detailed discussion involving the management of hard copy records, refer to *Guidelines on Best Practices for Climate Data Rescue* (WMO-No. 1182).

3.6 Step 5: Disaster recovery

The disaster recovery step involves ensuring that data can be recovered in the event of a disaster that causes loss or corruption of the data, in whole or in part. For databases or electronic data, this includes ensuring that there are multiple copies of the data and off-site backups, plus plans to ensure the efficient and complete restoration of the archive from these backup sources; the *Manual on the High-quality Global Data Management Framework for Climate* (WMO-No. 1238) provides further details of the requirements to ensure the security and recoverability of the NMHS climate record. For paper records, these requirements include appropriate storage and security (see above); it is strongly recommended that all paper documents be subject to imaging and digitization.

4. ROLE OF THE QUALITY ASSURANCE MANAGER

Noting the growing importance of reliable and trustworthy climate data, and with each NMHS having its own procedures and structures, management of the quality assurance process is very important. It is thus strongly recommended that each NMHS have a data QA manager responsible for ensuring that all aspects of the QA/QC process for climate data are properly followed and that any issues that arise be appropriately identified and addressed. The data QA manager should know the details of the process(es) along the whole data life cycle and have an overview of the state of the QA process throughout the NMHS even if various QA and QC tasks within the NMHS are distributed to different organizational areas. QA managers should regularly exchange their experiences internationally and attempt to stay abreast of emerging scientific developments pertinent to QC/QA. Ideally, processes within the QA/QC domain should aim to accord with a recognized ISO standard as this would provide added confidence in the processes and that the level of compliance with recognized quality standards is being met.

The following list describes the recommended responsibilities of the data QA manager to ensure quality in the recording, delivery and archiving of data. The tasks may be done by different sections of the NMHS or by external parties, but the data QA manager needs to ensure that they are carried out in such a way that the integrity of the climate record is not compromised. Observations metadata and documentation from these procedures should be available to the QC manager to ensure that there is an informed QC process.

Responsibilities:

- The data QA manager should ensure that instrument technology is fit for purpose by:
  - Ensuring that instruments are adequately tested, with performance and acceptance testing documented, ahead of field trials;
  - Running parallel field trials. For instruments that supersede older technology, parallel field trials need to be conducted for a minimum of one year (but whenever possible, for two or more years) to properly quantify the performance differences between the instruments. Ideally, these should be carried out before the new instruments are deployed operationally to ensure that the instruments perform reliably and consistently within set tolerances;
  - Ensuring that the results of the parallel field trials are retained and that they are available to data users/customers on request;
  - Ensuring that the instrument is sited operationally according to WMO or country specifications, with the instrument calibration tested against an instrument standard (where applicable) before deployment into service.
- For instruments already in operational service, the data QA manager should:
  - Ensure that instrument outputs are within tolerance through a regular inspection programme;
  - Ensure that observations are made at the required time intervals;
  - Regularly document and monitor the exposure of the instruments at the site, noting any significant changes that will affect the quality of the data (this information is transferred to the NMHS observations metadata system);
  - Check that routine maintenance is planned and takes place.

- For manually recorded observations, the data QA manager should:
  - Where observations are manually recorded in the station field book (record of observations), ensure that the field books[1] are checked for accuracy by the staff at the station. For sites with more than one observer, this responsibility usually falls to the officer in charge of the station;
  - Record metadata relating to any instrument or data quality issue in the field book or other such record of observations;
  - As far as practical, ensure that observers and observing networks employ standard practices and procedures (that is, consistent practices and observational methods) and that advice on any required changes in observational procedures is widely communicated and documented;
  - Check that refresher training is conducted and that new observers are trained when appointed.
- With respect to data ingest and archiving, the data QA manager should:
  - Ensure that adequate metadata and an audit trail for the data exist;
  - Ensure that the processes provide feedback to the site and/or observation supplier about systemic observation quality issues;
  - Ensure that checks on data flow are in place, addressing such potential issues as: data expected but not received, data corruption, time-stamping issues, and so forth (the data QA manager should have ongoing communication with the NMHS's ICT staff to ensure this);
  - Ensure that there are effective procedures for data recovery;
  - Ensure that backup and disaster recovery procedures are in place, documented, and fully understood (again, close liaison with the NMHS's ICT staff is required here);
  - Ensure that a hard copy of the data received is digitized and, as far as possible, imaged and that the hard copy records are archived.
- With respect to data quality, the data QA manager should:
  - Ensure that QC operators fully understand and can capably carry out the QC procedures of the NMHS (to ensure consistently high QC standards, it is suggested that periodic audits be conducted to verify that QC operators are adhering to standard and consistent processes);
  - Ensure the complete, continuous and timely performance of all QC procedures and tests;
  - Ensure that quality information is properly recorded as part of the metadata;
  - Ensure that all data are flagged in line with the flagging procedures of the NMHS;
  - Ensure that all procedural documentation is kept up to date and is accessible.

---

[1] This applies specifically to observations recorded on paper forms. At stations where manually observed data are entered into an electronic system, the data should also be checked in some manner.

ANNEX 1. QUALITY ASSURANCE AND QUALITY CONTROL TESTS: OVERVIEW

This annex provides a non-exhaustive high-level overview of the types of QA and QC tests that may be applied to the QA and QC of climate data, along with a suggested assessment of their relative importance. The tests are described in greater detail in Annex 2.

Table 2. Selected quality assurance and quality control tests and suggested importance

Legend: M = Mandatory, R = Recommended, O = Optional

Constraint tests: Tests to ensure that observations are technically and scientifically plausible based upon theoretical and climatological limits, sensor hardware specifications or database limits on data ingestion.

Name of test	*Short description*	*Notes*	*Suggested importance*
Sensor-based range test	Detects observations that are outside the range of theoretical limits or sensor hardware specifications		M
Database range test	Detects values that are outside the range of the ingestion criteria for the storage system	This test is carried out while the data are being ingested into the storage system.	M
Domain test	Determines whether the meteorological value is within the realm of scientific possibility, for example, T > 70 °C		M

Consistency tests: Tests utilizing comparisons with other parameters to ensure that inconsistent, unlikely or impossible combinations are either rejected or flagged as suspect. A manual investigation may then assess the validity of the suspect values.

Name of test	*Short description*	*Notes*	*Suggested importance*
Sub-daily test	This test is conducted to determine whether there are consistencies between values recorded at sub-daily observations, for example, three-hourly observations, and daily values.		M
Daily minimum versus daily ground minimum test	Compares the daily minimum and daily ground minimum temperatures		O
Hourly MSLP and SLP difference test	Tests to determine whether there is a significant change in the difference between mean sea-level pressure (MSLP) and station-level pressure (SLP) over two consecutive recordings	This test can be performed using hourly, three-hourly, etc. observations.	O
Precipitation multi-source comparison test	Checks whether data from one source is consistent with data from another source	The data sources may be different instruments or the same instrument from different communication paths.	R

Name of test	*Short description*	*Notes*	*Suggested importance*
Zero precipitation spatial test	Checks for instances when significant precipitation is recorded at one site but not at neighbouring sites, and vice versa. Note that the experience of QC operators may be required here to ensure that neighbouring comparison stations are sufficiently representative of the climate in the area.	Typically, this indicates precipitation being recorded on the wrong day or that the value is an accumulated total.	M
Insufficient neighbours test	Checks whether there is a sufficient number of stations within a reasonable distance of the candidate station to perform spatial tests	This test is generally performed on the first day of the month and for precipitation data.	R
Precipitation period test	Tests for overlap and underlap of rainfall accumulations (to check whether the period value of the record is inconsistent with the actual null precipitation reported dates)	Designed for rainfall precipitation records	R
Tracking test	Compares whether two or more elements, or instances of the same element at two neighbouring stations, rise and fall together	The two elements are expected to rise and fall together. This test is a very effective means of determining whether values should be rejected or flagged as suspect.	R
Maximum (minimum) air temperature consistency test	Checks for consistency between daily maximum (minimum) air temperature and sub-daily observations	This is an extension of the sub-daily tests referred to above.	R
Solar hour – astronomical test	Determines the difference between the sunshine duration and the calculated day length		M
Consistency test between air temperature and wet bulb temperature	Compares the air temperature to the wet-bulb temperature		M
Wet bulb versus dewpoint consistency test	Determines the difference between the wet bulb and dewpoint temperatures	The test fails if the difference is less than 0. Used with manual observations.	O
Wet bulb versus dewpoint test	Wet bulb temperature is tested against dewpoint temperature by recalculating the dewpoint temperature from the wet bulb air temperature	Used with manual observations	O
Dewpoint air temperature consistency test	Tests whether the dewpoint temperature is less than or equal to the air temperature		R
Visibility consistency test	Tests for visibility consistency with the present weather code against phenomena flags (fog, sandstorm, mist, dust storm)	Horizontal visibility. Used with manual observations	R
Total cloud amount consistency test	Tests for consistency of total cloud amount against various elements	Used with manual observations	R
Daily phenomena flag test	Checks the consistency of the daily phenomena flags with codes in the sub-daily tables and various other daily elements		R

Name of test	*Short description*	*Notes*	*Suggested importance*
Present weather consistency test	Checks the consistency of the present weather codes in the sub-daily tables against various other daily elements		R
Soil temperature test	Checks the consistency of soil temperatures at various depths		R
Message consistency test	Checks whether the messages all contain the same value for a given parameter where multiple messages are received for the same observation	Applies to all parameters	M
One-minute data (OMD) maximum minimum test	Compares the reported maximum or minimum temperature against the corresponding one-minute value at the time of maximum or minimum temperature to ensure that the values match	Daily maximum, daily minimum	R

Heuristic tests: Tests that rely on experience and knowledge of observation processes, techniques and instrumentation to detect inconsistent, unlikely or impossible values and flag them as suspect. A manual investigation may then asses the validity of the suspect values.

Name of test	*Short description*	*Notes*	*Suggested importance*
Relative humidity dry wet bulb test	Test to determine whether the wet bulb wick has dried out		R

Data provision tests: Tests to ensure that observations that do not match the expected schedule of observations are either rejected or flagged as suspect.

Name of test	*Short description*	*Notes*	*Suggested importance*
Observation received from the future test	Compares the local clock time of the observation to the time the observation was received		R
Period gap test	Determines whether the periods match the records that are present	Applies to daily observations	M
Large period test	Determines whether the periods are excessively large (many days) or less than one day	Applies to daily records	R
Future period overlap test	Test carried out to identify when an observation is sent late, and after receipt of a subsequent day's observation, to ensure consistency in the reported observation period		R
Rerun test	Checks whether the observation value, period or quality flag has changed since the last update	Applies to all parameters	R

Statistical tests: Tests that statistically analyse historical data to detect inconsistent, unlikely or impossible records and flag them as suspect. A manual investigation may then assess the validity of the suspect values.

Name of test	*Short description*	*Notes*	*Suggested importance*
Climate-based range test	Compares the meteorological value of the observation with the climatological upper and lower extreme values	Thresholds can be calculated to take into account seasonal variations in the observations	M
Flat line test	Checks the length of a run of meteorological values which are the same, that is, checks to determine whether the parameter is unchanged over time to an unlikely degree		R
Rapid change test	Checks that the difference between the previous and the current observation does not exceed a reasonable threshold		R
Spike test	Compares a given meteorological observation with the previous and following values	Similar to the Rapid change test; however, the Spike test looks for an unlikely rise then drop (or drop then rise)	R
Frequency (rounding) test	Checks for instances of excessive rounding of a value	Applies to manual observations where an operator rounds rather than interpolates values	O
Spatial variability test	Daily climatology comparing the differences between the station of interest and nearby stations		O
Mapping-based spatial test	Compares the meteorological value with the surrounding data values using a mapping analysis technique such as the Barnes analysis	The Mapping-based spatial test is used to weight the values from nearby stations to estimate a value at the location of the candidate station. Alternatively, the QC operator visually assesses the likelihood of a value at a suspect location against values at surrounding stations as shown on a map that includes topography.	R
Linear regression spatial test	Compares the meteorological value with the surrounding data values using linear regression		O
Linear regression spatial variability test	Uses the variability of the neighbouring stations rather than standard error estimates to calculate test limits	Variation on the Linear regression spatial test and the Spatial variability test	O
Linear regression multi-day test	Compares the climatology of the differences between the station of interest and nearby stations over a multi-day period	Similar to the Linear regression spatial test but applies to multiple days	R

Name of test	*Short description*	*Notes*	*Suggested importance*
Maximum/ minimum test	Checks to verify that the difference between the maximum and minimum temperatures is realistic (that is, greater than/equal to zero and less than an upper bound)	The upper bound is derived from a climatology using a minimum of 5 years (preferably 30 years) of data.	R

Sources: "QMS Test Specification" (Bureau of Meteorology, Australia (internal document)), *Climate Data Management System Specifications* (WMO-No. 1131), *Guide to Climatological Practices* (WMO-No. 100)

ANNEX 2. QUALITY ASSURANCE AND QUALITY CONTROL TESTS: DETAILS

1 Constraint tests

Tests to ensure that observations are technically and scientifically plausible based upon theoretical and climatological limits, sensor hardware specifications or database limits on data ingestion.

1.1 Sensor-based range test

Brief description: Detects observations that are outside the range of theoretical limits or sensor hardware specifications.

Parameters the test applies to: Temperature, humidity, barometric pressure, wind.

Detailed description: For automated observations, the Sensor-based range test is generally performed at the station. The limits are often set by the manufacturer, often with input from the NHMS. The sensor senses at one second intervals, and if the value exceeds the specified limit, the value is excluded. A minimum number of valid samples is required before a value is transmitted – the number of valid samples may vary from one manufacturer to another. From these valid samples, a maximum, minimum and mean are extracted. Testing also occurs for manual observations; for example, when resetting liquid in glass thermometers, the reset values are compared with the ambient air temperature. At field stations with both manual and automated stations, the liquid in glass thermometer readings is compared with temperature probes, applying reasonable tolerances.

Test frequency: As per observation frequency.

1.2 Database range test

Brief description: Detects observations that are outside the range of database ingest acceptance (database technical constraints).

Parameters the test applies to: All parameters stored in the database.

Detailed description: The Database range test can be performed at more than one point, such as at the point of ingest where the observed value first enters the NMHS, and subsequently when ingested into the CDMS. The aim is to avoid values created by an oversight, for example, during a correction or the ingest of data into database. Examples: temperature > 80 °C, negative pressure.

Test frequency: When the data are ingested into the CDMS and any upstream databases.

1.3 Domain test

Brief description: Determines whether the meteorological value is within the realm of scientific likelihood.

Parameters the test applies to: Temperature, humidity, barometric pressure, wind, and so forth.

Detailed description: This test checks whether an observed value lies between the 0.3 percentile and the 99.7 percentile of all values from the past 30 years. This test applies to both manual and automated observations. Where an observation is an accumulation value, the observation is divided by the period to test against one value per day.

Test frequency: Daily.

2 CONSISTENCY TESTS

Tests that ensure that inconsistent, unlikely, or impossible values are either rejected or flagged as suspect. A manual investigation may be conducted to assess the validity of suspect values.

2.1 Sub-daily test

Brief description: This test is conducted to determine whether there are consistencies between values recorded at sub-daily observations, such as three-hourly observations, and daily values.

Parameters the test applies to: Daily maximum and minimum air temperature, dry bulb temperature.

Detailed description: For daily minimum temperature, the difference between the recorded minimum value and the lowest three-hourly temperature over the previous 24 hours is calculated and tested.

For daily maximum temperature, the difference between the recorded maximum value and the highest three-hourly temperature over the next 24 hours is calculated and tested.

In both cases, a reading error of 0.5 °C is permitted to allow for instrument and observation tolerances in individual observations. Differences significantly less than zero should be flagged; very large differences should also be flagged.

Test frequency: Daily.

2.2 Daily minimum versus ground minimum test

Brief description: Compares the daily minimum and daily ground minimum temperatures.

Parameters the test applies to: Daily terrestrial minimum temperature, daily air minimum temperature.

Detailed description: This test checks the consistency of the daily minimum temperature, which is taken in the screen (for example, Stevenson screen), versus the ground minimum temperature, which is taken on the ground. The ground minimum will nearly always be less than the screen minimum. The test typically picks up issues in the ground minimum as that measurement can be problematic if not done correctly or if bubbles occur in the thermometer.

False positives can be generated by this test for sites that have a high level of heavy metals in the soil (for instance, iron oxide).

Test frequency: Daily.

Test order: The test is run after the air minimum temperature has been assessed.

2.3 Hourly MSLP and SLP difference test

Brief description: Tests to determine whether there is a significant change in the difference between mean sea-level pressure (MSLP) and station-level pressure (SLP) over two consecutive recordings.

Parameters the test applies to: Hourly MSLP and hourly SLP.

Detailed description: Tests to determine whether there is a significant change in the difference between MSLP and SLP over two consecutive recordings. The test can be performed using hourly, three-hourly, etc. observations. This test checks for transposition and calculation errors.

Note: At some stations which use the current temperature in the algorithm to derive MSLP from SLP, the SLP/MSLP relationship may change substantially through the diurnal temperature cycle. This issue arises mainly at higher-elevation stations where strong low-level inversions develop in winter.

Test frequency: At every observation.

2.4 Precipitation multi-source comparison test

Brief description: Checks whether data from one source are consistent with data from another source. The data sources may be different instruments or the same instrument from different data delivery paths or ingests.

Parameters the test applies to: Precipitation.*

Detailed description: This test checks to see that the data at a given data source are consistent with data from another co-located data source. The two data sources may be from different instruments or the same instrument from different data delivery paths or ingests.

In the case of data sources from the same instrument from different data delivery paths, it is expected that the two values will be the same. Where a discrepancy exists and it is not possible to make an objective decision as to which value is correct, the data are flagged accordingly. The same comment applies in the case of data sources from different instruments. In both cases, other tests and analyses (for example, spatial and temporal distribution) will aid in determining the most likely correct value. Other analyses outside of this test should be used to identify stations with consistent discrepancies and why such discrepancies exist.

Test frequency: Daily.

*Note: The same approach may be used for other elements measured using multiple sensors.

2.5 Zero precipitation spatial test

Brief description: Checks for instances when significant precipitation is recorded at one site but not at the neighbouring sites and vice versa. Typically, this indicates that precipitation is recorded on the wrong day or represents an accumulation over several days.

Parameters the test applies to: Daily precipitation.

Detailed description: This test checks for instances when there is a precipitation value recorded, but all the neighbours are zero and vice versa. This will typically be where the value is recorded on the wrong day. It may also represent an accumulated total at one or more stations; for instance, a three-day total may be recorded on a Monday if a manual station is unstaffed over a weekend. The discrepancy may not necessarily be picked up in spatial tests as the spatial tests may not adequately capture the spatial variability associated with showers. It should be noted that the efficacy of such checks may be affected if the “candidate” station is in a different climate regime compared to that of the so-called near neighbours.

This test is designed to have a low false failure rate instead of bringing up many potential suspects, so typically, the precipitation difference between the candidate and the surrounding stations would need to be significant.

Note: An addition to this test could be to test for zero values over multiple days for an extended period in situations in which neighbouring stations consistently report small amounts of precipitation not significant enough to trigger the test on any one day. This could occur, for instance, when a rain gauge becomes blocked. So, although the individual zero value may not deviate sufficiently from the neighbours to become suspect, the accumulated values do make the issue apparent.

Test frequency: Run daily but analysed in monthly blocks.

2.6 Insufficient neighbours test

Brief description: Checks whether there is a sufficient number of neighbouring stations to perform spatial tests. This test is generally performed on the first day of the month and for precipitation data.

Parameters the test applies to: Precipitation.

Detailed description: This test checks for cases in which there are insufficient neighbours to perform the suite of spatial tests. It is generally performed on the first day of the month (other days are not tested). If the result of the test is a failure, the QC operator will check all the days of the month.

Test frequency: The first day of the month only.

Test order: Run before other precipitation spatial tests.

2.7 Precipitation period test

Brief description: Consistency test assessing overlap and underlap for rainfall accumulations.

Parameters the test applies to: Daily precipitation.

Detailed description: This test assesses whether a record’s period value (the number of days over which the rainfall record is accumulated) is either greater or less than the actual number of days for which the precipitation data are reported as accumulated. This test is specifically designed for precipitation records. There are two parts in this test: overlap and underlap.

Overlap example: A record’s period value is greater than actual null precipitation reported dates. For example, if the period value is reported as three days on 4 May, but a non-zero rainfall value was last reported on 2 May, the actual period gap (rainfall accumulation) might be only two days – 3 May and 4 May. In this example, the rainfall value on 2 May would first be assessed for accuracy (the rainfall could have been recorded on the incorrect day). If the rainfall data on both 2 May and 4 May are assessed as correct, it is likely that there is an error regarding the period. To determine whether there is, in fact, a period error, the QC operator uses spatial data and other tools, such as radar, to confirm the analyses. The station record is also assessed to ensure that the data were correctly digitized.

Underlap example: A record’s period value is less than actual null precipitation reported dates. For example, if the period of rainfall accumulation is reported as two days on 6 May, but the last recorded non-null precipitation record is on 3 May, the actual period gap is three days.

Test frequency: Daily.

2.8 Tracking test

Brief description: Compares whether two or more elements, or instances of the same element at two neighbouring stations, exhibit expected synchronous behaviour (rise and fall together). This test is very effective and has low false error rates.

Detailed description: This test looks at how one element corresponds to one or more other elements, or how well observations of the same element correspond to those at a neighbouring site over time. It utilizes the very simple concept that two or more closely related elements are expected to show similar patterns of variation. In assessing the anomalies produced by this test, the QC operator must take into consideration the prevailing meteorological conditions, including local effects, and the spatial density of the network.

Combinations of elements to which tracking tests may be applied include:

- Mean sea-level pressure (MSLP) and station-level pressure (SLP);
- Terrestrial minimum temperature and minimum air temperature;
- Soil temperature at one depth and soil temperature at a different depth;
- Hourly dry bulb temperature, wet bulb temperature and dewpoint temperature (particularly for manually read thermometers);
- Daily evaporation and daily precipitation;
- Under 3 m wind run with 10 m wind run values.

Spatial correlations tracking between neighbours: Spatial correlations may be visually assessed by the QC operator by comparing the time series plot for a particular site to that of a neighbouring site on a graphic user interface in place of a computer algorithm. This approach may be applied to many parameters, particularly for networks with good spatial coverage. The following elements are generally assessed in this manner:

- Hourly dry bulb temperature;
- Hourly dewpoint temperature;
- MSLP and SLP;
- Wind speed.

Test frequency: In accordance with the observation frequency of both parameters.

2.9 Maximum (minimum) air temperature consistency test

Brief description: Checks for consistency between the daily maximum (minimum) air temperature and the values recorded at sub-daily observations. This is a specific application of Test 2.1 (Sub-daily test).

Parameters the test applies to: Maximum air temperature, minimum air temperature.

Detailed description: Checks for consistency between the daily maximum air temperature and the hourly or three-hourly air temperature, that is, the temperature that is recorded in the hourly or three-hourly records. In the case of three-hourly observations, no three-hourly value is allowed to exceed the daily maximum temperature recorded. A reading error of 0.5 degrees is permitted. This test can be applied if only one observation per day is available; however, the effectiveness of the test increases with more frequent sub-daily observations.

Notes:

1. This test is also applicable to minimum temperatures, as indicated in Test 2.1.
2. A variation of this test may be used to compare the maximum temperature with fixed-hour temperatures to ensure that the maximum is not higher than the fixed-hour temperatures by more than a specified amount. This variation is most effective for data series with a relatively high temporal resolution.

Test frequency: Daily.

2.10 Solar hour – astronomical test

Brief description: Determines the difference between the sunshine duration and the calculated day length in hours.

Parameters the test applies to: Sunshine duration.

Detailed description: This test determines the difference between the sunshine duration and the calculated day length. The test fails if the difference is greater than zero.

Test frequency: Daily.

2.11 Consistency test between air temperature and wet bulb temperature

Brief description: Compares the air temperature to the wet bulb temperature.

Parameters the test applies to: Dry bulb temperature and wet bulb temperature.

Detailed description: The dry bulb temperature and the wet bulb temperature are compared. If the wet bulb temperature, including a tolerance value, is greater than the air temperature, the test fails.

Note that while the tolerance in this test applies to manually read liquid in glass thermometers, the same algorithm is usually applied to electronic temperature sensor data. While it is recognized that the tolerance for electronic temperature sensors should be much lower, it is convenient in a hybrid network to use the same algorithm due to difficulties in discriminating instrument type at the QC stage.

Test frequency: Hourly but may be applied to any fixed-hour observation.

2.12 Wet bulb versus dewpoint consistency test

Brief description: Determines the difference between the wet bulb and dewpoint temperatures.

Parameters the test applies to: Wet bulb temperature and dewpoint temperature.

Detailed description: Determines the difference between the wet bulb and dewpoint temperatures. The test fails if the wet bulb temperature is recorded as below the dewpoint temperature. This test applies to manually read thermometer readings.

Test frequency: As per observation frequency.

2.13 Wet bulb versus dewpoint test

Brief description: Wet bulb temperature is tested against dewpoint temperature.

Parameters the test applies to: Wet bulb temperature and dewpoint temperature.

Detailed description: Wet bulb temperature is tested against dewpoint temperature by recalculating dewpoint temperature from the dry bulb and wet bulb air temperature. The observed and recalculated values of the dewpoint should agree to within one degree. This test is used for manual observations where the dewpoint temperature is calculated by the operator. It has greater sensitivity than Test 2.12 (Wet bulb versus dewpoint consistency test) and can be run after Test 2.12 to check for incorrectly calculated dewpoints that pass Test 2.12. For dewpoint values that fail the Wet bulb versus dewpoint test, the QC operator should recalculate the dewpoint using the observation method in use for the corresponding country.

Test frequency: As per observation frequency.

2.14 Dewpoint air temperature consistency test

Brief description: Tests whether the dewpoint temperature is less than or equal to the air temperature.

Parameters the test applies to: Dewpoint temperature and air temperature.

Detailed description: Air temperature should always be greater than or equal to dewpoint temperature, with dewpoint temperature equal to air temperature only for air that is saturated. Situations in which air saturation can occur include precipitation and fog events. It is important that the QC operator assess whether the dewpoint temperature is the result of saturated conditions and not due to observer error in calculating the dewpoint or due to a wet bulb wick drying out.

Test frequency: As per observation frequency.

2.15 Visibility consistency test

Brief description: Tests for visibility consistency with the present weather code.

Parameters the test applies to: Present weather and visibility.

Detailed description: Tests for visibility consistency with the present weather code against phenomena flags (fog, sandstorm, mist, dust storm). This test concerns horizontal visibility and is used with manual observations.

Examples of unlikely/impossible combinations include the following (to be adapted to different climates): Visibility above 1 000 m and fog, visibility above 1 000 m or below 200 m and slight or moderate sandstorm or dust storm, visibility above 10 km and mist, and so forth.

Test frequency: As per observation frequency.

2.16 Total cloud amount consistency test

Brief description: Tests for consistency of total cloud amount against various elements. Used with manual observations.

Parameters the test applies to: Total cloud amount, present weather.

Detailed description: Tests for consistency of total cloud amount against cloud height and present weather.

Examples of unlikely/impossible combinations: Total cloud amount less than low-level, middle-level, or high-level cloud amount; total cloud amount invisible, but there are observed precipitation trails (virga); cloudless, yet there is the occurrence of precipitation, and so forth.

Test frequency: As per observation frequency.

2.17 Daily phenomena flag test

Brief description: Checks the consistency of the daily phenomena flags with sub-daily observations.

Parameters the test applies to: Phenomena including hail, snow, thunderstorms, dust storms, strong wind, gales, frost, mist, haze, fog, and so forth.

Detailed description: The Daily phenomena flag test checks the consistency of the daily phenomena flags (days with hail, thunder, fog, and so forth) with codes in the sub-daily tables and various other daily elements. It detects instances where a phenomenon has been reported in a present weather synoptic code but has not been flagged in the midnight – midnight[1] phenomena categories and vice versa. It may also detect inconsistencies between phenomena and other variables, such as temperature, dewpoint, and so forth.

Examples of suspect combinations include the following (note: these may need to be adapted to different climate conditions): snow flag and minimum air temperature above 9 °C; no thunderstorm flag but thunderstorm observed during the day, no dust/haze/fog flag but visibility below related thresholds during the day; frost flag but minimum air temperature above 4 °C, and so forth.

Test frequency: Daily.

2.18 Present weather consistency test

Brief description: This test checks the consistency of the present weather codes in the sub-daily tables and various other daily elements.

Parameters the test applies to: Present weather and temperature and so forth.

Detailed description: The Present weather consistency test flags unlikely combinations of present weather and temperature, such as freezing precipitation or fog depositing rime and warm air temperatures. Other combinations can be included at the discretion of the NMHS.

One example of this test is a comparison of the air temperature with observations of rime deposits and freezing precipitation.

Test frequency: Daily and sub-daily.

2.19 Soil temperature test

Brief description: Checks the consistency of soil temperatures at various depths.

Parameters the test applies to: Soil temperature readings at depths of 5 cm, 10 cm, 20 cm, 50 cm, 1 m.

[1] Or another 24-hour period, depending on the NMHS

Detailed description: The Soil temperature test looks at various combinations of soil temperatures at different depths (these temperatures may need to be adapted according to different climates):

Target element	*Test expression (suspect if true)*
Soil temperature at a depth of 5 cm (SOIL 5 temp)	SOIL 5 temp = SOIL 10 temp AND SOIL 5 temp = SOIL 20 temp AND SOIL 5 temp = SOIL 50 temp AND SOIL 5 temp = SOIL 100 temp
	SOIL 5 temp – SOIL 10 temp > 16 OR SOIL 10 temp – SOIL 5 temp > 6
Soil temperature at a depth of 10 cm (SOIL 10 temp)	SOIL 10 temp – SOIL 20 temp > 10 OR SOIL 20 temp – SOIL 10 temp > 6
Soil temperature at a depth of 20 cm (SOIL 20 temp)	SOIL 50 temp – SOIL 20 temp > 6 OR SOIL 20 temp – SOIL 50 temp > 7
Soil temperature at a depth of 50 cm (SOIL 50 temp)	SOIL 50 temp – SOIL 100 temp > 7 OR SOIL 100 temp – SOIL 50 temp > 4
Note: Soil temperatures are in degrees Celsius.	

Test frequency: As per observation frequency.

2.20 **Message consistency test**

Brief description: Checks whether the messages all contain the same value for a given parameter where multiple messages are received for the same observation.

Parameters the test applies to: All parameters stored in the database.

Detailed description: The Message consistency test picks up instances where multiple messages are received for the same observation. The test checks to ensure that the messages contain the same value for a given parameter. If there are multiple messages with a different value, the test will fail.

Test frequency: As generated.

2.21 **One-minute data (OMD) maximum minimum test**

Brief description: Compares the reported maximum or minimum temperature against the corresponding one-minute value at the time of maximum or minimum temperature to ensure that the values match.

Parameters the test applies to: Maximum air temperature and minimum air temperature.

Detailed description: The OMD maximum minimum test compares the reported maximum or minimum temperature to the corresponding one-minute value at the time of maximum or minimum temperature to ensure that the values match. This test is designed to identify AWS clock drift and missing one-minute data. If the one-minute air temperature does not match the daily maximum air temperature or daily minimum air temperature, or if the one-minute data are missing at the time of the reported maximum or minimum temperature, the test will fail. This test also includes a coarse spike test that compares the one-minute value for both the minute before and the minute after the daily extreme to ensure that the difference is not greater than 0.5 °C for the minimum temperature and 1.0 °C for the maximum temperature.

Test frequency: Daily.

3. HEURISTIC TESTS

Tests (usually multivariate) that rely on experience and knowledge of observation processes, techniques and instrumentation to detect inconsistent, unlikely or impossible records and flag them as suspect. A manual investigation may then asses the validity of the suspect values.

3.1 Relative humidity dry wet bulb test

Brief description: Test to determine whether the wet bulb wick has dried out or is contaminated.

Parameters the test applies to: Wet bulb temperature, dry bulb temperature.

Detailed description: This test looks for a wet bulb wick that has dried out by examining the 3 p.m. observations where the relative humidity is > 90%. To reduce the false error rate, cases where there is precipitation and/or fog in addition to a relative humidity of > 90% are eliminated by examining past and present weather codes.

The test will be performed only if:

- The observation time is 3 p.m. (or the equivalent time when the maximum thermometer is read in the afternoon);
- The elevation is < 1 100 m (it is not possible to distinguish the effects of orographic clouds from the effects of drying wet bulbs using this test).

Test frequency: Daily at 3 p.m. (or the equivalent time when the maximum thermometer is read in the afternoon).

4 DATA PROVISION TESTS

Tests to ensure that observations that do not match the expected schedule of observations are either rejected or flagged as suspect.

4.1 Observation received from the future test

Brief description: Usually presented as a report of instances where the observer has sent the observation ahead of time.

Parameters the test applies to: Compares the local clock time of the observation to the time the observation was received.

Detailed description: An NMHS or a local regional entity will set a limit constituting the allowable ‘earliness’ of the observation. For instance, a 30-minute limit will accept data from observations reported up to 30 minutes earlier than the official observation time. It is recommended that a log identifying consistently early observations received from a station be maintained to highlight deficiencies in observational practice; these logs can be used to prioritize training requirements.

Test frequency: As generated.

4.2 Period gap test

Brief description: Determines whether the period (day accumulation) matches the records that are present.

Parameters the test applies to: Daily parameters.

Detailed description: The period gap test applies to daily observations and simply tests to see whether the periods (day accumulations) match the records that are present. For instance, a synoptic message is received for Monday morning for a station that has not recorded data on the weekend. The maximum and minimum temperature values in the database for that Monday will need to be checked as they may be multi-day values even though they are archived as a single day value. See also the earlier discussion regarding Test 2.5 (Zero precipitation spatial test).

Test frequency: Daily.

4.3 Large period test

Brief description: Determines whether periods (day accumulations) are excessively large (a period of several days) for parameters where a large gap between daily readings may affect the integrity of the data. For instance, if rainfall is not read for five days in a hot climate, the rainfall total is likely to be reduced through evaporation.

Parameters the test applies to: Daily rainfall and evaporation readings.

Detailed description: The Large period test picks up unreasonably large periods that may affect the validity of the reading. For rainfall and evaporation, a smaller period gap may be acceptable depending on the climate and user requirements of the NMHS. This test can be used by the NMHS to help set limits on the number of days beyond which rainfall and evaporation totals may be considered suspect.

Test frequency: Daily.

4.4 Future period overlap test

Brief description: Test carried out to identify when an observation is sent late, and after receipt of a subsequent day's observation, to ensure consistency in the reported observation period.

Parameters the test applies to: Daily parameters.

Detailed description: The Future period overlap test is carried out to identify when an observation is sent late, and after receipt of a subsequent day's observation, to ensure consistency in the reported observation period. If the period reported in the subsequent day's message does not match the expected period, the test will fail. An example of a situation which would lead to the failure of the Future period overlap test would be the following:

- The daily observation on the 8th is not sent on time.
- The daily observation on the 9th is sent with a period of two.
- After receipt of the daily observation for the 9th, the daily observation for the 8th is sent, which leads to an inconsistency in the period reported on the 9th.

Test frequency: Daily.

4.5 Rerun test

Brief description: Checks whether the observation value, period or quality flag has changed since the last update. If there have been any changes since the last update, the test will fail.

Parameters the test applies to: All parameters stored in database.

Detailed description: The rerun test checks whether the observation value, period or quality flag has changed since the last update. If there have been any changes since the last update, the test will fail. The test will not fail if the value, period or quality flag has been changed by a QC operator who is listed on the accepted users list. QC actions, therefore, will not trigger a test failure.

Test frequency: As generated.

5 STATISTICAL TESTS

Tests that statistically analyse historical data to detect inconsistent, unlikely or impossible records and flag them as suspect. A manual investigation may then asses the validity of the suspect values.

5.1 Climate-based range test

Brief description: Compares the meteorological value with the climatological upper and lower values. Different thresholds can be applied to take into account seasonal variations in the local climate. Values falling outside these bounds will be flagged as suspect, or at minimum, subject to further investigation.

Parameters the test applies to: Generally, daily observations. However, this test can also be used for sub-daily elements if a suitable QMP has been determined.

Detailed description: Compares the meteorological value with the climatological upper and lower values. The efficacy of this test is reliant on the temporal record being sufficiently long (at least 30 years is recommended) to ensure that the QMP represents full climate variability. Hence, this test will be unreliable for new sites and sites with incomplete temporal records; for these types of sites, the QMP of a nearby station with a similar climate but a longer temporal record may be used.

In place of a fixed QMP, the QMP may be based on a statistical test (for example, more than three or four standard deviations from the mean). However, care should be taken in setting such QMPs as some meteorological variables have frequency distributions which differ substantially from the normal (Gaussian) distribution.

Test frequency: Daily and sub-daily.

5.2 Flat line test

Brief description: Checks whether a run of meteorological values remains the same over a suspiciously long period, that is, checks to determine whether the parameter is unchanged over time.

Parameters the test applies to: Daily and sub-daily parameters.

Detailed description: Flat lines may occur due to faulty equipment or when manually entered data are incorrectly entered. It is important to note that this test is of limited value for some parameters that have small variability, such as soil temperatures taken well below the surface level.

Test frequency: As per observation.

5.3 Rapid change test

Brief description: Checks the difference between the currently observed value and the previous value for significance.

Parameters the test applies to: Sub-daily parameters with an observation frequency of three hours or less.

Detailed description: The Rapid change test detects problems in the data where there has been a rapid rise or fall in the data with time due to non-meteorological factors. Because this test assumes that the rapid change is not due to meteorological variability, it loses sensitivity for longer observation time intervals.

The parameter variability may not be constant for every station/month, and it may not be constant for every hour. This temporal and spatial variation can justify the calculation of discrete QMPs at the station, month and hour level.

When the time period between the observed value and the next or previous value is greater than three hours, this test is generally not performed.

Test frequency: As per observation.

5.4 Spike test

Brief description: Compares a given meteorological observation with the preceding and following values.

Parameters the test applies to: Sub-daily data. This test is particularly suited to high frequency AWS data.

Detailed description: The Spike test is similar to the Rapid change test; however, the Spike test looks for sudden jumps followed by sudden drops in a variable, or vice versa. Such variations may reflect, for instance, a voltage spike at an AWS. The QMP limits on the Spike test are smaller than those on the Rapid change test. When the time period between the observed value and the next or previous value is greater than three hours, this test is not performed. The Spike test performs best on high frequency data for observation periods shorter than 20 minutes.

Test frequency: As per observation.

5.5 Frequency (rounding) test

Brief description: Checks for instances of excessive rounding of a value. This test is applicable for manual observations where an operator rounds to the graduation marks on the instrument rather than estimating the value between the instrument graduations.

Parameters the test applies to: This test is usually used when thermometers and rain gauges are read manually.

Detailed description: The frequency of occurrence of the last digit (for example, the first decimal place for elements such as temperature) is examined over a number of days and stored as a QMP.

If the frequency of one of the values is 'high', for example, if there is a suspiciously large number of values ending in zero, the test is regarded as failed. The frequency is expressed in terms of a percentage of the values analysed. Note that rounded values are not assessed as suspect; they are given a flag to indicate a lower level of confidence. This test is important for identifying training issues.

The Frequency (rounding) test is performed for an entire month and assigned to a nominated element and date.

This test is not performed unless at least 75% of the values exist (and are non-zero) over the period in question.

Test frequency: Monthly.

5.6 Spatial variability test

Brief description: Differences between the station of interest and nearby stations are compared.

Parameters the test applies to: Daily and sub-daily parameters.

Detailed description: The difference between a station and its neighbours is calculated and compared against a reasonable threshold, for instance, the standard deviation of such differences derived from the historical record.

Test frequency: As per observation frequency.

5.7 Mapping-based spatial test

Brief description: Compares the value of the meteorological element with the surrounding data values. A mapping-based technique such as the Barnes analysis is used to weight nearby stations.[2]

Parameters the test applies to: Daily and sub-daily parameters.

Detailed description: The aim of the mapping analysis is to estimate a value at a target station based on the actual values at points in the general region of the station using an objective weighting function (nearby points should have more influence and should therefore be allocated a greater weighting).

The basic principle of the scheme is that a value of the variable under consideration is evaluated through mapping-based interpolation from a number of network points[3] – those network points corresponding to station locations. At each station location, there is the actually observed value and the interpolated value based on data from all other nearby observations (with an observation's influence on a point being determined by a weighting function). At the point P under analysis, the observed value is compared to the interpolated value estimated by the mapping.

Constraints:

1. This test loses sensitivity for comparison stations greater than a set distance from the target station. NMHSs will need to determine the maximum range for the test. In Australia, data-sparse regions are relatively flat, with a homogenous climate, thus the range (R) is set to 200 km, and stations beyond this distance are excluded from the analysis.

---

[2] Other types of spatially based mapping analyses include, for instance, analyses which incorporate elevation as a third dimension for variables such as temperature and precipitation that are topographically sensitive.

[3] In this description of the algorithm, the network is simply the location of the stations, so the interpolation is done directly to the station locations. Alternatively, if a gridded analysis of the field is available, the interpolation to station locations can be done from the grid. This alternative might be suited to, for instance, NWP-related tests.

2. The results of this test should be interpreted according to the local characteristics of the site. In particular, large gradients can occur in some meteorological variables in complex topography or near coastlines. A detailed interpretation of the results of this test may require some knowledge of the sites involved. For example, a site on a hilltop or slope may consistently record higher minimum temperatures than other sites in the region which are in valley or plain locations.

Note: Apart from statistical interpolation, the reasonableness or otherwise of values may be assessed via a visual inspection of the values at the suspect station against the values at the surrounding stations. In the case of a visual inspection, the same constraints apply as those stated above.

Test frequency: Daily and sub-daily but can be applied to other time periods such as monthly.

5.8 Linear regression spatial test

Brief description: Compares the meteorological value with the surrounding data values using linear regression.

Parameters the test applies to: Daily precipitation but can be adapted for other parameters.

Detailed description: Linear regression spatial analysis[4] produces an estimate for an element at a target station and time based on how well the neighbouring stations have correlated with that target station over time. The correlation is calculated by means of linear regression, and stations which better correlate with the target station are given a higher statistical weighting than those stations which do not correlate as closely.

Test frequency: Daily but can be adapted to other observation frequencies.

5.9 Linear regression spatial variability test

Brief description: Uses the variability of the neighbouring stations rather than standard error estimates to calculate test limits. This test is a variation on the Linear regression spatial test.

Parameters the test applies to: Daily rainfall but can be adapted to other parameters.

Detailed description: This test uses the variability of observations at neighbouring stations (at the same observation time), rather than standard error estimates, to calculate test limits. It is designed to assess the reliability of neighbouring stations in analysing variations. This test is well suited to precipitation when there are changes in the variability of a set of observations in a given region.

Constraints:

1. This test loses sensitivity for comparison stations that are too far away or too different in elevation from the target station. NMHSs will need to determine the maximum range for the test. In Australia, the elevation range for a "homogenous" climate is set to a maximum of 300 m.

2. The type of rainfall situation should also be taken into account as in situations where isolated thunderstorms occur; even stations which can be considered close to each other can differ significantly in their daily totals.

Test frequency: Daily but can be adapted to other observation frequencies.

[4] Adapted from a method described in Hubbard, K. G.; Goddard, S.; Sorensen, W. D.; Wells, N.; Osugi, T. T. Performance of Quality Assurance Procedures for an Applied Climate Information System. *Journal of Atmospheric and Oceanic Technology* **2005**, *22* (1), 105–112. https://doi.org/10.1175/JTECH-1657.1.

5.10 Linear regression multi-day test

Brief description: Compares the climatology of the differences between the station of interest and nearby stations over a multi-day period. This test is similar to the Linear regression spatial test but applies to multiple days. This test is sensitive for detecting potentially suspect rainfall stations where the observer regularly fails to read the rain gauge.

Parameters the test applies to: Daily rainfall but can be adapted to other parameters.

Detailed description: Although this test is similar to the Linear regression spatial test, the values it uses need to be defined differently. For accumulations, precipitation is the sum of the neighbours' daily readings, maximum temperature is the highest of the neighbours' daily readings, and minimum temperature is the lowest of the neighbours' daily readings.

Test frequency: Daily; however, can be adapted to other observation frequencies.

5.11 Maximum/minimum test

Brief description: Checks to verify that the difference between the maximum and minimum temperatures is realistic (that is, greater than/equal to zero and less than an upper bound). Bounds are derived from climatology using a minimum of 5 years (preferably 30 years) of data. This test is sensitive for detecting manually read misreads, which tend to be inaccurate by 5 or 10 °C.

Parameters the test applies to: Daily maximum and minimum air temperature values.

Detailed description: The difference between the maximum and minimum temperature is calculated. The test fails if the difference is less than zero (the maximum is less than the minimum) or exceeds an upper bound.

If the difference is less than zero, the test results in an error, but if it exceeds the upper bound, it results in a suspect value.

Test frequency: Daily.

____

For more information, please contact:

World Meteorological Organization

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland

Strategic Communications Office

Tel.: +41 (0) 22 730 87 40/83 14 – Fax: +41 (0) 22 730 80 27

Email: communications@wmo.int

public.wmo.int

JN 21373