DATA QUALITY CONTROL

(Dr. E.I. Sarukhanian, WMO)

INTRODUCTION

Meteorological observations are exchanged between countries on a worldwide basis. Users need to be confident that the observations they receive from other countries are made to a known standard as agreed upon and set by WMO. The accuracy of the data used is of primary importance to many kinds of analyses, computations and scientific investigations. The need for quality control of observational data, therefore, arises from the fundamental importance of obtaining data of the highest possible quality for all purposes, including the World Weather Watch programme, regional and national requirements and international research programmes.

The basic characteristics of quality control and the general principles to be followed both as standard and recommended practices are laid down in the Manual on the Global Observing System, Volume I (Annex V to the WMO Technical Regulations), Global Aspects, Part VI, (WMO-No. 544), the Manual on the Global Data Processing System, Volume I, (Annex IV to the WMO Technical Regulations), Global Aspects, Part II, (WMO-No. 485), and the Manual on the Global Telecommunication System, Volume I, Global Aspects, Part II, Attachment II-8 (WMO-No. 386). Detailed information on implementation of quality control procedures is also contained in the Guide on the GOS, Part VI (WMO-No. 488 revised version) and the Guide on the GDPS, Parts III, VI and VII (WMO-No. 305). The requirements at the observation site with regard to exposure, measurement and observation are set out in detail for each variable to be measured in the Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8). The latter contains in addition a special chapter outlining quality control procedures for such observations as a whole and dealing with the quality control of upper-air observations in more detail.

1. BASIC CHARACTERISTICS OF QUALITY CONTROL

The purpose of quality control of observational data shall be error detection, possible error correction and therefore error prevention, in order to ensure the highest possible standard of accuracy for the optimum use of these data by all possible users.

Within the framework of the GOS, quality control shall be a real-time activity which has to be performed prior to the transmission of the observational data on the GTS.

Quality control on a real-time basis shall also take place in the GDPS, prior to the use of the observational data in data processing (i.e. objective analysis and forecasting).

By meteorological quality control is meant the checking of the meteorological content of observational data. In contrast, telecommunications error-control, which is carried out as an inherent function of the GTS, aims solely at defecting errors introduced during transmission. As it is noted, procedures for error control in respect of data transmissions on the GTS are given in the Manual on the GTS, Volume I, Attachment II-8.

Quality control shall also be performed on a non-real-time basis, prior to forwarding the observational data for storage.

Quality control shall be applied to all observational data obtained from either the surface-based or the space-based sub-system. The observational data obtained from these sub-systems shall meet specific minimum standards and practices of quality control. Either manual and/or automatic methods shall be used in quality control.

2. GENERAL PRINCIPLES

2.1 Responsibility

The primary responsibility for quality control of all observational data shall rest with the Members from whose Services the observations originated.

Members should pay due attention to the quality control of observational data at the national level, aiming at the prevention of errors at the observational site, as well as the National Meteorological Centres. It is of the utmost importance, therefore, that Members should make adequate provision for quality control of data to ensure that they are as free from error as possible.

To detect errors which may escape the national quality-control systems and errors introduced subsequently, Members concerned should ensure that RMCs and WMCs also carry out appropriate quality control of observational data they receive.

Members shall inform the Secretary-General (for general dissemination) of any special features of their observing systems which may be important in the correct interpretation of the data provided.

2.2 Relay of data

Quality control of observational data needed for operational use shall not cause any significant delay in onward transmission on the GTS.

2.3 Minimum standards

According to the Manual on the GOS, Members are obliged to implement minimum standards of real-time quality control at all levels for which they are responsible (e.g. observing stations, NMCs, RMCs, WMCs) and according to the Manual on the GDPS, they are recommended to do so before data received via telecommunication links are processed. The minimum standards apply both to real-time and non-real-time processing and should lead to various records of quality control actions.

For the NMCs not capable of implementing these standards Members concerned should establish agreements with an appropriate RMC or NMC to perform the necessary quality control on an interim basis.

Quality-control standards may be introduced progressively at a GDPS centre using a modular approach. The general priorities for implementation of the minimum standards under the modular concept concern quality control of data, according to:

(a) sources (e.g. stations);

(b) type (e.g. SYNOP, TEMP); (e.g. 00 UTC, 12 UTC);

(c) time

(d) parameters and characteristics (e.g. pressure, wind, temperature, amount of precipitation).

WMCs having multiple responsibilities as an RMC and/or an NMC, and RMCs also having a responsibility as an NMC, should assume the minimum standards pertinent to all levels at which the centre operates.

Table 1 lists the minimum standards for real-time and non-real-time quality control at NMCs, RMCs and WMCs (see Annex A). Where applicable, regional associations and national Meteorological Services will set up similar quality-control standards for data exchanged only at regional or national levels.

The minimum standards specify which data are to be quality-controlled and how often. The detailed methods for performing the quality control are left to the Members to develop, but should conform to the minimum standards. The geographic area (zone) of responsibility for application of the minimum standards will correspond to that undertaken by each WWW centre for data processing, as laid down in Attachment III.2 of the Manual on the GDPS (see Annex B).

An essential part of the quality-control plan includes an exchange of information about data deficiencies between adjacent centres and between NMCs and observation points in order to resolve those deficiencies and minimize their recurrence.

The frequencies with which information is exchanged in order to improve the quality of data and products should correspond to the frequency with which monitoring reports are exchanged. These are given in the plan for Monitoring the Operation of the WWW, as given in Attachment II.14 of the Manual on the GDPS.

2.4 Advanced standards

The primary purpose of qualilty control is to detect data deficiencies and to attempt to correct them in real time. Thus, the WWW centres should perform quality-control operations as these are developed and as their technical capabilities allow. Centres which have high-speed computers can apply standards for quality control which are far beyond the minimum standards. These advanced standards should involve more real-time quality control, including correcting or flagging of more reports, parameters and levels than listed in Table 1. The Guide on the GDPS gives information on methods for more advanced quality control.

It is also the responsibility of automated centres to perform nearly continuous inspection and quality control of processing programmes that enable computers to identify, decode process and array data properly.

2.5 Minimum standards for processed data

Minimum standards for quality control of processed data should include:

(a) standards for presentation of processed data as they are given in Attachment II.4 to the Manual on the GDPS;

(b) spatial and temporal coherence in the meteorological structure of the product (that is, no impossible or contradictory atmospheric states).

3. LEVELS OF APPLICATION OF QUALITY CONTROL PROCEDURES

It is generally agreed that quality control of meteorological data begins with the installation of the instruments at the observation site and ends with the last stage of processing prior to final delivery of the data to the user. This means that observational data have to be quality controlled at different levels of data transfer in both real-time and non-real-time, using various procedures:

Levels of quality control procedure:

(a) quality control procedures at the observation site starting with data acquisition by manual or automatic meteorological stations;

(b) quality control procedures at collecting centres prior to the transmission of observational data on the GTS;

(c) quality control procedures for transmission at GTS-Centres (standard telecommunication procedures, e.g. error-detection and control);

(d) quality control procedures at GDPS Centres and other available facilities.

According to the Manual on the GOS, quality control should be performed at the following stages:

(a) during and after obtaining Level I data;

(b) during the reduction and conversion into Level II data;

(c) after obtaining Level II data;

(d) during and after coding of Level II data.

Note: Definition of data levels.

Level I: Primary data. In general these are instrument readings expressed in appropriate physical units and referred to Earth co-ordinates. Examples are: radiances or positions of constant-level balloons, etc., but not raw telemetry signals. Level I data still require conversion to the meteorological parameters specified in the data requirements.

Level II: Meteorological parameters. These are obtained directly from many kinds of simple instruments, or derived from the Level I data (e.g. average winds from subsequent positions of constant-level balloons).

Level III: Initial state parameters. These are internally consistent data sets, in grid-point form obtained from Level II data by applying established initialization procedures. At those centres where manual techniques are employed, Level III data sets will consist of a set of manually-produced initial analyses.

3.1 Quality control during and after obtaining level I data

At this stage, quality control should seek to eliminate:

(a) measurement errors (systematic or random);

(b) errors due to departure from technical standards;

(c) errors due to unsatisfactory exposure of instruments;

(d) subjective errors on the part of the observer.

The type of quality control necessary depends to a large extent on the type of the observation platform (surface or space-based, manned or automatic, etc). Details are given in the Guide to Meteorological Instrument and Methods of Observation (WMO-No. 8).

3.2 Quality control during the reduction and conversion into Level II data

Quality control during the reduction and conversion of Level I into Level II data should seek to eliminate errors resulting from conversion techniques used or the computational procedures involved.

(1) In modern automatic data-acquisition systems, the high sampling rate of observations and the generation of noise may necessitate checking of the data just prior to the conversion or thereafter. The quality of the reduced data may be improved by applying special data-processing procedures like interpolation, filtering and smoothing.

(2) Some meteorological instruments provide "raw" data in excess of what is needed to derive parameters of the state of the atmosphere. Such redundant systems (e.g. special rawinsonde observations) permit the conversion to Level II data through the use of alternative computational aprocedures. Observation values produced by different conversion formulae will have different standards of accuracy.

(3) In some observational systems, the deduced values may depend on computational procedures used to solve systems of equations that do not have a unique solution, for example, upper-air temperature derived from spectral radiances obtained by remote sensing from satellites. The quality of the reduced data will depend on the method selected to solve the governing equations.

3.2.2 Quality control after obtaining Level II data

Methods and rules to be used by the responsible centres should include one or more of the following general methods:

(a) internal consistency;

(b) logical inference;

(c) tests against limits;

(e) statistical rules;

(f) empirical rules

Members are free to choose methods of quality control they wish to use as long as these methods conform to the minimum standards. Details of these methods are described in the Guide on the Global Observing System (WMO-No. 488) and given below:

(a) Internal consistency is established by comparing various quantitative meteorological variables or by using qualitative information. Examples of internal consistency checks are:

- station pressure with mean sea level pressure, if appropriate, or station pressure with tendency;
- temperature with dew point depression;
- height of base of cloud with type of cloud;
- sea state with reported wind speed;
- high relative humidity with such weather phenomena as rain, snow, sleet, fog, dense mist and drizzle;
- check of the upper-air sounding to ensure that winds reported are less than, or equal to the maximum wind for the sounding.

(b) Logical rules can be used to check the consistency of qualitative meteorological information. A typical example is: present weather compared with type of cloud;

(c) Consistency in space may be accomplished by using one-dimensional, two-dimentional, three-dimensional checks or combinations. Examples are:

- hydrostatic checks

- horizontal check of observations of adjacent stations using interpolation methods and dynamical constraints

- complex horizontal and vertical checs (using both hydrostatic relation and statistical rules);

(d) consistency in time is established mainly by checks or a sequence of meteorological variables against tendencies or by comparison of the present value with that obtained previously. Examples are;

- check of pressure against pressure tendency

- check of reported air temperature with maximum and minimum temperature for the day as appropriate

- check on the reported position of a ship in relation to its previous position and its speed of movement.

In space-time quality control procedures use is made of physical and dynamical constraints which exist between non-simultaneous observations taken at adjacent stations or locations. Examples of quality control methods based on checking consistency in space and in time using statistical structure parameters within the interpolation scheme are given in the Guide on the GOS, Part VI, Appendix H.

(e) Tests may be performed against physical, climatological and absolute limits. Examples are:

- tests of actual lapse rate against dry adiabatic lapse rate in the free atmosphere

- tests against climatological parameters

- tests against absolute limits (e.g. wind direction within the range 0-360 degrees).

(f) Statistical rules are applied in several quality control procedures, for example:

- tests against statistical limits

- tests using statistical structure parameters within interpolation schemes.

(g) Empirical rules.

Some of the quality control procedures make use of empirical relationships between the meteorological variables concerned. For example: checking the empirical relationship between consecutive levels of a complete wind profile is applied.

3.4 <u>Quality control during and after coding of Level II data</u>

Since the purpose of quality control within the framework of the GOS is to provide the highest possible reasonable standard of observational data before they are distributed to the users, quality control should also be extended to the stage when the data are encoded. The control should include identification, code and format checks. Observing stations that receive the NMC output should check that the data for their station are in accordance with the observations in the station register.

4. ASPECTS OF IMPLEMENTATION

The quality control programme should be introduced progressively, using a modular approach. The modular concept shall apply in both the real-time and non-real-time modes. The progressive implementation of the quality control shall be consistent with the priorities given in Table I.

Members should designate quality-control specialists for the purpose of developing, implementing and maintaining their quality-control programmes and should maintain records of the results of quality control. Such records can be used to evaluate the effectiveness of the quality-control programme, to introduce new procedures when necessary, and to meet the requirements of the monitoring of the performance of the WWW. Quality-control records should cover the items listed in Table 1, Column 7, for both real-time and non-real-time.

5. QUALITY CONTROL OF DATA TO BE STORED

The purpose of non-real-time quality control is to ensure the highest possible standard of basic meteorological data, as welll as data for climatological, hydrological and other purposes, before they are stored and subsequently delivered to users.

The none-real-time quality control should be applied to all basic observational data, even those having been subjected to real-time quality control in connexion with the distribution of the data over the GTS. The primary responsibility for non-real-time quality control rests with the centres which store the data (see Annex B). Minimum standards of non-real-time quality control of data to be stored by WWW centres are given in Table 1. The control should be performed on a routine basis and begin as soon as possible after the data to be stored have been received at the data centre.

Methods for detecting errors in data being stored by WMCs, RMCs and NMCs may involved:

(a) testing the identification elements of all reports (e.g. testing marine reports to ensure that they are not landlocked);

(b) testing the various parameters against upper and lower limits, in order to delete obviously erroneous data at an early stage;

(c) testing the various parameters against statistically determined limits, which are normally functions of geographical position and time of the year and, in some cases, also depend on the values of other elements;

(d) checking the physical consistency between different elements in the same observation (e.g. that high relative humidity is observed in connexion with such weather phenomena as rain, snow, sleet, fog, dense mist and drizzle);

(e) testing the horizontal consistency, by comparison with simultaneous observations at adjacent stations or analysed fields;

(f) testing the vertical consistency (e.g. hydrostatic control) of certain upper-air and also of oceanographic data; and

(g) testing the time consistency using both preceding and subsequent observations from the same stations.

When detected by quality-control procedures, meaningless data should be eliminated and erroneous data corrected and marked, as appropriate.

It is desirable for all WWW centres to use the methods indicated above when performing non-real-time quality control. However, it is recognized that all centres may not be able to carry out the complete programme. Therefore, when data are exchanged internationally, information regarding the quality-control procedures used should accompany the exchanged data.

Table I

GDPS MINIMUM STANDARDS FOR QUALITY CONTROL OF INCOMING DATA

(RECEIVED VIA THE GTS OR OTHER MEANS)

(1)	STATION LIST (2)	TYPES OF REPORT (3)	TIMES OF OBSERVATIONS* (4)	PARAMETERS TO BE QUALITY CONTROLLED (5)	PROCEDURES FOR QUALITY CONTROL (6)	RECORDS TO BE MAINTAINED (7)	MINIMUM FREQUENCY FOR PERFORMING QUALITY CONTROL (8)
REAL-TIME	WMCs GLOBAL EXCHANGE LIST IN WMO No. 386 GUIDE ON THE GTS RMCs & NMCs REGIONAL BASIC SYNOPTIC NETWORK GIVEN IN WMO No. 217	SYNOP	00, 06, 12, 18	FM 12; All mandatory groups	Checking: -Detection of missing data at centres -Adherence to prescribed coding formats -Internal consistency -Time consistency -Space consistency -Physical and climatological limits Remedial Action -Before further processing, correct or flag erroneous or suspect data Notification: -Discrepancies and missing data should be made known to the appropriate centre or station Note: It is recognized that notification of not all errors or doubtful data can be done in real time by processing centre	-Information to identify source of data such as station, aircraft, ship -Type of deficiency (non-receipt, incomplete or incorrect reports, etc.) -Identification of deficient element (whole report, specific group, specific parameter, etc.) -Frequency of occurrence of data deficiencies (according to station type and element)	Preferably with each operational cycle; otherwise, with sufficient frequency to establish representative records.
		SHIP	00, 06, 12, 18	FM 13; All mandatory groups			
		PILOT PART A & B	00, 06, 12, 18	FM 32; Secs. 1, 2, 3, 4			
		PILOT SHIP PART A & B	00, 06, 12, 18	FM 33; Secs. 1, 2, 3, 4			
		TEMP PART A & B	00, 12	FM 35; Secs. 1, 2, 3, 4, 5, 6			
		TEMP SHIP PART A & B	00, 12	FM 36; Secs. 1, 2, 3, 4, 5, 6			
		SATEM SATOB	Asynoptic	FM 86; mean temperatures FM 88; cloud-motion winds			
		AIRCRAFT METEOROLOGICAL OBSERVATIONS	Asynoptic	-time and position -wind -temperature -flight level			
		CLIMAT**	Monthly	FM 71; Sec. 1			
		CLIMAT SHIP**	Monthly	FM 72; Sec. 1			
		CLIMAT TEMP**	Monthly	FM 75			
		CLIMAT TEMP SHIP**	Monthly	FM 76			
NON-REAL-TIME	WMCs GLOBAL EXCHANGE LIST IN WMO No. 386 GUIDE ON THE GTS RMCs & NMCs REGIONAL BASIC SYNOPTIC NETWORK GIVEN IN WMO No.217 AND FOR CLIMATE STATIONS, AS DETERMINED BY REGIONAL ASSOCIATIONS	SAME AS ABOVE PLUS:	Same as above plus:	Same as above plus:	Checking: Same as in real time and in addition: -Review of recorded data in comparison with observations taken before and after -Inter-comparison of parameters and calculations -Check of supplemental data -Check of extreme values Remedial Action: -Correct errors and flag data as appropriate Notification: Refer discrepancies to observing stations or WWW centre as follows: -once per month from NMCs -once every three months from RMCs -once every six months from WMCs	Summarize records developed in real time to include: Same as above with all data deficiencies found in real-time combined with additional ones found in non-real-time	With sufficient frequency to establish representative records
		PILOT PART C & D	00, 06, 12, 18	FM 32; Secs. 1, 2, 3, 4			
		PILOT SHIP PART C & D	00, 06, 12, 18	FM 33; Secs. 1, 2, 3, 4			
		TEMP PART C & D	00, 12	FM 35; Secs. 1, 2, 3, 4, 5, 6			
		TEMP SHIP PART C & D	00, 12	FM 36; Secs. 1, 2, 3, 4, 5, 6			
		ROCOB	Asynoptic	FM 39; Secs. 1, 2			

* Use observation time nearest to main synoptic hours when observation not taken at main synoptic hours.
** Monthly on receipt and prior to initial distribution or use.

ATTACHMENT III.2

DATA TO BE STORED AT RMCs

1. Members should ensure that their RMCs provide for the storage and retrieval of basic observational data received through the GTS and/or other means for the zones of responsibility as indicated below:

RMC	Zone of responsibility
Region I	
Antananarivo	To be decided later
Algiers	The zone of responsibility of RTH Algiers for collection of observational data
Cairo	The zone of responsibility of RTH Cairo for collection of observational data
Dakar	Zone of responsibility of RTH Dakar for collection of observational data
Lagos	To be decided later
Nairobi	The zone of responsibility of RTH Nairobi for collection of observational data
Tunis/ Casablanca	To be decided later
Region II	
Beijing	The zone of responsibility of RTH Beijing for collection of observational data
Jeddah	The zone of responsibility of RTH Jeddah for collection of observational data
Khabarovsk	The zone of responsibility of RTH Khabarovsk for collection of observational data
New Delhi	The zone of responsibility of RTH New Delhi for collection of observational data
Novosibirsk	The zone of responsibility of RTH Novosibirsk for collection of observational data
Tashkent	The zones of responsibility of RTHs Tashkent and Tehran for collection of observational data
Tokyo	The zones of responsibility of RTHs Tokyo, and Bangkok for collection of observational data

NOTE: Certain RMCs are not proposed for any specific zone of responsibility because of the need to avoid unwarranted duplication and to achieve the best compatibility of RMC zones with RTH zones, considering the capabilities and arrangements of the GTS.

Region III

Brasilia	The zones of responsibility of RTHs Brasilia and Maracay for collection of observational data
Buenos Aires	The zone of responsibility of RTH Buenos Aires for collection of observational data

Region IV

Washington	All Region IV (acting for RMCs Miami and Montreal)

Region V

Melbourne	The zones of responsibility of WMC/RTH Melbourne for collection of observational data (acting for RMC Darwin)
Wellington	The zone of responsibility of RTH Wellington for collection of observational data

Region VI

Bracknell	The zone of responsibility of RTH Bracknell for collection of observational data
Moscow	The zones of responsibility of WMC/RTH Moscow and RTHs Prague and Sofia for collection of observational data
Norrköping	The zone of responsibility of RTH Norrköping for collection of observational data
Offenbach	The zones of responsibility of the RTHs Offenbach, Vienna and Paris for the collection of observational data
Rome	The zone of responsibility of RTH Rome for collection of observational data

2. The types and frequency of basic meteorological data to be stored by the RMCs are as indicated below:

Type	Frequency
SYNOP	3 hourly
SHIP	6 hourly
PILOT/TEMP	6 or 12 hourly
PILOT SHIP/TEMP SHIP	6 or 12 hourly
Selected aircraft reports	
Selected satellite data	
DRIBU	

3. Members should ensure that their RMCs archive the following analyses for their zones of responsibility:

(a) Surface analyses twice per day;

(b) Upper-air analyses for at least four of the standard isobaric surfaces listed in paragraph 3.2.1 of Part *II* of this Manual.
