I. WORKSHOP REPORT

1. Opening of the meeting

The Commission for Basic Systems at its ninth session (Geneva, 1988) emphasized the importance of data quality control and monitoring of data quality especially in connection with development of the concept of data management. As follow-up to this, WMO/ECMWF Workshop on Data Quality Control Procedures was held at the European Centre for Medium-Range Weather Forecasts (Shinfield Park, Reading, U.K.) on 6-10 March 1989. The workshop was opened at 10.00 on 6 March by the Director of ECMWF, Dr. L. Bengtsson. Dr. E. Sarukhanian, Chief of Observing Systems Division of the WMO Secretariat, addressed the workshop on behalf of the Secretary-General of WMO.

The workshop was attended by 26 participants including experts from all WMO Regions nominated in consultation with Presidents of Regional Associations, and highly-qualified experts from advanced GDPS centres. Dr. D. Söderman (ECMWF) was the Scientific Director of the workshop. A list of participants is given in Annex A.

2. Objectives

The main objectives of the workshop were:

- to exchange information on status of real-time data quality control procedures used in WMO Regions;
- to provide participants with information and guidance on evolving methods of real-time data quality control at advanced GDPS centres;
- to discuss the current status of data quality monitoring and the organization and procedural aspects of a global data quality monitoring scheme;
- to produce recommendations on the future development of data quality control procedures and monitoring of data quality for forthcoming sessions of the CBS working groups on the GOS, GDPS and Data Management.

3. Organization of the meeting

During two days the meeting had an in-depth consideration of the present status of real-time data quality control at observation sites, at NMCs, RSMCs and WMCs. Presentations were made by representatives of all WMO Regions as well as experts from several advanced GDPS centres. The programme of the workshop is given in Annex B.

Representatives from leading GDPS centres (Washington, ECMWF, Bracknell) responsible for the implementation of the monitoring of data quality control, submitted information on the current status and procedural aspects of the monitoring quality of aircraft, satellite, radiosonde and surface marine data. Then the workshop was divided into informal working groups with group 1 (chairman Mr. R.J. Sowden) discussing real-time quality control and group 2 (chairman Dr. P. Julian) discussing the data quality monitoring and related procedures, organizational aspects and coordination.

4. Real-time quality control

4.1 Background

As global numerical weather prediction models become increasingly sophisticated and are operating at horizontal resolutions which will be below 100 km in the near future, they become more sensitive to variations in the quality of the initial data. Attention at many GDPS centres is now being focused on the improvement of quality control procedures and it is therefore important to examine the procedures currently in use and to make recommendations for the future.

The WMO guidelines for quality control procedures to be carried out by the data transport and the data management subsystems are currently not clearly separated in the documentation. National collecting centres have an obligation to perform quality control on data before it is disseminated on the GTS but the guidance on how to do this is insufficient.

Many developing countries have severe problems in meeting quality control requirements. One possible solution is to install micro-processors at observing sites or collecting centres to check on data formats and consistency, both internal and temporal.

The quality of the data assimilation systems and the related background fields (usually 6-hour forecasts) operated by the major GDPS centres, warrant the use of such systems for the quality control of the global observing system, in both real-time and delayed mode.

The quality control performed by GDPS centres need not be strictly standardized since it is a local operation and, at present, does not influence other centres. However, there is an increasing interest in the GDPS community to have an exchange of information on the implemented quality control procedures which are often based on the confidence placed on the observations or observing systems such as satellite soundings.

The feedback of data monitoring results to the data producers is considered to be of prime importance, not only in delayed mode but possibly also in near-real-time. If a GDPS centre, through its real-time quality control procedures, recognises an error in one or more data items, then this information could immediately be communicated back to the data producer so that it could be corrected. The BUFR representation of data, either in binary or tabular form, is considered to be the ideal vehicle for this type of message, although there could be complications with increased data volumes on the GTS.

4.2 Implications of recent code developments

The meeting noted that with the advent of a general code concept such as BUFR, which was approved by CBS for operational use as of November 1988 and the tabular equivalent (to be approved) which maps onto BUFR but presents the information to a human being rather than a computer, the meteorological community will remove the constraints imposed by the old codes for the exchange of observations and any associated information.

Modern real-time quality control procedures create to an increasing extent valuable information which is data associated and required by the user. BUFR and its tabular equivalent will allow the dissemination of the observations

together with any relevant quality control information. Together with basic information of the location of the station, the date and time, the type of instrument and observational procedures, such data can be disseminated regularly. This is particularly important for stations frequently changing the equipment or when new systems are brought into operation. Corrections to observations may be applied at the observing site or the collecting centre. These corrections should be specified in the message. Replacing original data without reference should be discouraged. BUFR provides the means of exchanging corrections for elements without having the need to repeat complete reports.

Thus in future it will be feasible and desirable to exchange in real-time:

(i) specification of instrument types and observing techniques;
(ii) information on corrections applied;
(iii) information on the level of quality control applied;
(iv) error estimates of the observations;
(v) confidence indicators for suspect observations;
(vi) feedback from data users.

This development allows and requires that new steps for quality control procedures and the related information exchange are considered and established.

4.3 Quality control at the observing site

Ground based observations are frequently still produced at sites where no automatic quality control can be employed and where the level of staff training is insufficient. This is in particular applicable to SYNOP and PILOT data while most upper-air ascents (TEMP) are produced automatically with stringent quality control in accordance with the CIMO Guide. It was felt that the "minimum" standards for quality control at manually operated sites are ill defined at present and should be more precise. The minimum quality control should be confined to a few but important parameters, such as surface pressure, temperature and wind. The WMO Secretariat should provide the means and guidance to perform such control e.g. on time series diagram paper to enable a "minimum" core checking procedure. It was also felt that such time series checks are only of limited use in the case of infrequent observations (e.g. PILOT) requiring spatial checks which cannot be performed at the observing site.

There are various levels of automation at the observing sites, such as fully automated systems, semi-automation requiring additional manual intervention and manual reading of instruments with automated data entry and processing facilities for quality control. There is a requirement to define (ref. Chapter 6, Guide on the GDPS) and provide modules of software to be used in micro-processor or PCs for the automatic data checking at the local site before the data is passed on to the collecting centre. The design, implementation and distribution of such modules should be arranged by the WMO Secretariat.

4.4 Quality control at National Meteorological Centres

It was noted that it is the practice at many collecting centres (NMCs) that the observations are injected onto the GTS without any quality control in order to achieve the fastest possible data dissemination. Quality control, however, is performed within a short period after the observation time allowing the NMC to refer any queries to the observing sites and to issue corrections with confidence. Such practices violate the current mandatory requirement laid down by the WMO (Manual on the GDPS, Vol. I, Global Aspects Part II). However, the meeting felt that an adjustment of the regulations to conform with the common practice should be considered.

At the national level (NMC) all data types should undergo the quality control procedure laid down in the manuals and guides of the GOS and GDPS. It was noted that this can only be performed to the extent possible at manual or semi-automated NMCs. However, little is known about the quality control check which are carried out at any NMC, manual or automated. The WMO Secretariat is encouraged to undertake a survey to establish which quality control is employed before injecting the data on the GTS.

According to the Manual on the GOS, part VI, para. 1.2, quality control shall be applied to all observational data obtained from either the surface-based or the space-based subsystems. However, in practice NMCs do not perform these tasks for satellite data as they are assumed to be the responsibility of the satellite operators. This was considered to be insufficient based on previous experience. Basic tests should be performed on the final meteorological products (soundings, wind).

4.5 Real-time feedback to data producer

NMCs may well be considered to be the first level of data users to provide feedback to the producer by means of their quality control. GDPS centres are the next level of users. Based on their quality control findings the requirement to send a message to the data originators will, at times, arise. However, it was felt that such queries (possibly also requests for missing data) should be sent according to agreed schedules so that operators interested in feedback and monitoring know when to expect results.

The meeting agreed that a pilot study possibly for TEMP data should be considered to gain experience and also establish which information should actually be fed back to the data producer.

4.6 Review of WMO Manuals and Guides

There is a need to review the contents of the WMO Manuals and Guides bearing in mind that a much wider range of information may now be exchanged with the new general codes. Adding quality control and related information for each data type is a requirement which should also be brought to the attention of the WG on Data Management. Chapter 24 of the CIMO Guide also needs reviewing in the light of automatic observing systems.

4.7 CIMO matters

The meeting noted with satisfaction that the last WMO Radiosonde Intercomparison in the series between RA II/RA VI will take place in autumn 1989. Thereafter over 95% of the global radiosonde network will have been evaluated. The results will allow the removal of cross-system biases at the data producer level.

At present, CIMO is not involved in the quality control procedures related to measurements from satellites. In view of the growing importance of satellite data it appears to be desirable to obtain guidance from CIMO with respect to the instruments and methods used to obtain space borne observations.

4.8 Summary of recommendations

(a) When it becomes feasible, the new general code concepts of BUFR in binary and tabular form should be utilised to exchange together with the observations

(i) instrument and observational procedure information

(ii) information on data corrections applied.

(iii) quality control information

(b) Minimum quality control procedures for key elements such as surface pressure, temperature and wind should be defined and the WMO Secretariat should provide standard means for implementation (diagram paper).

(c) Computer based quality control at the observational site or collecting centre should be supported by standard software modules. Their design, implementation and distribution should be arranged by the WMO Secretariat.

(d) A survey should be undertaken by the WMO Secretariat to establish the level of quality control performed at NMCs before injection of data onto the GTS. Data dissemination prior to quality control application should also be allowed with possible corrections to follow.

(e) The value of providing quality control feedback in real-time to data producers should be established in a preliminary study for radiosonde data.

(f) The WMO Manuals and Guides should be reviewed in the light of the new possibilities of information exchange provided by the general code concepts. Quality control procedures should be clearly separated from the data transport and matters relating to codes and formats.

5. Data monitoring

5.1 Background

Over recent years the major GDPS centres have developed systems to monitor the quality of observations based on statistics produced by the data assimilation systems. Several studies have shown that this method provides a very powerful tool for identifying problems with individual stations or observing systems.

The CBS Extraordinary Session (Hamburg, 1985) recommended that major GDPS centres should monitor the quality of surface-based observations and compile statistics by station index number or call-sign. The statistics should then be analysed to produce lists of stations or platforms believed to be persistently producing erroneous data. These lists should be based on data received over one month and should be exchanged monthly between participating centres.

CBS-IX (Geneva, 1988) recommended that lead centres be appointed to monitor the quality of each main type of observation. They should liaise with the participating centres and co-ordinate all the results and produce every six months a consolidated list of observations believed to be of consistently low quality. The presently nominated centres are:

RSMC Bracknell for marine surface observations;

RSMC ECMWF for radiosonde and pilot observations;

WMC Washington for aircraft and satellite observations.

5.2 Role of lead centres

The major topic of discussion concerned the role of the lead centres. In order to produce a consolidated list of stations the lead centre is required to take account of the monthly lists from all participating centres. However, it was clear that the various centres use different methods of producing the lists, particularly regarding the criteria used for assessing a station to be suspect. Also some centres made efforts to exclude observations containing 'gross' errors (e.g. transmission errors) while others did not. In some cases there is the additional problem that some GDPS centres make corrections to observations (e.g. day/night radiation corrections for radiosondes) before using the data in the assimilation system, resulting in further inconsistencies. It was felt that the lead centre should be responsible for recommending the criteria to be used, after liaison with the other participating centres, and that any data correction employed by the centre should be made clear in the monthly list. In addition, a gross error limit check should be agreed; data failing this check would not be included in the calculation of statistics but the monthly list would contain an indication of the number of gross errors detected.

Currently lead centres are only responsible for producing a consolidated list of stations every six months. The working group thought that there may be circumstances which would warrant a much quicker response time informing the WMO Secretariat of, say, a sudden deterioration in the performance of a station. This flow of information should occur as soon as possible after the detection of the problem.

5.3 Formats for the exchange of monitoring information

The current format of data exchange for lists of suspect radiosondes, ships etc. is not suitable for certain forms of station error, e.g. if a radiosonde station is apparently using a station height which does not agree with that specified in the WMO master list. GDPS centres should recognize that additional problems specific to certain data types can also occur and that the relevant information should also be exchanged as appropriate, without reference to the formats recommended by CBS.

The working group then addressed the problem of defining a common format for the monitoring of aircraft and satellite data. With respect to wind data from aircraft and geostationary satellites, wind vectors of observations minus first-guess field averaged over latitude/longitude boxes was thought to be most appropriate. The size of the boxes and the specific levels to be monitored would be determined by the appropriate lead centre. The problem of monitoring data provided by the polar-orbiting satellites was discussed at length. Over the next few years most major GDPS centres will be considering the option of using satellite radiances directly in their data assimilation schemes in some way. However, until these plans are more concrete it was considered to be inappropriate to attempt to define a monitoring method for radiances. In the mean time the conventional sounding data produced by NESDIS should be monitored using tools such as mean observed minus first-guess values of thickness averaged over latitude/longitude boxes for standard layers. Co-location statistics between the satellite soundings and radiosondes in the form of vertical profiles were also considered to be useful particularly if the set of radiosondes were limited, perhaps to one specific instrument type in order to overcome the problem of radiation correction incompatibility.

5.4 Distribution of monitoring information

While many data producers recognise the benefits of quality monitoring, there are some who remain to be convinced. Therefore, in order to retain credibility it was felt that, when compiling the consolidated lists of suspect stations,

lead centres should ensure that a station would only be included if a high degree of confidence could be attached to that decision.

With the consolidated lists as much evidence as possible should be passed on to the WMO Secretariat and, if possible, the exact problem should be defined so that the data producer is able to take the correct remedial action. The evidence could take the form of graphical output for example, and should attempt to compare the performance with neighbouring stations where applicable.

In order to avoid misinterpretation it was felt that the provisional monthly lists should only be circulated to GDPS centres as defined in para. 20 of the Plan for Monitoring the Operation of the WWW, Attachment II-14 of the Manual of the GDPS.

5.5 Summary of recommendations

(a) Concerning the role of the appointed lead centres,

(i) when compiling the consolidated lists of suspect stations, they should be strict in choosing only those stations where they are confident that the observations are of consistently low quality; where possible, clear evidence should be passed to WMO defining the problem;

(ii) information on problems with observing subsystems should also be passed to WMO;

(iii) recognising the fact that deteriorations in observation quality can be detected on time-scales much shorter than 6 months (the interval recommended by CBS for producing consolidated monitoring information), they should determine the appropriate response time for communicating suspect stations (or observing systems) to WMO and other GDPS centres;

(iv) they should define common methods and criteria to be used for compiling monthly statistics, preferably after liaison with the other participating centres, including the definition of a 'gross error check'.

(b) Concerning the formats for the monthly exchange of monitoring results,

(i) monthly lists of suspect stations and data platforms should contain an indication of the number of 'gross' errors detected (see rec. (a) (iv));

(ii) monitoring information for wind data from aircraft and geostationary satellites should be exchanged in the form of mean wind vectors of observed minus first-guess values averaged over latitude/longitude boxes for designated levels;

(iii) monitoring information for satellite sounding data should be exchanged in two forms:

- mean observed minus first-guess values of thickness averaged over latitude/longitude boxes for designated standard layers, and

- co-location statistics with radiosondes displayed as vertical profiles.

(iv) while standard formats have been agreed for the exchange of monitoring information, they should recognize that additional specific problems can occur with certain data types (such as incorrect practice of reporting pressure information in SYNOPs) and these should also be communicated to the WMO Secretariat.

(c) Recognizing the fact that the monthly lists of suspect stations could be misinterpreted if the methods of compilation are not completely understood, they should be circulated only to those centres familiar with the practices of data quality monitoring. In addition, they should contain a clear explanation of the criteria used and the limitations of the system.