QUALITY CONTROL TOOLS FOR METEOROLOGICAL DATA IN THE METEOSWISS DATA WAREHOUSE SYSTEM

Marc Musa, Estelle Grüter, Markus Abbt ,Christian Häberli, Esther Häller, Ursula Küng, Thomas Konzelmann, Rudolf Dössegger

MeteoSwiss, Zurich, Switzerland

ABSTRACT

There is an ongoing trend to further automate meteorological and climatological observations and to increase their temporal and spatial resolution. The increasing amount of data is a challenge for data processing, management and quality assurance. Within the MeteoSwiss Date Warehouse System a new quality control system was implemented. The different modules of this system allow to checking and treating data at a high temporal resolution (usually 10 min).

1 INTRODUCTION

The ongoing automation of meteorological and climatological measurements and observations is a challenge both for data processing and data management. The aim is usually an increase in temporal and to a lesser extent spatial resolution. This results in a much larger amount of data as well as new needs and possibilities for quality assurance. For the use in climatology the homogeneity of the data set has to be achieved.

Currently, MeteoSwiss operates four networks consisting of 117 automatic surface stations, 25 conventional climate stations (with three observation terms per day) and some 350 rain gauge stations. With the modernisation of some of the networks in the project SwissMetNet, the number of automatic stations will increase to 136.

To cope with these new challenges and needs, MeteoSwiss launched a project to design and implement a corporate software and database architecture for the management and processing of climatological and meteorological data (Haeberli and Tombros, 2001). Quality control (QC) tools are a crucial point within this architecture. The request for flexibility and expansion has led to a modular architecture, which consists within the MeteoSwiss data Warehouse System of five main parts (CC-module, PuMAB, PuMIB, VERA,THOMAS)

2 DATA TRANSFORMATION PROCESS AND CONCEPTUAL ARCHITECTURE OF THE SYSTEM

The vast majority of data from surface stations is transmitted via telecommunication links to MeteoSwiss. Only a small fraction of data is taken by observers on sheets and has to be digitised. Fig. 1 depicts the transformation process to successively improve the data quality according to defined criteria and methods. This process also creates aggregated values (e.g. daily values) and adds context information regarding the transformations applied and the measuring and observing conditions. Historical climate data sets are stored in the database in levels 3,4 and 5 depending on the source.

The different quality levels are the result of the various quality control modules. The tools to reach level 2 data will be developed in the project SwissMetNet. These tools are not part of this abstract and will not be discussed any further. Thus the application of the CC-module transforms level 2 data into those of level 3, while the tools of automatic and interactive processing and VERA are used between level 3 and 4. By applying the tool for homogenization level 5 is reached.

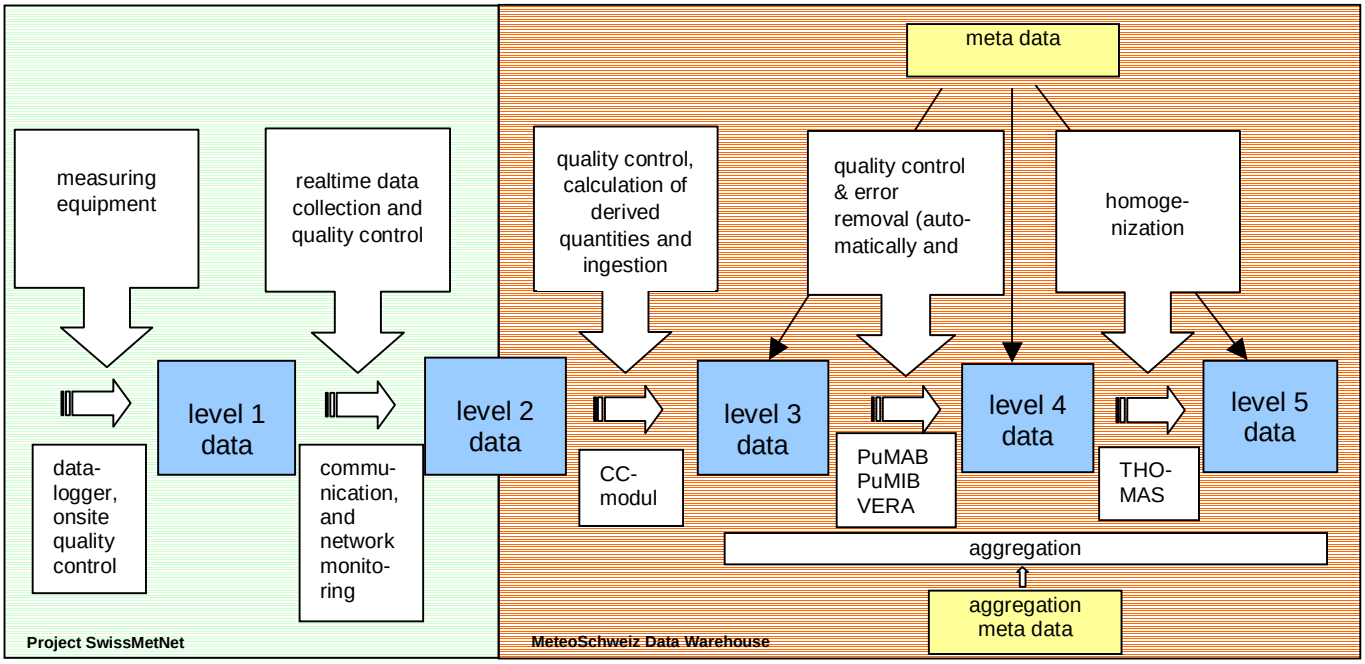

Fig. 1: Levels of data quality at MeteoSwiss

Fig. 2 shows the integration of the new quality control tools in the corporate, conceptual data warehouse architecture which is being realised at MeteoSwiss. All QC modules are implemented and are currently becoming operational.

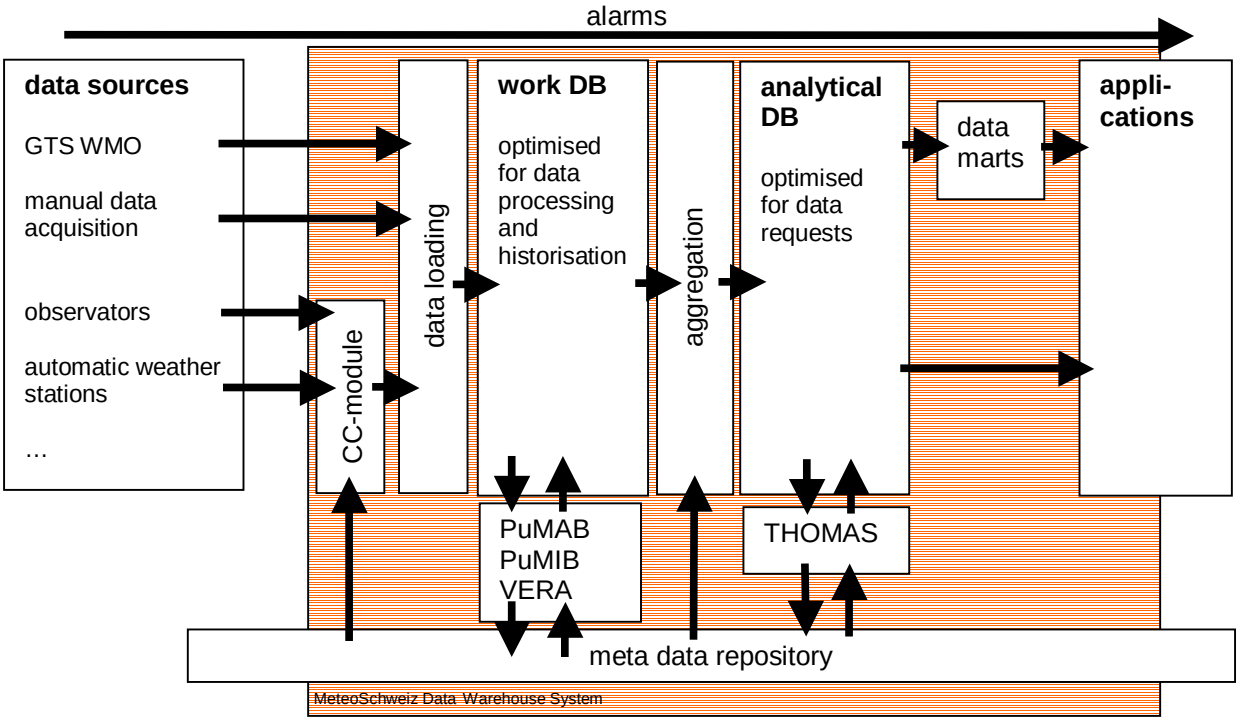

Fig. 2: Conceptual architecture of the MeteoSwiss data warehouse system. The 'analytical database' is the core 'data warehouse' of the system. The 'data marts' displayed here are not data marts in a strict sense (i.e. replicates of the 'analytical database') but rather independent, already existing databases for mainly analytical processing which are integrated into the architecture.

3 MODULES OF THE QC SYSTEM

The **CC-module** (calc&check module) checks the consistency of the data coming in from the stations for a single observation term and calculates as well derived parameters such as wet-bulb or dew point temperature in the same temporal resolution as the source data. All data are then loaded into the Work-DB. This is the first quality check of the data in the data warehouse system.

The module (**PuMAB**) verifies the values using an enlarged set of tests including tests for temporal variability. This process starts automatically once a day. PuMAB treats so-called meteoblocks which it extracts from the Work-DB. A meteoblock contains data of 33 h for each station and all parameters measured and observed at that station. In a first step, the data are tested in a similar way as in the CC-Module but within timeseries and a larger number of tests (e.g. variability tests). In a next step PuMAB automatically generate replacement values, if there are short gaps in the data.

Cases, which can not be handled automatically, are handed over to the third module (**PuMIB**) for interactive processing. The tool for interactive processing has a graphical interface which presents the data on a spreadsheet as well as time series plots. Corrections can be performed on either representation. The two parts of the interface are synchronized and altered values are updated immediately. At any stage of the work, the operator can test the corrections. As soon as all corrections are completed, the data are checked again. If there are no more tests violated, the case is considered as being solved. It will no longer appear in the list. If not then the case will immediately be given back to the editor. It is also possible to test and treat interactively any data from the work DB for which tests are defined.

The **VERA** Module (Vienna Enhanced Resolution Analysis) was developed at the Department of Meteorology and Geophysics at the University of Vienna. This system is used for spatial plausibility tests with nearby stations and was already applied in the context of the 'Mesoscale Alpine Programme' (Groehn et al., 2000), (Steinacker et al. 2000). The system gives information of the percentage of suspicious values, possible systematic errors, representativeness of a station and the bias in consequence of subscale effects (e.g. urban heat island). VERA could be successfully integrated in the MeteoSwiss Data Warehouse System.

The tool **THOMAS** (Tool for Homogenization of Monthly Data Series) was developed at MeteoSwiss. With this tool homogenization of monthly data series using standardised methods is possible to use. The tool has a graphical user interface which allows an interactive work with the data series. The homogenization tests are stored in a library which can be easily extended with new tests. Further information can be found in the final report of the project Norm90, which will be published this year (Begert et al. 2003). The tool THOMAS will be implemented in the MeteoSwiss Data Warehouse System this year.

4 OVERVIEW OF THE QUALITY CONTROL TESTS

The QC tests used at MeteoSwiss follow the recommendations of WMO (WMO, 1993). Additional tests were added to take into account the particularities of Alpine weather (e.g. Foehn, strong cold air pools). At present about 190 tests are implemented. They fall into four main categories:

1. **Limit tests**. Most variables are compared to physical ('hard'; i.e. 0 and 360 deg for wind direction) and climatological ('soft') limits. For the climatological limits, the 99.9% percentiles were determined for each month and station from 15-20 year time series of high temporal resolved data (usually 10 minute data). These tests are applied in the CC-module, PuMAB and PuMIB.

2. **Variability tests**. There are two different types of variability tests: one which tests the maximally allowed variability during a specified time interval (e.g. the difference between two subsequent values) and one which tests the minimal required variability during a certain period ('dead band'

range). For the minimal as well as for the maximal variability, physical and climatological limits are defined. This kind of tests is applied in PuMAB and PuMIB.

3. **Inter-parameter consistency tests**. Values measured at the same time and at the same place may not be inconsistent to each other (e.g. 8 octas of total cloud cover with bright sunshine). These tests are applied in the CC-module, PuMAB and PuMIB.

4. **Spatial consistency tests**. Values of the same parameter measured at the same time at nearby stations may not differ too much from each other. This kind of tests is applied in VERA.

In order to assure that all modules use the same set of tests and constants, the 'plausibility test server' (PTS) was implemented as the core of the QC system. The PTS is a set of C-libraries and is used by the CC-module, PuMAB and PuMIB.

5 REFERENCES

Begert, M., et al., 2003: Homogenisierung von Klimamessreihen und Berechnung der Normwerte 1961-1990, Schlussbericht des Projektes Norm90; Veröffentlichung der MeteoSchweiz.

Groehn, I. et al., 2001 : Qualitätskontrolle und Analyse von Beobachtungsdaten. DACH Meteorologentagung. Oesterreichische Beiträge zu Meteorologie und Geophysik ISSN Heft Nr. 27/Publ. Nr. 399

Häberli Ch., and D. Tombros, 2001: A Data Warehouse Architecture for MeteoSwiss: An Experience Report. *International Workshop on Design and Management of Data Warehouses (DMDW'01) in conjunction with the 13th Conference on Advanced Information Systems Engineering (CaiSE'01*). [available from Swiss Life, IT Research&Development, CC/ITRD, P.O. Box, 8022 Zurich, Switzerland].

Steinacker, R., Ch. Häberli, W. Pöttschacher, 2000: A transparent method for the analysis and quality evaluation of irregularly distributed and noisy observational data. *Mon. Wea. Rev.* **128**. 2303-2316.

WMO, 1993: *Guide on the Global Data-Processing System*. WMO 305.