Norwegian Meteorological Institute

DMI SMHI

NORDKLIM – Nordic co-operation within climate activities

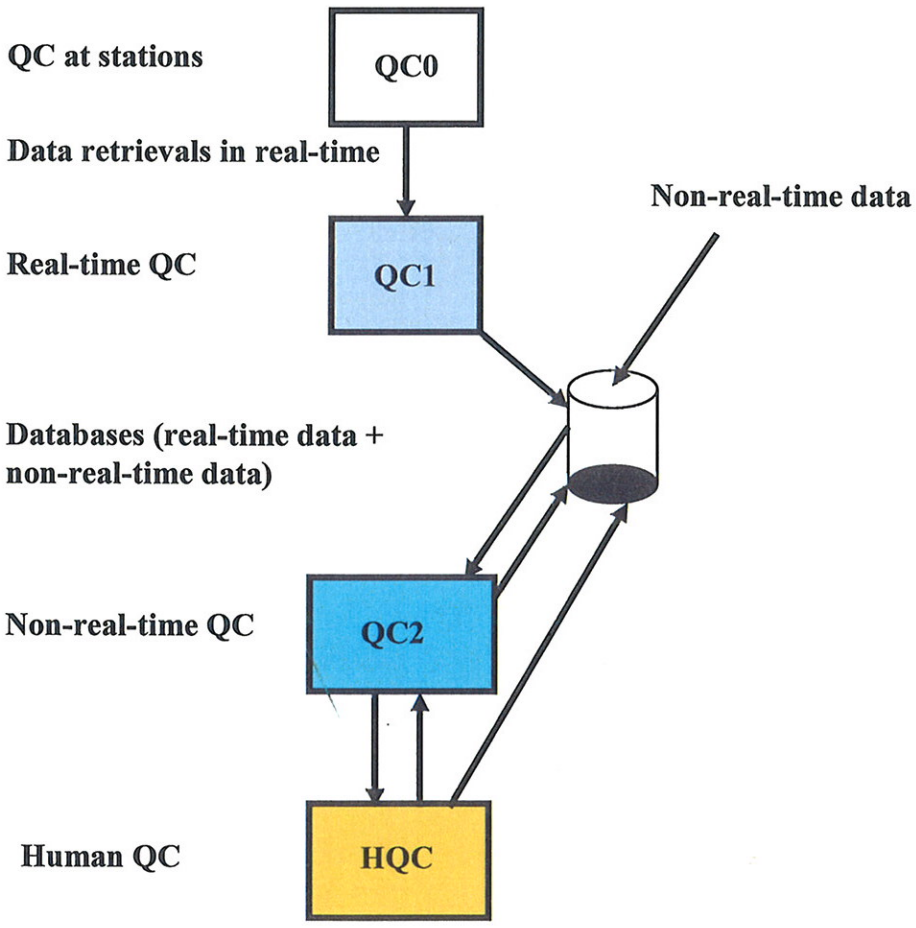
Quality Information of Meteorological Observations

Recommendations for a common Nordic end-user flagging system

Lars Andresen (ed), Ulf Fredriksson, Per-Ove Kjensli, Ola Pettersson, Pauli Rissanen, Antti Samuli, Flemming Vejen

Report 11/2003

KLIMA



NORDKLIM – Nordic co-operation within climate activities

Quality Information of Meteorological Observations

Recommendations for a common Nordic end-user flagging system

Lars Andresen (ed), Ulf Fredriksson, Per-Ove Kjensli,
Ola Pettersson, Pauli Rissanen, Antti Samuli, Flemming Vejen

November 2003

CLIMATE REPORT

ISSN 0805-9918

NORWEGIAN METEOROLOGICAL INSTITUTE
BOX 43 BLINDERN , N - 0313 OSLO, NORWAY

PHONE +47 22 96 30 00

REPORT NO.
11/2003 KLIMA

DATE
17.11.2003

TITLE

Quality Information of Meteorological Observations

Recommendations for a common Nordic end-user flagging system

AUTHORS

Lars Andresen[Ψ](ed), Ulf Fredriksson*, Per-Ove Kjensli[Ψ], Ola Pettersson*, Pauli Rissanen[§], Antti Samuli[§], Flemming Vejen[£]

[£]DMI, Denmark, *SMHI, Sweden, [§]FMI, Finland, [Ψ]met.no, Norway

PROJECT CONTRACTORS

NORDKLIM/NORDMET on behalf of the National meteorological services in Denmark (DMI), Finland (FMI), Iceland (VI), Norway (met.no) and Sweden (SMHI)

ABSTRACT

This report is prepared under Task 1 in the Nordic NORDKLIM project: Nordic Co-Operation Within Climate Activities. The NORDKLIM project is a part of the formalised collaboration between the NORDic METeorological institutes, NORDMET.

The report describes a system for coding quality information of meteorological observations to end-users. The user flags are based on control flags from the quality control process and comply with the WMO Class 33 BUFR-codes: 002, 003 and 035. A 5-digit code number secures information of: QC-level and corresponding quality controls, which are carried out, status, quality and action concerning the original data values and the most important checking method category, which the system has found and given as an argument for making changes to the data.

KEYWORDS

Quality information, Meteorological observations, NORDKLIM

SIGNATURES

Caje Jacobsson
NORDKLIM Task 1 Manager

Eirik J. Førland
Eirik J. Førland
NORDKLIM Activity Manager

Table of contents

1. Introduction

The Task 1 group in the Nordic NORDKLIM project works on issues concerning the quality control of meteorological observations. Different automatic quality control methods (QC) of meteorological observations are described in Rissanen et al. (2000) and Vejen et al. (2002). Manual quality control methods (HQC) are described in Andresen et al. (2003). In the quality control process control flags describe the results of the different checks during the process and are useful in the HQC work. Some meteorological institutes are not going to store permanently that huge amount of control flag information. Therefore simplified quality information is required, meant for end-users of meteorological data and for those who are interested in the end-quality status when data are to be used. We call such information for end-user flags or shortly user flags. Whether control flags are permanently stored or not, there is a need for user flags, which are derived from the control flags and other information following the original observations.

1.1. Who are end-users?

External customers who are ordering meteorological data are probably interested in very simplified information, like "Data quality controlled and found OK" or "Data suspect". For internal use concerning research and development of quality control systems more detailed information is needed. With a modernizing and automating of weather stations and data collecting systems, data will be available and accessible close to real-time, and may be used before the quality control process – QC1, QC2 and HQC – is finished. Then it is important also to be able to describe the quality control level that data have passed – and not passed.

1.2. Background

During the project work Finland, and later Norway, came up with a proposal for end-user flagging. Both proposals consisted of a 4-digit code number, where each digit had a special significance. The Finnish system is described in Vejen et al. (2002, Ch.12.4). In a work meeting in Oslo in August 2003, where the aim was to make a common Nordic recommendation on user-flags, the Task 1 group decided to extend and improve the Norwegian proposal. This report describes the end result of that process. Nevertheless it is possible to make almost similar quality information as the recommended one, from the Finnish flag code system.

2. End-user flags

To be able to satisfy the need for both simple and detailed quality information on end-user data, we have built a system where it is possible to combine information from different quality information groups. A minimum requirement to the system has been to comply with most of the information from the BUFR code/flag tables 0-33-002, 003 and 035 in WMO BUFR and Common Code tables (2001). Flagging of single observations and statistics is treated separately.

2.1. Flagging of single observations.

A 5-digit code number secures information on: QC-level and corresponding quality controls, which are carried out (Group 1), Status, Quality and Action concerning the original data values (Group 2-4) and the Checking method category (Group 5). The last group indicates what kind of methods has been used to assign flags to errors or suspicious values. Such

information makes it possible to estimate quite detailed statistics on the performance of the QC system and algorithm groups. Furthermore, Group 5 gives the reason why observations may have been modified. The last mentioned group should be optional when user flags are asked for. Each digit in the code number, U1-U5 is connected to the corresponding group information below in Group 1–5. The information in the groups is ordered with highest quality at the top and lowest quality or uncertain information at the bottom.

Group 1. QC level

1	**QC1 and QC2 and HQC carried out**
2	QC2 and HQC carried out (not QC1)
3	QC1 and HQC carried out (not QC2)
4	HQC carried out (not QC1, not QC2)
5	**QC1 and QC2 carried out (not HQC)**
6	QC2 carried out (not QC1, not HQC)
7	**QC1 carried out (not QC2, not HQC)**
8	**Only QC0 carried out**
9	Information of quality level not given

Group 1 gives some information about how comprehensive the controls have been. The highlighted values represent the main stages of the control procedure. The next three groups contain information related to the original value.

Group 2. Status

0	Regular observation period and observation time
1	Non-standard observation time
2	Non-standard observation period, shorter than normal
3	Non-standard observation period, longer than normal
4	Non-standard observation time, and period shorter than normal
5	Non-standard observation time, and period longer than normal
6	
7	
8	Original value missing
9	Status information not given

Group 2 gives information about deviations from standard observation procedures.

Group 3. Quality

0	Original value OK (certainly OK)
1	Original value slightly suspect (probably OK)
2	Original value highly suspect (probably erroneous)
3	Original value erroneous (certainly erroneous)
4	
5	
6	
7	
8	
9	Quality information not given

Group 3 gives information about the quality level of the original value.

The wording of Group 3 is partly similar to the wording of BUFR code 0 33 003. We talk about original value OK as code value 0 and erroneous value as code value 3, while the BUFR code talks about "data not suspect" and "data considered unfit for use", respectively. For the rest the wording is identical. The contents of the parentheses express the quality in terms of probabilities, and may be used optionally.

Group 4. Action

0	No action
1	Corrected manually
2	Interpolated manually
3	Corrected automatically
4	Interpolated automatically
5	Manually derived from accumulated value
6	Automatically derived from accumulated value
7	
8	Rejected
9	Quality information not given

Group 4 gives information about what type of action has been done to the original value. Except for the code values 0, 8 and 9, which give useful information about the original value, the other code values tell that the original value is erroneous, but nothing about the degree of gravity (given in Group 3). When corrections are done, Group 4 shows how it is done.

The following group is optional.

Group 5. Checking method category

0	Original observation checked and found OK
1	Range check (one parameter)
2	Internal consistency check (more than one parameter)
3	Step check (one parameter)
4	Step consistency check (more than one parameter)
5	Spatial consistency check, based on observed data
6	Spatial consistency check, based on time series
7	Spatial consistency check, based on model data
8	Spatial consistency check, based on statistics
9	Quality information not given

Group 5 gives information about the most important checking method category, which the system has used as an argument for doing changes to the original observation. When errors have been discovered, Group 5 shows how the errors were found (i.e. control method).

Groups 3-5 are connected: Group 3 and 5 give a diagnosis of the original value. Group 5 gives the explanation of the flags in Group 3. If an observation is corrected we would like to know how it is corrected (Group 4) and why (Group 5).

Different countries may have different ways of carrying out the HQC control. The most important HQC work is probably based on error lists, which take care of the results from the automatic QC1 and QC2 controls. When the HQC work is limited to these predefined cases, most of the observations never go through the HQC stage. Then value 5 of Group 1 and the 0-values of Group 2-5 represent the optimal situation, i.e. the best stage of checking level and quality of observation. The 5-digit code number is 50000. If all observations are "examined" by any HQC procedure, e.g. by looking at the spatial distribution of the values, the code number for the optimal situation is 10000, whether the original values are considered suspicious or not.

2.2. Flagging of statistics

It is possible to make statistics on data with a chosen quality standard (See different examples in Ch.3). E.g. if only original values are interesting for the statistics, corrected or interpolated values may be omitted, by testing on user flags. When quality standard is not chosen, statistics might include observations with different types of flag information. In both cases it is possible to give an additional Group 1 type of information like "Most of the data are checked at the QC1 and QC2 level, but not at HQC". There must, of course, be an explanation of what implications are hiding behind these abbreviations.

2.3. Flagging further details

As mentioned earlier some countries are interested in storing control flag information. Danish proposals tend to combine such information with the 5-digit user flag code, described above, in one number. Suppose the user flag code is $U_1U_2U_3U_4U_5$ and the control flag information is taken care of by a 16-digit control flag code: $c_1c_2c_3\ldots c_{16}$, this information can be combined in one decimal number:

$U_1U_2U_3U_4U_5,c_1c_2c_3\ldots c_{16}$.

By this means, detailed technical information, or other kinds of QC information needed by individual countries, can be attained in the flag without disturbing the 5-digit user flag code. If the total information is to be used, write the whole number (integer and decimals). If only the user flag is to be used, write the integer part of the number.

3. Examples

In the following, the end-user flags are tested on a lot of different situations, which may occur in real life and that may be requested. This flagging system makes it possible to store simplified QC information in a climate database. Such information is useful in many connections, and it can be utilized as a filter when an end-user wants to extract data of any given quality level that in turn corresponds to a chosen quality description. The examples can be looked at in two ways: (1) how quality information is kept as code; and (2) how a filter should be set up to extract just those observations that fulfil the quality requirements mentioned in the examples.

In the examples, parenthesis and the letter "x" are used to specify allowed flag values of the filter definition, when a certain quality description has been chosen. A parenthesis means that it is a must to check for all the specified flag values. "x" means that it is not necessary to check for the flag value, but it does not mean that all values are allowed. If a request is made

on data having "Original values OK", it is sufficient to check that U3=0. It is not necessary to check U2, U4 and U5.

Some digits initiate special digit values in other groups or exclude some digit values. If U3=0-3, then U1, U4 and U5 ≠9. If U3=0, then U4=U5=0 and probably U2=0. If U3=1-3, then U5=1-8. If U2=8, then U3=U5=9. Etc.

In some examples there might be other correct solutions depending on how the request is defined. Especially in example 9 other definitions of the quality of an accumulated value than given here, are possible.

1. Regular observation, quality controlled and found OK at the end of the control process (HQC included).

1 0 0 0 0

2. The observation is not suspect (the formulation refers to BUFR code/flag tables 0-33-002 and 003, code value 0). That means that the original observation is checked and found OK,

x x 0 x x

Or it is corrected (no accumulation),

x x (1-3) (1-4) x

3. Original observation is suspect (no correction has taken place).

x x (1-3) x x

4. Original observation is considered unfit for use. Or with other words: The observation is rejected.

x x x 8 x

5. Original observation has been changed automatically.

x x x (3, 4) x

6. Reason for correction.

x x x (1-4) (1-8)

7. Original observation is suspect, but not corrected.

x x (1-2) 0 x

8. Original observation is accumulated (observation period longer than standard).

x 3 x 9 x

9. A precipitation station that measures 24-hour precipitation, R, at 06 UTC, gives the last observation on Day(i). Next observation is coming on Day(i+2). The message from the observer is that the precipitation amount is accumulated during the last 48 hours. There are two alternatives for Day(i+1): a1) the observation is flagged as missing, no interpolation performed or a2) the observation is interpolated. a3) On Day(i+2) we will have the same alternatives for Day(i+1) until the whole missing period is examined. When the precipitation amount is distributed on the Day(i+1) and Day(i+2), the user flags will change.

Assume that QC1 and HQC are done (not QC2) on Day(i+1) and QC1, QC2 and HQC are done on Day(i+2). The accumulated value is interpreted as the original value, and here considered incorrect (too high) for the standard observation period, but considered correct for the accumulating period. In the first example (A) the quality information is related to the standard period (it is still an accumulation), in the second example (B) the quality information is related to the accumulating period.

A.

Day(i+1):	a1) R(i+1)	3 8 9 9 9	
	a2) R(i+1)	3 8 9 4 9	interpolation (QC1)
Day(i+2):	a3) R(i+1)	1 8 9 5 9	distribution (HQC)
	R(i+2)	5 3 3 0 (1, 5-8)	before HQC
	R(i+2)	1 3 3 5 (1, 5-8)	correction (HQC)

B1.

Day(i+1):	a1) R(i+1)	3 8 9 9 9	
	a2) R(i+1)	3 8 9 4 9	interpolation (QC1)
Day(i+2):	a3) R(i+1)	1 8 9 5 9	distribution (HQC)
	R(i+2)	5 3 0 0 0	before HQC
	R(i+2)	1 3 0 5 0	correction (HQC)

If R(i+2) is too high, also for the accumulating period, there will be some changes in the user flag code. a1) and a2) remain unchanged:

B2

Day(i+2)	a3) R(i+1)	1 8 9 (4/2) 9	interpolation (QC1/HQC)
	R(i+2)	5 3 (2-3) 0 (1, 5-8)	before HQC
	R(i+2)	1 3 (2-3) 1 (1, 5-8)	correction (HQC)

10. Observation is performed one hour late, but is else correspondent with the observation programme. After QC1, before QC2 and HQC, the observation is flagged (a1), after HQC, the observation is corrected (a2).

a1)	7 1 1 0 2
a2)	3 1 1 1 2

11. Observation is missing.

x 8 x x x

12. Status before HQC when QC1 and QC2 are passed.

5 x x x x

4. Conclusion

The above description is the Nordic recommendation for quality information of data to end-users. These guidelines will be acted on when exchanging data between the Nordic countries. In some cases there might be different interpretations of quality information, especially concerning accumulated values. I such cases it is important to give information about the chosen definition.

5. References

Pauli Rissanen (ed), Caje Jacobsson, Henning Madsen, Margareth Moe, Þóranna Pálsdóttir, Flemming Vejen (2000):
Nordic Methods for Quality Control of Climate Data. Nordklim, Nordic co-operation within climate activities,
DNMI Report 10/2000 KLIMA.

Flemming Vejen (ed), Caje Jacobsson, Ulf Fredriksson, Margareth Moe, Lars Andresen, Eino Hellsten, Pauli Rissanen, Þóranna.Pálsdóttir, Þordur Arason (2002):
Quality Control of Meteorological Observations. Automatic Methods Used in the Nordic Countries. Nordklim, Nordic co-operation within climate activities,
met.no Report 8/2002 KLIMA.

Lars Andresen (ed), Halldór Björnsson, Ulf Fredriksson, Knut Iden, Caje Jacobsson, Þóranna Pálsdóttir, Ola Pettersson, Pauli Rissanen, Antti Samuli, Flemming Vejen (2003):
Manual Quality Control of Meteorological Observations. Recommendations for a common Nordic HQC System. Nordklim, Nordic co-operation within climate activities,
met.no Report 9/2003 KLIMA.

World Meteorological Organization, Distributed databases – WMO BUFR and Common Code Tables – http://www.wmo.ch/web/ddbs/Code-tables.html – BUFR tables relative to section 3 / BUFR table B – Code tables and flag tables associated with BUFR table B.
7 November 2001.