DNMI
Det norske meteorologiske institutt

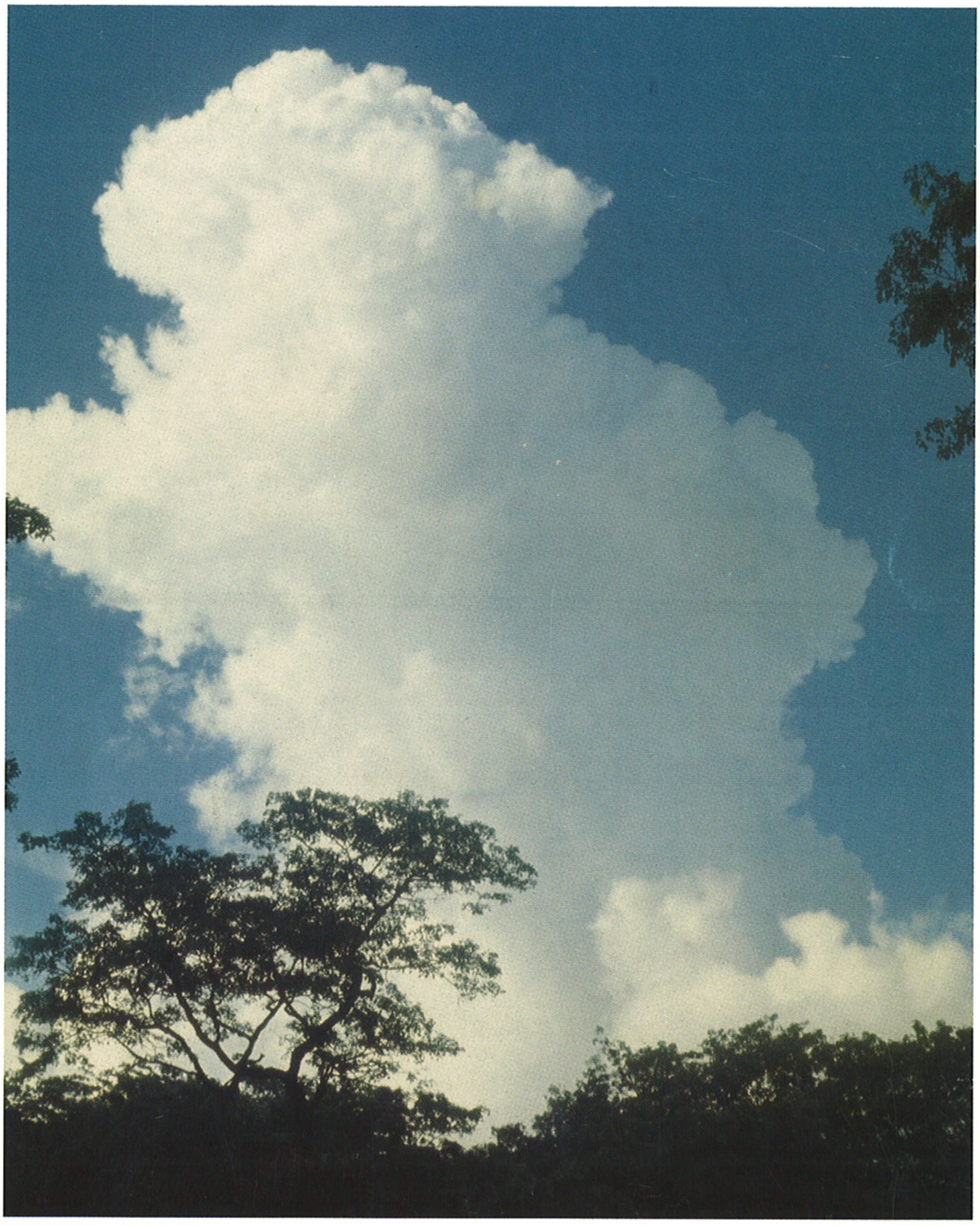
NORDKLIM: Nordic Co-Operation Within Climate Activities

Quality Control of Meteorological Real-Time Data: Results from the S-T-F system

Report no. 11/99

Petter Øgland

KLIMA



DNMI - REPORT

NORWEGIAN METEOROLOGICAL INSTITUTE
BOX 43 BLINDERN N-0313 OSLO

PHONE: +47 22 96 30 00

ISSN 0805-9918

REPORT NO.

11/99 KLIMA

DATE

Mar 13 1999

TITLE

Quality Control of Meteorological Real-Time Data: Results from the S-T-F system

AUTHOR

Petter Øgland

PROJECT CONTRACTOR

DNMI - Climatology Division / NORDKLIM

SUMMARY

This report is a part of the NORMET project NORDKLIM «Nordic Co-Operation Within Climate Activities» and is intended as a documentation of parts of the quality control documentation for project task no. 1.2 «Quality Control».

The S-T-F («Synop Temperature Faults») system is used for automatic and manual quality control of real-time data at DNMI. The report describes the development and performance of the system, with emphasis on how the quality checks relate to WMO guidelines.

The S-T-F program is a part of the Quality Assurance System at DNMI, and statistical process control is used for monitoring the the data processing of real-time data in order to find ways of improving the system.

Current analysis indicates that improvement of interpolation routines for filling in missing values is of greatest importance for the routine, and using output from numerical forecast models being the best way, at present, of achieving this goal.

KEYWORDS

1. Quality control
2. Statistical process control
3. Meteorological real-time data
4. NORDKLIM

SIGNATURE

Petter Øgland

Research Scientist

Bjørn Aune

Head of Climatology Division

Table of Contents

1. INTRODUCTION

In the NORDMET-project NORDKLIM: *Nordic Co-Operation Within Climate Activities*, there are two main tasks. Task 1 is dealing with **Climate Data** and has four subtasks (1.1 Network design; 1.2 Quality control; 1.3 Operational precpitation correction; 1.4 Long-term datasets). At a meeting in Oslo in January this year, the NORDKLIMA Advisory Commitee decided that highest priority within Task 1 should be given to subtask *1.2 Quality Control (QC)*. All Nordic countries have an urgent need for improved systems for controlling climate data and correcting suspect values. The main aim in 1999 for Task 1.2 is according to the NORDKLIMA project plans to work out a *joint report on Nordic algorithms for QC of climate data, and suggestions for QC routines on real-time data (incl. data from Automatic Weather Stations).*

This report is intended as a documentation of the S-T-F quality control system for real-time data developed at the DNMI. The name "S-T-F" is used for historical reasons and stands for "Synop Temperature Faults" (actually "Synop Temperatur Feilliste"), even though the present system is designed for detecting other faults than just those having to do with temperature. A detailed description of the development of the system until present state has therefore been included for future reference.

We begin with a discussion of technical details concerning quality control of meteorological real-time data. The first section defines elements of statistical process control applied here and is followed by sections on quality control of real-time data, including checks, methods, flags for marking manipulated observations and how the quality control must act as a part of a greater system. This is then followed by sections of quality control models of the S-T-F type, describing how this particular system has been developed and how it performs. The main results are summerised in the last section.

2. STATISTICAL QUALITY CONTROL

In order to make the data processing and quality control of real-time data fit in with the total system of data processing routines, guidelines from the quality assurance in the observations area of the UK Met Office (Shearman, 1992) have been of great interest. Shearman's emphasis is on quality as fitness for use, a definition shared by qualiy experts outside the circles of meteorology (Juran, 1995), and appears highly relevant when dealing with real-time quality control where there is little time for perfection as meteorological and climatological products have to be produced within certain deadlines.

2.1. Quality control and statistical process control

Mitra (1993) defines quality control as *a system that is used to maintain a desired level of quality in a product or service* and states that the general area of quality control can be subdivided into *statistical process control* and *acceptance sampling plans.*

Statistical process control involves the comparison of the output of a process or a service with a standard and the taking of remedial actions in case of a discrepancy between the two. It also involves the determination of the ability of a process to produce a product that meets desired specifications or requirements.

In the case of meteorological quality control this reads as regular comparison between meteorological observations (output of a process) with the standards defined by the quality control software, such as syntax checks, internal consistency checks, impossible values, extremes, outliers etc and then performing remedial action in terms of "correcting" the observations and, if possible, solving the problem that caused the observations to be in error in the first place. The determination of the ability of the process to produce a product that meets desired specifications or requirements would mean a determination concerning the station itself, why it is not producing results that fit the requirements.

Statistical process control may be devided into two main categories, *on-line* and *off-line*. *On-line process control* means that information is gathered about the process while it is functional, and in case of a difference between the output of the product or service from a determined norm, corrective action is taken in that operational phase. It is preferable to take corrective action if necessary on a real-time basis for on-line quality problems.

Off-line quality control procedures, on the other hand, deal with measures to select and choose controllable product and process parametres in such a way that the deviation between the product or process output and the standard will be minimised when the operation of the process takes place. Much of this task should be accomplised through product and process design. In other words, the goal is to come up with a design, within the contraints of resources and environmental parameters, such that when production of the component takes place, operation is at a desirable level.

Operative meteorological quality control is then mostly on-line, while if we plan to establish a new station, whether manual or automatic, and continue to test this until it seems to be working properly for producing the expected quality of observations, then we are performing off-line quality control. It is possible too, of course, to eliminate a certain station from the operative network and perform off-line quality control until it seems to again be producing according to specifications.

Both on-line and off-line quality control is relevant for the real-time quality control in this report. Off-line control is an important issue at DNMI (Aune, 1992).

2.2. Statistical concepts and quality definitions

The basic concept in statistical process control is the *Shewhart control charts* (Mitra, page 166) which is a graphical interpretation of a stochastic process representing vital characteristics of the process we want to control.

Normally the chart is presented with its central line (average) and upper and lower control limits, which define when the process is in or out of control. For the statistical applications of the real-time data control at DNMI the main focus has so far been on how to construct the mathematical representation of the process, therefore not yet showing too much concern with process limits as the process is presently out of control in terms of registering more defects than can be acceptable for practical use. It is therefore, at present phase, difficult and partly meaningless to define limits before the cause of variation in the process is reduced to chance causes alone.

A stochastic process for real-time quality control may be defined as

$$X(t)\colon \Omega \rightarrow R \qquad (1)$$

where Ω is the sample sample space of all possible outcome defined by suitable quality control checks. The process $X(t)$ can then be designed as a function counting the number of errors or warnings in the output. If the experiment of running the quality control and counting "defects" is performed on a regular basis, day by day for example, it is then fairly simple to produce a control chart by plotting the numerical results along the time line.

Several benefits seem to be realised by using control charts. Charts may help indicating the following:

1. *When to take corrective action.* Odd behaviour for the process may be easily detectible on a chart and may indicate that something is wrong with the quality control of real-time data, so that corrective action may be taken.
2. *Type of remedial action necessary.* A pattern in the chart may give an idea of what is going wrong with the system.
3. *When to leave a process alone.* When all assignable causes of variation have been removed from the process, the process will assumably still show variation due to chance causes. As warnings in meteorological quality control may or may not indicate that something is wrong, it seems reasonable to believe that variation will remain a part of the process.
4. *Process capability.* If the control chart shows a process to be in statistical control, one can establish the capability of the process and hence its ability to meet requirements. In other words, we should be able to estimate the number of warnings and errors the quality control software of real-time data will be producing.
5. *Possible means of quality improvement.* Less variation in the process means that the process is improving, the quality control system becomes better at identifying problems and correcing them, or the quality of observations are becoming better per se.
6. *How to set product specifications.* When designing new quality control software for other types of climatological data, the statistics provided by the current process may help giving numerical specifications in terms of how the new system and new process is expected to perform.

3. QUALITY CONTROL OF REAL-TIME DATA

At present there are 218 synop weather stations to be quality controlled in the TELE datatable in the KLIBAS database system at DNMI. Quality control is perfomred automatically twice a day and manually all day. At the Climatic Division there is one person who has the special responsibility of looking after these observations, helped by another person by the end of the month as climate statistics are to be produced.

Qualiy control of the same set of observations is also performed at the Forecast Division (region VØ), where the TELE table and S-T-F quality control software is accessed directly and at the Forecast Divison (region VNN) where special software is used for updating the data table.

3.1. Check structure

When in a later section the S-T-F quality control models are described, only a small set of qualiy control checks available for real-time data have been used as only a limited amount of time have been accessible for checking and modifying data, and too long check lists could easily have become demotivating for the people responsible for the manual checks.

The checks chosen for the S-T-F system are *basic quality control checks* in the WMO terminology used by Abbot (1986). For identifying errors the S-T-F system uses *internal consistency checks*. Since within one observation the parametres are somewhat interrelated, a simple check comprises making sure that the observation is consistent within itself.

A check for *impossible values*, or *extremes* rather, have also been included in the S-T-F system. Basically the S-T-F tests are corresponding to certain quality checks for daily values listed by Abbot in appendix II to the WMO/WCP report. The WMO checks are described below.

3.1.1. Maximum Temperature TX

Range check: -40 > TX > +45 (worldwide)

TX < TN

TX > Max(TT) + 3

where Max(TT) is the maximum dry bulb temperature measured at each hour.

TX > Max(TT) + 5

where Max(TT) is the maximum dry bulb temperature measured at every three hours.

3.1.2. Minimum Temperature TN

Range check: -50 > TN > +30 (worldwide)

TN > TX

TN < Min(TT) - 3

where Min(TT) is the minimum dry bulb temperature meastured at each hour.

TN < Min(TT) - 5

where Min(TT) is the minimum dry bulb temperature meastured at every three hours.

3.1.3. Rainfall Total RR

Range check: 0 > RR > 250 mm

3.2. Quality flags

At DNMI quality flagging has been considered an important issue during the development of the KLIBAS database system (Moe, M. et al, 1991; Kjensli, Lystand and Øgland, 1993; Kjensli, 1995). However, different kinds of quality flagging have been used for the different data processing routines. In the case of the quality flags in the TELE data processing routine significance and range of flags have grown out of necessity as the routine has been developing.

The present TELE table contains 13 columns used for marking quality of special meteorological parametres within the observation. All columns are defined in the Oracle data table of type RAW(1) and are given names as follows: FLTT for air temperature, FLTN for minimum air temperature, FLTX for maximum air temperature, FLRR for precipitation, FLSS for snow depth, FLFX for max wind, FLFG for wind gust, FLN for cloud cover, FLP for air pressure at sea level, FLP0 for air pressure at station level, FLPP for air pressure tendency, FLA for air pressure characteristic and FLOBS for flagging the complete observation. Each flag column is at present allowed to take on seven different values.

'0' Meaning that the element is unmanipulated. The column containing no value at all ('NULL') is used signifying the same state.

'1' Meaning that the observation has been corrected manually.

'2' Meaning that an estimate has been manually inserted where the original value was missing.

'3' Meaning that the element is similar to '2', but in this case we don't want the VNN routine to manipulate it as they would do be default if it had been of value '2'.

'4' Meaning that the element was orignally missing, but has then been automatically interpolated.

'5' Meaning that the element has been automatically corrected.

'6' Meaning that the element has been inserted from the METAR data table.

3.3. Data manipulation

Manipulation or altertation of values in the TELE data table is done both manually and automatically by a great many programs, see the section on *quality control in system* below.

How many manipulations that should be performed by the program responsible for identifying errors is something to be discussed. Due to experience from the TELE data processing routine so far, two types of updates have been necessary to implement.

1. Whenever automatically inserted values to TELE do not fit with the quality checks, the values are updated. Each automatically inserted value is flagged '4' by the interpolation programs, and when the value is removed by S-T-F, the flag column is updated with the value '5'.
2. As relative humidity on automatic weather stations mostly, but also on traditional weather stations, have a tendency of showing values greater than 100 percent, an automatic update to 100 is performed by the program. As there is no FLUU column, not flag is used for indicating in this case.

3.4. Quality control in system

Quality control software is a vital ingredience for making the data processing routine efficient. Equally important, perhaps, as to the internal list of checks in the quality control programs is the interface and interaction between different parts of the total quality system. The figure 1 below shows some of the main components of the present system.

Observations in the TELE data table are manipulated both manually and automatically. Manual updates are done through SQL*Forms (Øgland, Vidal and Aasen, 1994; Øgland, 1996b) whenever done by Klima or VØ. This gives direct access to the observations with relatively great degree of freedom for the person in charge of quality control.

A more restricted form of data manipulation is performed via VNN, where lists of missing values and lists of replacement values and sent back and forth via the SMS data exchange system (Øgland, 1998d).

The automatic data processing is run by the SYNO_KONTR program (Øgland, 1999a), which is responsible for running the interpolation programs INTERPOL_P0 (Øgland, 1999c), INTERPOL2, INTERPOL3 (Øgland, 1997) and HIRLAM (Øgland, 1999b) in correct order, and also to execute the S-T-F program which will remove interpolated values that are in conflict with the internal quality checks.

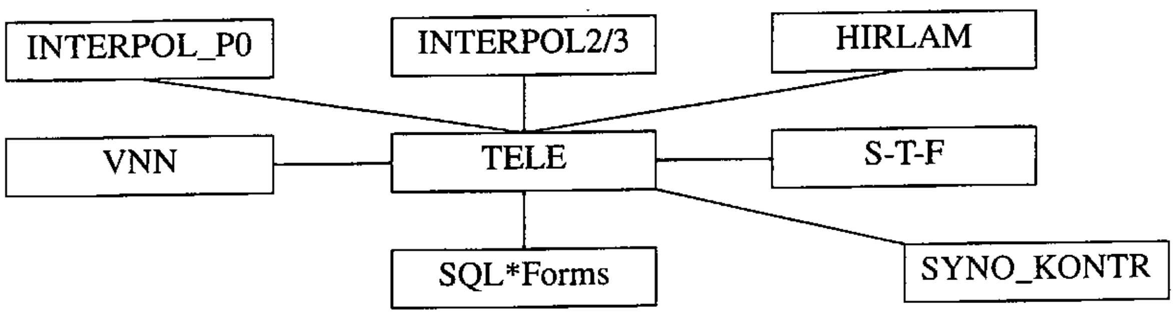


Figure 1. Rough description of the data manipulation routines in the TELE data processing

4. QUALITY CONTROL SYSTEMS: "CONTROL" AND "MDK"

As soon as the first version of the SYNO_INN data collection system (Øgland, 1994a) was reading observations from the GTS into the TELE table in the KLIBAS database system, it became imperative to make sure that the observations were of sufficient quality to be used for climatological purposes. An experimental quality control system CONTROL was therefore developed (Øgland, 1994b).

4.1. The CONTROL quality control model

The CONTROL system was designed to be run automatically and included functions both for quality control and interpolation. After some experimenting, however, it was desided that the system had a tendency of making "type 1 errors", i.e. assuming correct values to be in error, and consequently make alterations based on the wrong assumptions, and was finally terminated as there was not sufficient capacity either to monitor the system nor to develop the ideas further.

4.2. The MDK quality control model

Progress was made, however, when a manual routine for data processing of TELE observations was established in May 1995. Previous work on automatic quality control for automatic weather stations (AWS) and synoptic weather stations, the ADK quality control system (Lystad and Øgland, 1994; Øgland, 1994c), paved way for a manual quality control system MDK which consisted of modules for presenting and printing results from the ADK quality control performed on either SYNOP or AWS data (Øgland, 1995a).

5. QUALITY CONTROL SYSTEMS: THE "S-T-F" MODELS

The MDK system was put in operative use in June 1995, but proved impractical. Instead of improving this particular system, it was then decided to find out whether it was more easy to reimplement the S-T-F ("Synop Temperature Faults") quality control system that had been used on the ND-100 computer for preliminary real-time quality control of synop observations in 1987-1995.

In August 1995 a complete description of the original S-T-F system was made (Øgland, 1995). The original S-T-F system separated between to modes of messages; errors and warnings, and it contained three types of quality controls. Firstly an internal consistency check was performed for temperature, secondly a time consitency check was performed for temperature, not allowing the absolute difference between temperature TT and TX nor the difference between TT and TN be larger than 15 centigrades, and, thirdly, limit tests for temperature TN, TX, precipitation RR and snow depth SS.

5.1. The S-T-F quality control model: version 1.0

In August 1995, the first version of S-T-F was running on the newly established TELE data table (Øgland, 1995c). The TELE data table had been created in June 1995 and at the time of the release for this first version of S-T-F no quality flags were defined. Attempts at integrating S-T-F with Quality Assurance methods were done, however, publishing quality statistics in monthly KLIBAS-reports starting June 1995 (Øgland, 1995d).

5.1.1. Quality checks

The quality control checks used in the first version of S-T-F were based directly on the specifications, including some of the errors in the specifications, due to misreading of the FORTRAN code of the original ND-100 version of S-T-F. In addition to checks for missing values, the set of tests used in this version was based on the following:

$$TT(18, \textit{previous day}) > TX(06) \Rightarrow \textit{error} \quad \text{(A.1)}$$

$$TT(06) > TX(06) \Rightarrow \textit{error} \quad \text{(A.2)}$$

$$TT(06) > TX(18) \Rightarrow \textit{error} \quad \text{(A.3)}$$

$$TT(12) > TX(18) \Rightarrow \textit{error} \quad \text{(A.4)}$$

$$TT(18) > TX(18) \Rightarrow \textit{error} \quad \text{(A.5)}$$

$$TT(18, \textit{previous day}) < TN(06) \Rightarrow \textit{error} \tag{A.6}$$

$$TT(06) < TN(06) \Rightarrow \textit{error} \tag{A.7}$$

$$TT(06) < TN(18) \Rightarrow \textit{error} \tag{A.8}$$

$$TT(12) < TN(18) \Rightarrow \textit{error} \tag{A.9}$$

$$TT(18) < TN(18) \Rightarrow \textit{error} \tag{A.10}$$

$$TX(06) < TN(06) \Rightarrow \textit{error} \tag{A.11}$$

$$TX(18) < TN(18) \Rightarrow \textit{error} \tag{A.12}$$

$$15.0 < |TX(06) - TT(06)| < 100.0 \Rightarrow \textit{warning} \tag{A.13}$$

$$15.0 < |TX(18) - TT(12)| < 100.0 \Rightarrow \textit{warning} \tag{A.14}$$

$$15.0 < |TN(06) - TT(06)| < 100.0 \Rightarrow \textit{warning} \tag{A.15}$$

$$15.0 < |TN(18) - TT(06)| < 100.0 \Rightarrow \textit{warning} \tag{A.16}$$

$$TX(t) > 30.0, t = 06, 18 \Rightarrow \textit{warning} \tag{A.17}$$

$$TX(t) < -35.0, t = 06, 18 \Rightarrow \textit{warning} \tag{A.18}$$

$$TN(t) > 25.0, t = 06, 18 \Rightarrow \textit{warning} \tag{A.19}$$

$$TN(t) < -35.0, t = 06, 18 \Rightarrow \textit{warning} \tag{A.20}$$

$$RR(t) > 30.0, t = 06, 18 \Rightarrow \textit{warning} \tag{A.21}$$

$$SS(06) > 150 \Rightarrow \textit{warning} \tag{A.22}$$

5.2. The S-T-F quality control model: version 1.4

After the initial version 1.0 there were several corrections and improvements of the program, S-T-F being redesigned in order to interface with VNN and VØ. The first of these versions to be documented on its own was the version that was made when the program was expected to work both manually and automatically in a gradual automation of the TELE data processiong routine (Øgland, 1998a).

The program was by now interacting with the quality control flags that were now a vital part of the TELE data processing routine, and the quality checks A.1-A.22 had been slightly changed due to experience with the data set, although the structure remained the same.

5.3. The S-T-F quality control model: version 2.0

After the old version 1.4 of the quality control computer program S-T-F was build into the KLIBAS automatic quality control routine, SYNO_KONTR, worked commenced on a new version 2.0 of S-T-F, built completely from scratch (Øgland, 1998b).

This new version of the program was run in parallell with the old one and included functions for comparing the results from both programs. This new version also included additional *internal consistency tests* similar to those suggested by Abbot (1986) dealing with precipitation RR and present weather WW, W1 and W2.

In the formulae below -1 is used to indicate no precipitation as $RR = 0$ is used for encoding very small amounts of rainfall.

$RR(t)<0.0$ and $5 \le W1(t) \le 8,\ t=6,18$ => *warning* (A.23)

$RR(t)<0.0$ and $5 \le W2(t) \le 8,\ t=6,18$ => *warning* (A.24)

$RR(t)<0.0$ and $20 \le WW(t) \le 27,\ t=6,18$ => *warning* (A.25)

$RR(t)<0.0$ and $50 \le WW(t) \le 99,\ t=6,18$ => *warning* (A.26)

5.4. The S-T-F quality control model: version 2.1

In the revised version 2.1 of the quality control program S-T-F functions for automatically removing interpolated values that are assigned with flagg '4' were added (Øgland, 1998c).

As realtive humidity was becoming a problem of greater concern, *impossibe value tests* were added in this version.

$UU(t)<5,\ t=6,12,18$ => *warning* (A.27)

$UU(t)>100,\ t=6,12,18$ => *warning* (A.28)

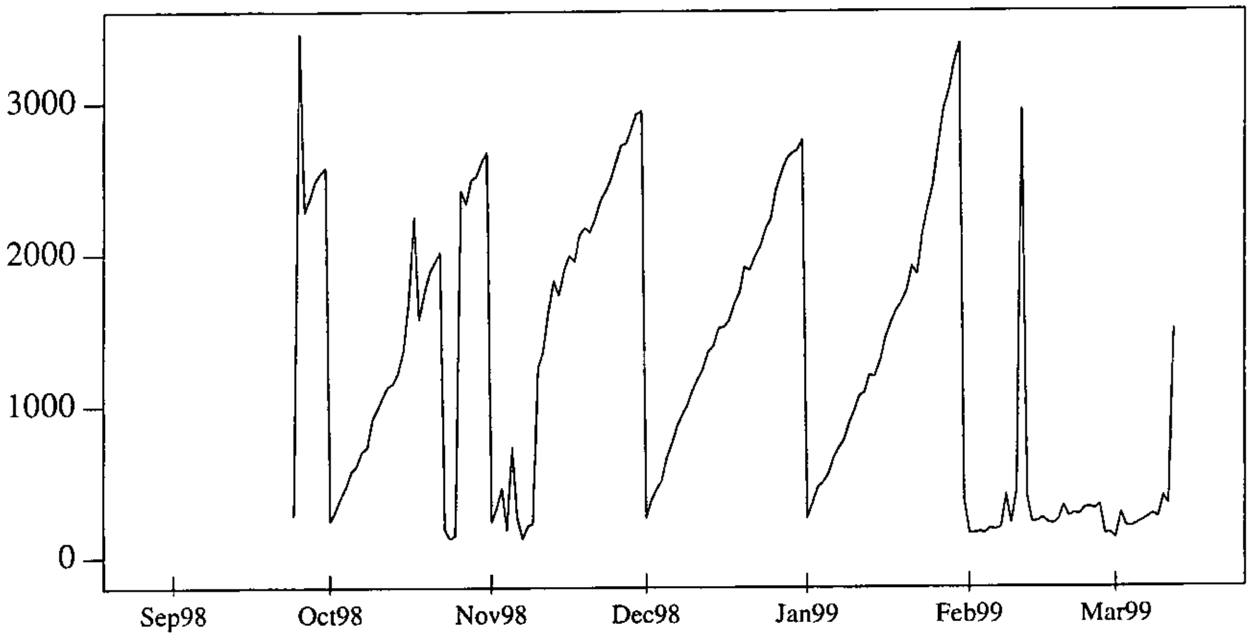
5.5. The S-T-F quality control model: version 2.2

In the 2.2 version of S-T-F, daily results from the quality control were counted and logged on files in order to be displayed as curves related to statistical process control in the monthly statistics reports. The problems related to the worst case station was also displayed among these statistics as will be illustrated below (Øgland, 1999d).

An other feature implemented in version 2.2 was an automatic action of updating relative humidity UU to 100 whenever the test A.28 notified that UU was greater than 100.

The third new feature in this version was that the flag '5' was put in use to indicate a type '4' interpolation that had been removed or altered by the program.

Below is an example of the type of *Shewhart control chart* that has been updated by the S-T-F system. The plot shows the day by day sum of problems, the number of messages concerning missing values, warnigns and errors generated by S-T-F that is. The plot is updated by each run of the program and checked daily. Each month it is published as a part of the *Interpolation and Quality Control Status Report* (Øgland, 1999f).



Day by day sum of problems recorded by S-T-F

Statistics have been recorded since September 24th 1998, using the total number of stations in TELE as a random population for most of the time. In order to get better insight on how the specific stations that were being used for the montly climate statistics were performing acording to S-T-F, a limited range of stations were used from January 30th to March 10th 1999, but as S-T-F has a responsibility of altering observations, the complete set of TELE stations was used as domain again from then on.

A prominent feature of the control curve is the systematic growth in the number of recorded problems as we reach the end of the month. As S-T-F is defined to be checking the last 31 days when running in automatical mode, it is rather surprising that we find a pattern of this kind.

The process control functions implemented in S-T-F also makes a list of the five worst stations selected from the period defined by the curve. These statistics are also presented in the montly status report. The present list consist of references to the three stations nos. 75600 LEKA, 76930 NORNE and 79530 RANA - BÅSMOEN.

1. 1999.03.12: Number of problems recorded by S-T-F for station no. 79530 = 113
2. 1998.11.30: Number of problems recorded by S-T-F for station no. 76930 = 90
3. 1999.03.08: Number of problems recorded by S-T-F for station no. 79530 = 83
4. 1998.11.26: Number of problems recorded by S-T-F for station no. 75600 = 81
5. 1998.11.25: Number of problems recorded by S-T-F for station no. 75600 = 81

In addition to presenting these general statistics, the process control functions also makes an extract from the S-T-F output-file, presenting the situation according to program for the currently worst station. This is also presented in the montly status report.

79530 RANA - BÅSMOEN is the station where the most problems have been recorded at the present, 113 problems. Output from S-T-F for this station looks like this:

```
79530 Mangler  TN(06,01)
79530 Mangler  TN(18,01)
79530 Mangler  TT(06,02)
79530 Mangler  TN(06,02)
79530 Mangler  TX(06,02)
79530 Mangler  RR(06,02)
79530 Mangler  UU(06,02)
79530 Mangler   N(06,02)
79530 Mangler  TT(12,02)
79530 Mangler  TT(18,02)
79530 Mangler  TN(18,02)
79530 Mangler  TX(18,02)
79530 Mangler  RR(18,02)
79530 Mangler  UU(18,02)
79530 Mangler   N(18,02)
79530 Mangler  TT(06,03)
79530 Mangler  TN(06,03)
79530 Mangler  TX(06,03)
79530 Mangler  RR(06,03)
79530 Mangler  UU(06,03)
79530 Mangler   N(06,03)
79530 Mangler  TT(12,03)
79530 Mangler  TT(18,03)
79530 Mangler  TN(18,03)
79530 Mangler  TX(18,03)
79530 Mangler  RR(18,03)
```

```
79530 Mangler  UU(18,03)
79530 Mangler   N(18,03)
79530 Mangler  TT(06,04)
79530 Mangler  TN(06,04)
79530 Mangler  TX(06,04)
79530 Mangler  RR(06,04)
79530 Mangler  UU(06,04)
79530 Mangler   N(06,04)
79530 Mangler  TT(12,04)
79530 Mangler  TT(18,04)
79530 Mangler  TN(18,04)
79530 Mangler  TX(18,04)
79530 Mangler  RR(18,04)
79530 Mangler  UU(18,04)
79530 Mangler   N(18,04)
79530 Mangler  TT(06,05)
79530 Mangler  TN(06,05)
79530 Mangler  TX(06,05)
79530 Mangler  RR(06,05)
79530 Mangler  UU(06,05)
79530 Mangler   N(06,05)
79530 Mangler  TT(12,05)
79530 Mangler  TT(18,05)
79530 Mangler  TN(18,05)
79530 Mangler  TX(18,05)
79530 Mangler  RR(18,05)
79530 Mangler  UU(18,05)
79530 Mangler   N(18,05)
79530 Mangler  TT(06,06)
79530 Mangler  TN(06,06)
79530 Mangler  TX(06,06)
79530 Mangler  RR(06,06)
79530 Mangler  UU(06,06)
79530 Mangler   N(06,06)
79530 Mangler  TT(12,06)
79530 Mangler  TT(18,06)
79530 Mangler  TN(18,06)
79530 Mangler  TX(18,06)
79530 Mangler  RR(18,06)
79530 Mangler  UU(18,06)
79530 Mangler   N(18,06)
79530 Mangler  TT(06,07)
79530 Mangler  TN(06,07)
79530 Mangler  TX(06,07)
79530 Mangler  RR(06,07)
79530 Mangler  UU(06,07)
79530 Mangler   N(06,07)
79530 Mangler  TT(12,07)
79530 Mangler  TT(18,07)
79530 Mangler  TN(18,07)
79530 Mangler  TX(18,07)
79530 Mangler  RR(18,07)
79530 Mangler  UU(18,07)
79530 Mangler   N(18,07)
```

```
79530 Mangler  TT(06,08)
79530 Mangler  TN(06,08)
79530 Mangler  TX(06,08)
79530 Mangler  RR(06,08)
79530 Mangler  UU(06,08)
79530 Mangler   N(06,08)
79530 Mangler  TT(12,08)
79530 Mangler  TT(18,08)
79530 Mangler  TN(18,08)
79530 Mangler  TX(18,08)
79530 Mangler  RR(18,08)
79530 Mangler  UU(18,08)
79530 Mangler   N(18,08)
79530 Mangler  TT(06,09)
79530 Mangler  TN(06,09)
79530 Mangler  TX(06,09)
79530 Mangler  RR(06,09)
79530 Mangler  UU(06,09)
79530 Mangler   N(06,09)
79530 Mangler  TT(12,09)
79530 Mangler  TT(18,09)
79530 Mangler  TN(18,09)
79530 Mangler  TX(18,09)
79530 Mangler  RR(18,09)
79530 Mangler  UU(18,09)
79530 Mangler   N(18,09)
79530 Mangler observasjon for 10.03.1999 kl 06
79530 Mangler observasjon for 10.03.1999 kl 12
79530 Mangler observasjon for 10.03.1999 kl 18
79530 Mangler observasjon for 11.03.1999 kl 06
79530 Mangler observasjon for 11.03.1999 kl 12
79530 Mangler observasjon for 11.03.1999 kl 18
79530 Mangler observasjon for 12.03.1999 kl 06
```

The case of 79530 RANA - BÅSMOEN is fairly typical for the current status of the S-T-F meteorological quality control of real-time data at DNMI, non-operative stations topping the problem statistics. The cases of NORNE and LEKA in the list above where quite similar when these stations were not transmitting data during a period of time.

6. DISCUSSION AND CONCLUSION

Methods of Statistical Process Control (SPC) have been used for monitoring the automatic quality control of real-time data. Results from the past six months show an average of about 1000 defects per day and a pattern of growing number of problems as we reach the end of the month. As the quality control program should be using observations from past 31 days, this is not a pattern that one might expect and it seems rather likely that is an error in the specification part of the automatic system which should then be investigated and corrected.

The main type of problems, when running S-T-F on the complete set of real-time data at DNMI, is missing values that have not been interpolated or ill-chosen interpolates that have been removed by the S-T-F program. Before any more elaborate advances can be made on the quality check structure in the S-T-F program, methods of filling gaps in the data series should be researched and implemented. In addition to manual interpolation, there are currently three interpolation programs running over the TELE data table (Øgland, 1997;1999b;1999c), filling in missing observations according to different methods.

At the present, interpolation with forecast data seems like a natural first approach as validation projects at DNMI (Jensen and Ødegaard, 1997; Haakenstad and Jensen, 1998) are concerned with measuring the forecast errors in order to improve the numerical models. The use of output from a numerical model to monitor the qualiy of real-time data could then be a natural next step in the development of the S-T-F system (Hall, Ashcroft and Wright, 1991).

The fact that missing data is the most frequent type of defect recorded by the S-T-F program is not, however, a fact derived from investigating the statistics, but rather an observation registered from running and trying to improve the system. It seems rather obvious, though, that the type of errors should be recorded by the S-T-F process control in order to gain better insight to how the system is functioning.

In order to record the type of errors and warnings generated on a daily basis, it is necessary to alter the format of the output so that any new type of warning or error message will be automatically recorded and made into a statistical attribute. A natural way of doing this would be to use the first six columns, say, for station indication, the next 18 columns or so for dating the error or warning and the rest of the line for the actual message.

By introducing a format like this it will become fairly simple to record the number of specific types of problems for the complete set of stations and for each station in particular. A useful statistical method for error analysis would then be frequency diagrams and frequency matrices.

It should be noticed that there is no statistical treatment of the quality flags or updates of flags or observations done by the S-T-F program. Simple flag statistics are done by the SYNO_KONTR program after the flag columns in TELE may have been updated by several computer programs. In the next version of S-T-F, however, it seems reasonable that one should at least record the number of updates that are performed by the program itself as this information would be impossible to record by SYNO_KONTR or oth er parts of the system. Statistical analysis of updates of quality flags in the TELE table may prove useful when searching for errors and explanations in the quality control system (Øgland, 1999e).

Finally, in the statistical process control, there are vital statistical parametres missing, such as process average and standard deviation. These parametres are important for describing process capacity. Process capability is a vital key prerequisite for measuring process improvement.

REFERENCES

Abbot, P.F., 1986: *Guidelines on the Quality Control of Surface Climatological Data*, WMO, WCP-85, TD-No. 111, Geneva, May 1986.

Aune, B., 1992: Note to the databasegroup 17.11.1992 (appendix to:) *Kontroll av data - spesifikasjonsrapport*, KLIBAS-report no. 01/93, DNMI, Oslo, May 1993.

Førland, E.J., B. Dahlström, R. Heino, T. Jonsson, H. Madsen, 1998: *NORDKLIM. A Nordic project on Nordic Cooperation with Climate Activities*, DNMI/SMHI/FMI/VI/DMI, Oslo/Norrköping/Helsinki/Reykjavik/Copenhagen, 08.06.1998.

Førland, E.J., 1999: *Minutes from 1st NORDKLIM Steering Committee Meeting*, Oslo, Norway, 27.-28. January 1999.

Haakenstad, H. and Jensen, M.H., 1998: *Validering av operasjonelle modeller for september, oktober og november 1998*, Research Note No. 21, DNMI, Oslo, December 1998.

Hall, C.D., Ashcroft, J. and Wright, J.D., 1991: The use of output from a numerical model to monitor the quality of marine surface observations, *Meteorological Magazine*, **120**, August 1991, 137-149.

Jensen, M.H. and Ødegaard, V., 1997: *Quality in Model Forecasts at Norwegian Synop Stations*, Research Report No. 44, DNMI, Oslo, February 1997.

Juran, J.M., 1979: *Managerial Breakthrough*, rewised edition, McGrawhill, New York.

Kjensli, P.O., Lystad, S.L. and Øgland, P., 1993: *Kontroll av data - spesifikasjonsrapport*, KLIBAS-report no. 01/93, DNMI, Oslo, May 1993.

Kjensli, P.O., 1994: *Flaggstruktur for nedbørdata*, KLIBAS-report no. 13/94, DNMI, Oslo, December 1994.

Lystad, S.L. and Øgland, P., 1994: *ADK/ADL: Datakontroll for EDAS-automatstasjoner*, KLIBAS-report no. 36/94, DNMI, Oslo, November 1994.

Mitra, A., 1993: *Fundamentals of Quality Control and Improvement*, Macmillan Publishing Company, New York, 1993.

Moe, M., Iden, K.A., Kjensli, P.O., Kristiansen, S., Lystad, S.L., Nordin, B., Vidal, Å.M. and Aasen, T., 1991: *Database/maskin prosjektet i Klimaavdelingen 1990-1991. Informasjonsmodell, flagging og kontroller. Status pr 30.06.91*, KLIMA-report no. 32/91, DNMI, Oslo, July 1991.

Shearman, R.J., 1992: Quality Assurance in the observations area of the Meteorological Office, *Meteorological Magazine*, **121**, 1992, 212-216.

Øgland, P., Vidal, Å.M. and Aasen, T, 1994: *Data inn - spesifikasjonsrapport*, KLIBAS-report no. 01/94, DNMI, Oslo, January 1994.

Øgland, P., 1994a: *Innlasting av synoptiske data til arbeidslager*, KLIBAS-report no. 05/94, DNMI, Oslo, February 1994.

Øgland, P., 1994b: *Eksperimentering med enkel kvalitetskontroll og interpolasjon av værstasjonsdata med varierende tidsoppløsning*, KLIBAS-report no. 18/94, DNMI, Oslo, June 1994.

Øgland, P., 1994c: *Implementasjon av datakontroll ADK på Typhoon*, KLIBAS-report no. 35/94, DNMI, Oslo, December 1994.

Øgland, P., 1995a: *MDK: Manuell datakontroll av SYNOP-data.*, KLIBAS-report no. 17/95, DNMI, Oslo, May 1995.

Øgland, P., 1995b: *Hurtigkontroll av SYNOP-data 1987-1995.*, KLIBAS-report no. 19/95, DNMI, Oslo, July 1995.

Øgland, P., 1995c: *Programmer S-T-F og UTI for hurtigkontroll av SYNOP-data.*, KLIBAS-report no. 20/95, DNMI, Oslo, July 1995.

Øgland, P., 1995d: *Driftsrapport juni 1995.*, KLIBAS-note no. 09/95, DNMI, Oslo, July 1995.

Øgland, P., 1996a: *Automatisering av kvalitetskontroll av geofysiske data ved Klimaavdelingen*. KLIBAS-report no. 04/96, DNMI, Oslo, February 1996.

Øgland, P., 1996b: *Beskrivelse av synop-rutine. Versjon 1.1.* KLIBAS-report no. 09/96, DNMI, Oslo, March 1996.

Øgland, P., 1997: *Automatic interpolation of SYNOP weather observations: Interpol1 and Interpol2.* KLIBAS-report no. 61/97, DNMI, Oslo, August 1997.

Øgland, P., 1998a: *Computer program S-T-F used for automatic quality control.*, KLIBAS-report no. 08/98, DNMI, Oslo, February 1998.

Øgland, P., 1998b: *Quality control computer program S-T-F v.2.0.*, KLIBAS-report no. 11/98, DNMI, Oslo, February 1998.

Øgland, P., 1998c: *Quality control computer program S-T-F v.2.1: Automatic removal of bad INTERPOL2 values.*, KLIBAS-report no. 17/98, DNMI, Oslo, May 1998.

Øgland, P., 1998d: *Updating data in TELE by VNN v.2.1*, KLIBAS-report no. 62/98, DNMI, Oslo, November 1998.

Øgland, P., 1999a: *SYNO_KONTR v.2.1: Daily interpolation and quality control*, KLIBAS-report no. 06/99, DNMI, Oslo, January 1999.

Øgland, P., 1999b: *Interpolating in TELE with Hirlam10 forecast data*, KLIBAS-report no. 08/99, DNMI, Oslo, February 1999.

Øgland, P., 1999c: *Interpolating air pressure in TELE with INTERPOL_P0 v.1.2*, KLIBAS-report no. 11/99, DNMI, Oslo, February 1999.

Øgland, P., 1999d: *Real-time quality control by S-T-F v.2.2.*, KLIBAS-report no. 18/99, DNMI, Oslo, March 1999.

Øgland, P., 1999e: Problems with the AWS at Finsevatn causing it to top the error statistics for CHECK_H_STAT in the TELE quality control routine, *KLIBAS research notes volume 3*, KLIMA-report no. 08/99, DNMI, Oslo, March 1999, 32-34.

Øgland, P., 1999f: *Interpolation and Quality Control status for February 1999*, KLIBAS-note no. 05/99, DNMI, Oslo, March 1999.