

# **Thermometer Screen Intercomparison in De Bilt (the Netherlands)**

## **Part I: Understanding the weather-dependent temperature differences**

Short title: Thermometer screen intercomparison Part I

***J.P. van der Meulen, T. Brandsma***

*Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands*

*International Journal of Climatology (accepted February 2007)*

Corresponding author: Dr Theo Brandsma  
KNMI  
PO Box 201  
3730 AE De Bilt  
The Netherlands

Telephone: +31 30 220 66 93  
Telefax: +31 30 221 04 07  
E-mail: [theo.brandsma@knmi.nl](mailto:theo.brandsma@knmi.nl)

## ABSTRACT

Temperatures of ten thermometer screens have been studied for particular weather conditions during a 6-year field experiment in De Bilt (the Netherlands). The comparison comprised two versions of an aspirated Young screen, 4 naturally ventilated round-shaped multi-plate screens (KNMI, Vaisala, Young, Socrima), a slightly aspirated version of the KNMI screen, a synthetic Stevenson screen (both aspirated and naturally ventilated) and a naturally ventilated wooden Stevenson screen. All screens were equipped with fast-responding sensors. A simple method is presented to obtain inter-sensor accuracies of about  $0.03^{\circ}\text{C}$  under field conditions. The response time of the screens is studied by making a daily comparison of the timestamps of the minimum and maximum temperatures. The analysis shows that the response of the naturally ventilated Stevenson screens is about 7–8 minutes slower than for the other screens. The screens have been compared for conditions of rainfall, wind, clear sunny days, days with snow cover and days with fog. It is demonstrated how these weather conditions affect the temperature measurements of the screens. The results show that the screens can roughly be classified into three distinct groups: (1) the round-shaped multi-plate screens, (2) the naturally ventilated Stevenson screens, and (3) the strongly aspirated screens. Due to the complexity of reducing climate and siting dependent environmental impacts on temperature measurements, it is not possible to design one particular screen as a world-wide reference. For each climate a special screen has to be developed as the best balance between the application of impacts reduction techniques and sensing the real air temperature.

KEY WORDS: temperature, thermometer screens, field experiment, the Netherlands

## 1. INTRODUCTION

The measurement of air temperature in operational meteorology and climatology is still a difficult task that demands continuous attention of the scientific community. Adequate air temperature measurements, representative for the area the temperature station represents, require (among others) minimization of local effects related to the situation of the station (e.g. trees and buildings) and minimization of environmental effects that influence the actual sensing of the temperature at the position of the sensor. For the latter, thermometer screens (also known as radiation screens) are used. Although international guidelines have been specified to obtain uniformity in the measurements (WMO, 1996), no standard thermometer screen has been defined. Consequently, many different designs of thermometer screens are in use throughout the world each with its own specific characteristics.

Internationally, temperature of the air nearby the earth's surface is defined as “the temperature indicated by a thermometer exposed to the air in a place sheltered from direct solar radiation” (WMO, 1992). Alternatively, in this paper we define air temperature as “the temperature of the air at the position of the sensor if no measurement equipment would be installed”. This latter definition stimulates a proper design of thermometer screens and may eventually result in the development of new methods of temperature measurements that use no screens at all, such as the use of sonic anemometers for indirectly measuring air temperature.

The main functions of thermometer screens are to protect the sensor (or thermometer) from direct or indirect radiation from the sun during the day and from radiation from the sensor to the sky at night (see e.g. Mawley, 1897) and to protect the sensor from wetting. Wetting of the sensor and screen is caused by fog, drizzle or rain, in combination with wind, and may result in a negative bias as the sensor starts acting like a wet-bulb thermometer. However, protection against radiation and wetting conflicts with the requirement of sufficient ventilation. Due to insufficient ventilation a microclimate develops within the screen, with its strength strongly dependent on the bulk of the screen. In an overview of recent changes in thermometer screen design, Barnett *et al.* (1998) note that a problem with the traditional wooden screen is its bulk, causing a large thermal inertia. In fact, all screens develop to some extent their own microclimate and the difference with the ambient climate depends on screen type and design. In general

it may be stated that the larger the bulk of a screen, the stronger the microclimate within the screen and the more the sensed temperature deviates from the real air temperature.

From the above, it follows that thermometer screens should have a high reflectivity to minimize heating of the plates (or louvres) and subsequent warming of the air as it flows over the plates to the sensor. To decrease the time the air is in contact with the plate, the plate radial distance should be as small as possible. To prevent the development of microclimates within screens, also the blockage of ambient airflow by the screen should be minimized by making the plate spacing as large as possible (taking into account the requirement that wetting of the sensor should be prevented). Consequently, any design of a thermometer screen is a result of compromises and consensus. Finding the optimal design is a serious challenge that may require the use of different methods for studying the behavior of screens.

Among the methods for comparing the behavior of thermometer screens are e.g.: wind tunnel experiments, numerical simulations, development of analytical equations, and field experiments. For instance, Gill (1983) compared the radiation errors above a simulated snow surface of seven thermometer screens in a wind tunnel. The results show for all screens a strong increase of radiation errors with decreasing ventilation rates, especially for rates smaller than 2 m/s. Brock *et al.* (1995) performed a wind tunnel experiment to compare two thermometer screens with respect to temperatures and vertical wind profiles within the screen, compared to ambient wind speed and temperature. Although the laboratory or wind tunnel provides controlled environments, they are not always representative for field conditions (e.g. Lin *et al.*, 2001). Richardson and Brock (1995) developed analytical equations that relate direct and indirect radiational heating errors to shield and sensor parameters. They used their analytical expressions to show the importance of applying small and reflective temperature sensors. Many field experiments have been carried out (Barnet *et al.*, 1998; Hatton, 2002; Larre *et al.*, 2002; Lefebre, 1998; Leroy *et al.*, 2002; Sparks, 1972 and 2001; Spetalen *et al.*, 2000; Zanghi, 1987). Most of these experiments concentrate on average monthly or seasonal temperature differences between screens (e.g. Chenoweth, 1992; Quayle *et al.*, 1991). Andersson and Mattisson (1991) and Perry *et al.* (2007) compare several screens for a period of about one year. They discuss the differences between the screens for several weather conditions.

In the field, the actual differences between the air temperature in the screens and the ambient air temperature depend mainly on the prevailing weather conditions. The magnitude of these weather-dependent inter-screen temperature differences for commonly used screen types has not yet been sufficiently studied. Furthermore, the state of the ground and humidity of the soil will also affect the air temperature in the screen as radiation or reflection will cause screen-dependent biases. This can easily occur in areas with snow cover and in deserts. As a consequence, when designing a screen both its behavior under different weather conditions and for different local exposures should be considered. Therefore, the Commission for Instruments and Methods of Observation (CI-MO) recommended to organize an international intercomparison of thermometer screens in different climatological regions (WMO, 2003). This work is planned to start in 2007.

It follows that for each climatological region a different design of thermometer screen will be the most appropriate and no single screen is recommended as a standard reference for global practices. Nevertheless, there have been efforts to design and recommend a standard for the measurement of temperature and humidity. An example is the WMO Reference Psychrometer (Wylie and Lalas, 1992). This reference is, however, not commonly accepted as a suitable standard for intercomparisons. Another approach is published by the International Standardization Organization (ISO, 2004), which should become the standard procedure for screen intercomparisons. The best way to select the optimal thermometer screen for a particular climate region is to intercompare sets of screens in the field for a prolonged period of time. Comparing the behavior of the screens for specific weather conditions will give a good understanding of their qualities and of the technology involved.

Here we compare a set of screens for the climate conditions in the Netherlands. We extend the approach of Anderson and Mattisson (1991) for a six-year field experiment in the period

1989–1995 at the instruments test site at KNMI in De Bilt. Although Van der Meulen (1998, 2000) presented initial results of the experiment, a thorough analysis of the data was still lacking. In the experiment 10 screens are compared. All screens operated for at least two years in parallel with a reference screen. The original reason for the experiment was to provide a basis for choosing an appropriate thermometer screen design for automatic measurements of temperature and relative humidity. In particular the possible choice for any round shaped screen to replace the traditional Stevenson screen was of interest. Recently, we reanalyzed the data in order to obtain weather-dependent corrections for the long-term daily De Bilt temperature time series.

The emphasis in this paper will be on understanding the weather-dependent temperature differences between the screens. This is mainly related to the interests of the operational meteorologist, who wants to measure air temperature as close as possible to the real air temperature. This type of analysis should always precede a statistical analysis of the data. In Part II (Brandsma and Van der Meulen, 2007) we concentrate on the description and modeling of the mean temperature differences and extremes. The interest of the climatologist is to prevent artificial breaks due to screen changes or to have the ability to correct for these breaks. In Section 2 of this paper, we first describe the setup of the experiment, the calibration and the methodology. Section 3 compares the screens for conditions of rainfall, clear-sky, fog and snow. Section 4 closes with a discussion and conclusions.

## 2. DATA AND METHODS

### 2.1. Experimental setup

The field experiment was performed in a 6-year period between 9 January 1989 and 11 February 1995 at the testing-site of KNMI in De Bilt (WMO 06260, 52.099°N, 05.176°E), situated in the center of the Netherlands, about 55 km east from the North Sea coast. Ground surface of the testing-site is at 2 m above mean sea level and the soil consists of a mixture of clay and sand, with groundwater levels 50–80 cm below ground surface in summer and < 40 cm in winter. The climate is typical for the mid-latitudes with prevailing westerly winds and precipitation for 7% of the time amounting to about 800 mm per year. Annual mean air temperatures are around 10°C, with the monthly average daily minimum temperature varying between –0.1°C (February) and 12.5°C (July) and the monthly average daily maximum temperature varying between 5.2°C (January) and 22.3°C (August). The measurements were performed above a flat open terrain with short cut grass cover and are sufficiently far removed from major obstacles like buildings, forests or lakes, which may give unpredictable impacts.

The experiment compares 10 different screens. Figure 1 shows an overview of the test-site with the screens and Figure 2 shows detailed pictures of the screens. In November/December 1991 the whole setup was moved about 50 m southeast due to a redivision of the measuring field. All screens are placed within an area of about 7 by 20 m in two parallel rows and equally spaced. The mutual distance between the screens is about 2.5 m, large enough to avoid mutual influencing. The temperature measurements in the screens are performed at a height of 1.5 m (the observational height in the Netherlands). To avoid any impact due to heating of the stands, the screens (except the Stevensons) were placed at the ends of horizontally positioned “H” or “I”-shaped supports (see Figure 1). Only the KNMI multi-plate screen operated during the whole 6-year period, the other screens operated generally for periods of at least two years. Table I presents some details of the screens and the sensors. The abbreviations for the screens given in this table are used throughout this paper. The color of the screen interior of Knmi.ref, Knmi.asp and Vaisala is black, for all other screens the interior color is white. Maintenance and the relocation of the testing-site caused data gaps of several months.

To examine mutual differences in behavior, both artificially and naturally ventilated screens were selected. Although artificial ventilation seems to be beneficial, amplified airflow may also introduce negative impacts, *e.g.* in the case with air saturated with moisture.

All of the presented screens (or comparable ones) are still in operation in different parts of the world. In general, the round-shaped multi-plate screens have become the default screen in automatic weather stations (AWS), which are mainly found in the industrialized countries. In the developing countries, manned stations are still the backbone of the national synoptic networks and Stevenson screens are widely used there. In climate observing networks, Stevenson screens are still commonly used worldwide. The recent introduction of AWS for climatological practices will result in a conversion from Stevenson screens to round-shaped multi-plate screens.

The sensors used in the test are identical 4-wired platinum (Pt) resistance thermometers, with an absolute measurement uncertainty  $< 0.05$  °C during a period of three years or more. To minimize the self-heating of the sensor below an acceptable value ( $< 0.01$  °C), a resistance value of  $500 \Omega$  (at  $0.00$  °C) was chosen (a so-called Pt500). The overall measurement uncertainty (sensor and equipment during the period of the intercomparison) of the sensors is  $0.1$  °C. The temperature scale defined for the measurement of air temperature is in line with WMO regulations (i.e. ITS-90).

Temperature is sampled and archived instantaneously every 15 seconds with a resolution of  $0.01$  °C. Unless stated otherwise, we use 10-minute mean temperature values. This averaging period was chosen to minimize the local spatial differences of the true air temperature due to small-scale turbulence (sometimes larger than  $0.5$  °C per minute). To study the effects of other meteorological elements on the temperature differences between screens, the following operationally measured elements at the KNMI-terrain (WMO 06 260) have also been used: wind speed  $u$  (at 10 m until 26 June 1993 and thereafter at 20 m), global radiation  $K\downarrow$ , cloud cover  $N$ , precipitation  $P$ , and relative humidity  $rh$ . Most elements are measured within a range of about 30 m from the test-site and can be considered representative for that site. Windspeed and direction are measured (operationally) about 200 m east of the experimental site. At the beginning of the intercomparison, the elements are available on an hourly basis (10 minutes before whole hours UTC) but from 27 September 1989 onwards wind speed, precipitation, humidity and incoming global radiation are also available on a 10-minute resolution.

There are two factors that complicate the intercomparison of the screens. First, sensors may be subject to unpredictable degradation causing drift or instable read-outs. The sensors were therefore recalibrated at intervals of about 2 years in a calibration lab at KNMI and against references with approved traceability to the international temperature standards. If, after recalibration, any sensor did not satisfy the stated uncertainty of  $< 0.05$  °C, it was replaced. Second, also the screens themselves may cause drift of temperature due to contamination or degradation. For example, dark pollution will increase the absorption of the short-wave radiation resulting in self-heating. To minimize these kinds of effects, the screens were inspected at least once a week and cleaned up if necessary.

## 2.2. Calibration of the sensors under field conditions

In fact the observed temperature differences between screens should be determined with an uncertainty which is an order of magnitude smaller than the required target uncertainty for operational air temperature measurements ( $0.1$  °C). To meet this requirement, we calibrated the sensors against the sensor of the reference screen for conditions where there are no changes in temperature, no precipitation and with a constant air flow (no cooling and no warming), usually during nighttime (see Figure 3 for an example). For those conditions we may expect the screens (and sensors) to have the same temperature. Since there was no international reference screen defined that may serve as a real scientifically understood reference, we used the naturally ventilated KNMI multi-plate screen (Knmi.ref) as a relative reference. That was the only screen that operated for the whole 6-year period and is the current operational screen in the Netherlands. For the purpose of calibration, temperature observations from Knmi.ref were selected that sat-

isfy the criterion that the root mean squared deviations in the period from one hour before to one hour after the observation, are smaller than a predefined threshold  $\delta$ :

$$\sqrt{\frac{1}{13} \sum_{j=-6}^{j=6} (T_{i+j} - \bar{T}_i)^2} < \delta \quad (1)$$

where  $T_i$  is the  $i$ th 10-minute temperature observation of Knmi.ref and the average running mean temperature  $\bar{T}_i = 1/13 \sum_{j=-6}^{j=6} T_{i+j}$ . The choice of  $\delta$  is somewhat subjective: large  $\delta$  results in a noisy signal whereas small  $\delta$  may not leave enough observations to determine the corrections. After comparing several values of  $\delta$  by visual inspection, we chose  $\delta = 0.03^\circ\text{C}$ . A total of 0.7% of the total number of observations satisfies this criterion, of which 81% are nighttime observations.

Figure 4 presents the temperature differences  $\Delta T$  (screen – Knmi.ref) for which  $T_i$  of Knmi.ref satisfies the criterion in Equation (1). The figure clearly shows biases with respect to Knmi.ref. The biases are generally up to about  $0.2^\circ\text{C}$ . The plots for Vaisala and Young (upper row middle and right panel) indicate malfunctioning of sensors for a part of the series (Dec 1990 – Oct 1991). Recalibration of the sensors at 23 October 1991 revealed a significant temperature-dependent bias. This bias corresponds to the observed bias. We decided to omit the period from the analysis. Young.aspI shows an erratic course, indicating that something might be wrong with this screen/sensor. We decided therefore to leave out this screen for further analysis. Young.aspII shows regular negative departures (colder than Knmi.ref). We decided to analyze these departures further (Section 3) but we did not take them into account in the calculation of the correction. On the basis of Figure 4, we calculated corrections for each screen and applied these to the original data. Figure 5 presents  $\Delta T$  (screen – Knmi.ref) after applying the corrections. The resulting bias is about  $0.03^\circ\text{C}$ , which is an order of magnitude smaller than the original bias of  $0.2^\circ\text{C}$ .

### 2.3. Methodology

In the remainder of the paper we will focus on understanding the weather-dependent temperature differences between the screens. The point of departure is the bias-corrected 10-minute average temperatures. In section 3.1 we study the response times of the screens. This is an important measure that may help to explain observed temperature differences between screens. We consider the response time of the complete system, including the sensor and the screen. As the response time of the sensors itself is in the order of a few seconds, the response time of the whole system will in fact mainly be determined by the interaction of the air with the sensor inside the screen. First, we compare the times  $t_n$  and  $t_x$  of the occurrence of minimum and maximum temperatures, respectively, and thereafter we will compare the standard deviation of 15-sec temperature samples in the 10-minute intervals for several weather conditions. In section 3.2 we study the behavior of the screens for special conditions like sunrise and sunset, warm and sunny days, snow days and for conditions of rainfall and fog.

### 3. RESULTS

#### 3.1. Response times of screens

For each screen we calculated the daily values of  $t_n$  and  $t_x$  in the summer half year (Apr-Sep) when the diurnal temperature cycle is most pronounced. Differences with respect to Knmi.ref were calculated satisfying the following conditions: (a)  $t_n$  occurs in the morning between 2 and 8 UTC, and (b)  $t_x$  occurs in the afternoon between 12 and 18 UTC. Figure 6 presents histograms of the differences. The differences are expressed in 10-minute intervals. For instance, a difference of 2 signifies that the screen under consideration responds on average 20 minutes slower than Knmi.ref. The figure shows that for most screens  $t_n$  and  $t_x$  are in the same 10-minute interval as for Knmi.ref. Exceptions are Stev.pvc and Stev.wood. During  $t_n$  and  $t_x$  both these screens clearly respond more slowly than Knmi.ref and the other screens.

Table II presents some statistics of the differences in  $t_n$  and  $t_x$  between the screens and Knmi.ref. Only days with absolute values of the differences  $\leq 4$  were considered in the table (to minimize the effect of multiple minimums or maximums). The table shows that the differences in the  $t_n$  and  $t_x$  are small and insignificant for most screens. Stev.pvc and Stev.wood respond about 7 and 8 minutes slower, respectively, than Knmi.ref during  $t_n$ , while during  $t_x$  these differences are about 3 minutes. Stev.pvc.asp is the only screen that responds significantly faster than Knmi.ref; for  $t_n$  this amounts to about 2 minutes while for  $t_x$  there is no significant difference. Figure 7 shows the differences in  $t_n$  and  $t_x$  between Stev.pvc and Stev.wood and the reference screen Knmi.ref as a function of windspeed. The corresponding number of days used for each of the four curves is given in Table II. For  $t_n$  the figure shows an obvious decrease of the differences with increasing wind speed for both Stev.pvc (left panel) and Stev.wood (right panel). For both lines the decrease is significant at the 0.01 level, though the explained variances are small, 4.2% and 4.9%, respectively. For  $t_x$  the decrease for both Stev.pvc and Stev.wood is small and not significant.

#### 3.2. Behavior of screens for special conditions

##### 3.2.1 Rainfall

Rainfall combined with wind and/or artificial ventilation may cause wetting of the sensors and consequently wet-bulb effects that may impede the measurement of the real air temperature. Here we study this effect by comparing temperature measurements for hours with continuous rainfall and by looking at a special period in detail.

Figure 8 presents  $\Delta T$  (screen – Knmi.ref) for rainfall conditions (all 10-min intervals with rainfall depth  $> 0$  mm). The figure shows a large negative bias for Young.aspII ( $-0.229^\circ\text{C}$ ) and only a small and negligible bias for the other screens. Table III compares the statistics of  $\Delta T$  for rainfall conditions with those for dry/cloudy conditions (cloud cover = 8 octa). The table shows that Young.aspII is on average  $0.145^\circ\text{C}$  colder for wet conditions than for dry/cloudy conditions. For all other screens these differences are negligible. The anomalous behavior of Young.aspII is also apparent from the standard deviation of  $\Delta T$ . For Young.aspII the standard deviation for wet conditions is  $0.71^\circ\text{C}$  larger than for dry/cloudy conditions. In contrast, for all other screens the standard deviation of  $\Delta T$  for wet conditions almost equals that for dry/cloudy conditions and is much smaller than the standard deviation for Young.aspII.

Further analysis of the wet 10-min intervals of Young.aspII (6625 values) shows a small negative correlation between  $\Delta T$  and windspeed ( $r = -0.46$ ). There are also negative correlations of  $\Delta T$  with rainfall amount and the difference between the air temperature and wet-bulb temperature, but these correlations are much weaker ( $r = -0.10$  and  $r = -0.15$ , respectively). Due to the large number of values, all correlation coefficients are significant at the 0.001 level.

To study the behavior of Young.aspII for wet conditions somewhat further, we selected a six-day period with rainfall in December 1993. Figure 9 presents the temperature differences simultaneous with rainfall intensity and windspeed. The figure clearly illustrates the cooling effect of rainfall on the temperatures of Young.aspII, and the amplifying effect of windspeed on the cooling is also visible. We furthermore calculated the standard deviation of the 15-sec samples for this six-day period in each 10-minute interval (40 samples each). For the selected period there is no obvious relationship between this standard deviation and  $\Delta T$  (not shown). The average standard deviation equals 0.065°C. For the other screens operating in this period, the average standard deviation ranged between 0.038°C for Socrima and 0.050°C for Stev.pvc.aspII. For Knmi.ref the average standard deviation is close to the latter and amounts to 0.049°C.

From the analysis above, it is evident that for conditions of rainfall the temperature measurements of Young.aspII are too low because the sensor has been directly exposed to raindrops.

### 3.2.2 Clear sky

It is well known that differences between (naturally ventilated) screens are most obvious on clear days with low wind speeds. Increasing cloudiness and wind speed cause inter-screen differences to dissipate. Here we discuss the differences between the screens for a selection of four representative pairs of consecutive clear days, two pairs in winter and two in summer. Furthermore, we present summary statistics of temperature differences for high radiation levels. A clear day is defined here as a day with mean hourly cloud cover  $\leq 1$  octa. The temperature differences are presented together with the standard deviation of the 15 sec samples  $sT$ ,  $T$  of Knmi.ref,  $u$ ,  $K\downarrow$ , and  $rh$ . Before calculating the standard deviation  $sT$  for a 10-min interval, the daily cycle has been removed using a LOESS smoother (Cleveland *et al.*, 1988) with a span of 1 hour and a linear fit. Because the screens did not all operate at the same time, each pair of days concerns only a limited set of screens.

#### 8–9 February 1989

Figure 10 compares the screens for 8–9 February 1989. The figure shows a distinct behavior of the two Stevenson screens Stev.pvc and Stev.wood.  $\Delta T$  of these screens shows a noisy pattern during nighttime hours (especially in the night from 8–9 February) and mainly positive values during daytime hours and the early evening. The multi-plate Vaisala and Young screens behave more or less in the same way as the multi-plate reference screen Knmi.ref. The distinction between the Stevenson screens and the multi-plate screens is also reflected in  $sT$ . The average values of  $sT$  for the multi-plate screens (only that of Knmi.ref is shown) range between 0.075 and 0.086°C for Young and Knmi.ref, respectively, and for the Stevenson screens (only that of Stev.wood is shown) between 0.036 and 0.047°C for Stev.wood and Stev.pvc, respectively. This corresponds with the results in Section 3.1.

The noisy  $\Delta T$  pattern of the Stevenson screens during the night is related to the smoothness of the air temperature curve of Knmi.ref (third panel from above). Irregularities in  $T$  are reflected in  $\Delta T$  of Stev.pvc and Stev.wood. Further inspection of  $dT/dt$  for all screens (not shown) reveals that the  $dT/dt$  of the Stevenson screens almost always crosses the zero line after the  $dT/dt$  of the multi-plate screens. In other words, (local) maximums and minimums of the multi-plate screens precede those of the Stevenson screens (see also Section 3.1).

Wind speed is an important factor in explaining the shape and magnitude of  $\Delta T$ . In Section 3.1 we already showed how the response times of the Stevenson screens depend on  $u$ . Wind speed exhibits a clear daily cycle with maximum values during the day and minimum values during the night. The low values of  $u$  during the night of 8–9 February, combined with variable wind direction, are probably responsible for the erratic behavior of  $T$  and (indirectly) of  $\Delta T$ . Furthermore, it should be realized that  $u$  is measured at a height of 20 m and that during stable nights this wind speed is uncoupled with that at screen level. Consequently,  $u$  at screen level will be much smaller than the presented values of  $u$ . During the day, the differences in heating

of the screens by the sun also become important. It is clear from Figure 10 that *Stev.pvc* and *Stev.wood* become warmer during the day than the multi-plate screens. Around  $t_x$  this amounts to about 0.4°C for this particular situation. The daytime  $u$  values of the first day are smaller than those on the second day. This is reflected in the somewhat larger  $\Delta T$  values on the first day compared to the second.

Here and in the following figures, the effect of  $rh$  is not clear. It may be that  $rh$  becomes important when it approaches 100%. In that case small droplets may form on the sensors causing wet-bulb effect, but this effect could not be detected (see also Section 3.2.3 on fog conditions).

#### 5–6 July 1989

Figure 11 compares the screens for 5–6 July 1989. Both days are warm summer days and concern the same screens as in Figure 10. Compared to Figure 10,  $u$  in Figure 11 is much larger, especially for 5 July. For the latter day this results in  $\Delta T$  values close to zero for all screens. Due to the somewhat reduced values of  $u$  on 6 July, the  $\Delta T$  patterns resemble to some extent those in Figure 10. Note that on 6 July, just after sunrise,  $\Delta T$  of *Stev.pvc* and *Stev.wood* are negative. Due to the low values of  $u$  at that time those screens cannot follow the rapid temperature rise. A counteracting effect may result from the typical flat shape of the vertical oriented panels of these types of traditionally south oriented screens. The panels will catch sunshine at maximum during sunrise and sunset, *i.e.* when the sun is just above horizon and direct radiation is coming in horizontally and caught by one of the side panels. Apparently, the latter effect is relatively small.

The daytime values of  $sT$  are much larger than the corresponding values in Figure 10. This reflects the effect of increased convection on summer days compared to winter days. On warm summer days, convective cells of varying space and time scales develop. Within the cells temperature varies in space. When these cells pass along the screens, the faster responding screens are better able to adapt to the cell temperature than the slower ones.

#### 1–2 January 1993

Figure 12 compares the screens for 1–2 January 1993. Both days are cold winter days with temperatures below 0°C. Compared to Figs. 10–11, *Stev.pvc* is replaced by *Stev.pvc.asp* and the remaining screens are added. Especially on 2 January the differences between the screens are small, probably due to the high values of  $u$ . The behavior of *Young.apsII* is more or less opposite to that of the naturally ventilated *Stev.wood*. During nighttime  $\Delta T$  of *Young.apsII* is mainly positive while during daytime negative values prevail. Note again that the multi-plate screens behave more or less the same, though *Socrima* seems to respond somewhat slower than the other multi-plate screens. This is also evident from the  $sT$  values of *Socrima* which are smaller than the *Knmi.ref* values. Note also that *Stev.pvc.asp* tends to follow *Young.apsII*, though the deviations from *Knmi.ref* are smaller. This is also reflected in the  $sT$  values of these screens, the mean  $sT$  values for this two-day period are 0.073 and 0.084°C, for *Stev.pvc.asp* and *Young.apsII*, respectively.

#### 11–12 July 1994

Figure 13 compares the screens for 11–12 July 1994. Both of these days are warm summer days with maximum temperatures around 30°C. Only *Young.apsII*, *Socrima* and *Stev.pvc.asp* were in operation, together with *Knmi.ref*. The most notable feature is the behavior of *Young.apsII*. During daytime hours *Young.apsII* is about 0.5–0.7°C cooler than *Knmi.ref* while during nighttime hours it is the opposite. As in Fig 12, *Stev.pvc.asp* tends to follow *Young.apsII* though with smaller values of  $\Delta T$ . The  $\Delta T$  values on 11 July are somewhat larger than that on 12 July, probably due to smaller  $u$  values on 11 July. *Socrima* deviates only marginally from *Knmi.ref*.

The  $sT$  values of all four screens show large differences between the screens. The average  $sT$  values are  $0.090^{\circ}\text{C}$  for Knmi.ref,  $0.171^{\circ}\text{C}$  for Young.aspII,  $0.061^{\circ}\text{C}$  for Socrima and  $0.126^{\circ}\text{C}$  for Stev.pvc.asp.

The distinct behavior of Young.aspII on these summer days is probably only partly a result of its faster response to temperature changes. During daytime hours, the artificial ventilation is probably directly responsible for the relatively low temperatures. On the other hand, during nighttime hours the artificial ventilation may also indirectly have an effect on the measured temperatures. This ventilation may then disturb the development of a stable layer around the screen. As a result, temperatures may be slightly higher than for the naturally ventilated screens. To a lesser extent, these effects are also important for Stev.pvc.asp.

#### *Temperature differences for high radiation levels*

For screen design, the behavior of the screens for clear-sky conditions with high radiation levels may especially be important. For  $K\downarrow > 750 \text{ W/m}^2$ , Table IV presents  $\Delta T$  for four categories of windspeed (corresponding to the quartiles of windspeed in the selected intervals). The table shows that  $\Delta T$  ranges between  $-0.536^{\circ}\text{C}$  for Young.aspII and  $0.174^{\circ}\text{C}$  for Stev.pvc for the lower quartile of windspeed and that the absolute  $\Delta T$  values becomes smaller with increasing windspeed. The individual values range between  $-1.22^{\circ}\text{C}$  for Young.aspII and  $0.83^{\circ}\text{C}$  for both Stev.pvc and Stev.wood.

#### *3.2.3 Fog*

Fog may have an effect on temperature measurements due to condensation of droplets on the sensors. To study this effect we used the hourly synoptic data to select all hours with fog and visibility  $< 200 \text{ m}$ . A total of 1.7% of all hours satisfies this criterion. Figure 14 presents  $\Delta T$  (screen – Knmi.ref) for these fog conditions. Although there are inter-screen differences in the variability of  $\Delta T$ , the means of  $\Delta T$  are small and range between  $-0.012^{\circ}\text{C}$  for Knmi.asp and  $0.047^{\circ}\text{C}$  for Stev.pvc. Differences between wet and dry bulb temperatures decrease with increasing relative humidity. This is probably the reason that for high relative humidity conditions like fog ( $rh$  on average 99.5%), condensation of small droplets on the sensors (or on the inner panels of the screen) are of little effect.

#### *3.2.4 Snow*

On clear days with snow cover, effects of differences in screen design on the temperature measurements may be extremely large (see e.g. Gill, 1983). Due to the large albedo of snow (between 0.40 and 0.95 for old and fresh snow, respectively) compared to grass (about 0.2), there is a strong increase in reflection of radiation for conditions with snow cover. Screen design determines the extent to which the reflected radiation directly or indirectly reaches the sensor. Unfortunately, there are only a few days per year with significant snow cover in the Netherlands (on average 4 days/year with snow cover  $> 5 \text{ cm}$  in De Bilt). The number of clear days with snow cover is, of course, even smaller.

There were only two serious periods with uninterrupted snow cover during the experiment: (1) a long extended period from 8–22 February 1991, when only Knmi.ref and Knmi.asp were in operation, and (2) a short period from 6–8 January 1995, when Knmi.ref, Young.aspII, Socrima and Stev.pvc.asp were in operation. For the first period Table V presents some statistics of  $\Delta T$  for both daytime and nighttime hours and compares these with the corresponding values for all winter (DJF) days. Average cloud cover is 5.4 and 5.7 octa in the snow-cover and DJF-days period, respectively. The average  $u$  in the snow-cover period (2.4 m/s) is somewhat lower than in the DJF-days period (3.6 m/s). There is, however, no statistical relationship between the magnitude of  $u$  and  $\Delta T$  for that period. Despite the similarity between the screens, the table shows some interesting differences in the temperature measurements. Firstly, there is a relative coolness of Knmi.asp both during day and night and this coolness is larger in the snow-

cover period. For the snow-cover period this coolness exceeds the predicted measurement uncertainty of  $0.03^{\circ}\text{C}$ . Secondly, the minimum values of  $\Delta T$  are in the snow-cover period ( $-1.39$  and  $-0.74^{\circ}\text{C}$  for night and day, respectively). And, thirdly, the standard deviation of  $\Delta T$  is much larger during the snow-cover period than for all days.

The screens in the second period are compared in Figure 15. The figure shows some interesting differences between the screens, especially on 7 January when there is a period with reduced cloud cover (not shown in the figure) and low wind speeds from just before sunrise until about sunset. Around noon, the Young.apsII and Stev.pvc.asp are up to  $1.3^{\circ}\text{C}$  and  $0.5^{\circ}\text{C}$  cooler than Knmi.ref, respectively, while Socrima is up to about  $0.6^{\circ}\text{C}$  warmer than Knmi.ref.

#### 4. DISCUSSION AND CONCLUSION

We compared 10 different thermometer screens with respect to their weather-dependent temperature differences. An important aspect of the study is the calibration of the sensors under field conditions. Although the temperature sensors were calibrated at regular intervals, a stated operational measurement uncertainty of  $0.1^{\circ}\text{C}$  is not good enough for such a thermometer screen intercomparison. We introduced a simple method to obtain inter-sensor accuracies of about  $0.03^{\circ}\text{C}$  and to reveal malfunctioning of the instruments. The method compares the sensors in the screens with the sensor in the reference screen Knmi.ref for situations where  $dT/dt \sim 0$  for an extended period of time around the measurement. The resulting corrections are imposed on the data. Similar corrections can be obtained by comparing the screens for conditions of high wind speed and overcast skies. Following the calibration, we had to exclude some data from the analysis and Young.aspI was completely skipped. The proposed method may also be used to reduce inter-sensor accuracies in situations where no calibration values are known or where there are discrepancies between recalibration and pre-calibration values.

It would have been preferable to have a double or triple setup of the screens in the inter-comparison. Although this may not exclude all possible errors, it may serve as an extra check to guarantee that the temperature data are correctly measured. WMO also advocates the use of multiple setups: in the year 2007 they start a screen intercomparison study with a double setup in Algeria. In practice, however, multiple setups require extra equipment and manpower and may therefore not always be feasible.

The results in this paper show that the screens can roughly be classified into three groups with distinct behavior: (1) the round multi-plate screens (Knmi.ref, Knmi.asp, Vaisala, Young, Socrima), (2) the naturally ventilated Stevenson screens (Stev.wood, Stev.pvc), and (3) the strongly aspirated screens (Young.aspII, Stev.pvc.asp). The round multi-plate screens have similar response times (Section 3.1) and there are only minor deviations from the temperatures of Knmi.ref in the presented examples (Section 3.2). There is, however, some indication that Socrima responds somewhat slower than the other multi-plate screens (lower standard deviation  $sT$  of the 15 s temperature samples and larger response time at  $t_n$ ). This may be related to the somewhat more closed construction of Socrima compared to the other multi-plate screens. The (slight) aspiration of Knmi.asp apparently has little effect because  $\Delta T$  (Knmi.asp–Knmi.ref) is generally close to zero. An exception is the case with uninterrupted snow cover in the period 8–22 February 1991. For this (for the Netherlands exceptional) situation, there is some evidence that Knmi.asp is cooler than Knmi.ref ( $\sim 0.05^{\circ}\text{C}$  on average).

Stev.wood and Stev.pvc respond significantly slower to temperature changes than all other screens. Their response times with respect to Knmi.ref depend on wind speed  $u$ , and are significantly larger than for the other screens (on average about 7 min on  $t_n$  and 3 min on  $t_x$ ). This matches the low  $sT$  of these screens. Consequently, part of the temperature differences  $\Delta T$  (screen–Knmi.ref) for these screens result from differences in response times. For instance, on clear nights with low wind speed and an erratic course of the ambient temperature this may cause fluctuations in  $\Delta T$  of about  $1^{\circ}\text{C}$ . During the day, the effect is a lowering of  $\Delta T$  for periods

of strongly positive  $dT/dt$  (e.g. after sunrise) and an increase in  $\Delta T$  for periods of strongly negative  $dT/dt$ . This effect is, however, dominated by the direct and indirect heating of the screens by the sun during the day (radiation error). This heating effect is most obvious around  $t_x$  (when  $dT/dt \sim 0$ ) and can easily amount to  $0.5^\circ\text{C}$ . It should be noted that there are only minor differences between the wooden and synthetic version of the Stevenson screen.

The third distinct group contains the strongly aspirated Young.aspII and Stev.pvc.asp screens. Although they have a completely different design, both screens have in common that they have a fast response to ambient temperature changes (large values of  $sT$ ) and tend to be cooler during the day and somewhat warmer during the night. As expected, the differences with respect to the other screens become smaller as  $u$  increases. The most important advantage of artificial ventilation of screens is the reduction of the daytime radiation error. Young.aspII is better equipped to attain this than Stev.pvc.asp. In the presented examples, Young.aspII can easily be  $0.5\text{--}0.8^\circ\text{C}$  cooler during the day than Knmi.ref while Stev.pvc.asp is only about  $0.2^\circ\text{C}$  cooler. For the studied period with snow cover these values were about 2 times as large. The relative warmth of both screens during clear nights (especially Young.aspII), may be caused by disturbance of the stable stratified layer around the screen by the artificial ventilation. However, more research is needed to confirm this.

A peculiarity of the Young.aspII is the relatively low temperatures during rainfall conditions ( $\Delta T = -0.23^\circ\text{C}$ ).  $\Delta T$  for these conditions is negatively correlated with  $u$ , suggesting that aspiration of small droplets to the sensor may be responsible for the effect (causing wet-bulb effects). This may be an important disadvantage when Young.aspII is to be used in regions with frequent rainfall.

Comparison of the results with other studies is not easy because different reference screens have been used and other screens have been compared. Nonetheless, studies like those of Anderson and Mattison (1991) also indicate that Stevenson screens respond much more slowly than the modern round multi-plate screens, causing large temperature differences, especially around sunrise and sunset. The relatively small inter-screen differences between the round multi-plate screens are evident in other studies as well.

From the results in this paper, it appears that  $u$  is a major factor in determining interscreen temperature differences. The magnitude of  $u$  determines to a large extent the differences in response times between screens, the magnitude of radiation errors, the effects of rainfall on Young.aspII and the potential effect of artificial ventilation on the build up of a stable layer on clear nights. In the present experiment  $u$  was measured at a height of 10 m (and later 20 m) at a distance of about 200 m from the screens. For the development of transfer functions it would be much better to measure  $u$  at screen height close to the screens with high-accuracy anemometers (e.g. sonic anemometers) especially in the range  $0\text{--}3$  m/s.

Together with Part II (Brandsma and Van der Meulen, 2007), this paper should give an adequate basis for assessing the impacts of screen changes in countries with climates comparable to those of the Netherlands.

## ACKNOWLEDGEMENTS

We are grateful to David Murphy for proofreading the manuscript. We also thank the reviewers for comments.

## REFERENCES

- Andersson T, Mattisson I. 1991. *A Field test of thermometer screens*. Swedish Meteorological and Hydrological Institute. Report No.: RMK 62. 41pp.
- Barnett A, Hatton DB, Jones DW. 1998. *Recent changes in thermometer screen design and their impact*. Instruments and observing methods, Report No. 66. (WMO/TD-No. 871). 12pp., WMO, Geneva).
- Brandsma T, Van der Meulen JP. 2007. Thermometer screen intercomparison in De Bilt (the Netherlands), Part II: Climatological effects. This issue.
- Brock FV, Semmer SR, Jirak C. 1995. *Passive solar radiation shields: wind tunnel testing*. Preprints, 9th Symposium on Meteorological Observations and Instrumentation; American Meteorological Society, Charlotte, North Carolina, March 27–31. 179–183.
- Cleveland WS and Devlin SJ. 1988. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* **83**: 596-610.
- Chenoweth M. 1992. A possible discontinuity in the U.S. historical temperature record. *Journal of Climate* **5**: 1172–1179.
- Gill GC. 1983. *Comparison Testing of Selected Naturally Ventilated Solar Radiation Shields*. Report to NOAA Data Buoy Office for Development Contract #NA-82-0A-A-266, 15pp., 15 figs.
- Hatton DB. 2002. *Results of an Intercomparison of Wooden and Plastic Thermometer Screens*, paper no. P1.1(19) presented at TECO-2002 (Bratislava), Instruments and Observing Methods Report No. 75, WMO/TD–No. 1123, Geneva.
- ISO (International Standardization Organization). 2004. *Test Methods for Comparing the Performance of Thermometer Shields/Screens and Defining Important Characteristics* (ISO/DIS 17714), Geneva.
- Larre MH, Hegg K. 2002. *Norwegian National Thermometer Screen Intercomparison*, paper no. P1.1(1) presented at TECO-2002 (Bratislava), Instruments and Observing Methods Report No. 74, WMO/TD–No. 1028, Geneva.
- Lefebvre G. 1998. *Comparison of Meteorological Screens for Temperature Measurement*, paper presented at TECO-98 (Casablanca), Instruments and Observing Methods Report No. 70, WMO/TD–No. 877, pp. 315, Geneva.
- Leroy M, Tammelin B, Hyvönen R, Rast J, Musa M. 2002. Temperature and humidity measurements during icing conditions, paper no. 1.1(1) presented at TECO-2002 (Bratislava), Instruments and Observing Methods Report No. 74, WMO/TD–No. 1028, Geneva.
- Lin X, Hubbard FG, Meyer, GE. 2001. Airflow characteristics of commonly used temperature radiation shields. *Journal of Atmospheric and Oceanic Technology* **18**: 329–339.
- Mawley E. 1897. Shade temperature. *Quarterly Journal of the Royal Meteorological Society* **XXIII** (102):69–87.
- Perry MC, Prior MJ, Parker DE. 2007. An assessment of the suitability of a plastic thermometer screen for climatic data collection. *International Journal of Climatology* **27**: 267–276.
- Quayle RG, Easterling DR, Karl TR, Hughes PY. Effect of recent thermometer changes in the cooperative station network. *Bulletin American Meteorological Society* **11**: 1718–1723.
- Richardson SJ, Brock FV. 1995. Passive solar radiation shields: energy budget – optimizing shield design. Preprints, 9th Symposium on Meteorological Observations and Instrumentation; American Meteorological Society, Charlotte, North Carolina, March 27–31. 259–264.
- Sparks WR. 1972. *The Effect of Thermometer Screens on the Observed Temperature*. WMO-No. 315 (WMO, Geneva)
- Sparks WR. 2001. *Field Trial of Metspec Screens*. Met Office/OD, Technical report TR19, Wokingham.
- Spetalen, A, Lofseik, C and Nordli, PØ, 2000. A Comparison of air temperature radiation screens by field experiments and Computational Fluid Dynamics (CFD) simulations. In: Papers presented at the WMO Technical Conference on Meteorological Instruments and Methods of Observation (TECO-2000), Beijing, China, 23-27 October 2000. Instruments and Observing Methods report No.74 (WMO/TD – No 1028, WMO, Geneva, 2000).
- Van der Meulen. 1998. A thermometer screen intercomparison. In: Papers presented at the WMO Technical Conference on Meteorological Instruments and Methods of Observation (TECO-1998), Casablanca, Morocco 13–15 May 1998. Instruments and Observing Methods report No.70 (WMO/TD – No 877, WMO, Geneva, 1998).

- Van der Meulen. 2000. Temperature measurements: some considerations for the intercomparison of radiation screens. In: Papers presented at the WMO Technical Conference on Meteorological Instruments and Methods of Observation (TECO-2000), Beijing, china, 23-27 October 2000. Instruments and Observing Methods report No.74 (WMO/TD – No 1028, WMO, Geneva, 2000).
- WMO (World Meteorological Organization). 1992. *International Meteorological Vocabulary* (WMO-No. 182). WMO, Geneva.
- WMO (World Meteorological Organization). 1996. *Guide to Meteorological Instruments and Methods of Observation (WMO No. 8)*, Chapter 2, *Measurement of Temperature*. WMO, Geneva.
- WMO (World Meteorological Organization). 2003. Commission for instruments and methods of observation. 13<sup>th</sup> session, Bratislava, 25 September – 3 October 2002. WMO-No.947. WMO, Geneva.
- Wylie RG, L alas T. 1992. Measurement of temperature and humidity: specification, construction, properties and use of the WMO reference psychrometer. Technical note no. 194 (WMO-No.759). WMO, Geneva.

## TABLES

Table I. Details of screens and sensors. The Stevenson screens are of KNMI design.

Screen	Abbreviation	Start date	End date	Diameter (m)	Ventilation	Sensor
KNMI multi-plate	Knmi.ref	89/01/09	95/02/01	0.30	naturally	Pt500
KNMI multi-plate aspirated	Knmi.asp	90/12/13	93/02/20	0.30	aspirated (1 dm <sup>3</sup> /min)	Pt500
Vaisala multi-plate DTR11	Vaisala	89/01/09	93/02/20	0.30	naturally	Pt500
Young Gill multi-plate 41002	Young	89/01/09	93/02/20	0.12	naturally	Pt500
Young aspirated type I 43408	Young.aspl	90/12/13	92/08/18	0.15/0.04	aspirated (0.1 dm <sup>3</sup> /s)	Pt1000 <sup>1</sup>
Young aspirated type II 43408	Young.asplII	92/08/18	95/02/01	0.15/0.025	aspirated (0.1 dm <sup>3</sup> /s)	Pt1000 <sup>1</sup>
Socrima multi-plate BMO 1167A	Socrima	91/03/08	95/02/01	0.20	naturally	Pt500
Stevenson PVC	Stev.pvc	89/01/09	91/03/06	0.70	naturally	Pt500
Stevenson PVC aspirated	Stev.pvc.asp	91/03/07	95/02/01	0.70	aspirated (1 dm <sup>3</sup> /s)	Pt500
Stevenson wood	Stev.wood	89/01/09	93/02/20	0.70	naturally	Pt500

<sup>1</sup> For the Young aspirated screens the sensor is an integral part of the measuring device. The first value for diameter for these screens refers to the diameter of the radiation shield, the second to the tube that houses the sensor.

Table II. Differences (screen – Knmi.ref) in times  $t_n$  and  $t_x$  of occurrence of the minimum and maximum temperatures in the summer half year (Apr-Sep).

Screen	$\Delta t_n$ (minutes)			$\Delta t_x$ (minutes)		
	mean	st.err	days	Mean	st.err	days
Knmi.asp	0.27	0.72	184	-0.17	0.58	175
Vaisala	-0.18	0.48	283	-0.62	0.51	273
Young	0.04	0.51	278	0.34	0.66	265
Young.asplII	-0.33	0.92	150	-1.50	1.17	140
Socrima	0.98*	0.43	317	0.16	0.42	306
Stev.pvc	7.94*	0.75	175	3.31*	0.94	163
Stev.pvc.asp	-1.93*	0.53	316	0.00	0.52	299
Stev.wood	6.65*	0.78	185	2.60*	0.89	173

\* values differ more than 2 times the standard error from zero.

Table III. Statistics of the temperature differences  $\Delta T$  (screen – Knmi.ref) in wet intervals compared with dry overcast intervals. Note that the standard deviation is presented instead of the standard error. All mean values differ more than 2 times the standard error from zero.

Screen	$\Delta T$ (wet intervals)			$\Delta T$ (dry and overcast intervals)		
	mean	stdev	intervals	mean	stdev	intervals
Knmi.asp	-0.001	0.023	9406	-0.011	0.031	40172
Vaisala	-0.005	0.023	7637	-0.012	0.029	33173
Young	-0.018	0.041	7637	-0.001	0.035	33173
Young.aspII	-0.229	0.245	5895	-0.084	0.174	34339
Socrima	0.010	0.042	4489	-0.005	0.047	25933
Stev.pvc	-0.013	0.083	9535	-0.003	0.076	41435
Stev.pvc.asp	-0.026	0.047	4489	-0.004	0.041	25933
Stev.wood	0.013	0.089	8781	0.002	0.082	39097

Table IV. Mean temperature differences (screen – Knmi.ref) for intervals with high values of global radiation ( $> 750 \text{ W/m}^2$ ) for the four quartiles of windspeed.

Screen	$u < 2.9 \text{ m/s}$	$2.9 \leq u < 3.7$	$3.7 \leq u < 4.9$	$u \geq 4.9 \text{ m/s}$
Knmi.asp	-0.086	-0.042	-0.050	-0.029
Vaisala	-0.097	-0.047	-0.028	-0.007
Young	-0.127	-0.075	-0.026	-0.003
Young.aspII	-0.536	-0.368	-0.260	-0.202
Socrima	0.066	0.043	0.020	0.030
Stev.pvc	0.174	0.078	0.050	0.060
Stev.pvc.asp	-0.157	-0.133	-0.100	-0.027
Stev.wood	0.120	0.048	0.030	-0.003

\* values in italics are not significantly different from zero ( $2 \times \text{se}$ )

Table V. Statistics of nighttime and daytime  $\Delta T$  (Knmi.asp–Knmi.ref) for an extended period of uninterrupted snow cover (8–22feb91) and for all winter days (DJF, 176 days) with measurements for both Knmi.asp and Knmi.ref). Standard error information is given for the mean only.

Period	Nighttime $\Delta T$ (°C)				Daytime $\Delta T$ (°C)			
	mean	min	max	stdev	mean	min	max	stdev
8–22feb91	-0.061*	-1.390	0.250	0.157	-0.041*	-0.740	0.360	0.114
DJF	-0.006*	-1.390	0.310	0.057	-0.015*	-0.740	0.770	0.054

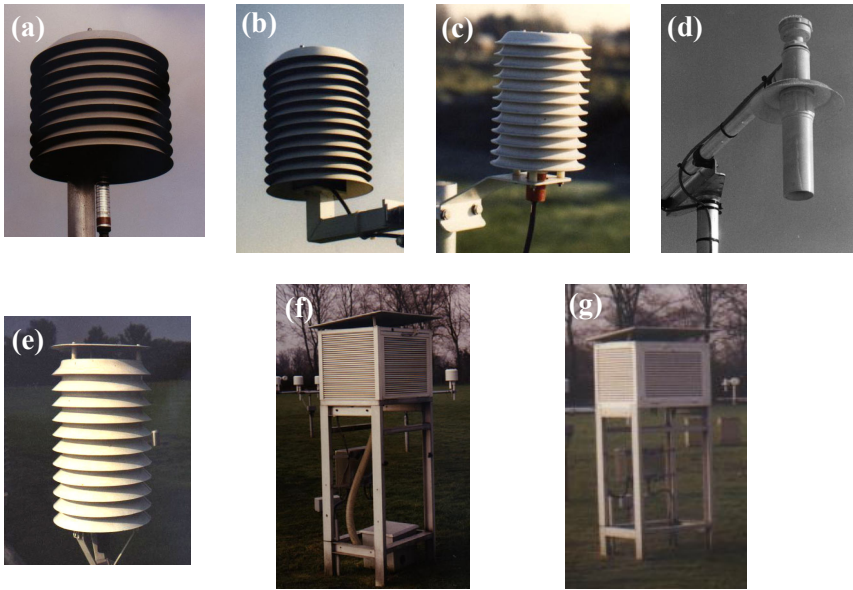
\* values differ more than 2 times the standard error from zero.

**FIGURES**



Figure 1. Experimental set-up, (a) for the period 1989-1991, (b) for the period 1992-1995.

Figure 2. Overview of screens: (a) KNMI multi-plate, (b) Vaisala multi-plate, (c) Young Gill



multi-plate, (d) Young aspirated (type I and type II), (e) Socrima multi-plate, (f) Stevenson PVC, and (g) Stevenson wood. Screens (a) and (f) consist of an aspirated and natural ventilated version.

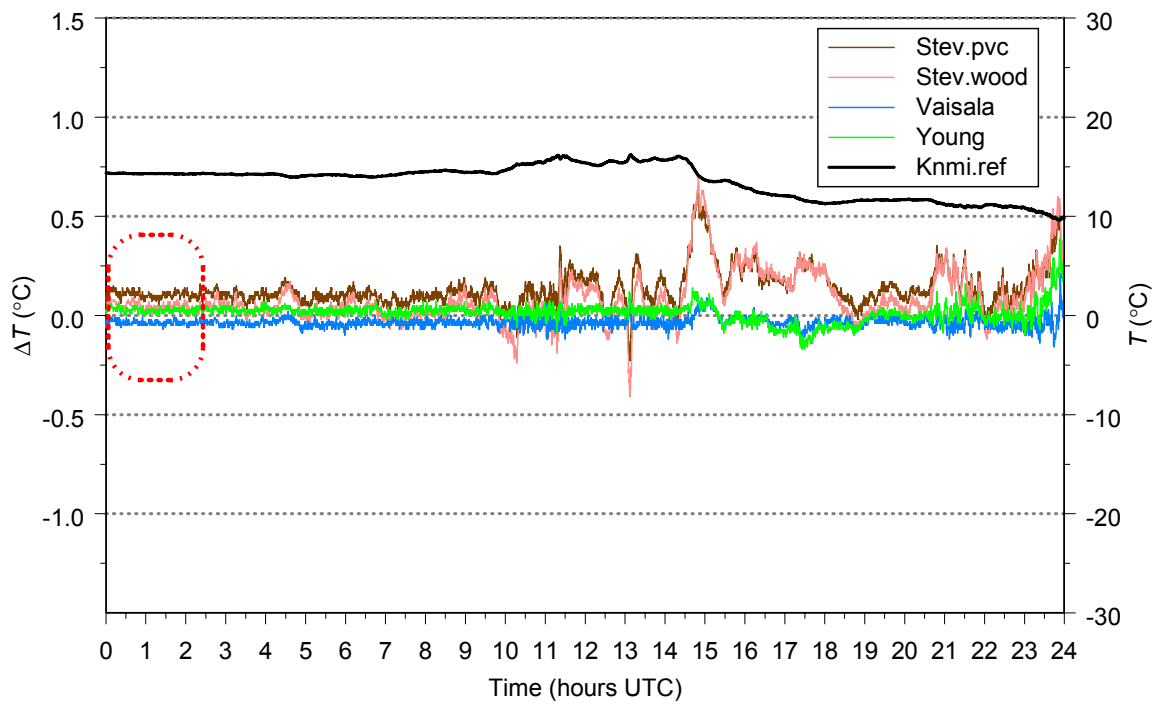


Figure 3. Example of conditions suitable for inter-sensor calibration (dotted box) on 13 October 1989. Data are 15 sec samples. The left axis gives the differences between the screens and Knmi.ref, the right axis gives actual air temperature of Knmi.ref. Local solar time = UTC + 20 minutes. The weather conditions for the data in the box are overcast skies and strong winds of 5 m/s.

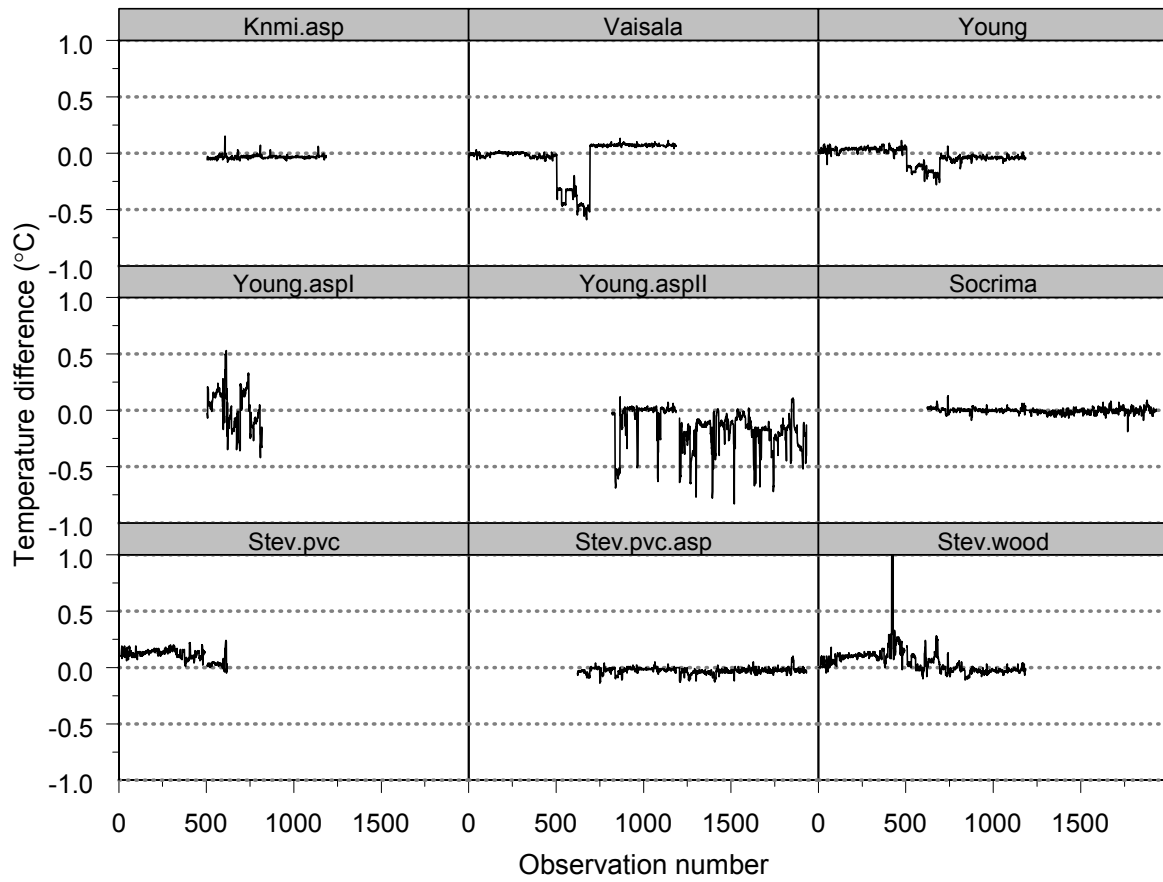


Figure 4. Temperature differences  $\Delta T$  (screen – Knmi.ref). Only observations are selected that satisfy the criterion that for Knmi.ref the root mean squared deviations in the period from one hour before to one hour after the observation are  $< 0.03^{\circ}\text{C}$ .

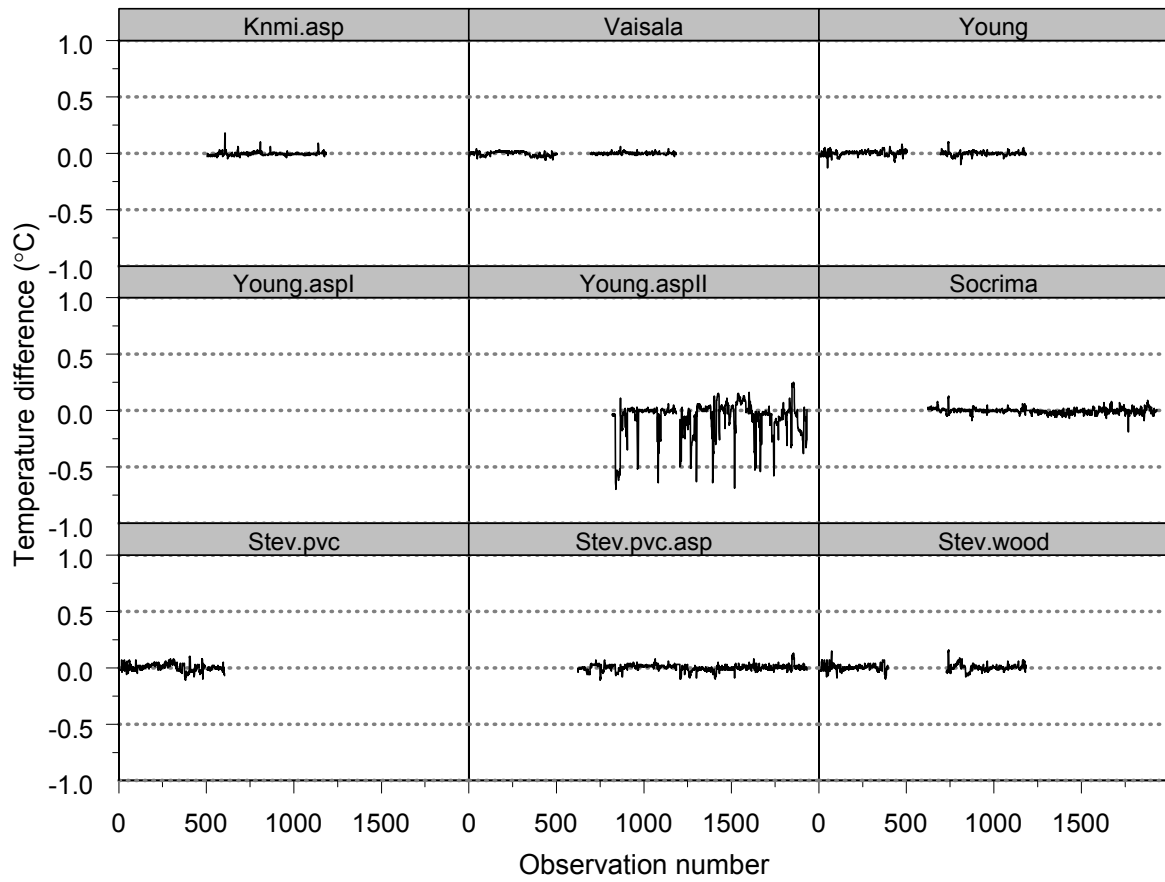


Figure 5. As Figure 4 but after bias corrections. The panel for Young.aspl1 is empty because no serious bias corrections could be applied.

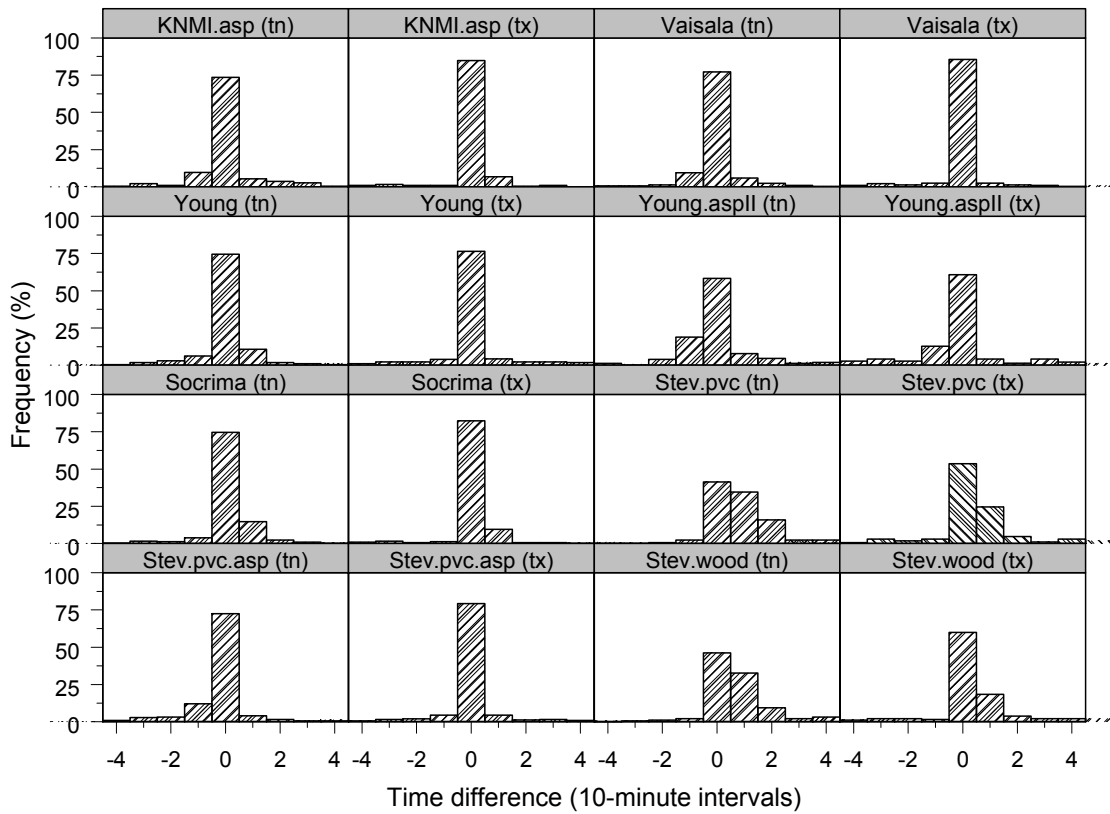


Figure 6. Histograms of the differences (screen – Knmi.ref) in times  $t_n$  and  $t_x$  of occurrence of the minimum and maximum temperatures in the summer half year (Apr-Sep). The number days that constitutes each histogram is given in Table II.

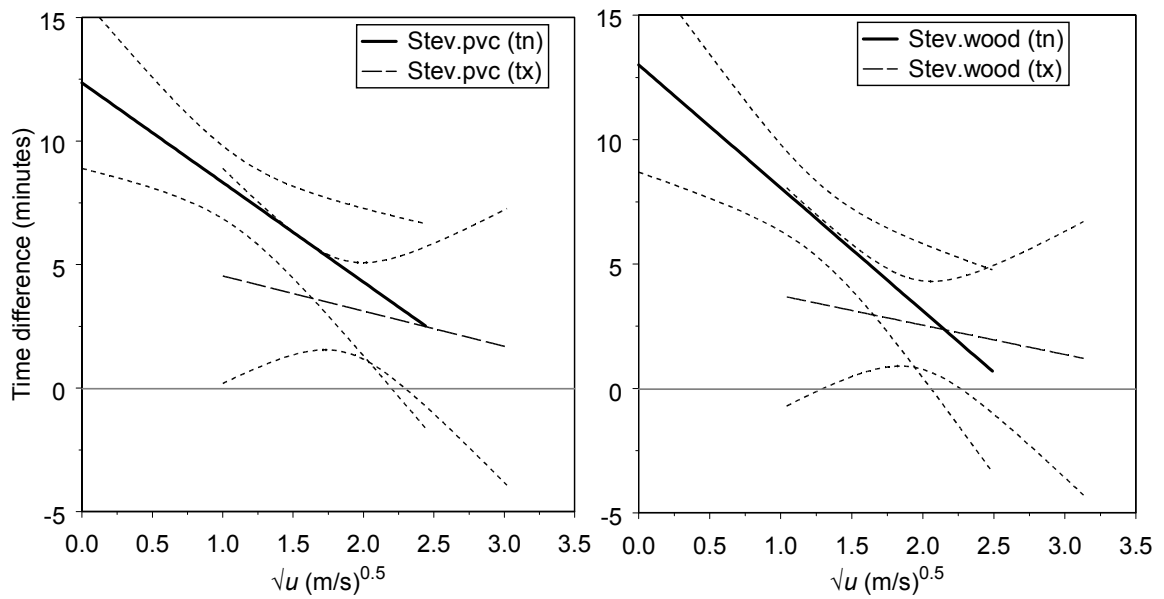


Figure 7. Relationship between the differences (screen – Knmi.ref) in times  $t_n$  and  $t_x$  of occurrence of the minimum and maximum temperatures and the square root of windspeed for Stev.pvc (left) and Stev.wood (right) in the summer half year (Apr-Sep). The lines represent linear least squares fits together with 95% confidence bands (short dashes).

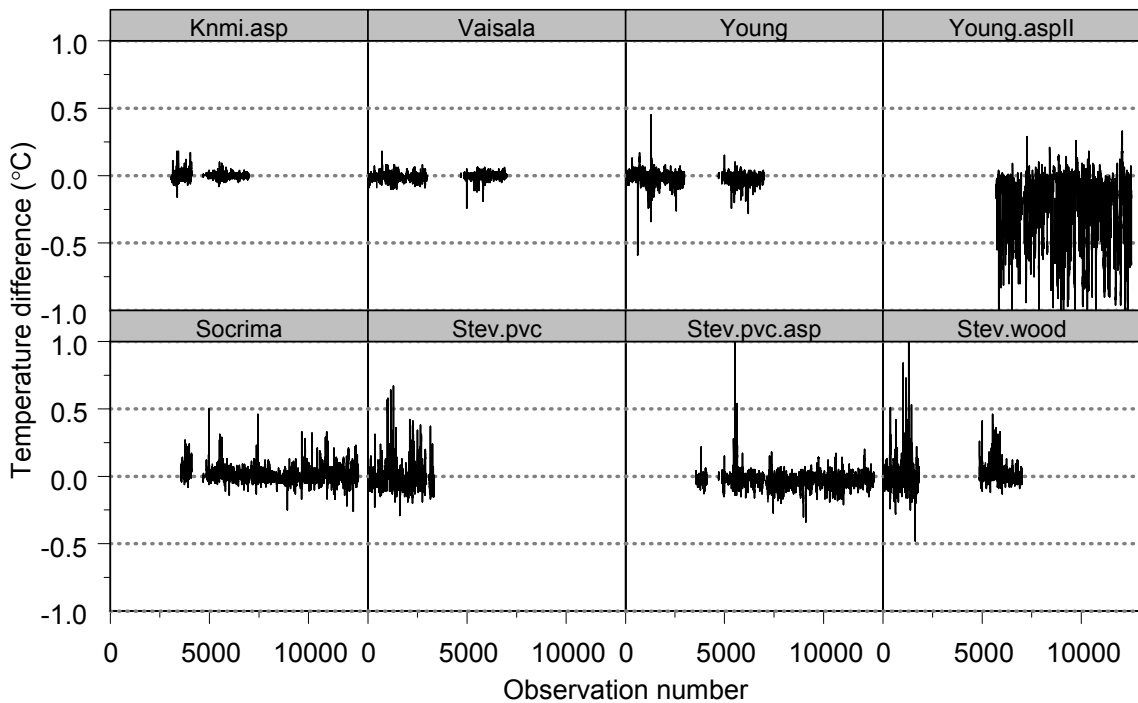


Figure 8. Temperature differences  $\Delta T$  (screen – Knmi.ref) for rainfall conditions. Only observations are selected with uninterrupted rainfall in the measurement interval.

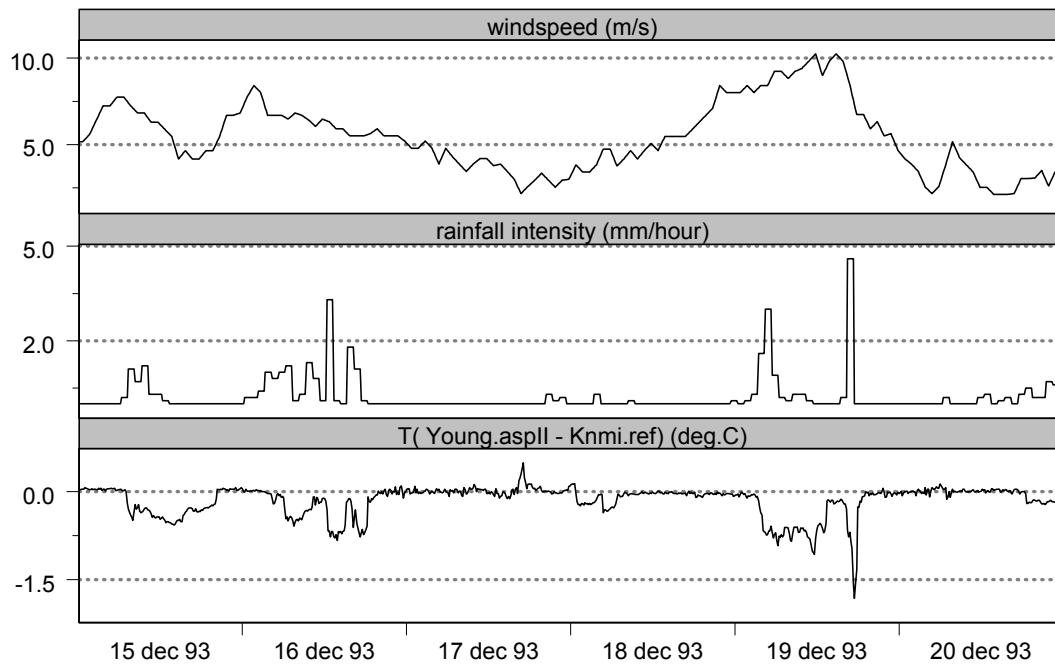


Figure 9. Temperature differences  $\Delta T$  (Young.asplI – Knmi.ref), rainfall intensity and wind speed during the period 15–20 December 1993.

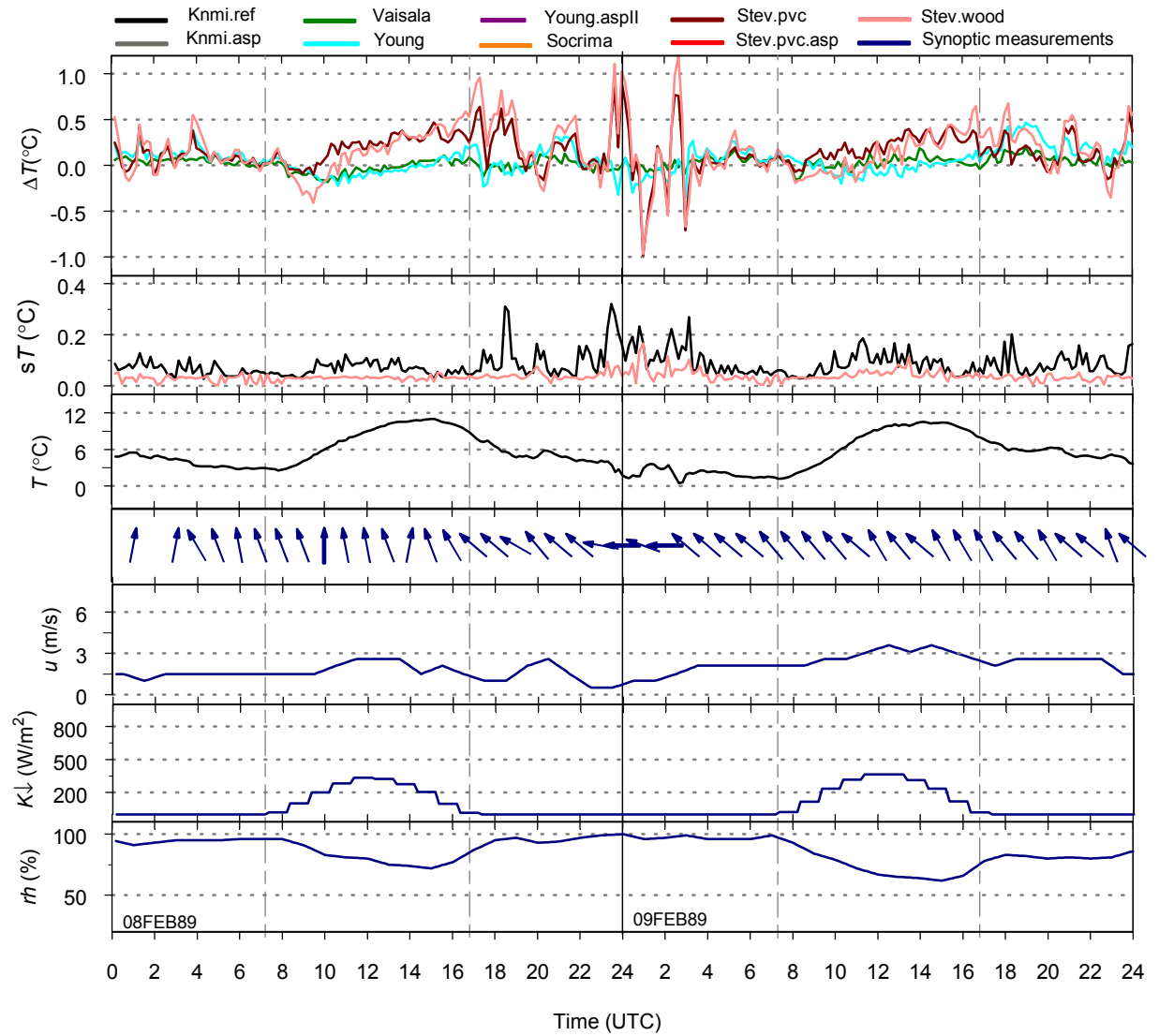


Figure 10. Temperature differences  $\Delta T$  (Screen – Knmi.ref) for the two consecutive clear days 8–9 February 1989. The figure also presents the standard deviation of the 15 sec samples  $sT$ , temperature  $T$  of Knmi.ref, wind direction (arrows pointing to the direction where the wind blows) and wind speed  $u$ , global radiation  $Q_g$  and relative humidity  $rh$ .  $\Delta T$ ,  $T$  and  $sT$  are 10-min mean values, the synoptical measurements are derived from hourly values. The dashed vertical lines denote the times of sunrise and sunset.

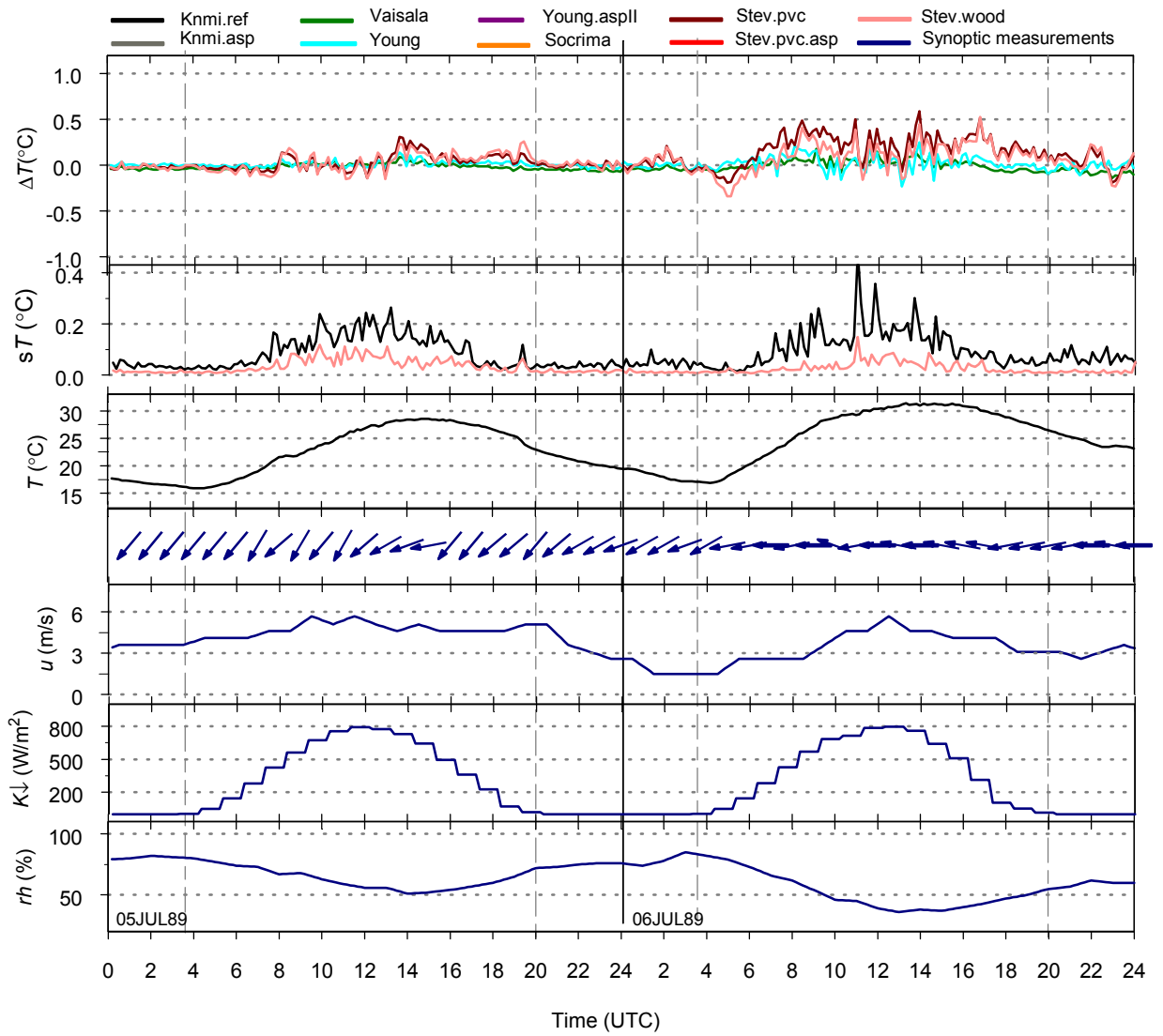


Figure 11. As Figure 10 but now for 5–6 July 1989.

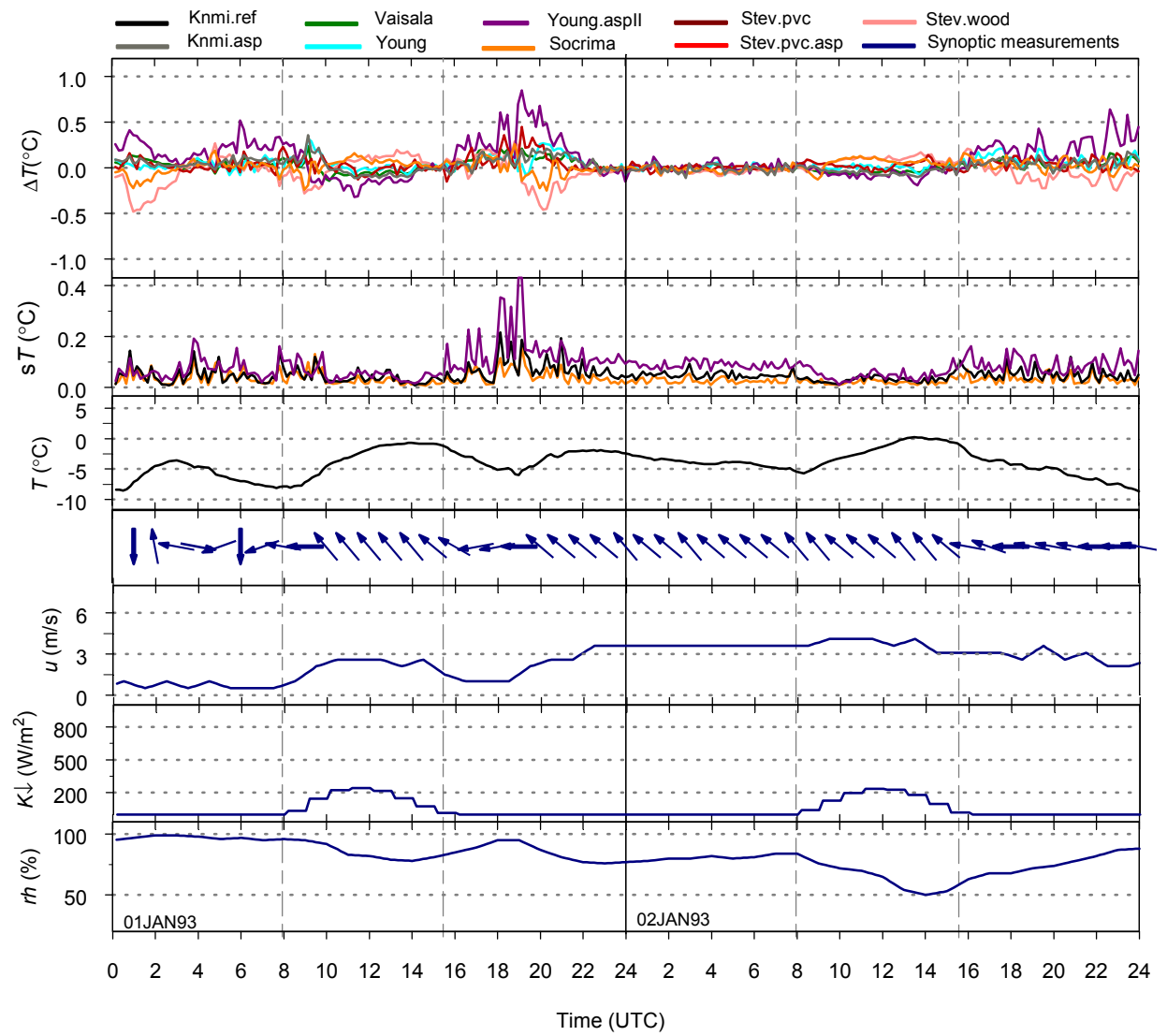


Figure 12. As Figure 10 but now for 1–2 January 1993.

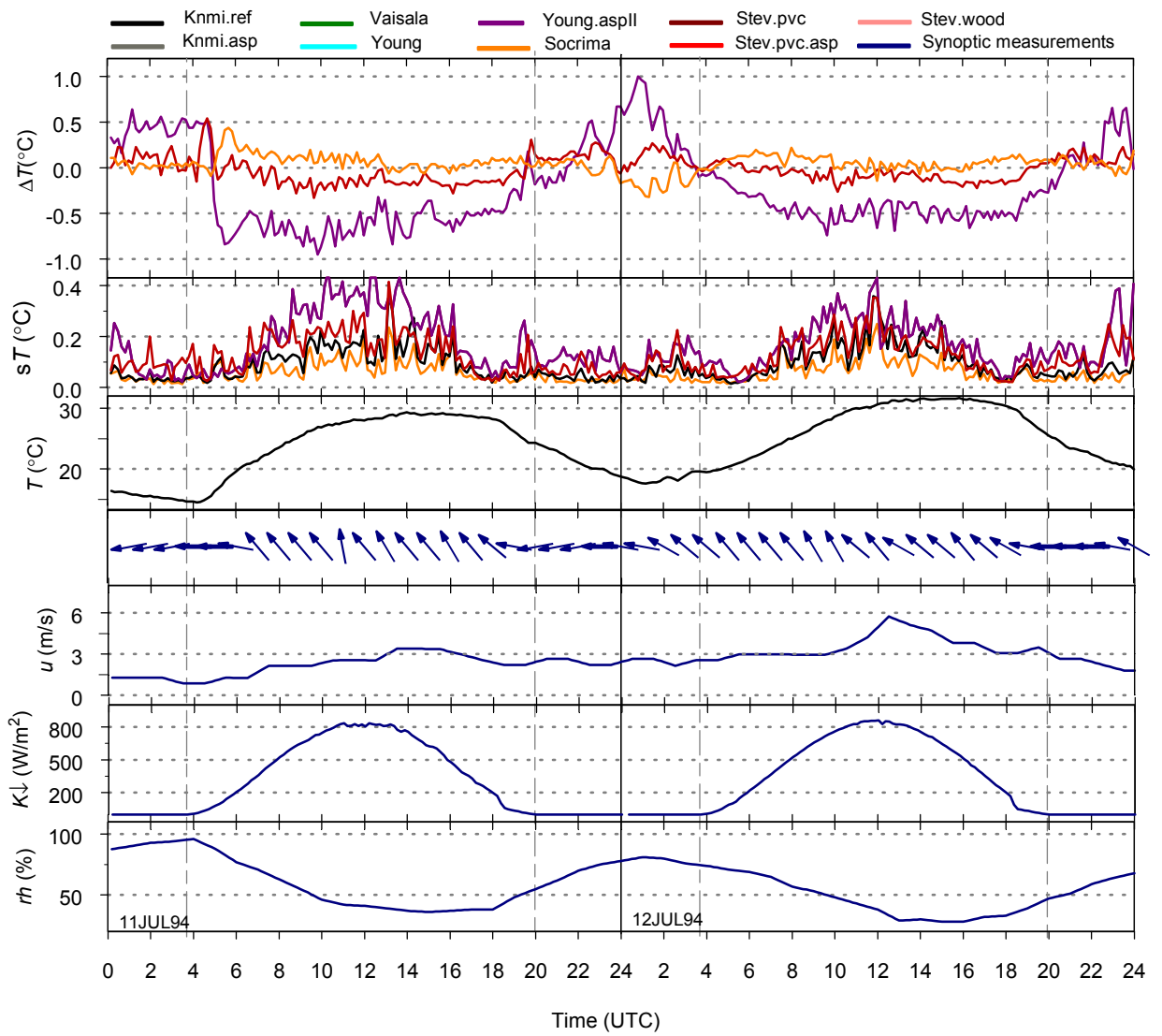


Figure 13. As Figure 10 but now for 11–12 July 1994.

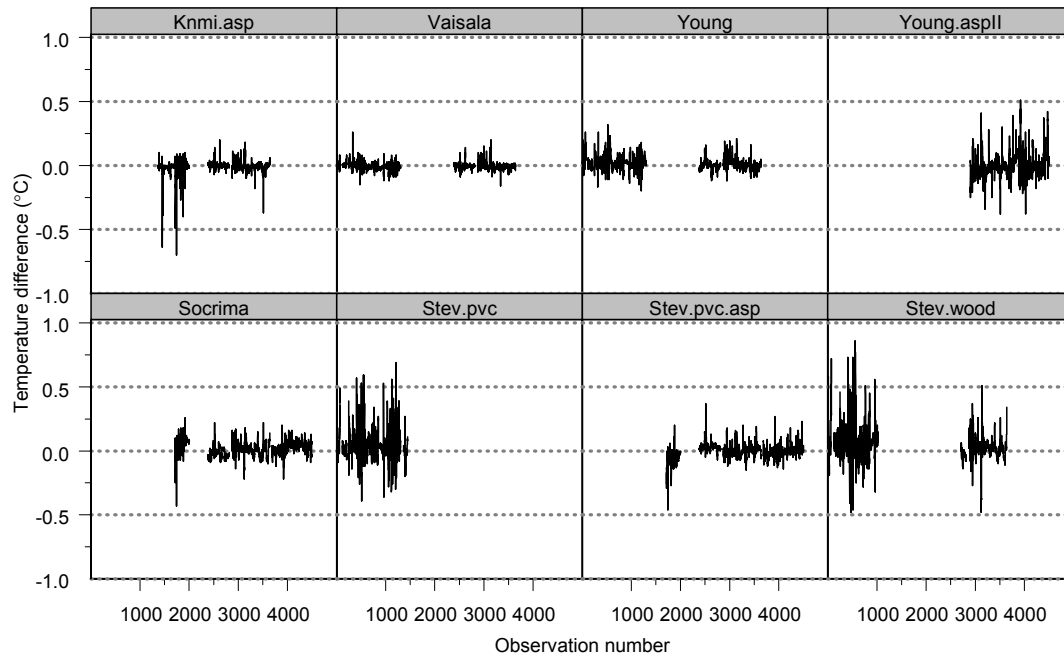


Figure 14. Temperature differences  $\Delta T$  (screen – Knmi.ref) for fog conditions. Only observations are selected from hours where the synoptic hourly measurements indicated fog with visibility < 200 m.

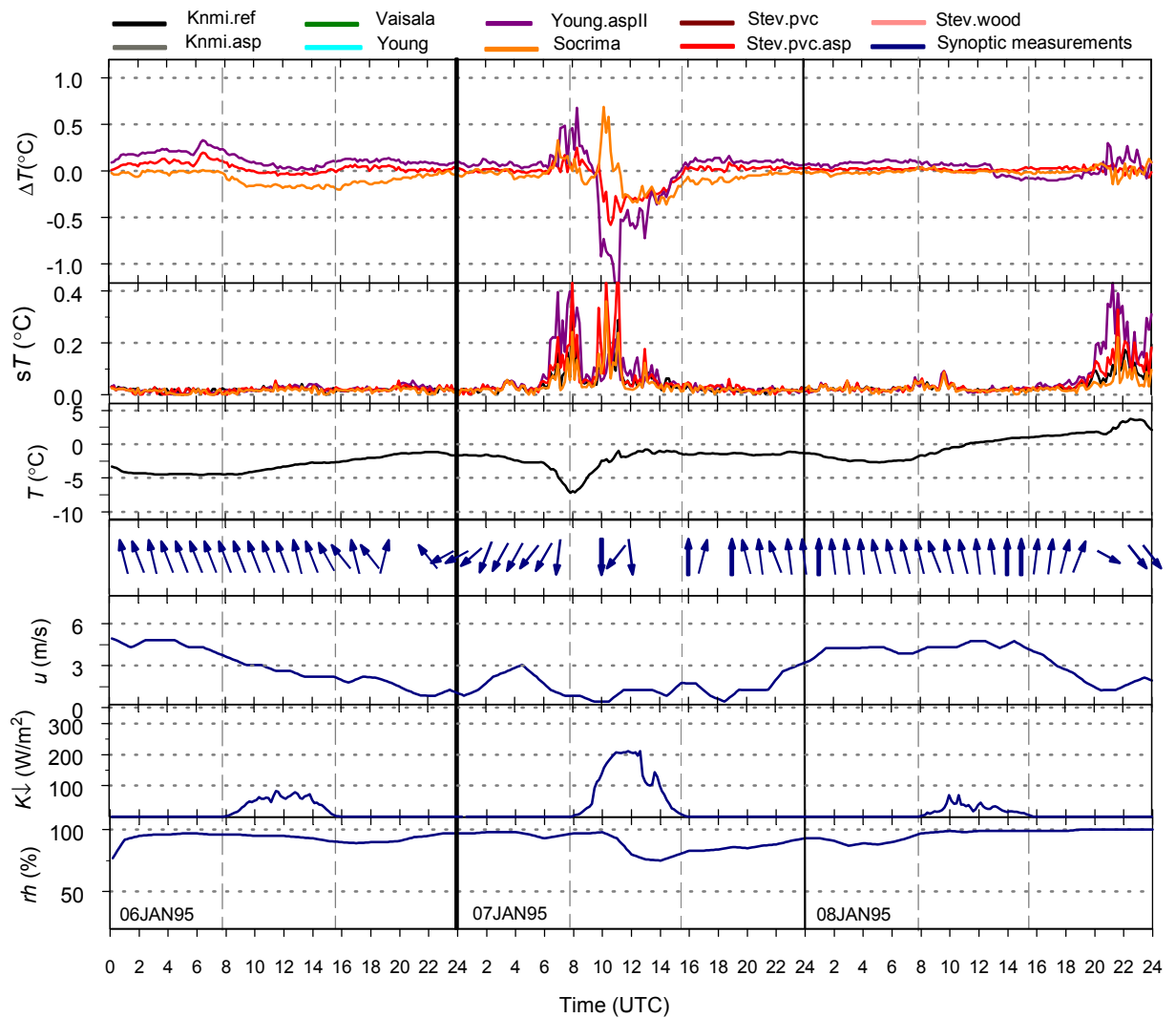


Figure 15. As Figure 10 but now for the three day period 6–8 January 1995 with uninterrupted snow cover. From 0–10 UTC on 6 January snow falls to a depth of 10 cm. Thereafter completely cloudy skies prevail till 5 UTC on 7 January, followed by partly clear skies till sunset that same day. Then there are again completely cloudy skies until 9 UTC on 8 January and thereafter it started raining until 19 UTC. The last part of 8 January has partly clear skies. The snow cover depth at the end of the period is 5 cm.