ERA report series

7 Data usage and quality control for ERA-40, ERA-Interim and the operational ECMWF data assimilation system

Christina Tavolato and Lars Isaksen, Research Department

Series: ERA Report Series

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

This report investigates and interprets time series of data usage, quality control decisions and departure statistics for ERA-40, ERA-Interim and the operational ECMWF data assimilation system. This is done with special emphasis on the performance of the (variational) quality control for conventional data. While ERA-40 and ERA-Interim are mainly affected by the changes in the available observation datasets, the changes in the use of observations are also due to developments within the assimilation system. ERA-40 used 3D-Var whereas ERA-Interim used 4D-Var, so the intercomparison shows elements of the impact of the assimilation system upgrade. Another interesting comparison is done of the ERA-Interim and the operational assimilation system. Whereas ERA-Interim uses a fixed 4D-Var assimilation system over the whole period, operations updated the 4D-Var assimilation system several times during the period investigated. Not all those updates relate to changes in the data usage, but some differences can be seen and evaluated. These differences are presented with selected examples, as well as tables of data usage statistics to give a general overview.

1 Introduction

The use and rejection of observations is an essential part of any data assimilation system. This report takes a closer look at the time series of data usage and quality control decisions for three different assimilation systems, focusing on conventional data: the ERA-40 reanalysis (1958-2001: Uppala et al. (2005, 2004)); the ERA-Interim reanalysis (from 1989 onwards: Simmons et al. (2007); Uppala (2007)) and the operational ECMWF analyses (1999-2010). The main period of interest is 1989-2010. This report will focus on:

- Data usage plots for all conventional observations for the longest possible time period where assimilation system diagnostics has been produced for the three assimilation systems investigated. Typically quality control (QC) rejected data accounts for a small fraction of the total amount of data, so in order to highlight changes in quality control decisions, rather than the total amount of the data, time series have been plotted using a logarithmic scaling for the data counts. To present some data usage changes clearer some plots were scaled as the percentage of the used data.

  Most of the data usage figures in this report have a similar format. The colours of the curves in the data usage plots are as followed: dark blue shows the total amount of available data, green the amount of used data, black the blacklisted data, red the rejected (first-guess (fg) rejected) data, and magenta the amount of data rejected by the variational quality control (VarQC, Andersson and Järvinen (1999)). This reflects the various QC decisions performed by the ECMWF data assimilation system. We will describe the QC method applied in the ECMWF analysis in some detail in chapter 2.

- In addition to the data usage we show and discuss some time series of departure statistics as well. In those plots standard deviation and bias for innovations ("observation minus background" (obs-bg)) are plotted in red and for "observation minus analysis" (obs-an) in blue.

- For some specialized investigations geographical maps of data count distribution were included.

The layout of this report is as follows. The design of the ECMWF analysis QC decisions is discussed in chapter 2. In chapter 3 we investigate the data usage time series for ERA-40. Chapter 4 looks at ERA-Interim, including a detailed comparison with ERA-40 performance. In section 5 we investigate the operational assimilation system and perform an intercomparison with ERA-Interim.

2 Quality control decisions in ECMWF's assimilation system

This section will briefly describe the quality control decisions performed by the ECMWF assimilation system (see Fig. 1). The quality control is a vital part of an operational data assimilation system and a lot of effort has gone into developing this at ECMWF over the years. The observations received from the GTS and other data sources are usually not quality controlled and therefore contain gross errors due to instrumental and human errors, when the measurement is recorded and during the transmission of the data.

Various checks are performed to eliminate gross errors, i.e., it is checked if routes for ships and aircraft are unrealistic, if the vertical profile of a radiosonde is hydrostatically consistent and if an observation is received more than once. Erroneous and redundant data are rejected and not presented to the analysis.

Because the resolution of the analysis is coarser than the measurement density, especially for many satellite observing systems, the data used by the analysis is thinned horizontally. For satellite data this is also done because observations are assumed to be uncorrelated in the assimilation system, which is questionable for dense data. Aircraft data and DRIBU (drifting and moored buoys) data is also thinned due to high spatial and temporal resolution.

The observation minus background biases and standard deviation of most conventional observations are evaluated on a monthly basis at an individual station level to check if they have deteriorated or improved. This information is used to blacklist data that systematically deviates from the background fields to an unacceptable degree. Similarly data sources that have improved can be whitelisted and allowed into the analysis again. The latter is one of the reasons why also rejected data is compared on a routinely basis against analysis and background fields. It is a very important part of the data assimilation procedure to monitor the quality of observations and flag and store the quality control decisions in feedback files. Most investigations in this report relies on these stored assimilation statistics. For satellite data the bulk of rejections are due to blacklisting. This is typically done at channel level, i.e., to eliminate use of channels that are contaminated by a radiance contributions from the surface or clouds.

The next step is the first guess check that compares individual measurements against the background model fields at appropriate time and space location. If the observation deviates more than 5-6 standard deviations from the typical departure value the observation is considered to be wrong and it is flagged as rejected. The observation will not be allowed to influence the analysis. With a typical quadratic observation cost function it is safer to avoid using suspect data, because it can cause an erroneous analysis.

The observations that pass the gross error checks, systematic error checks, satellite channel blacklisting, data thinning and first guess check is called the "active data". This data will contribute to the solution of the analysis problem. Finally, during the variational analysis minimization process the variational quality control (VarQC), described in Andersson and Järvinen (1999) is applied. In September 2009 a revision to use a Huber norm based VarQC (Tavolato and Isaksen (2010)) was introduced in operations. This allows a relaxed first guess QC for most conventional data. This assures that active data that deviates considerably from an analysis based on other observations in the vicinity (buddy-like quality control) is given a lower weight in the analysis. The VarQC weights are updated dynamically during the analysis step. Until January 2003 VarQC was used in the M1QN3 minimization (Gilbert and Lemaréchal, 1989) framework, after that in the conjugate gradient purely quadratic inner loop formulation (Andersson et al., 2004).

A brief overview of the QC decisions in the three assimilation systems investigated is given in Table 3 and 4 in chapter 6. The tables show the overall statistics for data volumes and quality control decisions during 1995 and 2005, respectively. The tables give the information for all the conventional observing systems used by the analysis and show interesting differences in the data usage. This will be discussed further in chapter 6.

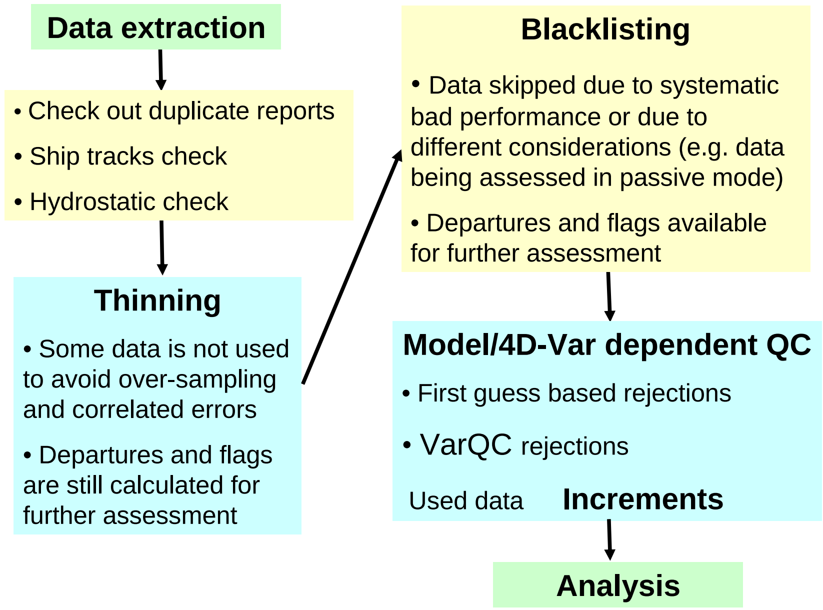


Figure 1: Quality control decisions in the ECMWF data assimilation system

3 ERA 40

ERA-40 is a second-generation reanalysis carried out in 2000-2003 by ECMWF for the 45 year period from September 1957 to August 2002. It is documented in detail in Uppala et al. (2005) and Uppala et al. (2004). The starting point was chosen due to the extension of the observation system at the end of 1957 in preparation for the IGY (International Geophysical Year). During the 45 year assimilation period conventional observations as well as satellite observations (from 1973 onwards) were used. The assimilation system was based on the 6 hour 3D-Var system which was operational at ECMWF from January 1996 to November 1997 (Andersson et al., 1998) with a spectral resolution of T159 (around 125km) and a vertical resolution of 60 levels, which became operational in October 1999 and was used until February 2006. The version of the forecast model was the one used operationally from June 2001 to January 2002 (CY23R4) with some modifications specifically for ERA-40:

- As mentioned above the horizontal resolution was T159 (this is a lower resolution than the operational resolution at that time which was T511). This resolution was chosen to meet the available computational resources for the 45 year assimilation.
- A 'first-guess at appropriate time' approach was used to compare observations against model values at the time they were measured rather than at the synoptic analysis time.
- The forecast timestep used in the assimilation was reduced to 30 minutes from 1 hour to improve the handling of tidal waves.

- The background error variances were increased by about 30% in order to fit the observations better than the operational analysis. This was especially beneficial for extreme events.

Statistics of data usage are available for the whole period of ERA-40 (1958-2002). Time series of the QC decisions for this time period is examined and discussed in this section and interesting results related to data counts as well as quality control decisions are shown.

3.1 Changes in data usage

A summary of the changes in data usage and when they happened is given in Table 1 for an overview on the data usage changes during the 45 year period of the ERA-40 assimilation. Possible explanations for the data count changes are given as well. Most of the described changes in the used data can be seen in the examples in Figs. 2 and 3, showing data usage time series for an upper-air (radiosonde zonal wind) and a surface (surface pressure measurements over land) observation type. The information in Table 1 are mainly based on Uppala et al. (2005), Uppala (2007) and Uppala and Dee (personal communication, 2007).

3.2 Performance of variational quality control

The variational quality control was introduced in ECMWF's variational assimilation scheme in September 1996. It was used in ERA-40 during the whole assimilation period.

The time series of different observation types used in the ERA-40 assimilation identifies two common features in the overall performance of the variational quality control (as an example see the data count time series for radiosonde humidity in Fig. 4):

- An increase of VarQC rejected data until the early 1970s for the upper-air data. This increase can be related to the fact that more and more radiosonde data is retrieved from higher, more unpredictable levels. That data will more likely be rejected by VarQC than data from lower levels.
- A decrease in VarQC starting in the 1980s (the exact beginning depends on the observing system) and continuing until the end of the dataset. It is known (Uppala et al., 2005) that the data quality as well as the quality of the background fields of the forecast model improves over time, leading to an overall improved assimilation system.

Another way of looking at these time series plots is shown in Fig. 5 where the data from Fig. 3 is plotted as percentage of the used data. These time-series are especially interesting when the number of used observations gradually changes over the years. This figure shows that the increase of rejected surface pressure data from 1990 onwards as seen in Fig. 3 is not visible any more when looking at the percentage plot (Fig. 5). This shows that the increasing number of rejections is due to the increasing number of used observations. Fig. 6 is an example of obs-bg/obs-an time series for radiosonde temperatures over the Northern Hemisphere that shows the improved performance from 1980 onwards.

By looking at the VarQC time seris for relative humidity from SYNOP observations (Fig. 7) gaps in the time series can be noticed (from the beginning until 1959, 1973, 1989-1994). ERA-40 was run as a number of individual reanalysis streams which were put together in the end. Those gaps appear at the beginning of each stream of ERA-40.

Year	Data change	Possible explanations
1967-76	More surface data	Additional NCAR/NCEP surface dataset USAF is included in ERA-40.
1973-78	More upper-air data	Additional NCAR/NCEP radiosonde and pilot dataset USAF and ON29 is included in ERA-40.
1974-76	More upper-air data	Additional radiosondes in the tropics due to the GATE (GARP (Global Atmosphere Research Programme) Atlantic Tropical Experiment) experiment are included in ERA-40.
1975-78	More upper-air data	Additional data received from the JMA.
until 1977	More upper-air data	The data was used twice because the encrypted station-IDs in observations received from US-NAVY were not identified by ERA staff. Usually the quality control will identify and blacklist redundant data from the same station. The effect of this mistake is not dramatic: the weight on good radiosonde data over USA was increased during the period.
1979	More upper air data Less surface data	Additional upper-air data was available due to the FGGE (First GARP Global Experiment: Bengtsson et al. (1982)). Additional surface data was also measured but it got thinned during that year, because it at that point was considered an excessive amount of data for the assimilation systems of the early 1980ies. It should be considered if it is possible to extract more of the thinned FGGE observations for future reanalyses.
1980-97	More upper-air data	Additional data received from the JMA.
1988-94	More upper-air data More surface data	Additional NCEP operational GTS data is included in ERA-40.
until 1998	More ship data More buoy data	NCAR COADS dataset is available and used.

Table 1: Changes in data usage of conventional data within ERA-40

3.3 Departure statistics

Taking a closer look at the departure statistics we see some quality control aspects appearing in the time series of the standard deviation and the bias of innovations.

Fig. 8 shows the departure statistics and the data count for radiosonde temperature data in the Northern Hemisphere at 100hPa. Looking first at the data count in the lower panel we see the increase and decrease during the 1970s as already noted above (Fig. 2 and Fig. 6). Taking a closer look at the departure statistics reveals two unusual events in the bias characteristics:

- 1975-1977: A change in the bias due to the VTPR bias correction for NOAA-4 (Uppala et al. (2005), Uppala and Dee 2007, personal communication).
- 1988: A jump in the bias as two independent streams of ERA-40 with different bias correction was joined together.

Looking at the standard deviation an improvement can be seen, especially in the innovation statistics. This is a clear sign that the quality of ERA-40 analysis improved during the 45 years, especially when satellite data became widely available in 1979.

Another dataset where a big improvement in the departure statistics can be seen is for the surface pressure data in the tropics. Fig. 9 shows the clear downward trend in the standard deviation from 1991 onwards in obs-fg and obs-an departure statistics. Also the variability of the bias has decreased over the ERA-40 time period (Uppala et al., 2005).

4 ERA-Interim including an intercomparison with ERA-40

ERA-Interim (Simmons et al., 2007; Uppala et al., 2008) is an interim reanalysis with an improved assimilation system compared to ERA-40 for the data rich period from 1989 onwards. It is now continuing as an ECMWF climate data assimilation system (ECDAS) until it is superseded by a new extended reanalysis.

The increased computer power allowed ERA-interim to be run with a 12 hour 4D-Var assimilation system. The benefit of this system compared to 3D-Var used in ERA-40 is systematically better forecasts, especially in the Southern Hemisphere. The horizontal resolution was also increased to T255 with 60 vertical levels. The version of the forecast model used was the one operational from September 2006 to June 2007 (CY31R2). The main additional differences compared to ERA-40 are:

- A better formulation of the background error constraint.
- Improved model physics and a new humidity analysis.
- A data quality control that draws on experience from ERA-40 and JRA-25 (25 year Japanese reanalysis).
- Variational bias correction (VarBC) of satellite radiance data (Dee, 2005; McNally et al., 2006) and improvements in radiosonde and surface pressure bias handling (Haimberger, 2007) and (Vasiljevic et al., 2006).
- More extensive use of radiances and an improved radiative transfer model.

4.1 Changes in data usage

Time series of ERA-Interim were examined and the results were compared to ERA-40 by taking a closer look at the overlap period 1989-2001.

The major changes in data usage for ERA-Interim are similar to the observed changes seen for ERA-40 over those twelve years, mainly because ERA-Interim primarily used the ERA-40 observation datasets as input data. Data usage changes during this time period are:

- 1995: Less available and used data due to the termination of the NCEP dataset (see Table 1).
- 1998: Less available and used ship and buoy data due to the termination of the NCAR COADS dataset (see Table 1).

The only observing system showing completely different behaviour of data usage in ERA-Interim compared to ERA-40 is the relative humidity from SYNOP observations. Fig. 10 shows the data usage for ERA-40 (top) and ERA-Interim (bottom) for 1989-2001. Focusing on the used data the big difference is the annual cycle in the ERA-Interim dataset compared to no seasonal difference in the data count in ERA-40. This is due to a change in the use of SYNOP humidity observations that occured in the operational system in September 2004 and was applied to the ERA-Interim assimilation. With the effect from cycle 28r3 SYNOP humidity data was blacklisted in operations at local nighttime. This blacklisting leads to an annual cycle, due to more observations used in the Northern Hemisphere summer because of the unequal distribution of the SYNOP observation network (with more observations on the Northern Hemisphere than on the Southern Hemisphere). The gap in VarQC in the ERA-40 dataset has already been discussed above. The reason for no VarQC visible in this dataset is that it got switched off for SYNOP two metre relative humidity in the operational system in early 2003 (see chapter 5) by mistake. Therefore the cycle used for ERA-Interim does not include any VarQC for this data type.

4.2 Performance of VarQC and other quality control aspects

In ERA-Interim the total data count has not changed significantly compared to ERA-40, however differences to ERA-40 in VarQC and first-guess rejections can be seen in ERA-Interim.

- When upper-air data of both datasets are compared a difference in the behaviour of VarQC can be noticed. ERA-Interim rejects more data in the VarQC than ERA-40. Nevertheless both datasets show a downward trend in VarQC rejections over the investigated period (not shown).

- Surface pressure observations (especially over land) show quite a different behaviour of VarQC and first-guess rejections in ERA-Interim compared to ERA-40. This can be illustrated by plotting the data count as percentage of the used data for both datasets (Fig. 11). Concentrating on the performance of VarQC first, more VarQC rejections can be seen in the ERA-40 dataset. By taking a closer look another difference can be spotted quite clearly: There are more first-guess rejections and blacklistings in the ERA-Interim assimilation and a smaller fraction of the available data is used. However comparing numbers (not shown) almost the same amount is used in both assimilations. This is because more observations are available in ERA-Interim.

  It is not clear where this additional data comes from and why a larger fraction is rejected or blacklisted. This needs further investigation.

- Another interesting fact related to surface pressure observations is that with the change in the number of total and used data in 1995 (see Fig. 3) the behaviour of the VarQC changes (Fig. 11). This change is not only due to the change in the data count since it can be seen in the time series, plotted as percentages of the used data, as well. There is not only a jump in VarQC in 1995, more interestingly the structure changes from a clearly annual cycle with more rejections in northern wintertime to a more random structure (taking the logarithmic scaling into account, less data count should show any structure even clearer and with a larger amplitude). This might be due to the different distribution of the observations before and after 1995.

  Comparing the distribution for one day (two cycles) for ERA-Interim in 1994 to the same day in 1995 some differences can be spotted. Fig. 12 shows the distribution of the data rejected by the VarQC on the 1st of February in 1994 (left) and 1995 (right). As seen in the time series there is less data rejected by the VarQC in 1995 than in 1994. Also the rejections seem to be more equally distributed (concerning the Northern and Southern Hemisphere) in 1995. In 1994 most of the rejections occur over Asia. Knowing that the total amount of data changed between 1994 and 1995 a closer look can be taken on the distribution of the used data for the same day (Fig. 13). At first the distribution seems quite similar in both years but looking closer reveals some differences. The distribution is almost similar when we look at the Southern Hemisphere but quite different (concerning the scale) on the Northern Hemisphere. For a closer look the data count difference of Fig. 13 is plotted in Fig. 14. The left picture shows the positive values (more data in 1994 than in 1995) and the right picture the negative values (more data in 1995 than in 1994) of this difference. An unequal distribution of positive and negative values is clearly visible. There are less VarQC rejections over Japan, middle Asia and over North-America leading to a more equal distribution of data counts on the Northern and Southern Hemisphere in 1995 than in 1994 (still the data distribution is not equal between Northern and Southern Hemisphere).

To illustrate the annual cycle in the VarQC rejections before 1995 seen in Fig. 11 the VarQC rejected data distribution was also plotted for 1994 and 1995 on the 1st of July (northern summer) in Fig. 15. Comparing those two month in 1994 (left pictures of Figs. 12 and 15) there is less VarQC rejection over Asia during the northern summer and more rejection over South-America, but also less rejection over Australia. So generally there are fewer rejections in the Northern Hemisphere in northern summer while rejections on the Southern Hemisphere are not changing that much. This leads to a clear annual cycle in VarQC before 1995. Starting with 1995 (right pictures of Figs. 12 and 15) the distribution of the rejections is similar in northern summer and winter, as seen in the time series of data usage. This can be shown more clearly when the difference of data distribution is plotted for the different seasons in 1994. Fig. 16 shows the difference between February and July. The positive values in the top left picture show more VarQC rejections in February 1994 while the negative values in the top right picture show more rejection in July 1994. The unequal distribution can be seen quite clearly. If we compare these plots to the same plots for the year 1995 (bottom left and right of Fig. 16) a much more equally distributed pattern can be seen.

4.3 Departure statistics

Looking at different departure statistics and comparing ERA-Interim to ERA-40 an improvement in the standard deviation can be seen in ERA-Interim. Also the bias in ERA-Interim compared to ERA-40 is reduced - especially for the innovation departures. So ERA-Interim is clearly performing better than ERA-40.

As an example the departure statistics for the surface pressure in the tropics can be compared (those showed a nice improvement over the ERA-40 period as presented in chapter 3.3). Fig. 17 shows in the top panel the ERA-40 statistics compared to the ERA-Interim statistics in the bottom panel. Especially the standard deviation of innovations, an important quality indicator, has improved in ERA-Interim compared to ERA-40. The larger difference between o-bg and o-an standard deviations for ERA-40, compared to ERA-Interim, is due to the 30% increased background error applied in ERA-40 (as discussed in Chapter 4).

Discuss Fig. 18 and Fig. 19 here.

In addition to the improvement also interesting changes in the departure statistics can be found. As an example Fig. 20 (bottom panel) shows the departure statistics of the 10 metre wind speed for drifting and moored buoys. Until the beginning of 1998 the NCAR COADS dataset is available and used. After that the count of used data drops significantly and the behaviour of the bias and the standard deviation changes. This might well be due to different distribution of the observations before and after 1998. To take a closer look at this, the data distribution of the used data of wind observations from drifting and moored buoys was plotted for one day (two cycles) in 1997 and 1998. Fig. 21 shows the 1st of February in 1997 (left) and 1998 (right). The first thing that can be seen is the difference in the total number of observations (the scale is the same in both pictures). Especially in the tropics more observations in general can be seen in 1998. Focusing on the Northern Hemisphere there are more observations in the Northern Pacific as well as in the Northern Atlantic in 1997. For a clearer picture the difference of the distribution is plotted as well. Fig. 22 shows the positive (left) and negative (right) difference in the number of used data and their distribution. The positive values show the amount of more data used in 1997 while the negative values show the additional data in 1998. As already discussed, more positive values can be found. Due to the fact that drifting and moored buoys are moving the positive/negative patterns are not completely uniform, even with the big data reduction at the beginning of 1998.

To see if there is any seasonal influence in the distribution of the observations also the 1st of July was plotted for those two years (Fig. 23 shows the 1st of July 1997 (left) and 1998 (right)). Comparing those distributions (also plotted on the same scale) we notice even more used data on the Northern Hemisphere during northern summer in 1997. In 1998 there is again less data, especially in the tropics.

5 Operational data

The operational assimilation System at ECMWF has improved considerably since 1997, resulting in significantly better forecast performance (Simmons and Hollingsworth, 2002). A 4D-Var assimilation system was introduced in November 1997 (Rabier et al., 2000). We will focus on the main changes during the period from 1999 to 2010. These are listed in Table 2.

Date	Year	Cycle	Changes
October	1999	21r4	At the start of this evaluation 6 hour 4D-Var was operational with a horizontal resolution of T319 and 60 vertical levels.
September	2000	23r1	The operational system changed to 12 hour 4D-Var.
November	2000	23r3	Horizontal resolution changed to T511 with a T159 inner loop.
January	2002	24r3	The observation time slot was reduced from 1 hour to 30 minutes.
January	2003	25r3_en	Multi-incremental (T95/T159) 4D-Var was introduced.
June	2004	28r2	The early delivery suite was introduced (Haseler, 2004).
February	2006	30r1	Change in horizontal resolution to T799 (T95/T255 inner loops) with 91 vertical levels. The top model level was raised from 1hPa to 0.01hPa.
June	2007	32r2	A third outer loop within 4D-Var (T95/T159/T255) was introduced.
September	2009	35r3	Huber Norm VarQC introduced for conventional data.
January	2010	36r1	Change in horizontal resolution to T1279.

Table 2: Changes in the operational ECMWF assimilation system

Data usage statistics and departure statistics of operational data were available from November 1999 to May 2010. This time period includes several cycle changes of the analysis system and the forecast model, as mentioned above, and a significant increase in the number of observations used. A number of changes of QC decisions have also occurred during this period. Some of these changes are described in this section. At the end of this section an intercomparison with ERA-Interim data for the period 1999-2006 is presented.

5.1 Changes in data usage

During this ten year period the following important changes in used observations occured:

- Upper-air data show an increasing trend of more data available and used throughout the whole period. This is not due to an increase in the number of radiosonde launches, but because modern radiosondes reach higher in the atmosphere before the balloon explodes, so more observations are available above the troposphere.
- 2008: From October to December American wind profiler were blacklisted due to not recognised new identifiers.
- 2008-05-20: Active assimilation of GRAS GPSRO bending angles.
- 2004-09-28: From cycle 28r3 onwards SYNOP humidity data was blacklisted at local nighttime (see Fig. 24). This leads to less used data and to an annual cycle in the used data count due to the distribution of the SYNOP observations (more observations on the Northern Hemisphere than on the Southern Hemisphere). This also led to a different treatment of SYNOP humidity data in ERA-Interim compared to ERA-40 as discussed in the previous section.
- 2003-04-29: The change to cycle 26r1 was a technical change only, however at this date more data in total can be seen for surface pressure observations over land (the number of used data for this observation type does not change, see Fig. 25). Part of the cycle change involves direct writing of observations into the Observation Data Base (ODB) so an assumption would be that more unused data is archived since that date.

5.2 Performance of VarQC and other quality control aspects

Changes in VarQC in the operational data assimilation system are typically related to cycle changes. Most of the mentioned changes can be found by looking, for example, at the data usage time series of radiosonde zonal wind (Fig. 26), radiosonde temperatures (Fig. 27), SYNOP ship 10 metre wind (Fig. 28) and time series for European profiler wind (Fig. 29).

- 2000-06-27: An increase in VarQC rejections in upper-air data.
  - Cycle 22r1: More satellite data were used and a new radiation scheme was introduced.
- 2000-09-12: An increase in VarQC rejections in all conventional datasets.
  - Cycle 23r1: Change from 6h 4D-Var to 12h 4D-Var.
- 2003-01-14: A decrease in VarQC rejections in all conventional datasets.
  - Cycle 25r4: Revised multi-incremental (T95/T159) 4D-Var algorithm. By mistake VarQC was switched off for SYNOP two metre relative humidity observations (see Fig.24).

- 2005-04-05: An increase in VarQC rejections of surface pressure data.
  - Cycle 29r1: A revised use of surface pressure observations: all surface pressure data are subject to an adaptive bias correction scheme and get a higher weight in the assimilation (reduced observation error applied).
- 2007-11-06: A decrease in rejections of profiler wind data until 2008-03-11 (cycle 32r3V).
  - Cycle 32r3: New radiosonde temperature and humidity bias correction.
- 2008-06-03: An increase in rejections for SYNOP ship wind observations.
  - Cycle 33r1: Use of four wind solutions for QuikSCAT. Extended coverage and increased resolution of limited area wave model. Improved shallow water physics and modified advection for ocean wave model.
- 2009-09-08: An increase in VarQC rejections of all conventional data sets.
  - Cycle 35r3: Huber Norm VarQC is introduced for all conventional datasets (excluding humidity observations). Within this change the VarQC weights are changed and the 25% margin does not characterise rejected data anymore. Data showing up as rejected might still influence the analysis with weights up to 25%. VarQC for SYNOP two metre relative humidity observations is switched on again.

Another interesting feature concerning data count can be found by looking at the data usage time series of ship surface pressure observations (Fig. 30). From 2005 to 2007 more observations tend to be first-guess rejected at the beginning of each month. Looking closer by zooming into the last two years and finally into three selected months (Fig. 31) shows that more first-guess rejection appears at the 1st of each month at 12:00. For a detailed description see Isaksen and Tavolato (2007).

This problem was forwarded to the operations department for further investigation. As a result the problem was identified to be due to an inconsistance in the observation handling software and it is fixed now. Results of the beginning of the next month (December 2007) shows this is the case (see Fig. 32). This is also clear from Fig. 30 that this problem now has been resolved.

5.3 Departure statistics

One example of a change in the bias correction can be seen in the time series of the departure statistics: Fig. 33 shows the departure statistics for 100hPa radiosonde temperature. A significant improvement of the bias can be seen clearly at the beginning of 2006, another one (which is smaller than the first one) at the end of the same year. These steps are related to the new variational bias correction method (VarBC, Dee (2005)). The VarBC method was first used in Feb 2006 to retune the bias predictors of the old bias correction method (Kelly, 2006) and shortly after VarBC completely replaced the old satellite bias correction method. The bias and standard deviations increased from September 2009, with the introduction of cycle 35r3. This is possibly due to an increase in the observation errors specified for radiosondes in the stratosphere.

- 2006-02-01 - Cycle 30r1: Change in bias correction scheme: Satellite radiance regression bias predictors were tuned. Change in horizontal resolution to T799 (T95/T255 inner loops) with 91 vertical levels.

- 2006-09-12 - Cycle 31r1: Satellite radiance regression bias predictors are retrieved from VarBC.

Note the gradual increase in data volume since 2005, predominantly due to the improved radiosonde technology which results in more measurements above 100hPa (see lower panel of Fig. 19). Also continuously more drifting buoys are operated and therefore data count increases (Fig. 20). Another interesting fact found by looking at the departure statistics of operational data is an increase in aircraft wind data (zonal wind at 850hPa) over the last years. Fig. 34 shows the continuing increase from the autumn of 2004 (this is more clearly seen here than in the data count time series due to a non-logarithmic scale). The wind data increased to twice the amount since early 2004. The bias got slightly smaller and the standard deviation improved as well over the last eight years. The data volumes have fallen slightly since 2008, most likely due to selective thinning by the data providers in order to reduce redundancy. The quality has improved in 2009.

5.4 Detailed intercomparison of operational data to ERA-Interim

An intercomparison between the ERA-Interim dataset and the operational assimilation system has been performed for the period 1999-2010. We expect the same amount of available data, however we would expect differences in data usage. ERA-Interim uses a fixed assimilation system (cycle 31r2, see Chapter 4) whereas during this period multiple cycle changes took place in operations (see table 2).

5.4.1 Data usage differences between ERA-Interim and operational data

Examples are given in Figs. 35 and 36. Fig. 35 shows the data usage plots for surface pressure observed by ships. The overall data count looks similar but differences can be seen in the behaviour of blacklisting, rejection and VarQC rejection. There is less blacklisting in ERA-Interim (top panel) and the VarQC time series looks constant over the whole period. In the operational assimilation system (bottom panel) changes in VarQC can be clearly seen at the beginning of 2003, 2005 and in 2009, as discussed in chapter 5. The Huber norm QC (Tavolato and Isaksen, 2010) introduced in September 2009 leads to an increase in the number of rejections. More observations are though given a weight between 0-25%, which is classified as rejected. A different VarQC behaviour is also present for the land surface pressure data (Fig. 36). Again changes in VarQC happening with model cycle changes can be seen in the operational assimilation system (top panel). This figure also shows a difference in total data count. ERA-Interim clearly shows an increase of all, used and rejected data at the beginning of 2002 (top panel). However no such change in data count can be found in the operational system. Further investigation of this difference is recommended. The impact of the Huber norm on rejections is also visible in Fig. 36.

5.4.2 Intercomparison of departure statistics

Another way to compare ERA-Interim and the operational system is to look at departure statistics. Figs. 37 and 38 show detailed comparisons of two different observing systems. Fig. 37 shows radiosonde temperature observations at 100hPa over the Northern Hemisphere (top panel: ERA-Interim, bottom panel: operations). Once again the reanalysis dataset seems to be much smoother. A big jump in the bias is visible in 2006 in operations when VarBC based predictor estimates were introduced, as discussed in chapter 5.3. Since VarBC is used for the whole period of ERA-Interim no jumps in the bias can be seen. Another interesting feature is visible in Fig. 38. Here statistics for aircraft wind observations at 850hPa over the Northern Hemisphere are shown. These time series show similar behaviour for the departure and bias statistics, but the data count of used data is different in those two assimilations. We recommend that this difference is further investigated.

6 Summary

This report has highlighted the changes in data availability, data usage and quality control in the 45 year ERA-40 reanalysis, the more recent ERA-Interim reanalysis (1989-2010) and the ECMWF operational assimilation system for the period 1999-2010. This report highlights assimilation features encountered from time series for various observing systems, with our attempts to explain the reasons for the changes. Overall changes in data usage for conventional data can be associated with improved quality of the observing system and improved assimilation systems. Examples for each of those categories are:

- The overall usage of more radiosonde data: due to improved radiosondes observations from high altitude passed the quality control and are used in the analysis. The improved analysis also results in more correct background statistics that will be in closer agreement with observations.
- Increase of aircraft observations: With the move from 3D-Var to 4D-Var more asynoptic observations are used by the assimilation system.

This report shows several examples of changes in the way conventional data is used by the three data assimilation systems discussed. But it is only possible to show a limited set of examples, so we have made two tables to summarize quality control decisions for all conventional data in 1995 and 2005. Table 3 compares the data usage of ERA-40 to ERA-Interim for the year 1995. The first three columns describe the observation type, the observed value and the pressure level of the observation. The following columns are all available observations, percentage of first-guess rejections, percentage of VarQC rejections and percentage of used observations. Note that these numbers do not have to add up to 100% since blacklisting and other forms of rejection are not considered in this table. The final columns give values for the approximate limits of first-guess and VarQC rejection in both datasets and the overall rejection ratio which is the sum of first-guess and VarQC rejections. Since ERA-Interim to a large extend used the observations from in ERA-40 it is not surprising that data counts are fairly similar.

Table 4 compares ERA-Interim and operational data usage for 2005. Since it is an reanalysis compared against the operational analysis the data usage differs significantly. One example is the use of land surface pressure observations, where almost 50% of the data is used by operations but just 20% within ERA-Interim. Another difference can be found looking at Japanese and American wind profiler observations close to the ground (900hPa - 1000hPa). These are blacklisted in operations and none of them are used, whereas ERA-Interim uses around 95% of those observations. The rejection limits seem to be similar, which is not surprising, since a 4D-Var data assimilation system is used for both data sets.

Finally we conclude that all the examinations made in this report benefitted from the use of the logarithmical scaled time series of the data usage. This is an excellent monitoring tool that should be used routinely to detect differences and changes in the data usage and quality.

Acknowledgements

We very much appreciate support from and discussions with many ECMWF collegues, with special thanks to Erik Andersson, Dick Dee, Sakari Uppala, Paul Poli and Elias Holm.

Acronyms

3D-Var	3 Dimensional Variational analysis
4D-Var	4 Dimensional Variational analysis
an	analysis
bg	background
COADS	Comprehensive Ocean Atmosphere Data Set
ECDAS	ECMWF Climate Data Assimilation System
ECMWF	European Centre for Medium-Range Weather Forecast
ERA	ECMWF Re-Analysis
ERA-40	A 45-year ERA from September 1957 to August 2002
ERA-Interim	Currently running ERA from 1989 onwards
fg	first-guess
FGGE	First GARP Global Experiment
GARP	Global Atmosphere Research Programme
GATE	GARP Atlantic Tropical Experiment
GTS	Global Telecommunication System
IFS	Integrated Forecasting System
IGY	International Geophysical Year
JRA	Japanese Re-Analysis
M1QN3	Quasi-Newton minimization technique
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
obs	observation
ODB	Observation Data Base
QC	Quality Control
SYNOP	Surface Synoptic Observation
VarBC	Variational Bias Correction
VarQC	Variational Quality Control
VTPR	Vertical Temperature Profile Radiometer

References

Andersson, E., Cardinali, C., Fisher, M., Hólm, E., Isaksen, L., Trémolet, Y., and Hollingsworth, A. (2004). Developments in ECMWF's 4D-Var system. In *Symposium on Forecasting the Weather and Climate of the Atmosphere and Ocean*, 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction. American Meteorological Society. Paper J1.4.

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasilijevic, D., Brancovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., Lanzinger, A., Miller, M., Rabier, F., Simmons, A., Strauss, B., Thépaut, J.-N., and Viterbo, P. (1998). The ECMWF implementation of three-dimensional variational assimilation (3D-Var). III: Experimental results. *Quart. J. Roy. Meteor. Soc.*, 124:1831–1860.

Andersson, E. and Järvinen, H. (1999). Variational quality control. *Quart. J. Roy. Meteor. Soc.*, 125:697–722.

Bengtsson, L., Kanamitsu, M., Kållberg, P., and Uppala, S. (1982). FGGE 4-dimensional data assimilation at ECMWF. *Bull. Am. Meteorol. Soc.*, 63:29–43.

Dee, D. P. (2005). Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, 131:3323–3343.

Gilbert, J.-C. and Lemaréchal, C. (1989). Some numerical experiments with variable-storage quasi-Newton algorithms. *Mathematical Programming*, B25:407–435.

Haimberger, L. (2007). Homogenization of radiosonde temperature time series using innovation statistics. *J. o. Climate*, 20(7):1377–1403.

Haseler, J. (2004). Early-delivery suite. *Technical Memorandum, ECMWF*, 454.

Isaksen, L. and Tavolato, C. (2007). Erroneous rejection of ship observations. *Research Department Memorandum, ECMWF*. File: R43.2/LI/0790.

Kelly, G. (2006). Retuning of the four predictor satellite radiance bias model using a long VarBC data assimilation. *Research Department Memorandum, ECMWF*.

McNally, T., Auligné, T., Dee, D., and Kelly, G. (2006). A variational approach to satellite bias correction. *ECMWF Newsletter*, 107:7–17.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, 126:1148–1170.

Simmons, A., Uppala, S., Dee, D., and Kobayashi, S. (2007). ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, 110:25–35.

Simmons, A. J. and Hollingsworth, A. (2002). Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 128:647–677.

Tavolato, C. and Isaksen, L. (2010). Huber norm quality control in the IFS. *ECMWF Newsletter*, 122:27–31.

Uppala, S., , Kållberg, P., Hernandez, A., Saarinen, S., Fiorino, M., Li, X., Onogi, K., Andrae, U., and Bechtold, V. D. C. (2004). ERA-40: ECMWF 45-year reanalysis of the global atmosphere and surface conditions 1957-2002. *ECMWF Newsletter*, 101:2–21.

Uppala, S. (2007). From ERA-15 to ERA-40 and ERA-interim. In *Procedings of the ECMWF/GEO Workshop on Atmospheric Re-analysis, 19-22 June 2006*, pages 17–22, RG2 9AX Shinfield Park, Reading, U.K. ECMWF.

Uppala, S., Dee, D., Kobayashi, S., Berrisford, P., and Simmons, A. (2008). Towards a climate data assimilation system: status update of ERA-interim. *ECMWF Newsletter*, 115:12–18.

Uppala, S., Kållberg, P., Simmons, A., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R., Andersson, E., Arpe, K., Balmaseda, M., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B., Isaksen, L., Janssen, P., Jenne, R., McNally, A., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N., Saunders, R., Simon, P., Sterl, A., Trenberth, K., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. (2005). The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, 131:2961–3012.

Vasiljevic, D., Andersson, E., Isaksen, L., and Garcia-Mendez, A. (2006). Surface pressure bias correction in data assimilation. *ECMWF Newsletter*, 108:20–27.

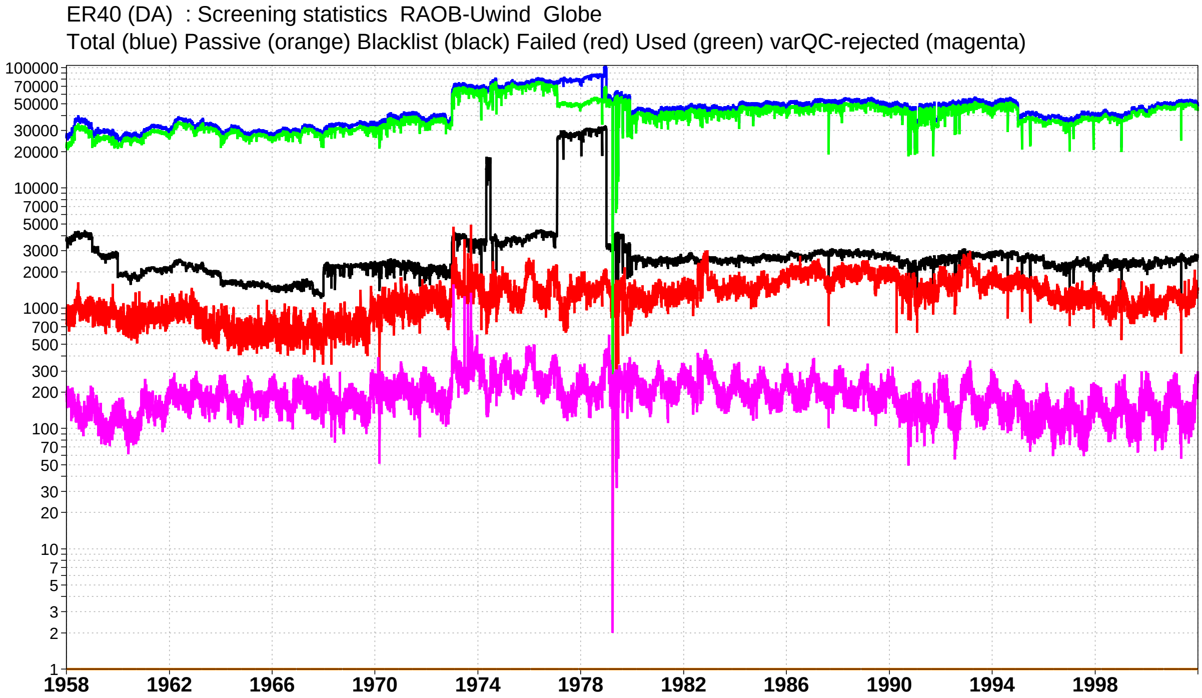


Figure 2: Time series of data usage for radiosonde zonal wind observations (all levels) from ERA-40 data.

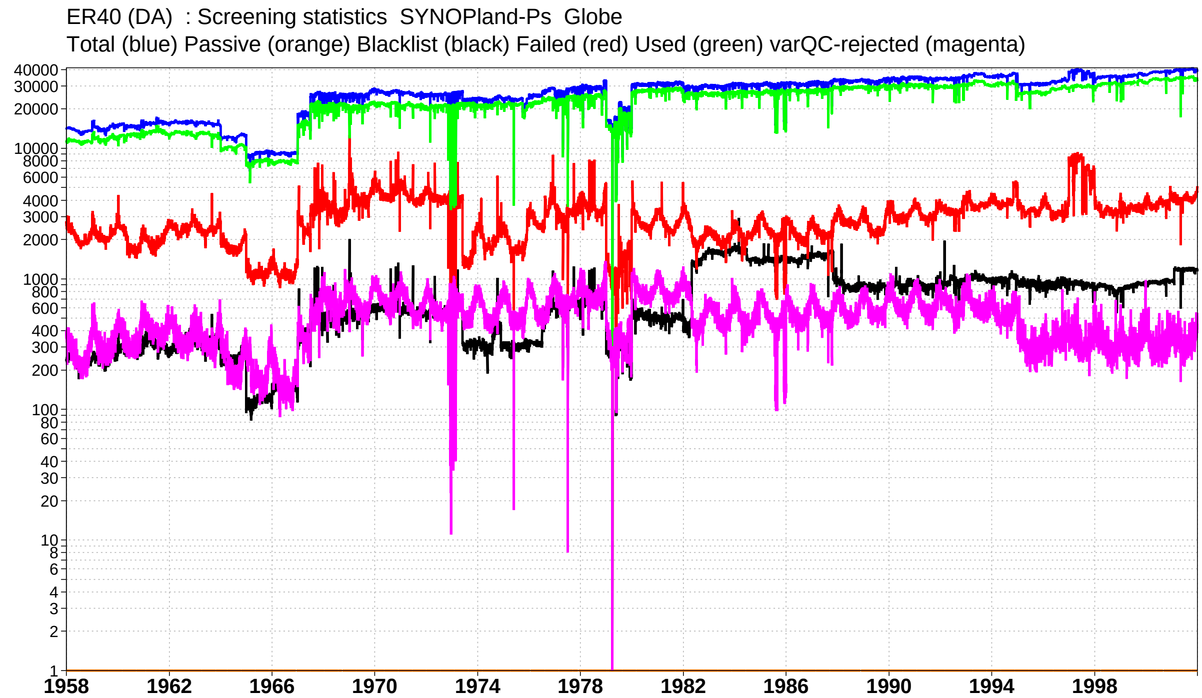


Figure 3: Time series of data usage for SYNOP surface pressure observations over land from ERA-40 data.

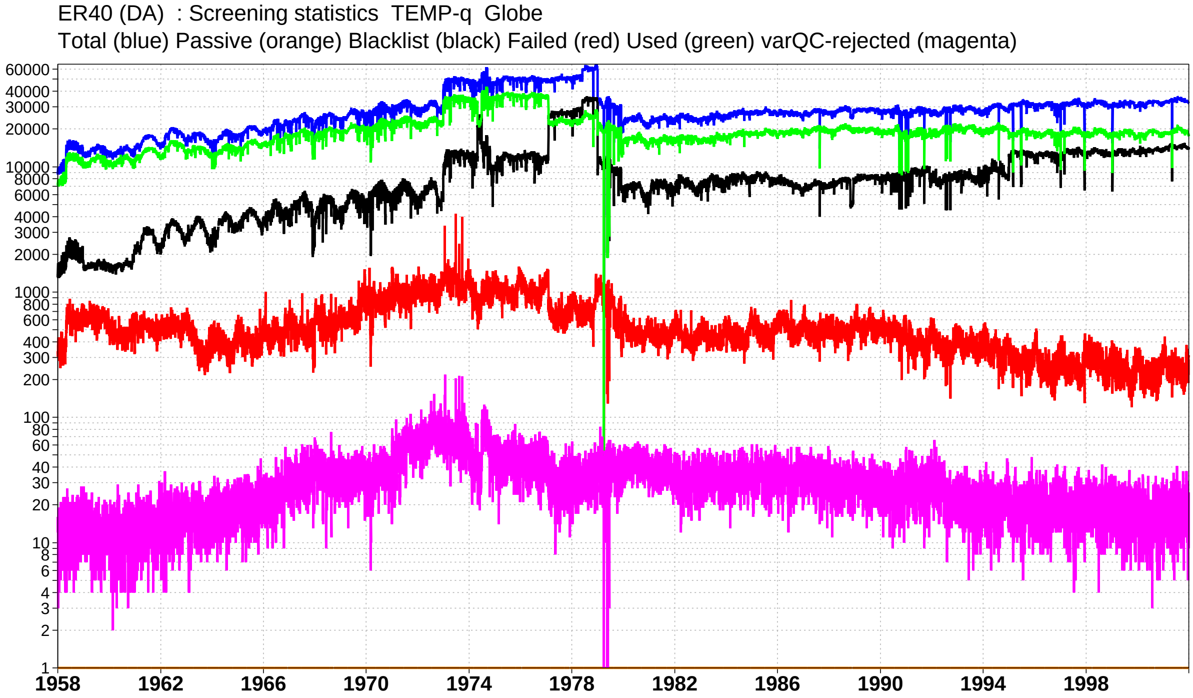


Figure 4: Time series of data usage for TEMP humidity observations (all levels) from ERA-40 data.

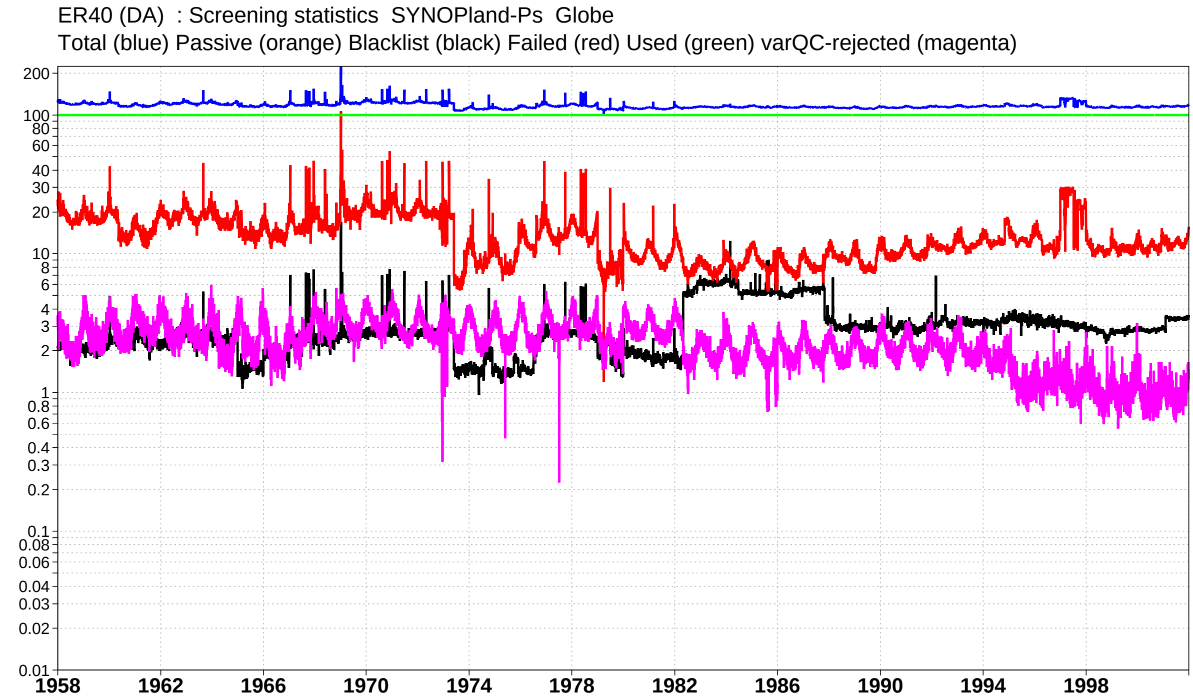


Figure 5: Time series of data usage for SYNOP surface pressure observations over land from ERA-40 data; data counts are plotted as percentage of used data.

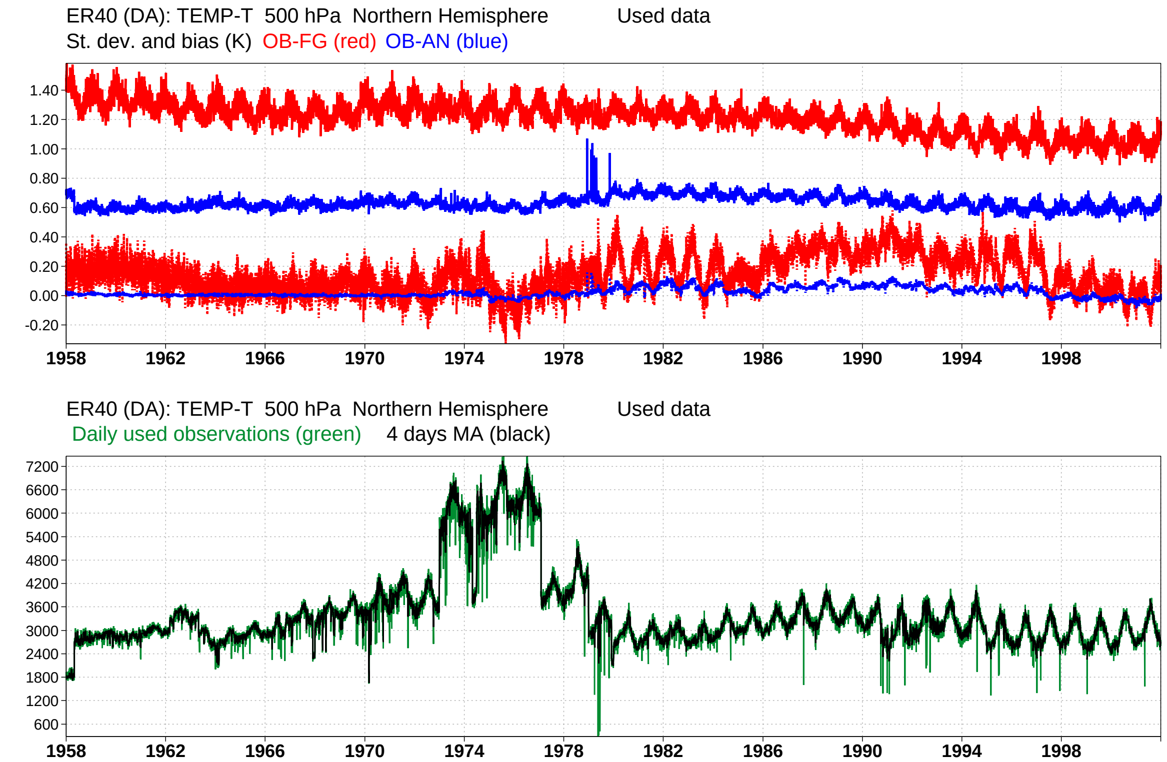


Figure 6: Time series of ERA-40 background and analysis departures for radiosonde temperature at 500hPa over the Northern Hemisphere. The lower panel shows used data counts.

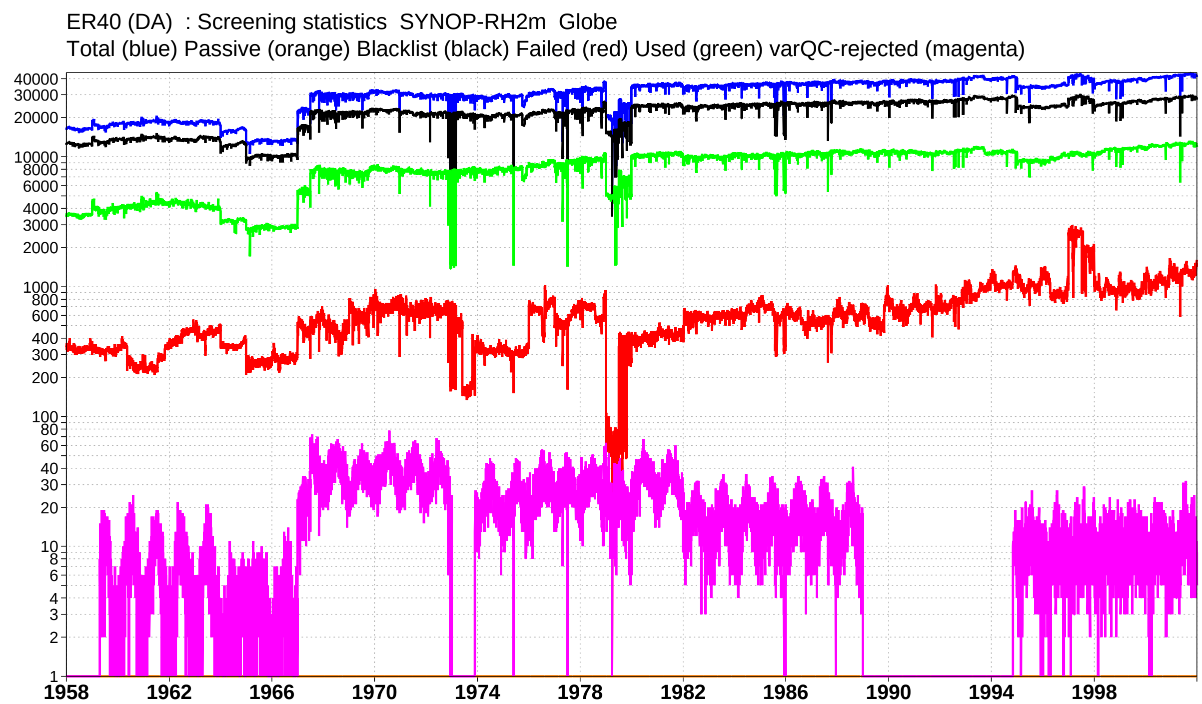


Figure 7: Time series of data usage for SYNOP relative humidity observations for ERA-40.

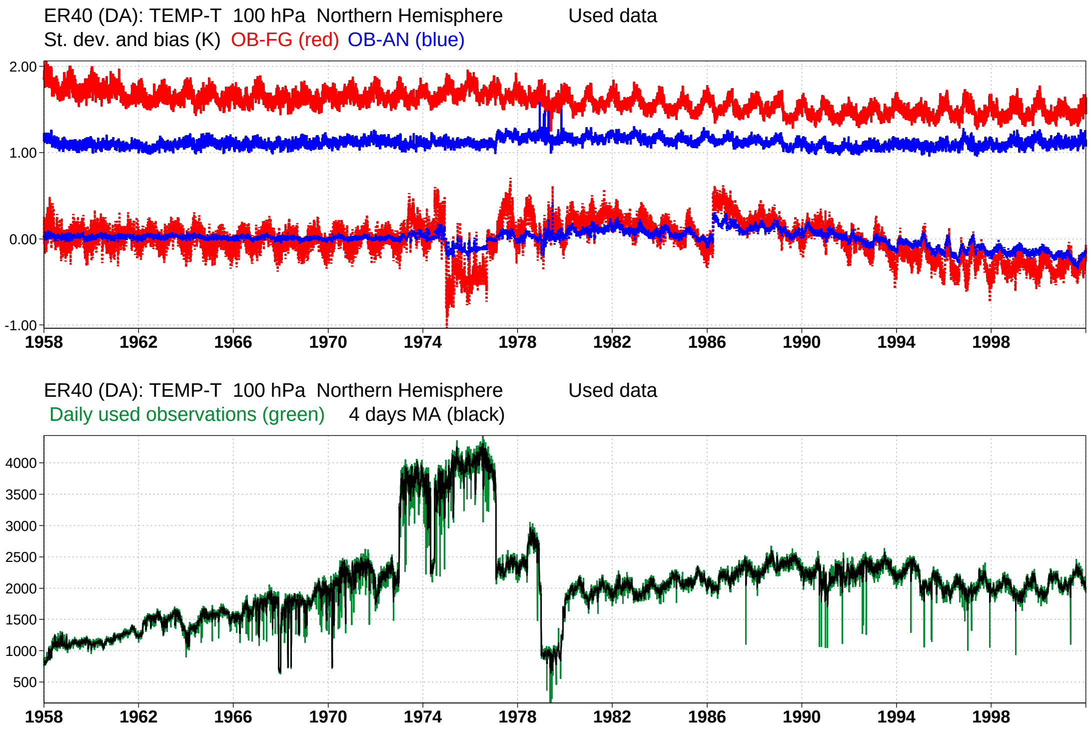


Figure 8: Departure statistics for radiosonde temperature at 100hPa on the Northern Hemisphere for ERA-40. The lower panel shows the used data counts.

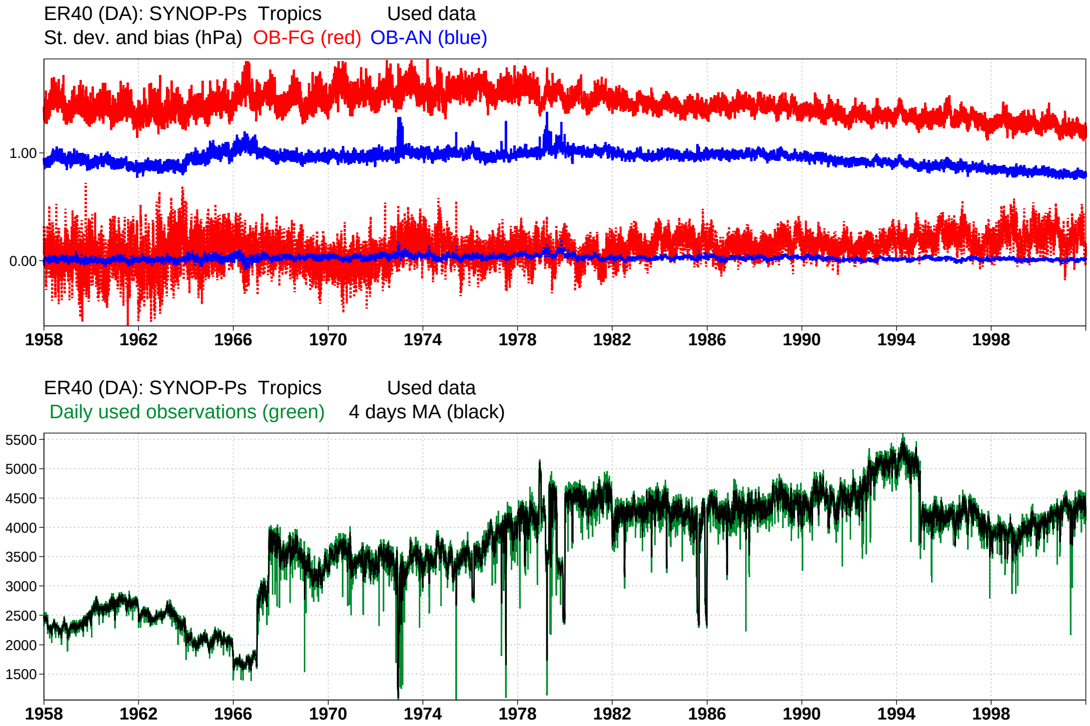


Figure 9: Departure statistics for SYNOP surface pressure in the tropics for the ERA-40 dataset. The lower panel shows used data counts.

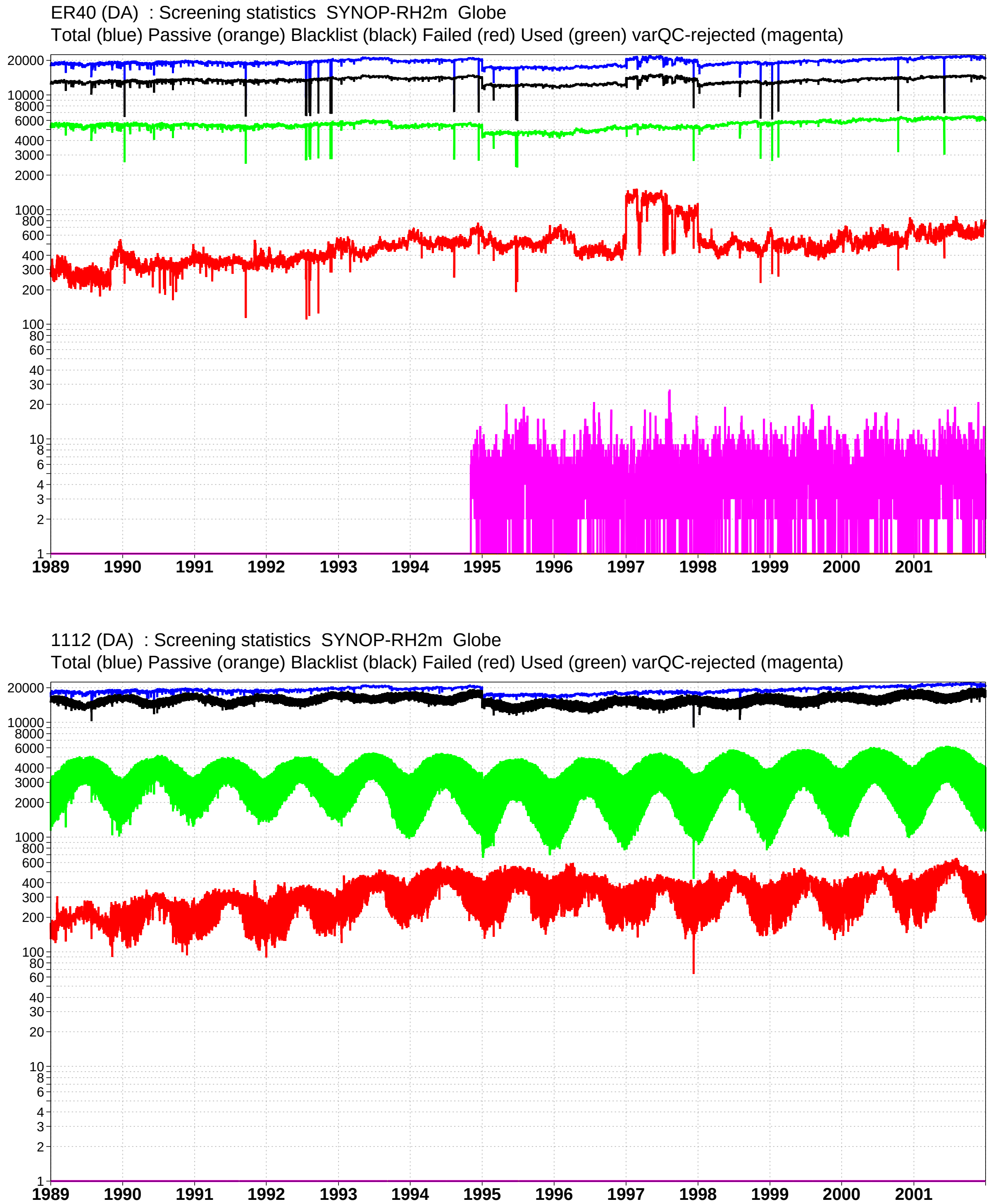


Figure 10: Time series of data usage for SYNOP relative humidity observations from reanalyses data. Top panel: ERA-40, Bottom panel: ERA-Interim

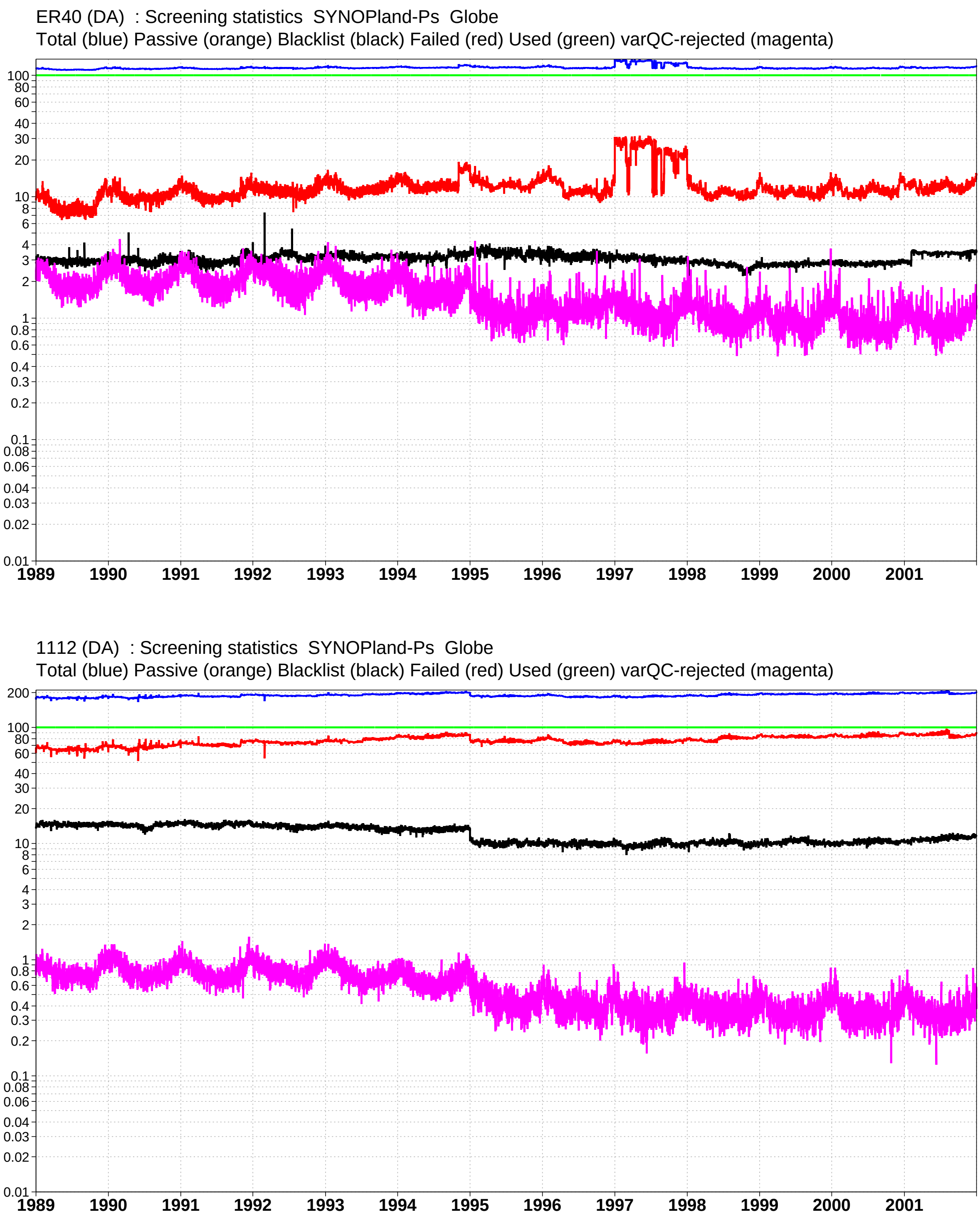


Figure 11: Time series of data usage for SYNOP land surface pressure observations from reanalysis data (Top panel: ERA-40, Bottom panel:ERA-Interim). Data counts are shown as percentage of the used data.

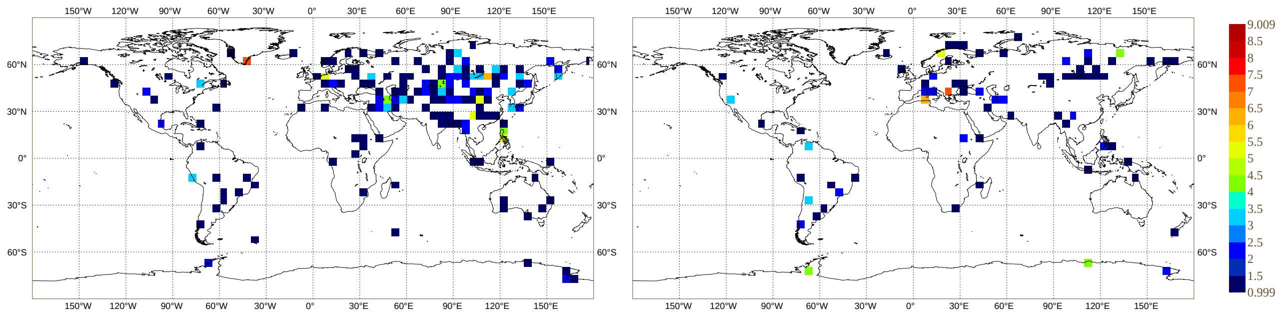


Figure 12: Data distribution for VarQC rejected SYNOP surface pressure observations over land on the first of February 1994 (left) and 1995 (right) of ERA-Interim data.

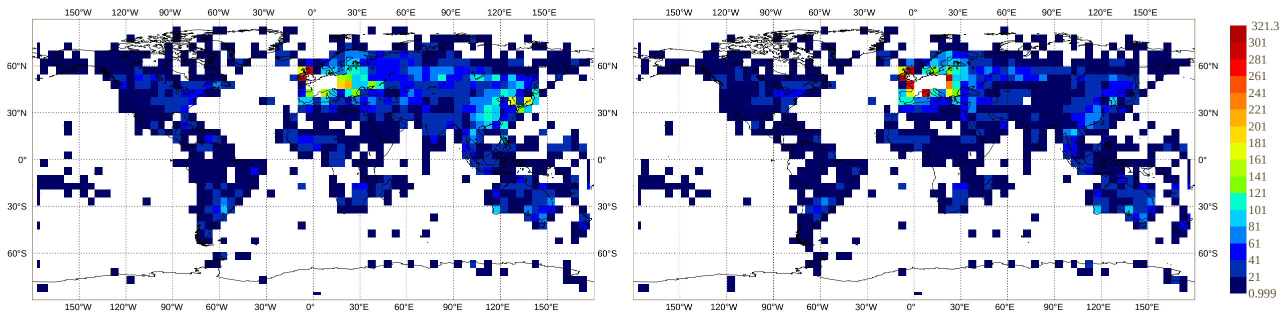


Figure 13: Data distribution for used SYNOP surface pressure observations over land on the first of February 1994 (left) and 1995 (right) of ERA-Interim data. White areas over Europe indicate an overshooting of the scale due to a high density of data and not a data sparse area.

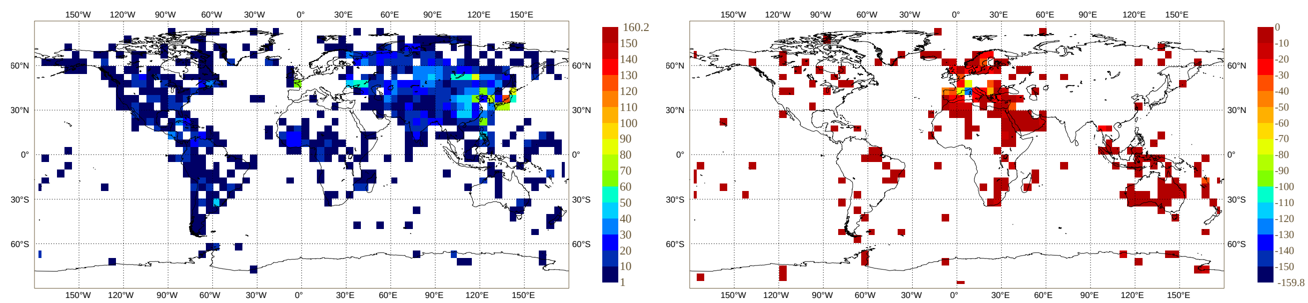


Figure 14: Difference of the data distribution in figure 13. Left: positive difference showing the additional data in 1994, right: negative difference showing the additional data in 1995.

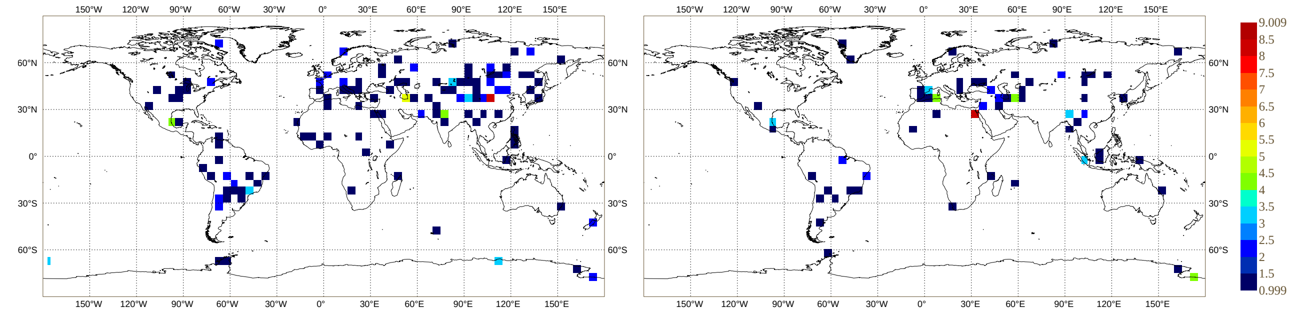


Figure 15: Data distribution for VarQC rejected SYNOP surface pressure observations over land on the 1. July 1994 (left) and 1995 (right) of ERA-Interim data.

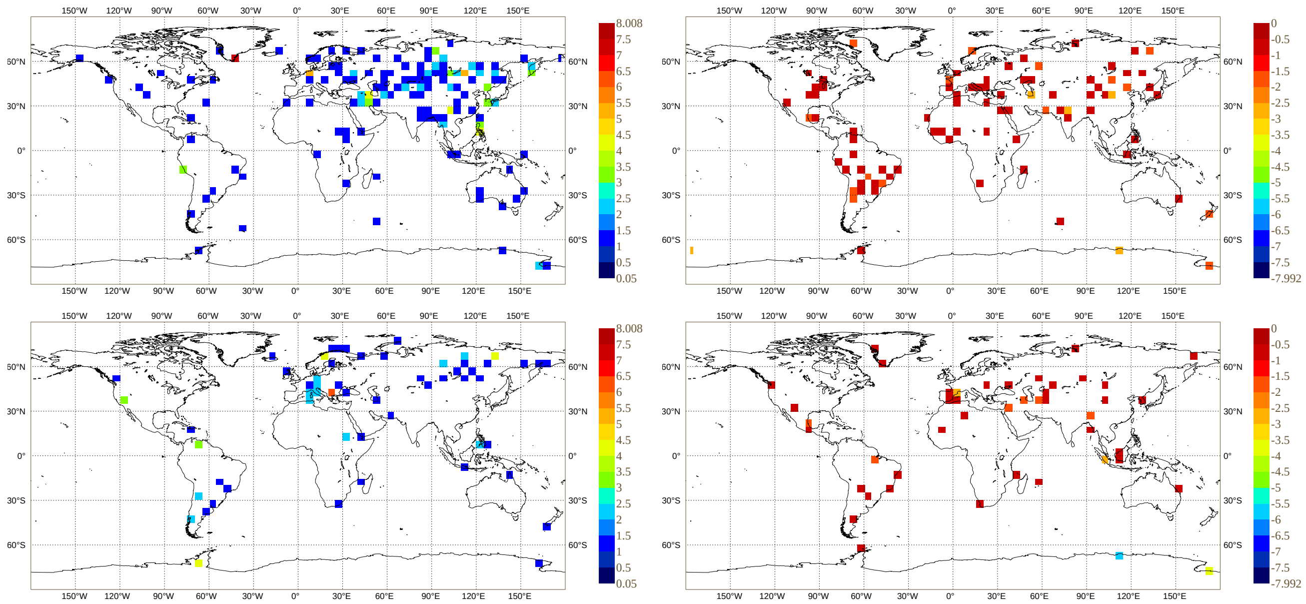


Figure 16: Top: Difference of the VarQC rejected data distribution between February and July 1994. Left: additional rejected data in February 1994, right: additional rejected data in July 1994. Bottom: Difference of the VarQC rejected data distribution between February and July 1995. Left: additional rejected data in February 1995, right: additional rejected data in July 1995.

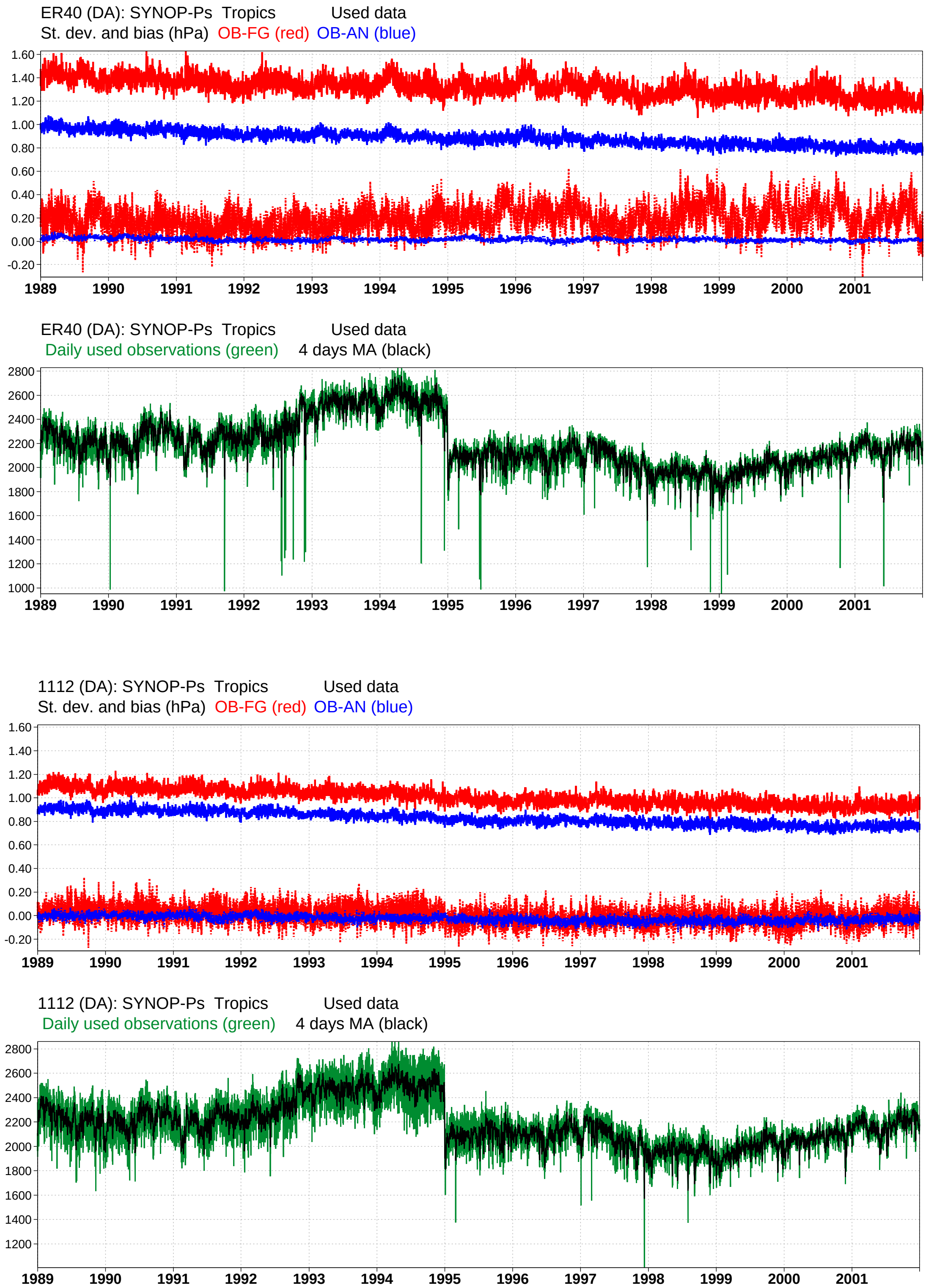


Figure 17: Time series of departure statistics for SYNOP surface pressure observations in the tropics from reanalysis data (Top panel: ERA-40, bottom panel: ERA-Interim). The lower panel shows the counts for used data during the time period (1989-2001).

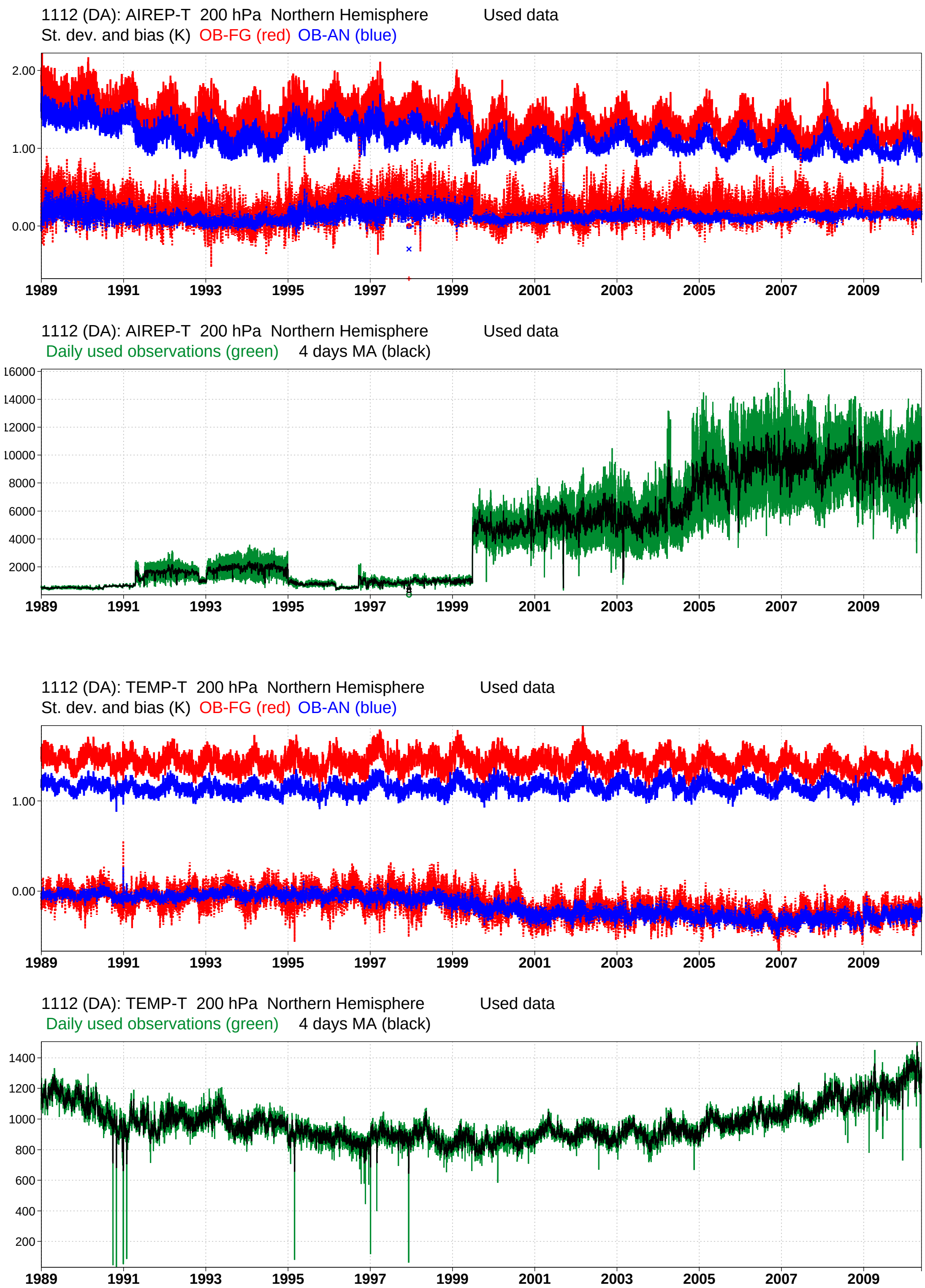


Figure 18: 1989-2010 time series of Northern hemisphere extra-tropics departure statistics from ERA-Interim data. Top panel: Aircraft Temperature departures near 200hPa. Bottom panel: Radiosonde temperature departures near 200hPa.

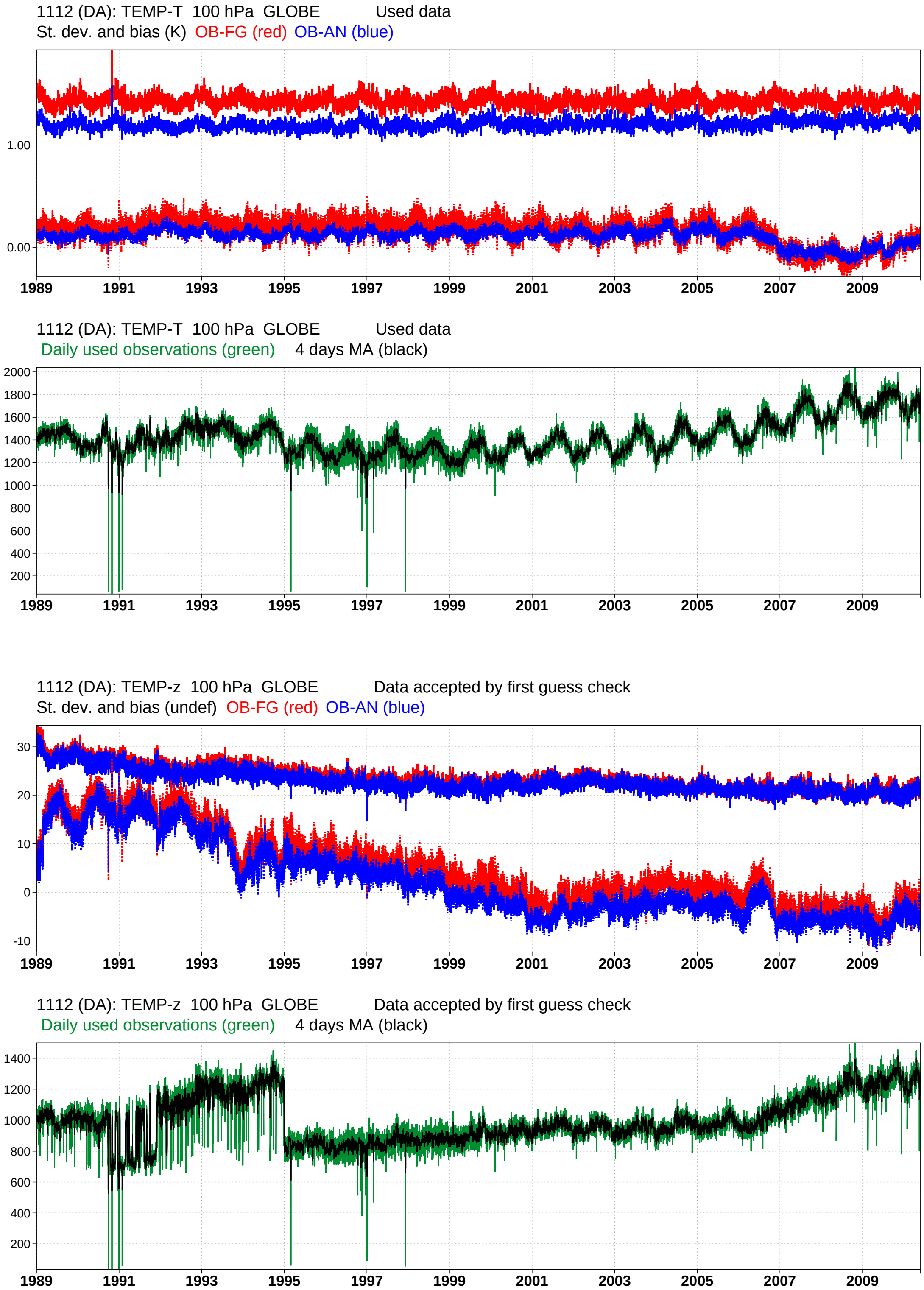


Figure 19: 1989-2010 time series of global departure statistics from ERA-Interim data. Top panel: Radiosonde temperature departures near 100hPa. Bottom panel: Radiosonde geopotential height departures at 100hPa. Note that radiosonde geopotential height data is passive. The dataset consists of observations that has passed the first guess quality control.

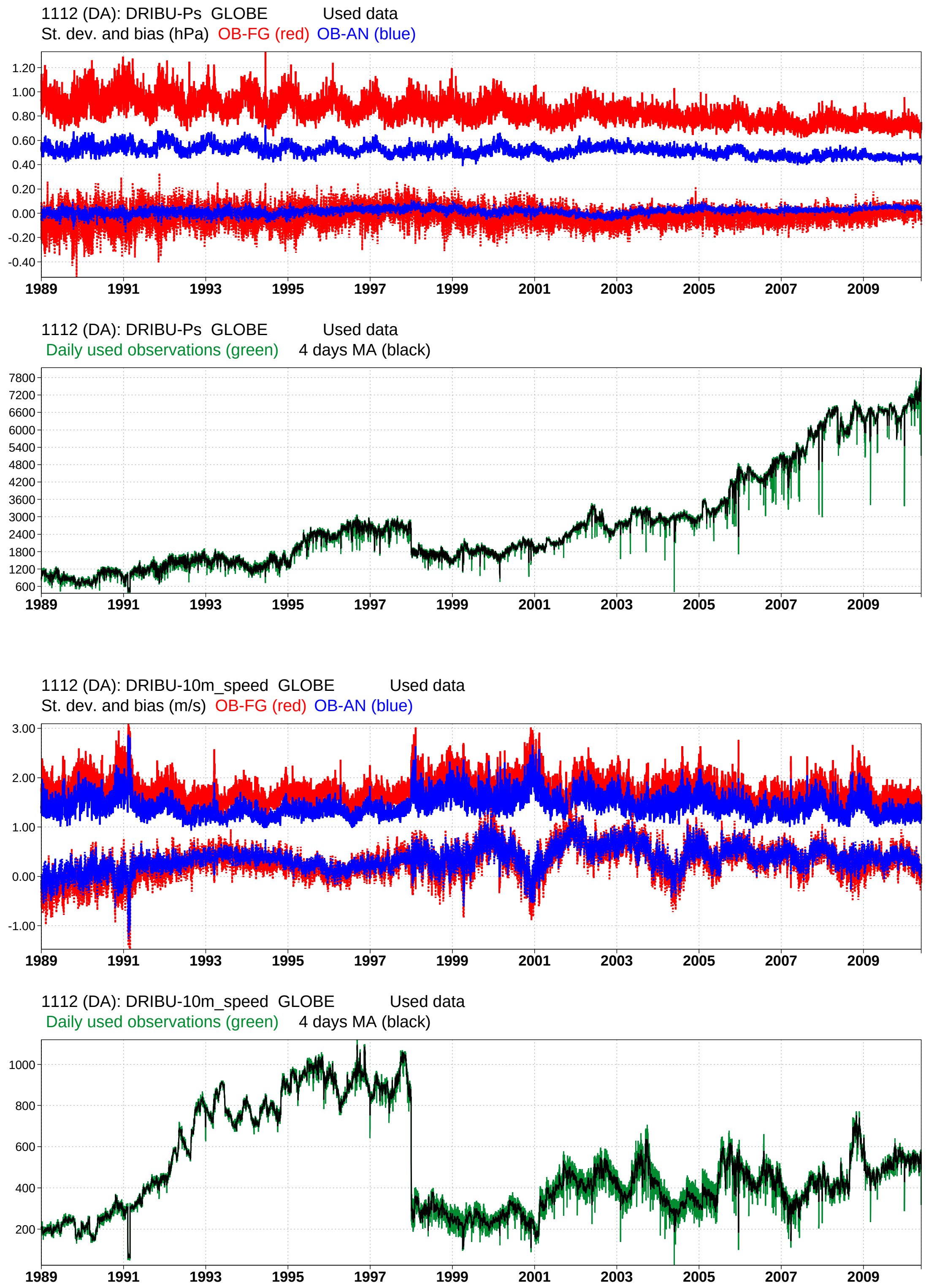


Figure 20: 1989-2010 time series of departure statistics from ERA-Interim data. Bottom panel: Surface pressure observations from drifting and moored buoys. Bottom panel: 10 metre wind speed from drifting and moored buoys.

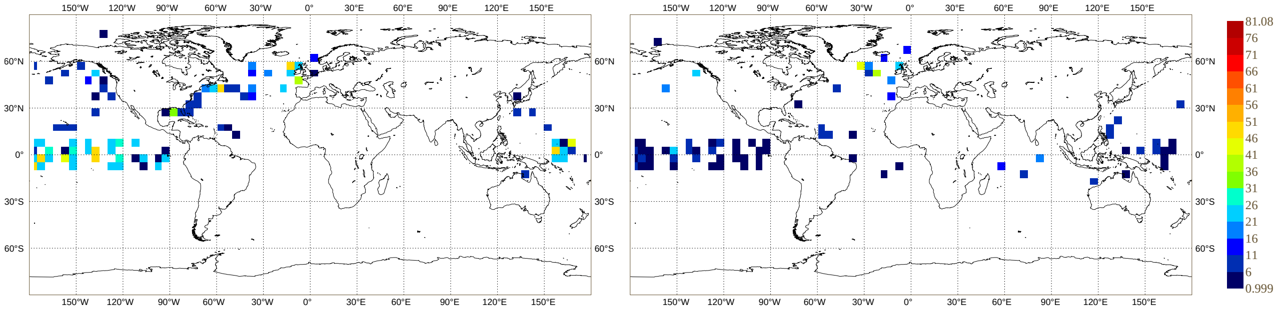


Figure 21: Data distribution for used drifting and moored buoys wind observations on the first of February 1997 (left) and 1998 (right) of ERA-Interim data.

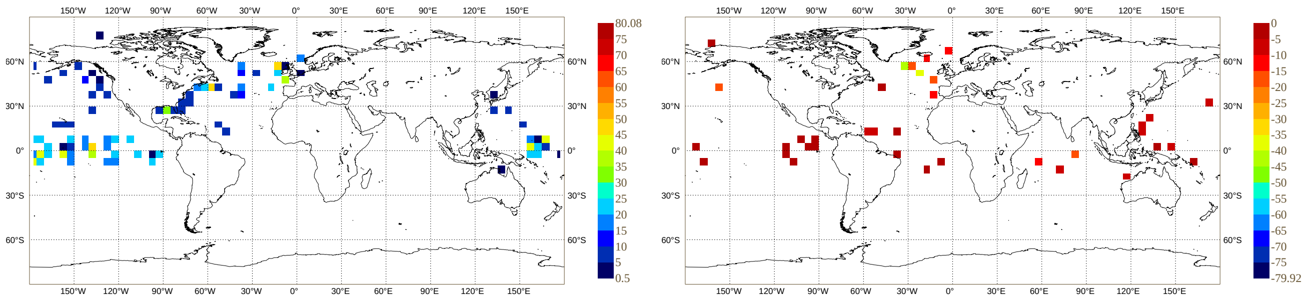


Figure 22: Difference of the data distribution in figure 21. Left: positive difference showing the additional data in 1997, right: negative difference showing the additional data in 1998.

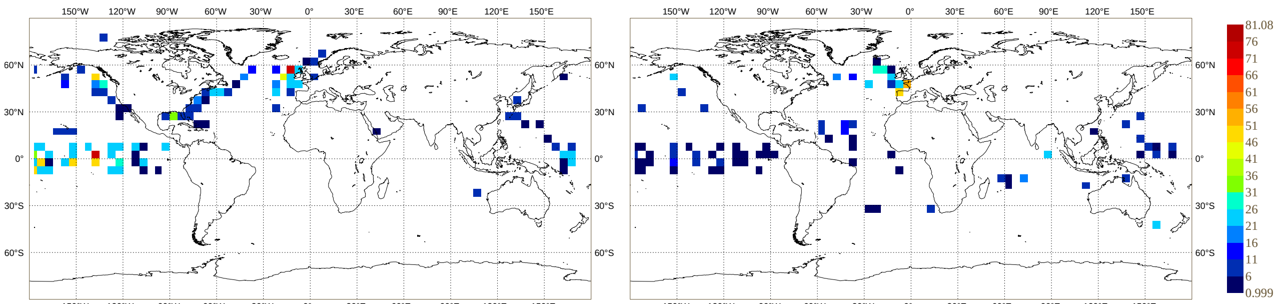


Figure 23: Data distribution for used drifting and moored buoys wind observations on the 1. July 1997 (left) and 1998 (right) of ERA-Interim data.

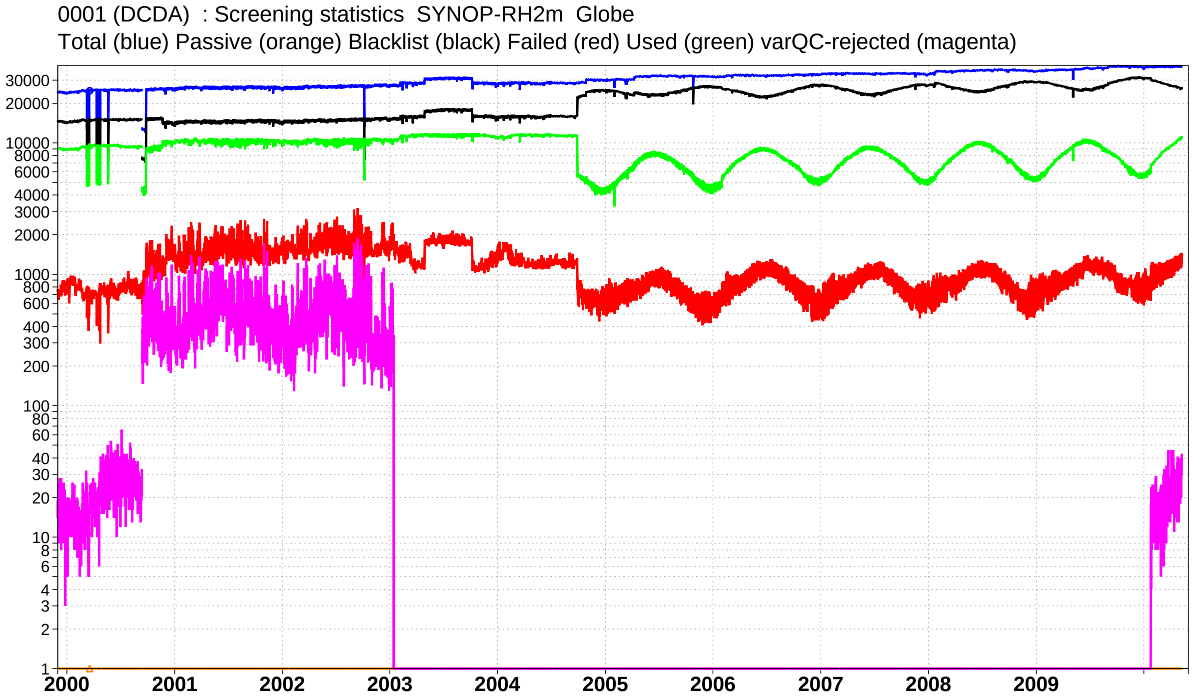


Figure 24: Time series of data usage for SYNOP relative humidity observations from operational data.

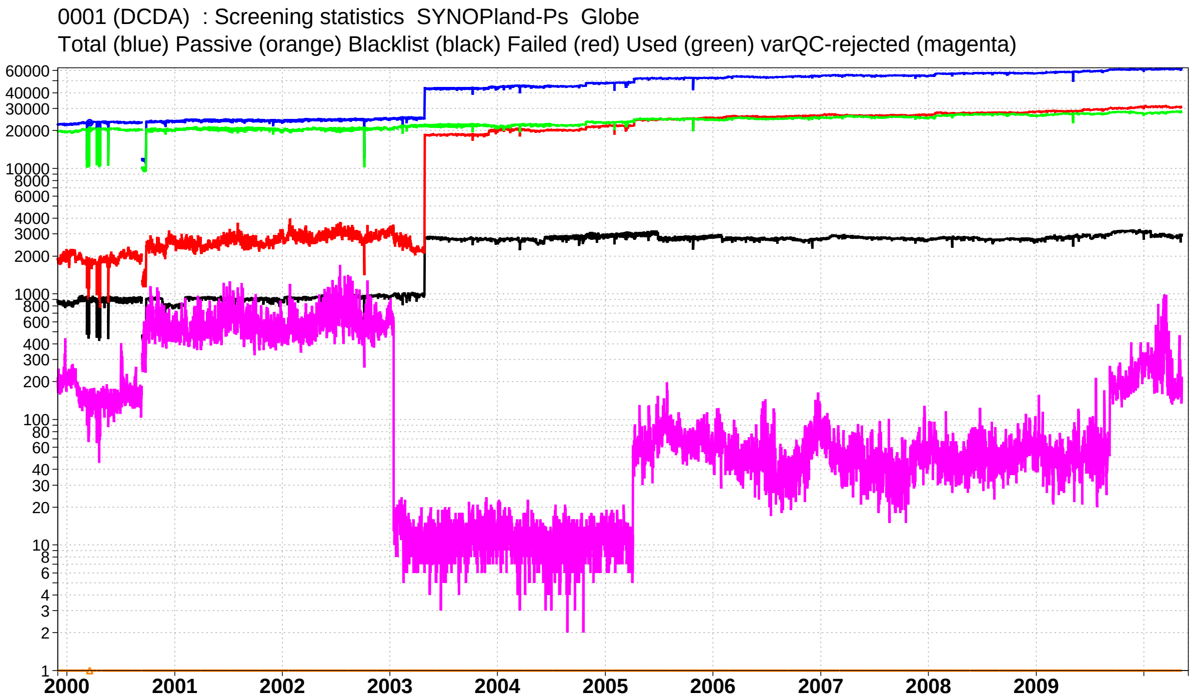


Figure 25: Time series of data usage for SYNOP surface pressure observations over land from operational data.

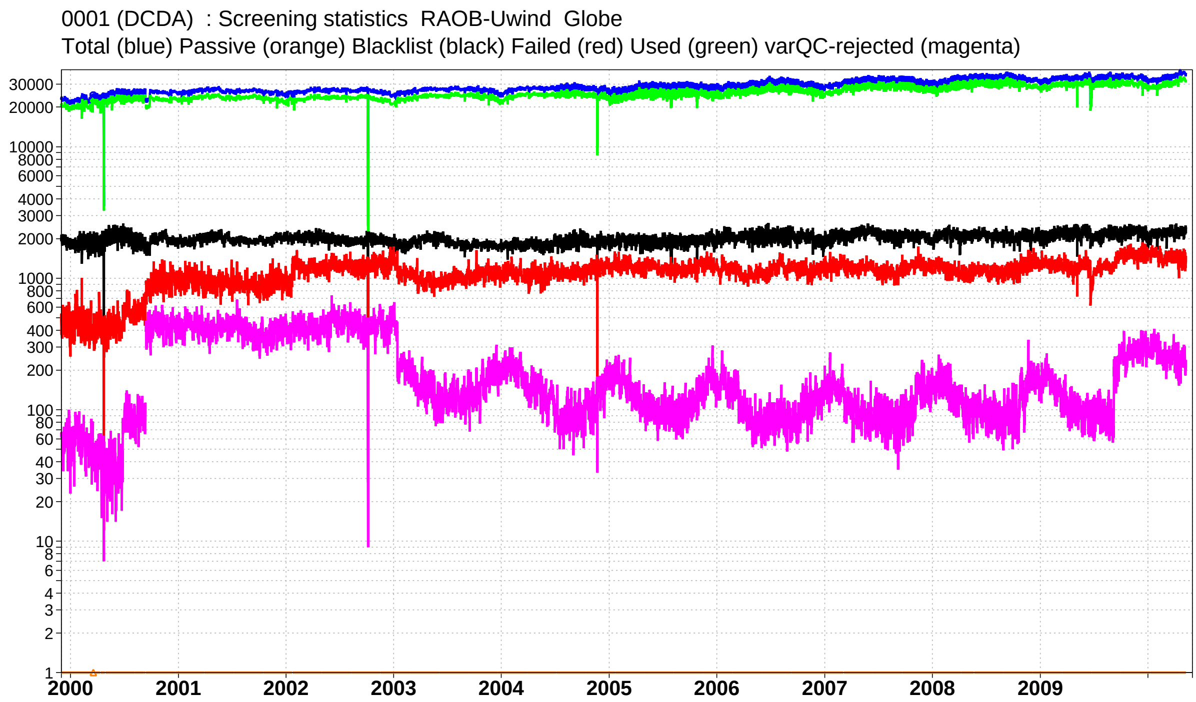


Figure 26: Time series of data usage for radiosonde and pilot zonal wind (all levels) observations from operational data.

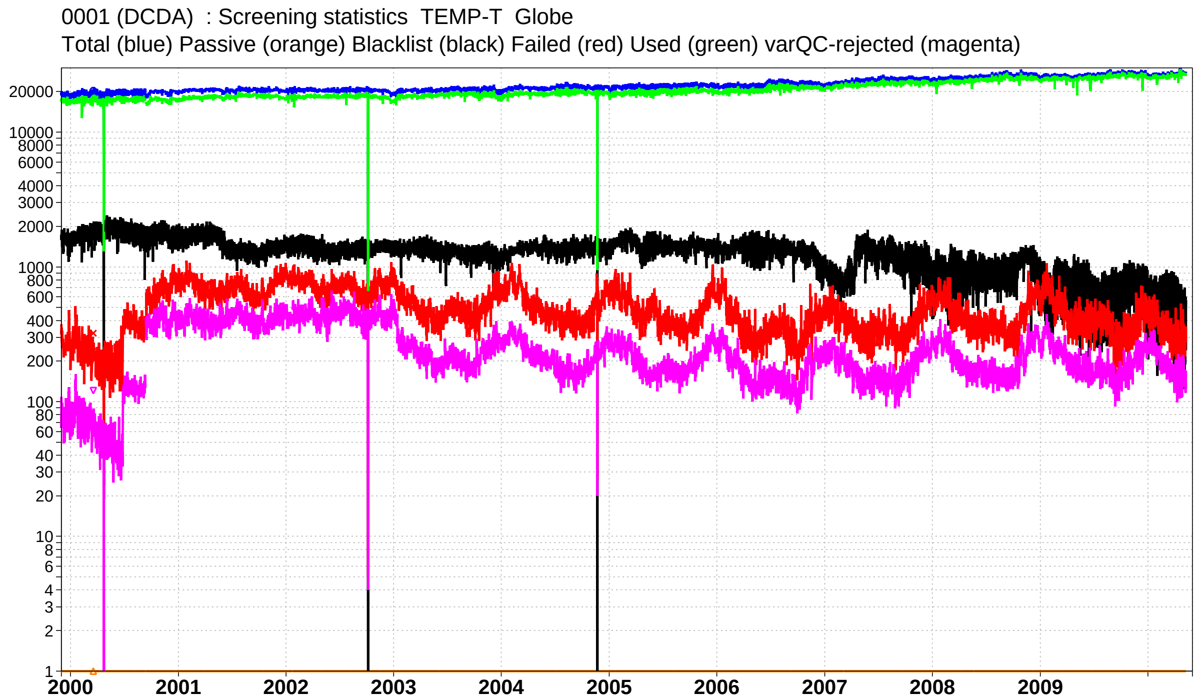
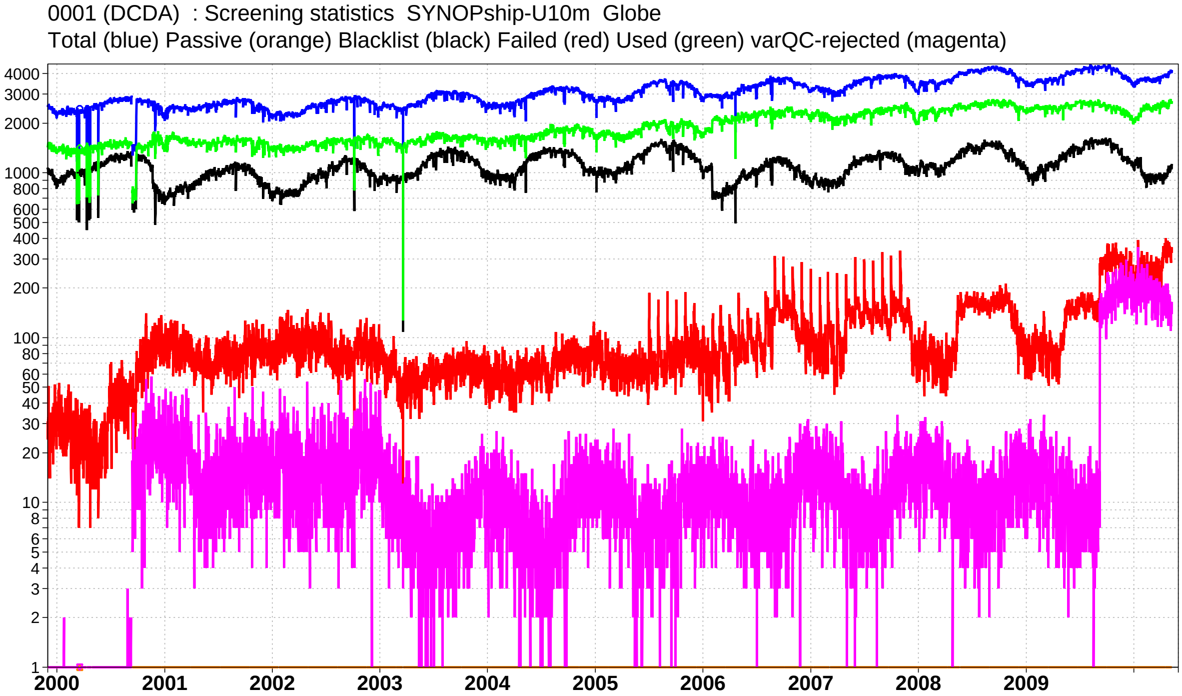


Figure 27: Time series of data usage for radiosonde temperature (all levels) observations from operational data.



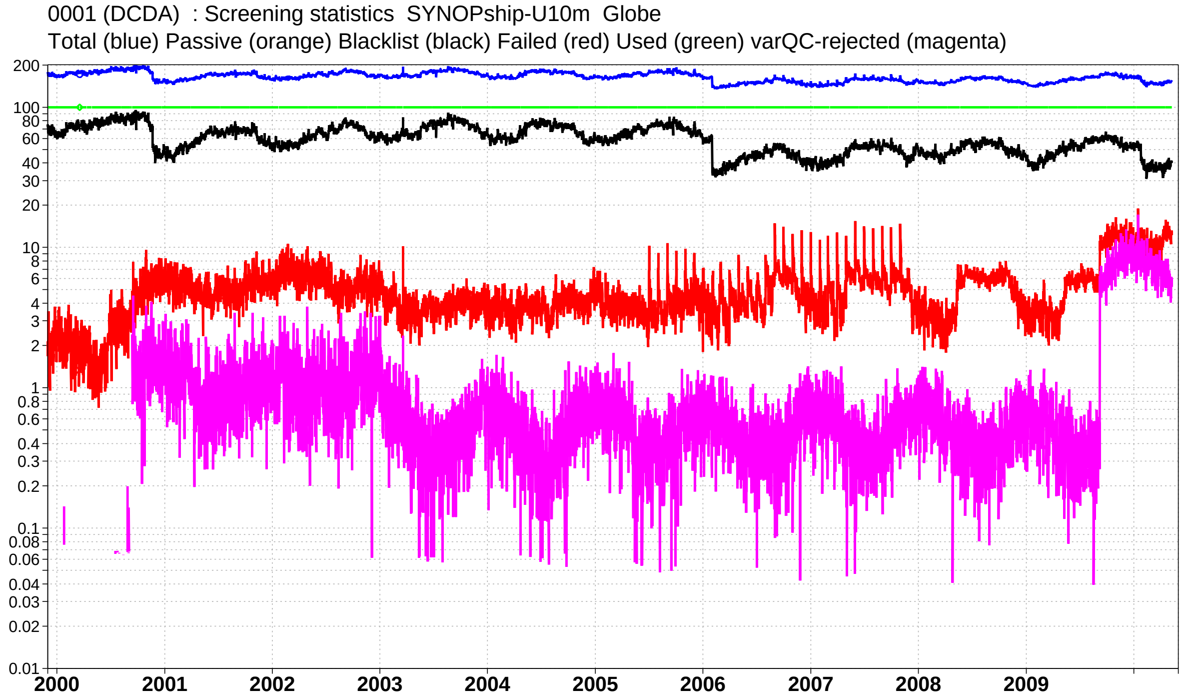
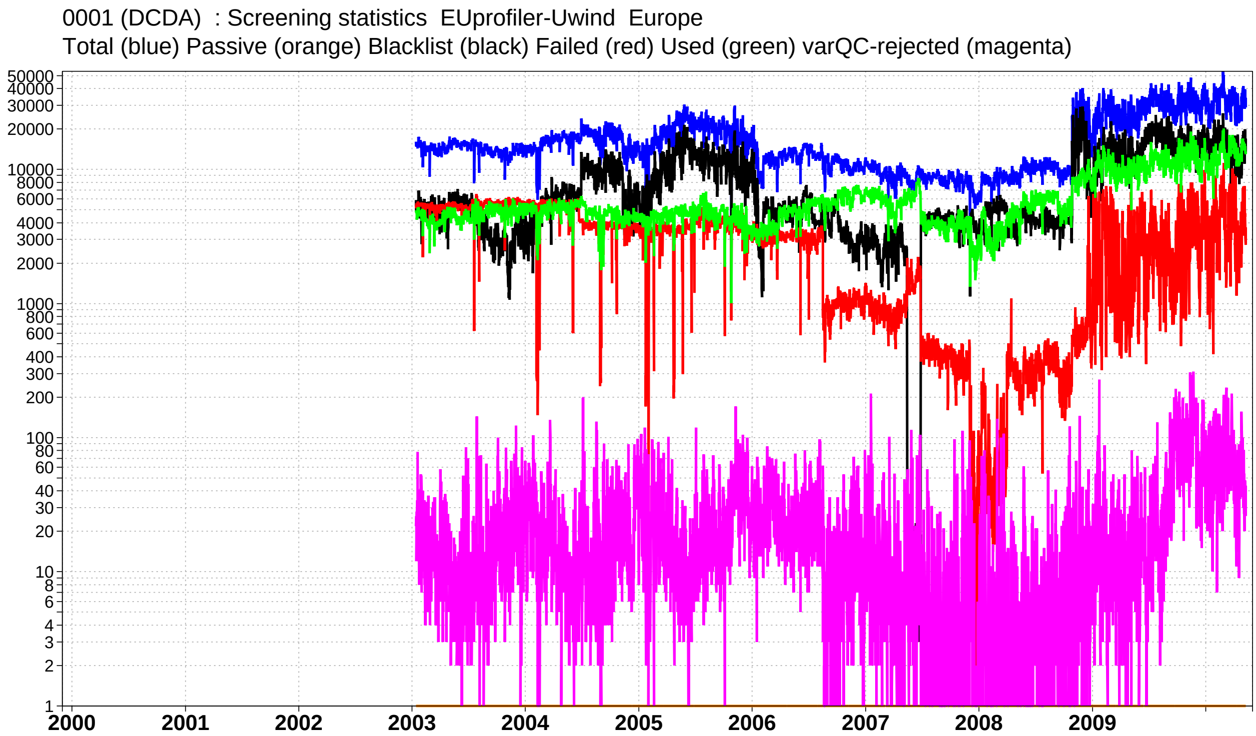


Figure 28: Data usage time series for 10 metre winds observed from ships. Top panel: total number of data counts, bottom panel: data plotted as percentague of used data.



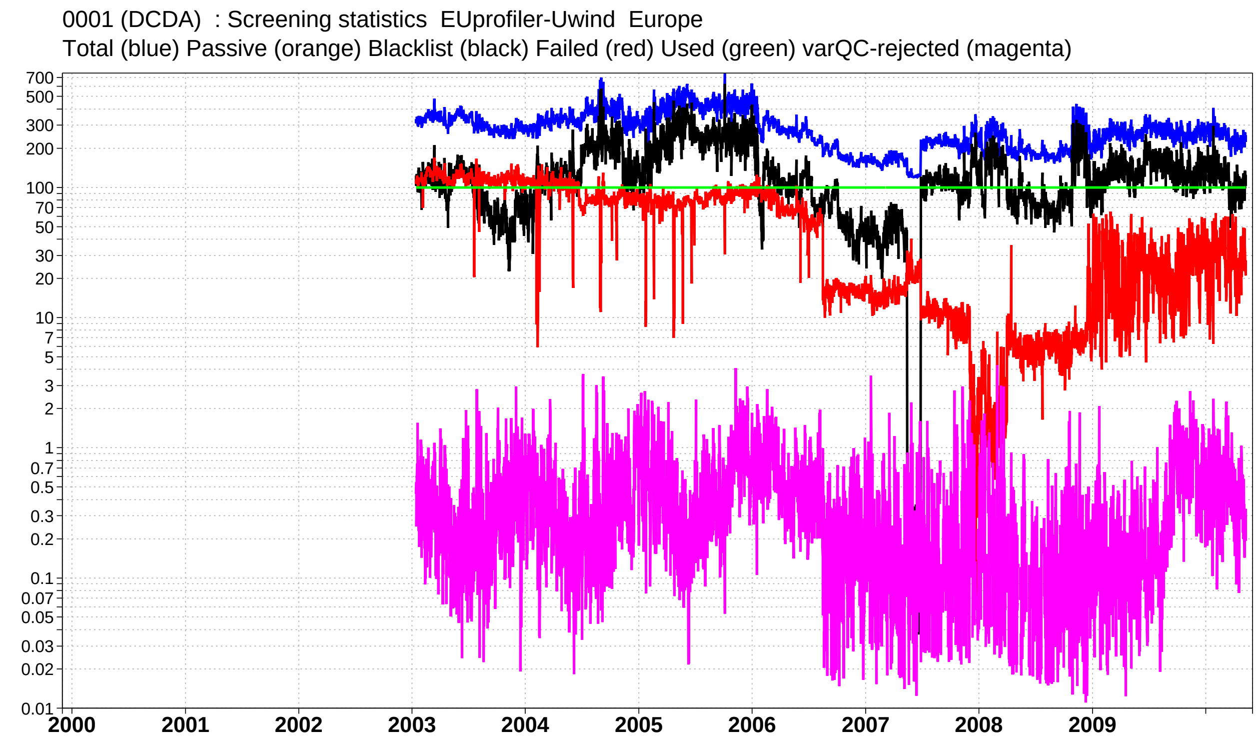


Figure 29: Data usage time series for wind observations from European wind profilers. Top panel: total number of data counts, bottom panel: data plotted as percentague of used data.

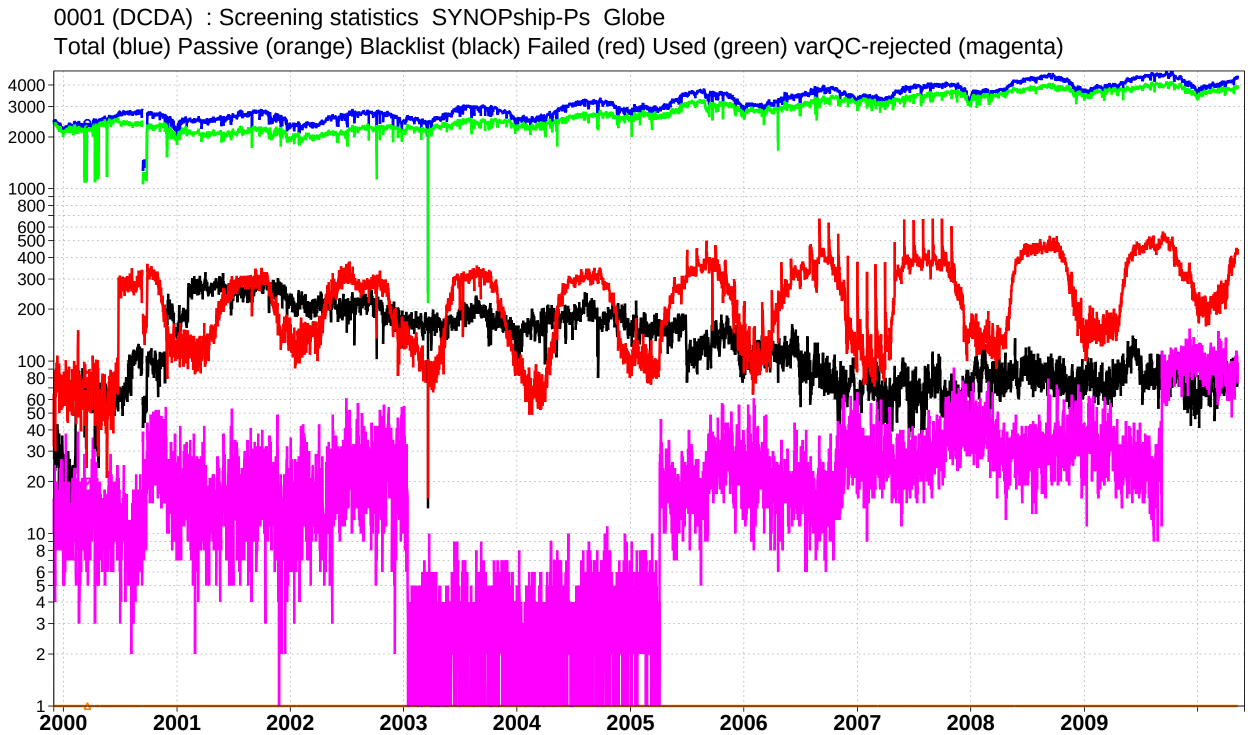


Figure 30: Time series of data usage for SYNOP ship surface pressure observations from operational data.

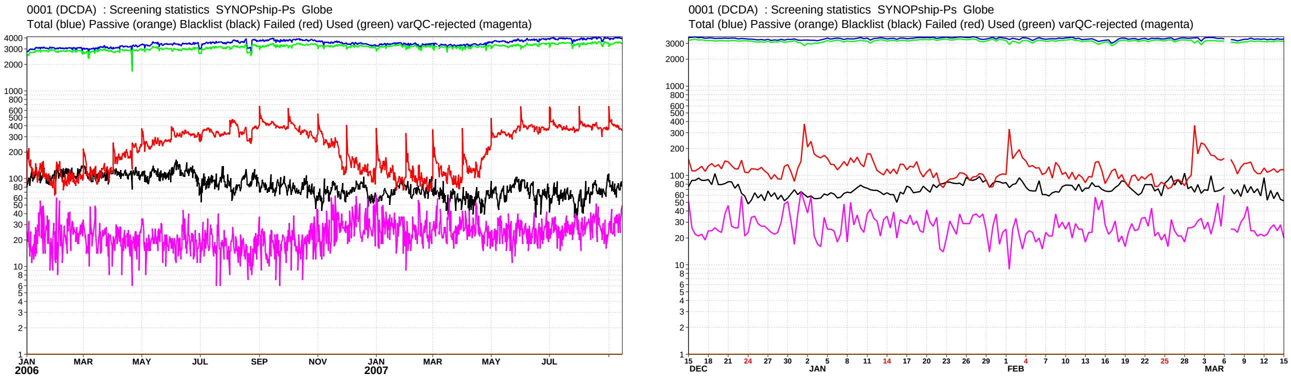


Figure 31: Time series of data usage for SYNOP ship surface pressure observations from operational data with special emphasis on the data rejection at the beginning of each months.

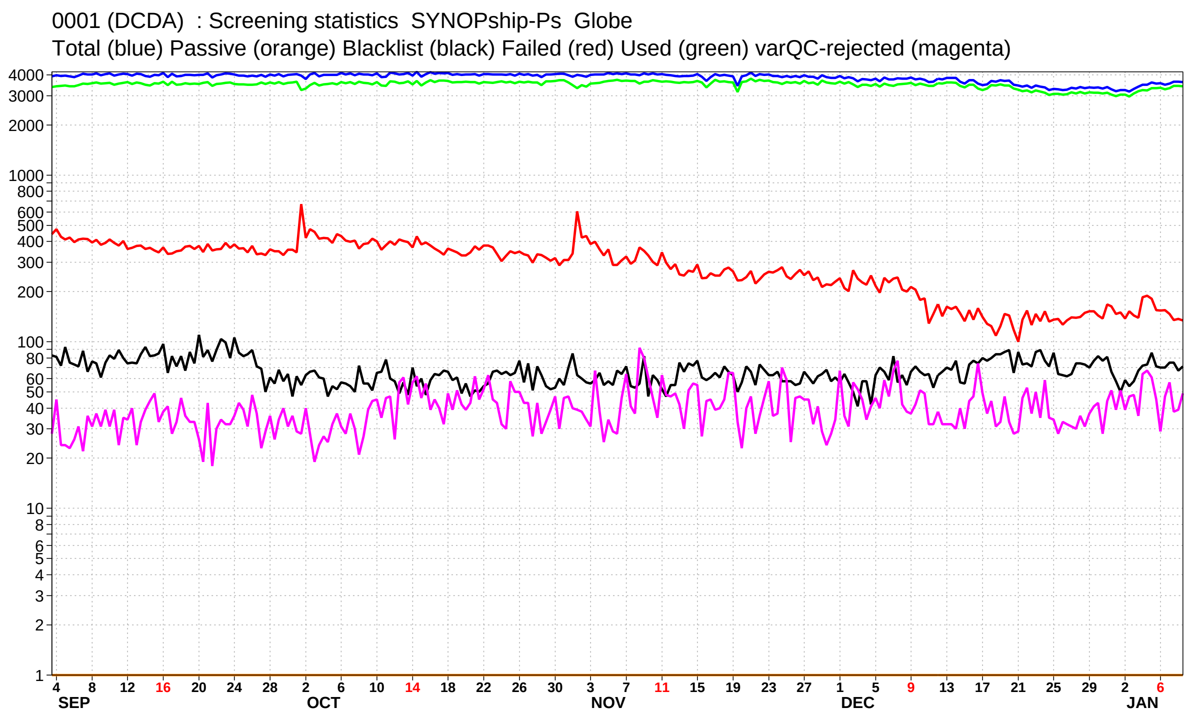


Figure 32: Time series of data usage for SYNOP ship surface pressure observations from operational data with the good result at the 1. December 2007 after solving the rejection problem.

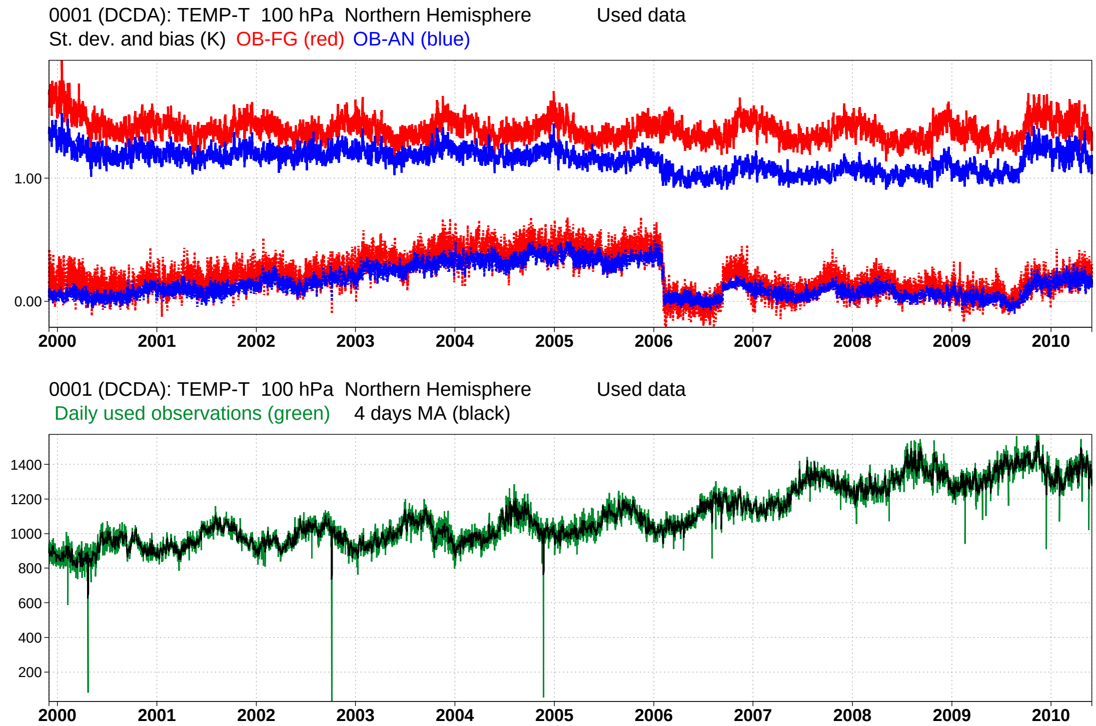


Figure 33: Departure statistics for radiosonde temperature at 100hPa on the Northern Hemisphere for the operational data. The lower panel shows used data counts.

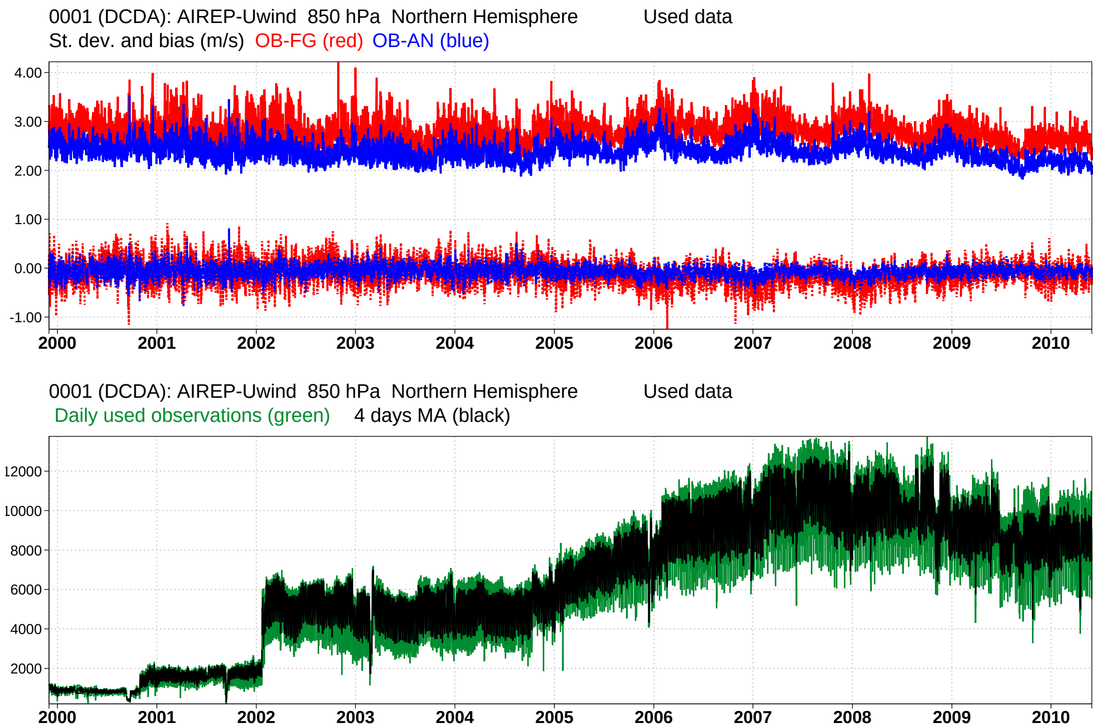


Figure 34: Departure statistics for zonal aircraft wind at 850hPa on the Northern Hemisphere for the operational data. The lower panel shows the count of used data during the time period.

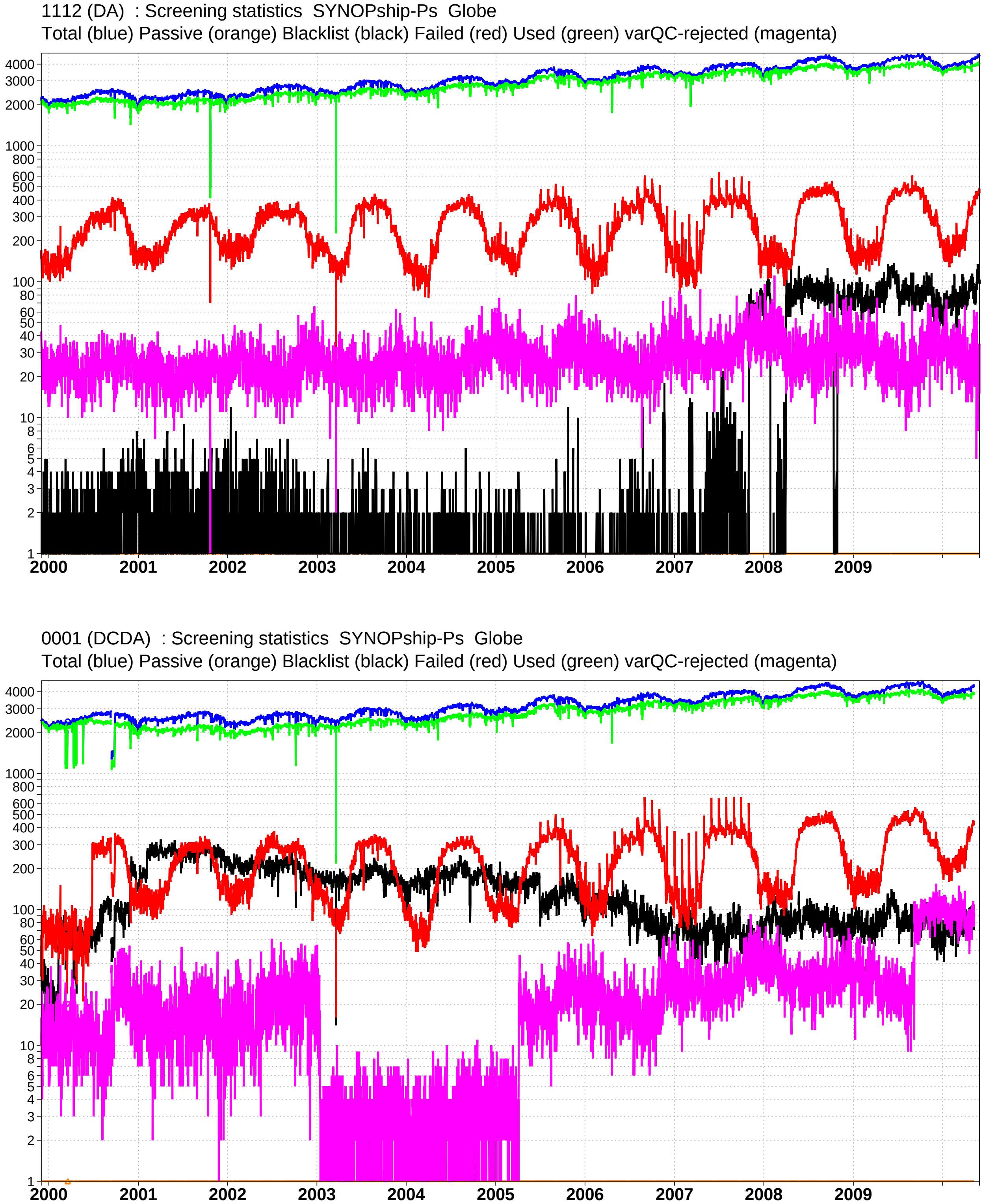


Figure 35: Time series of data usage for SYNOP ship surface pressure observations over the Northern Hemisphere. Top panel: ERA-Interim, bottom panel: operations

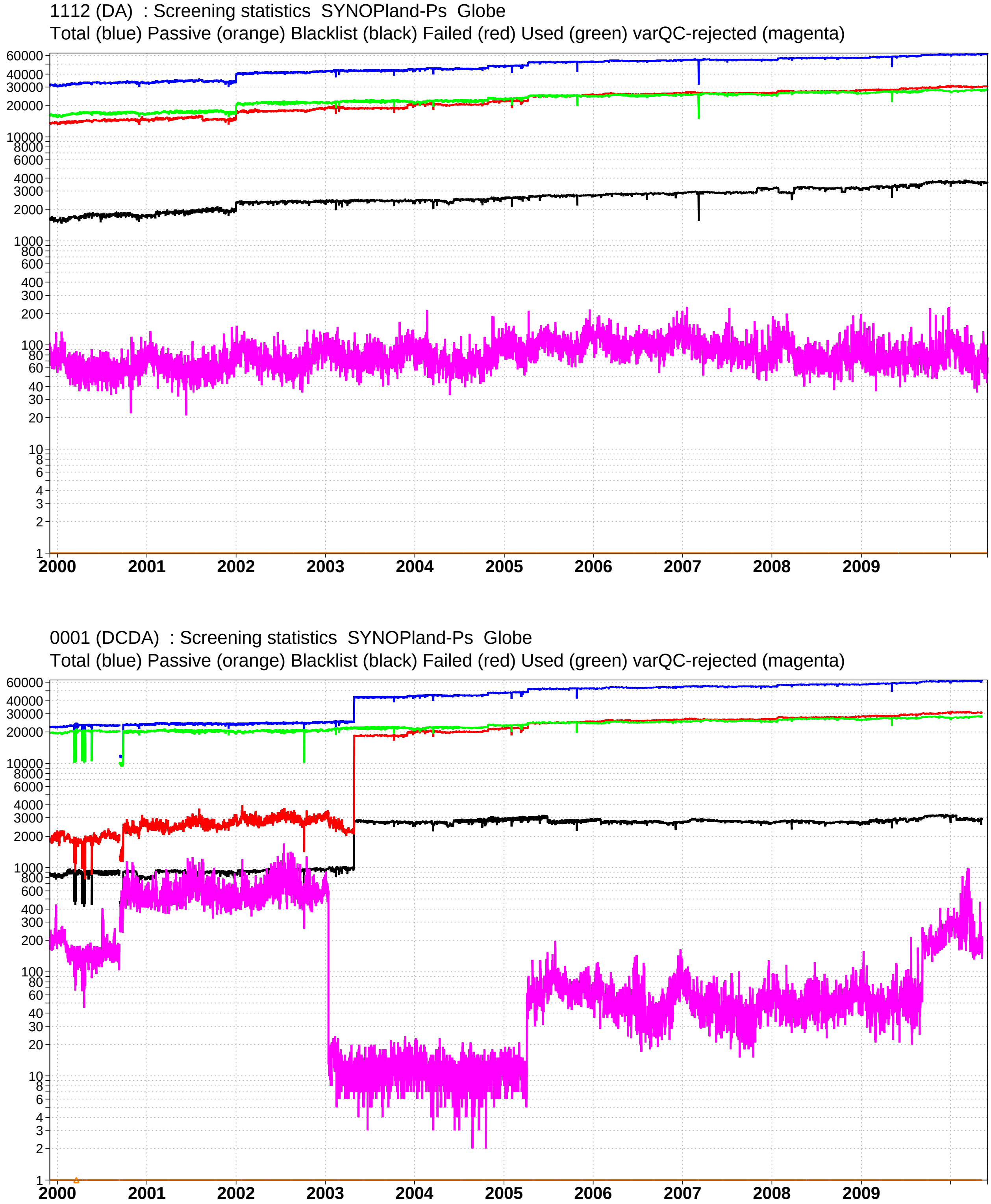


Figure 36: Time series of data usage for SYNOP land surface pressure observations over the Northern Hemisphere. Top panel: ERA-Interim, bottom panel: operations

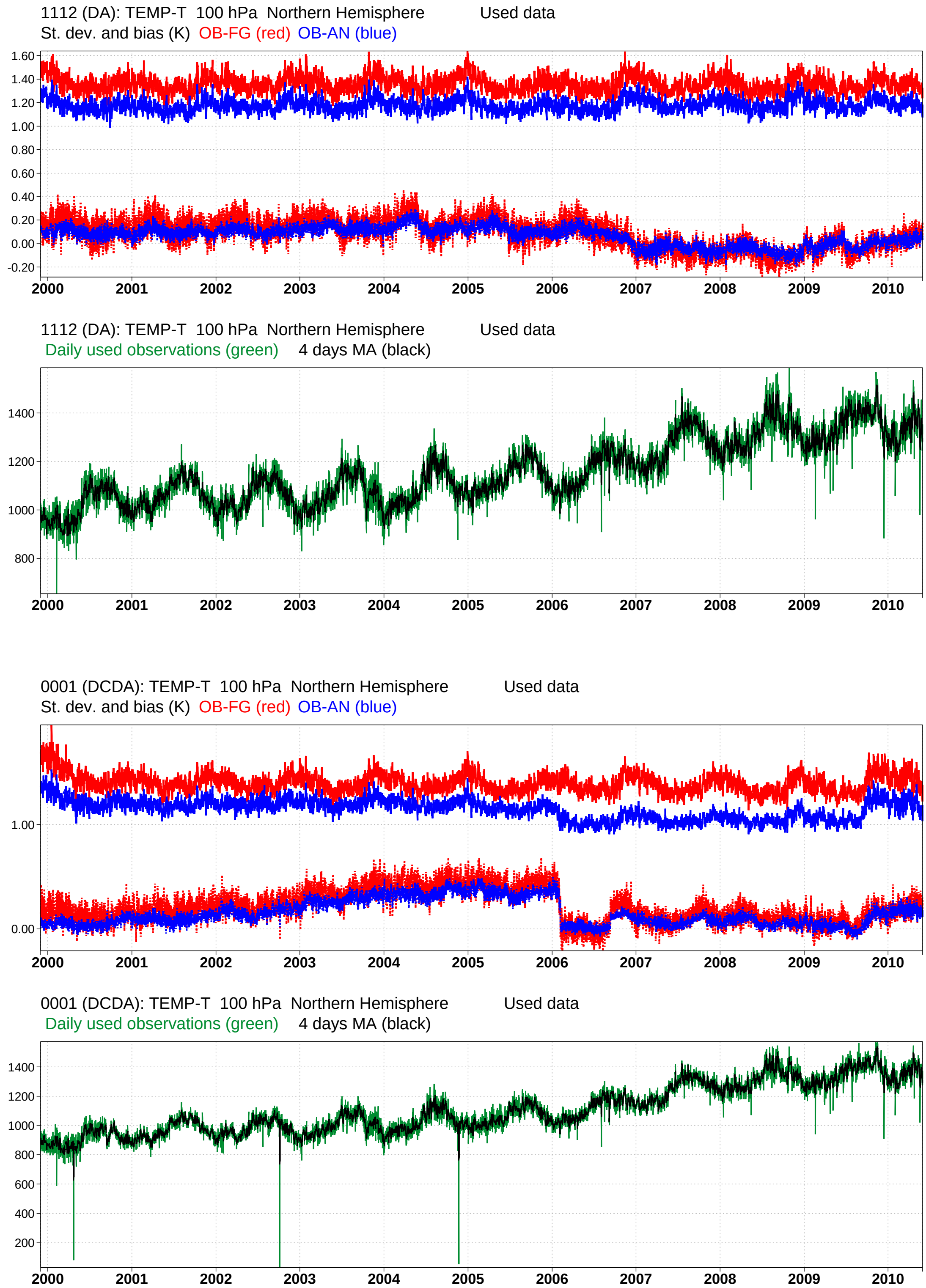


Figure 37: Departure statistics for radiosonde temperature at 100hPa on the Northern Hemisphere for the operational data. The lower panel shows the count of used data during the time period. Top panel: ERA-Interim, bottom panel: operations

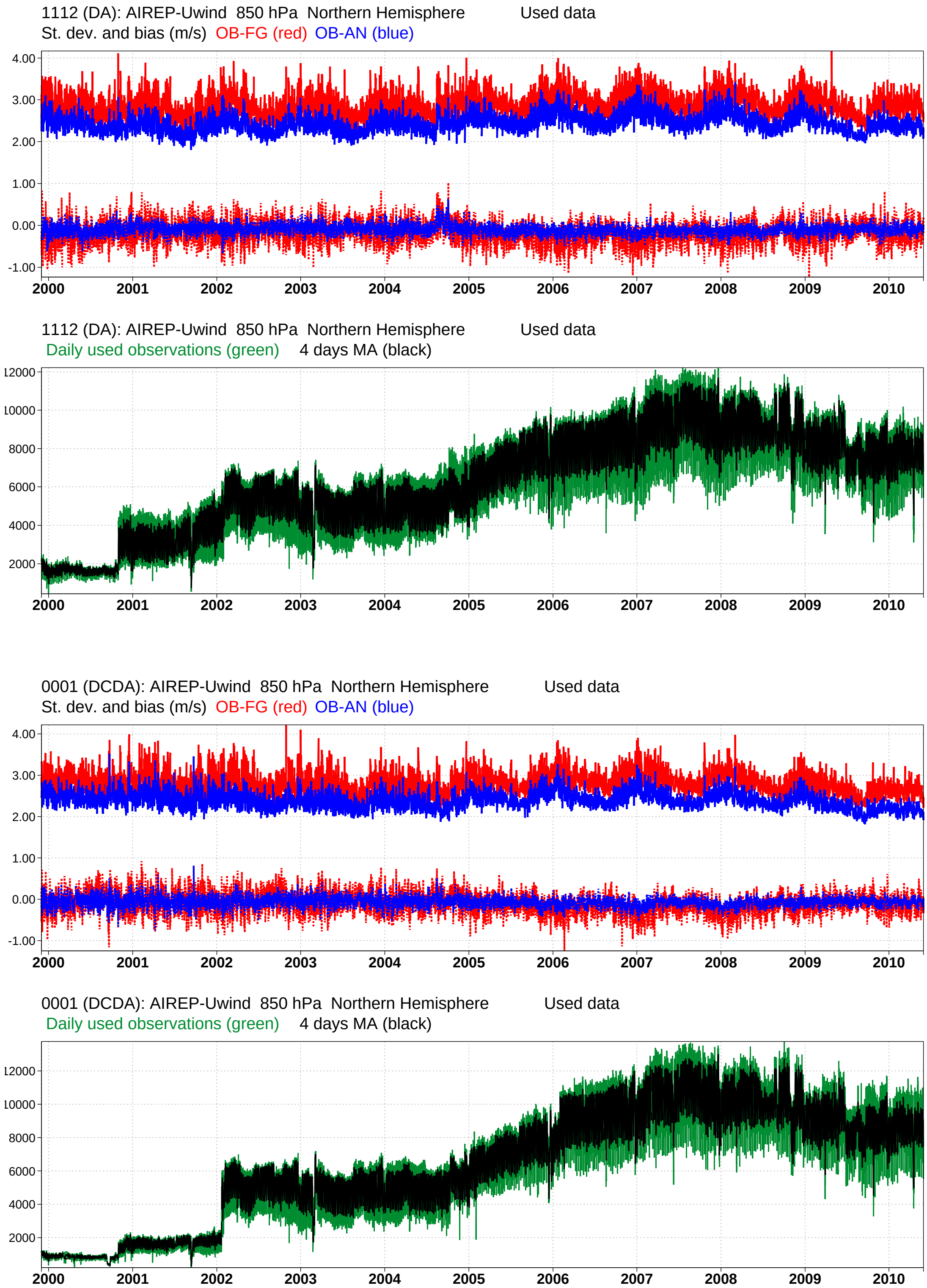


Figure 38: Departure statistics for zonal aircraft wind at 850hPa on the Northern Hemisphere for the operational data. The lower panel shows the count of used data during the time period. Top panel: ERA-Interim, bottom panel: operations

Data count and quality control decisions in ERA-40 and ERA-interim 1995

			All obs		FG rej		VarQC rej		Used Obs		Bg QC Limits			VarQC Limits			Rejection Ratio	
Obstype	Obsvalue	Level	E4	EI	E4	EI	E4	EI	E4	EI	E4	EI		E4	EI		E4	EI
SYNOP	Ps	surf	11168251	18041532	1.85%	1.25%	1.02%	0.25%	86.15%	53.61%	320.0	240.0	Pa	varying	200.0	Pa	2.87%	1.50%
SYNOP	rh	surf	11021987	11045165	0.02%	0.11%	0.03%	0.00%	30.67%	18.16%	52.0	n.a.	%	28.0	n.a.	%	0.05%	0.11%
SHIP	Ps	surf	2418295	2412324	1.12%	1.76%	0.70%	0.88%	78.00%	79.79%	380.0	260.0	Pa	varying	200.0	Pa	1.82%	2.64%
SHIP	U/V10m	surf	2633672	2628644	0.57%	0.67%	0.10%	0.30%	41.62%	50.42%	11.6	11.4	m/s	10.8	10.8	m/s	0.67%	0.97%
DRIBU	Ps	surf	2183944	3075928	1.88%	2.68%	1.14%	0.29%	64.52%	50.81%	380.0	320.0	Pa	varying	240.0	Pa	3.02%	2.97%
DRIBU	U/V10m	surf	1419484	1424686	0.46%	0.43%	0.25%	0.52%	43.11%	48.84%	10.9	10.7	m/s	7.4	7.4	m/s	0.71%	0.95%
METAR	Ps	surf	0	0														
TEMP	T	all	14173985	14101291	1.20%	1.06%	0.74%	1.00%	93.04%	93.37%	3.5	3.0	K	2.5	2.5	K	1.95%	2.06%
TEMP	T	0-100	3341643	3315190	1.10%	1.52%	1.43%	1.53%	94.07%	94.05%	4.8	4.7	K	3.6	3.6	K	2.53%	3.06%
TEMP	T	100-900	9617377	9574844	1.02%	0.90%	0.52%	0.71%	93.98%	94.06%	3.5	3.0	K	2.5	2.5	K	1.54%	1.61%
TEMP	T	1000-900	1214961	1211253	2.95%	1.10%	0.60%	1.77%	82.74%	86.02%	3.9	4.7	K	3.6	3.6	K	3.55%	2.87%
TEMP	U/V	all	10821439	10751160	0.71%	0.70%	0.33%	0.63%	91.60%	91.24%	10.5	11.0	m/s	9.1	9.1	m/s	1.04%	1.34%
TEMP	U/V	0-100	2789733	2768702	1.06%	1.00%	0.55%	0.74%	96.26%	95.67%	15.7	13.0	m/s	10.2	10.1	m/s	1.60%	1.75%
TEMP	U/V	100-900	6756365	6710221	0.61%	0.62%	0.26%	0.59%	95.77%	95.27%	11.7	11.5	m/s	9.1	9.1	m/s	0.87%	1.21%
TEMP	U/V	1000-900	1275243	1272139	0.48%	0.49%	0.24%	0.62%	59.34%	60.39%	10.5	11.0	m/s	9.1	9.1	m/s	0.73%	1.11%
TEMP	q	all	11519732	11477986	0.71%	0.46%	0.06%	0.17%	58.91%	58.62%	varying			varying			0.77%	0.63%
TEMP	q	100-900	8216284	8182711	0.67%	0.54%	0.08%	0.21%	70.92%	69.97%	varying			varying			0.75%	0.75%
TEMP	q	1000-900	1624186	1621401	1.61%	0.53%	0.05%	0.16%	59.08%	61.91%	varying			varying			1.67%	0.70%
AIREP	T	all	12023988	12066985	1.21%	1.18%	0.09%	0.08%	41.94%	39.56%	4.6	4.2	K	3.7	3.8	K	1.31%	1.26%
AIREP	T	100-900	10495862	10533445	1.20%	1.24%	0.11%	0.08%	45.14%	40.90%	4.6	4.2	K	3.7	3.8	K	1.30%	1.32%
AIREP	T	1000-900	1525185	1530593	1.30%	0.74%	0.02%	0.06%	19.92%	30.33%	5.5	6.2	K	5.1	5.2	K	1.32%	0.79%
AIREP	U/V	all	10410626	10449261	1.26%	1.36%	0.16%	0.16%	48.44%	44.00%	~15.0	~14.5	m/s	12.7	12.8	m/s	1.43%	1.52%
AIREP	U/V	100-900	9048325	9082108	1.38%	1.49%	0.19%	0.17%	52.37%	45.75%	~15.5	~15.5	m/s	12.8	12.8	m/s	1.57%	1.66%
AIREP	U/V	1000-900	1359187	1364033	0.46%	0.48%	0.02%	0.02%	22.30%	32.37%	~15.0	~14.5	m/s	11.7	12.8	m/s	0.48%	0.50%
PILOT	U/V	all	3928818	3939870	0.68%	0.81%	0.30%	0.63%	84.67%	82.55%	10.7	10.4	m/s	9.1	9.1	m/s	0.98%	1.44%
PILOT	U/V	0-100	447567	450149	1.07%	1.63%	0.36%	0.60%	63.10%	61.19%	15.7	13.0	m/s	10.5	10.2	m/s	1.44%	2.23%
PILOT	U/V	100-900	2856343	2863414	0.62%	0.72%	0.30%	0.65%	89.14%	86.66%	12.1	11.0	m/s	9.1	9.1	m/s	0.92%	1.37%
PILOT	U/V	1000-900	624904	626303	0.65%	0.65%	0.24%	0.56%	79.69%	79.11%	10.7	10.4	m/s	9.1	9.1	m/s	0.89%	1.21%
profiler	U/V	all	1497290	1502501	2.51%	2.80%	0.34%	0.61%	96.62%	95.91%	12.9	11.8	m/s	9.2	9.1	m/s	2.85%	3.41%
profiler	U/V	0-100	20978	21354	12.68%	13.49%	1.23%	1.39%	86.00%	85.03%	18.9	16.8	m/s	11.2	10.5	m/s	13.91%	14.88%
profiler	U/V	100-900	1440230	1444911	2.39%	2.66%	0.32%	0.57%	96.74%	96.07%	12.9	11.8	m/s	9.3	9.1	m/s	2.71%	3.23%
profiler	U/V	1000-900	36082	36236	1.60%	2.02%	0.45%	1.74%	97.72%	95.95%	12.9	12.4	m/s	9.2	5.5	m/s	2.05%	3.75%
EU-profiler	U/V	all	0	0														
EU-profiler	U/V	0-100	0	0														
EU-profiler	U/V	100-900	0	0														
EU-profiler	U/V	1000-900	0	0														
JP-profiler	U/V	all	0	0														
JP-profiler	U/V	0-100	0	0														
JP-profiler	U/V	100-900	0	0														
JP-profiler	U/V	1000-900	0	0														
US-profiler	U/V	all	1497290	1502501	2.51%	2.80%	0.34%	0.61%	96.62%	95.91%	12.9	11.8	m/s	9.2	9.1	m/s	2.85%	3.41%
US-profiler	U/V	0-100	20978	21354	12.68%	13.49%	1.23%	1.39%	86.00%	85.03%	18.9	16.8	m/s	11.2	10.5	m/s	13.91%	14.88%
US-profiler	U/V	100-900	1440230	1444911	2.39%	2.66%	0.32%	0.57%	96.74%	96.07%	12.9	11.8	m/s	9.3	9.1	m/s	2.71%	3.23%
US-profiler	U/V	1000-900	36082	36236	1.60%	2.02%	0.45%	1.74%	97.72%	95.95%	12.9	12.4	m/s	9.2	5.5	m/s	2.05%	3.75%

Table 3: Table of data count and quality control overview for conventional data in ERA-40 (E4) and ERA-interim (EI) for the year 1995. FG rej are first-guess rejected data; VarQC rej are variational quality control rejected; Bg QC limits are QC limits used by the background quality control and VarQC limits are QC limits used by the variational QC control.

Data count and quality control decisions in ERA-interim and the operational ECMWF assimilation system 2005

			All obs		FG rej		VarQC rej		Used Obs		Bg QC Limits			VarQC Limits			Rejection Ratio	
Obstype	Obsvalue	Level	Oper	EI	Oper	EI	Oper	EI	Oper	EI	Oper	EI		Oper	EI		Oper	EI
SYNOP	Ps	surf	38193760	37237480	0.68%	0.79%	0.13%	0.20%	47.05%	20.87%	260.0	260.0	Pa	200.0	200.0	Pa	0.81%	0.99%
SYNOP	rh	surf	21919122	21540624	0.08%	0.08%	0.00%	0.00%	21.74%	17.55%	n.a.	n.a.	%	n.a.	n.a.	%	0.08%	0.08%
SHIP	Ps	surf	2404746	2376857	0.77%	1.18%	0.69%	0.94%	88.87%	91.87%	280.0	280.0	Pa	200.0	200.0	Pa	1.46%	2.12%
SHIP	U/V10m	surf	2327477	2297341	0.64%	0.59%	0.32%	0.28%	59.84%	46.31%	10.8	11.6	m/s	10.8	10.8	m/s	0.96%	0.87%
DRIBU	Ps	surf	7078349	6349051	1.81%	2.26%	0.29%	0.31%	38.81%	40.42%	340.0	340.0	Pa	240.0	240.0	Pa	2.10%	2.57%
DRIBU	U/V10m	surf	419727	396648	1.76%	1.68%	1.70%	1.56%	86.30%	80.93%	10.3	9.9	m/s	7.4	7.4	m/s	3.46%	3.24%
METAR	Ps	surf	15551667	15453014	0.06%	0.23%	0.04%	0.16%	89.39%	91.02%	380.0	360.0	Pa	340.0	340.0	Pa	0.10%	0.38%
TEMP	T	all	16042720	15940932	0.84%	0.82%	0.92%	1.05%	91.50%	93.51%	2.9	3.0	K	2.5	2.5	K	1.76%	1.87%
TEMP	T	0-100	4118642	4087865	1.37%	1.08%	1.50%	1.48%	90.34%	95.37%	4.4	4.6	K	3.6	3.6	K	2.87%	2.56%
TEMP	T	100-900	10667340	10602075	0.62%	0.68%	0.65%	0.78%	92.28%	93.77%	2.9	3.0	K	2.5	2.5	K	1.27%	1.46%
TEMP	T	1000-900	1256738	1250990	1.01%	1.12%	1.23%	1.94%	88.73%	85.17%	4.6	4.9	K	3.6	3.6	K	2.24%	3.06%
TEMP	U/V	all	14570453	14340834	0.47%	0.54%	0.42%	0.60%	93.50%	92.77%	9.7	10.8	m/s	9.1	9.1	m/s	0.89%	1.14%
TEMP	U/V	0-100	4708088	4589855	0.58%	0.58%	0.49%	0.61%	96.45%	96.18%	11.9	12.9	m/s	10.1	10.1	m/s	1.07%	1.19%
TEMP	U/V	100-900	8534351	8430361	0.41%	0.51%	0.37%	0.57%	95.79%	95.23%	10.7	11.4	m/s	9.1	9.1	m/s	0.78%	1.08%
TEMP	U/V	1000-900	1327990	1320602	0.49%	0.57%	0.53%	0.78%	68.30%	65.25%	9.7	10.8	m/s	9.1	9.1	m/s	1.02%	1.35%
TEMP	q	all	13533114	13418561	0.47%	0.49%	0.16%	0.19%	59.07%	59.88%	varying			varying			0.64%	0.68%
TEMP	q	100-900	9364838	9280919	0.58%	0.59%	0.21%	0.24%	73.54%	75.29%	varying			varying			0.79%	0.83%
TEMP	q	1000-900	1666555	1660797	0.58%	0.62%	0.17%	0.19%	66.45%	63.12%	varying			varying			0.75%	0.81%
AIREP	T	all	69664448	66671628	0.22%	0.37%	0.06%	0.11%	58.39%	63.20%	4.0	4.1	K	3.8	3.8	K	0.28%	0.48%
AIREP	T	100-900	58595984	56113624	0.19%	0.33%	0.06%	0.11%	59.51%	66.33%	4.0	4.1	K	3.8	3.8	K	0.25%	0.44%
AIREP	T	1000-900	11067772	10557288	0.33%	0.54%	0.10%	0.13%	52.47%	46.55%	5.7	5.9	K	5.1	5.0	K	0.43%	0.67%
AIREP	U/V	all	69749344	66602336	0.40%	0.45%	0.10%	0.13%	64.91%	63.69%	13.6	14.7	m/s	12.5	12.5	m/s	0.50%	0.58%
AIREP	U/V	100-900	58935220	56307792	0.40%	0.46%	0.10%	0.14%	64.93%	66.43%	13.6	15.7	m/s	12.6	12.6	m/s	0.50%	0.60%
AIREP	U/V	1000-900	10813077	10293823	0.40%	0.41%	0.07%	0.07%	64.84%	48.71%	13.8	14.7	m/s	12.5	12.5	m/s	0.47%	0.48%
PILOT	U/V	all	6233195	6123265	0.52%	0.65%	0.46%	0.66%	78.42%	83.46%	9.9	10.9	m/s	9.1	9.1	m/s	0.99%	1.31%
PILOT	U/V	0-100	1326562	1245501	0.73%	0.81%	0.50%	0.62%	83.01%	84.43%	12.3	13.0	m/s	10.2	10.2	m/s	1.23%	1.43%
PILOT	U/V	100-900	4114306	4083615	0.48%	0.63%	0.48%	0.70%	79.82%	85.94%	9.9	11.6	m/s	9.1	9.1	m/s	0.96%	1.33%
PILOT	U/V	1000-900	792327	794149	0.37%	0.49%	0.33%	0.50%	63.49%	69.17%	10.7	10.9	m/s	9.1	9.1	m/s	0.71%	0.99%
profiler	U/V	all	27913468	28743658	0.34%	1.24%	0.21%	0.47%	48.54%	76.26%	10.9	10.4	m/s	9.1	9.1	m/s	0.55%	1.71%
profiler	U/V	0-100	584501	600806	3.09%	3.56%	1.52%	1.48%	48.88%	50.64%	12.2	14.5	m/s	10.1	10.2	m/s	4.61%	5.04%
profiler	U/V	100-900	24977224	25660696	0.30%	1.14%	0.20%	0.42%	52.82%	78.29%	10.9	10.4	m/s	9.1	9.1	m/s	0.50%	1.56%
profiler	U/V	1000-900	2351745	2482156	0.05%	1.71%	0.01%	0.77%	3.01%	61.45%	11.4	10.7	m/s	9.1	9.1	m/s	0.06%	2.48%
EU-profiler	U/V	all	13216667	13885356	0.27%	1.62%	0.14%	0.40%	24.00%	52.85%	10.9	10.4	m/s	9.1	9.1	m/s	0.42%	2.03%
EU-profiler	U/V	0-100	404936	424900	3.72%	3.70%	2.05%	1.85%	32.40%	31.79%	12.2	14.5	m/s	10.1	10.2	m/s	5.77%	5.54%
EU-profiler	U/V	100-900	11212697	11732330	0.18%	1.53%	0.09%	0.34%	26.49%	54.59%	10.9	10.4	m/s	9.1	9.1	m/s	0.27%	1.87%
EU-profiler	U/V	1000-900	1599034	1728126	0.07%	1.78%	0.02%	0.46%	4.43%	46.20%	11.4	10.7	m/s	9.1	9.1	m/s	0.09%	2.24%
JP-profiler	U/V	all	3503045	3481487	0.21%	0.53%	0.36%	0.80%	71.72%	98.44%	11.6	11.6	m/s	9.1	9.1	m/s	0.58%	1.33%
JP-profiler	U/V	100-900	3012748	2993565	0.25%	0.52%	0.42%	0.73%	83.39%	98.61%	11.6	12.0	m/s	9.1	9.1	m/s	0.67%	1.25%
JP-profiler	U/V	1000-900	490297	487922	0.00%	0.61%	0.00%	1.22%	0.00%	97.41%	n.a.	11.6	m/s	n.a.	9.1	m/s	0.00%	1.83%
US-profiler	U/V	all	11139485	11302865	0.46%	1.00%	0.24%	0.46%	70.60%	98.31%	11.4	11.1	m/s	9.7	9.1	m/s	0.70%	1.45%
US-profiler	U/V	0-100	179565	175906	1.67%	3.23%	0.31%	0.59%	86.05%	96.17%	13.9	14.5	m/s	11.2	10.8	m/s	1.98%	3.82%
US-profiler	U/V	100-900	10697506	10860851	0.45%	0.90%	0.24%	0.42%	72.07%	98.43%	11.4	12.2	m/s	9.7	9.1	m/s	0.69%	1.32%
US-profiler	U/V	1000-900	262414	266108	0.00%	3.28%	0.00%	1.97%	0.00%	94.62%	n.a.	11.1	m/s	n.a.	9.1	m/s	0.00%	5.25%

Table 4: Table similar to Table 3, but of data count and quality control overview for conventional data in ERA-Interim (EI) and the operational ECMWF assimilation system (Oper) for the year 2005 (Operational data counts from 5. April 2005 to 4. April 2006).