Statistics of the arrival times of conventional observations for HIRLAM at FMI

Kalle Eerola
Finnish Meteorological Institute

1 Introduction

The term "cut-off" time is used for the time interval between the nominal assimilation time and the time when the collection of observations ends and the data assimilation starts. So it implicitly defines, how complete observational data set we have in the time window of the data assimilation cycle.

Quite a long cut-off time, 150 minutes, is used currently in the HIRLAM configuration at FMI. It has been used for many years without checking, what the typical arrival times of different observation types are at the moment. When planning the RCR environment, the cut-off time becomes more critical, and a balance must be found between several opposite requirements:

- The duty forecasters require the HIRLAM products as quickly as possible. The time should not be longer than at the moment, which means that the forecasts should be ready three hours after the observation time.
- The number of grid points in RCR runs is much larger than in the present operational HIRLAM configuration at FMI, which means much longer computing time than at the moment.
- For a good forecast it is essential to have a good data coverage, which clearly favors a long cut-off time.

Some NWP centers use a re-assimilation cycle in order to get the best possible background field for the next cycle: the data assimilation cycle is re-run afterwards using a longer cut-off time, which means that new data have arrived meanwhile and the most complete observational data set can be used.

This report tries to answer the following three questions. First of all, can the current cut-off time of 2h 30min be shortened and how much, if a good data coverage is required? Secondly, how much new data can be received, if the planned RCR configuration uses a re-assimilation cycle with a cut-off time of the order five hours? The third question concerns the sounding data: can a reasonable data coverage be achieved with a short cut-off time (of the order two hours) and how much it is improved, when the cut-off time is increased.

Section two describes the data and methods of this study. Then, in section three the results are shown. Section four discusses the results and the conclusions from this study are presented in section five.

2 Data and method

In this study we look at the data, which are used in the current HIRLAM system at FMI. The following data types are considered: AIREP, AMDAR, BUOY, PILOT, SATOB, SHIP, SYNOP and TEMP. Of these, the soundings (TEMP and PILOT) are the most important data types, because they give vertical profiles of atmospheric conditions. Therefore special attention is given to them. The problem worth mentioning of the TEMP soundings is that their number is worldwide decreasing all the time. All observations from the proposed RCR area considered. For general interest, the proposed RCR area is shown in Figure 1.

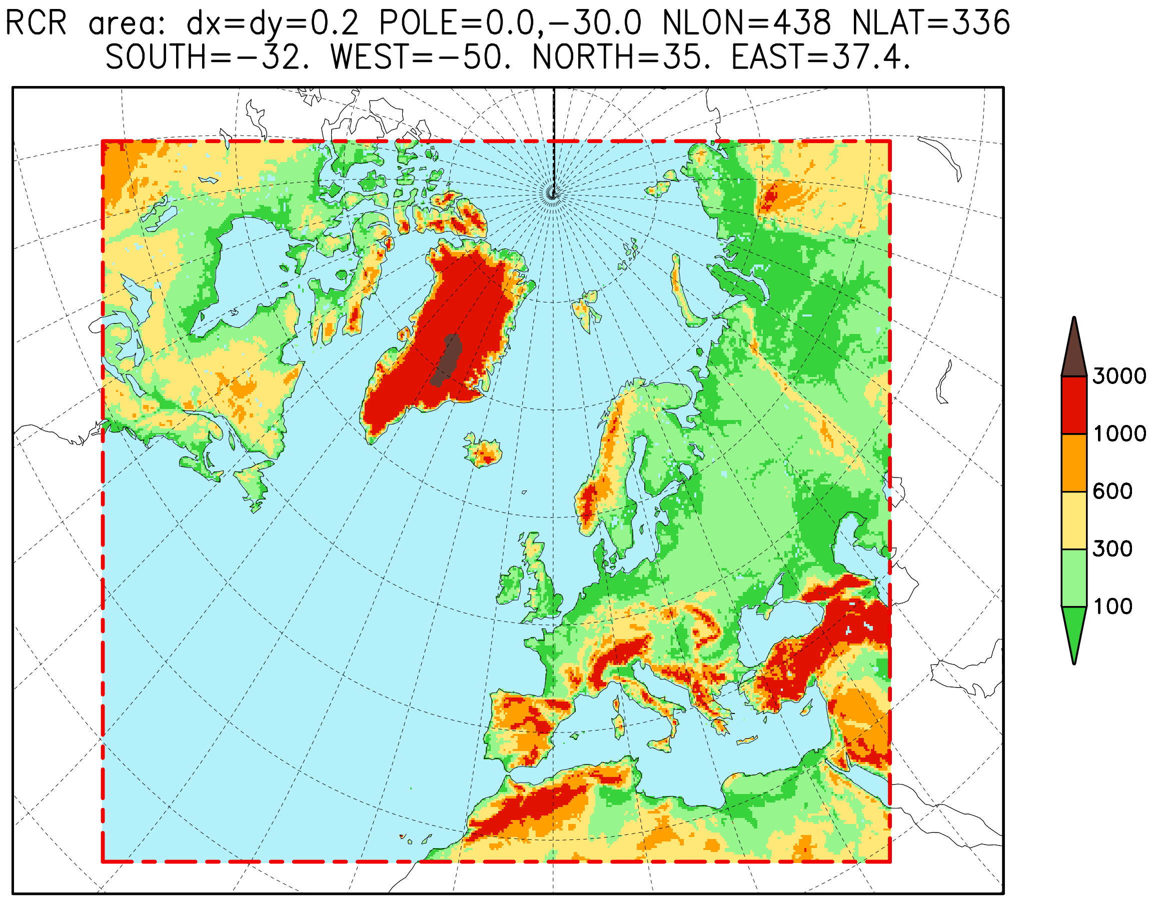

Figure 1: The proposed RCR area together with the HIRLAM version 6.1.2 orography.

In the following the term "arrival time" refers to the time, when an observation is available to the HIRLAM system. The data flow from GTS is a continuous process. The software to handle the incoming data for HIRLAM purposes is repeated every few minutes, depending on the data type. Every time a new file is created for observations, with the creation time in the filename. In this way it is possible to check at what time any single observation is added to the HIRLAM database. The running frequency of this software determines the time resolution of the arrival times. It depends on the observation type. Table 1 gives the running frequency for different observation types.

Table 1: The repeating frequency of the observation processing software for different observation types. It gives the time resolution, at which it is possible to measure the arrival times of observations to the HIRLAM system.

	Time resolution
SYNOP	2 min
SHIP	2 min
AIREP	15 min
AMDAR	15 min
BUOY	15 min
SATOB	60 min
TEMP	15 min
PILOT	60 min

The data of this study contain the time period starting at 6 June 2003 at 06 local time and ending at 20 June at 23 local time.

The data were processed to follow closely the way they are used in the proposed RCR system. We suppose a three hour data assimilation cycle. This means that data are always used from a three hour data window, centred at the nominal assimilation hour (00, 03, 06, ...). So observational data are used from a time period from ±90 min compared to the nominal observation hour.

The further processing of the observations tries to mimic the way the observations are selected in the HIRLAM 3DVAR system. First, if there are duplicates, only the most complete or the one arrived first is accepted. If there are several observations from the same station, only the one closest to the nominal assimilation hour is taken into account. An exception to the actual 3DVAR observation processing is that in this study the screening and quality control of observations were not possible.

Concerning soundings (TEMP and PILOT) the most complete observation was selected. This means the report with most parts (A, B, C and D) and maximum number of levels. Because making a complete PILOT or TEMP sounding takes quite a long time, it is a common practice to send parts of the observation already before the complete observation is ready. In practice there normally exists several versions from the same sounding with different number of parts (A, B, C, D) and levels. Therefore the arrival statistics for TEMP and PILOT reports were computed separately for the first arrived report and for the most complete report. The difference of these statistics were used to estimate, if the first arrived reports already contain enough information for HIRLAM purposes.

3 Results

The following figures (Figures 2 - 8) show the statistics of the arrival times of different observation types during the two-week period in June 2003. They show the percentage and cumulative percentage from the total data volume, received in 15 minute time slots, relative to the nominal cycle hour. *Note that the time scale varies from figure to figure.* In the following discussion we consider especially three possible scenarios for cut-off times: 90 min, 120 min and the current

150 min. Also the possibility to re-assimilate the four main synoptic cycles (00, 06, 12 and 18 UTC) with a longer cut-off time will be kept in mind.

3.1 Other than sounding data

Concerning SYNOP reports (Figure 2), almost all, 97%, observations are already available in one hour. Also SHIP observations (Figure 3) arrive quickly: 90% of all data is available in 90 minutes and 95% in two hours. So a very short cut-off, such as 90 minutes, would be adequate for SYNOP and SHIP reports.

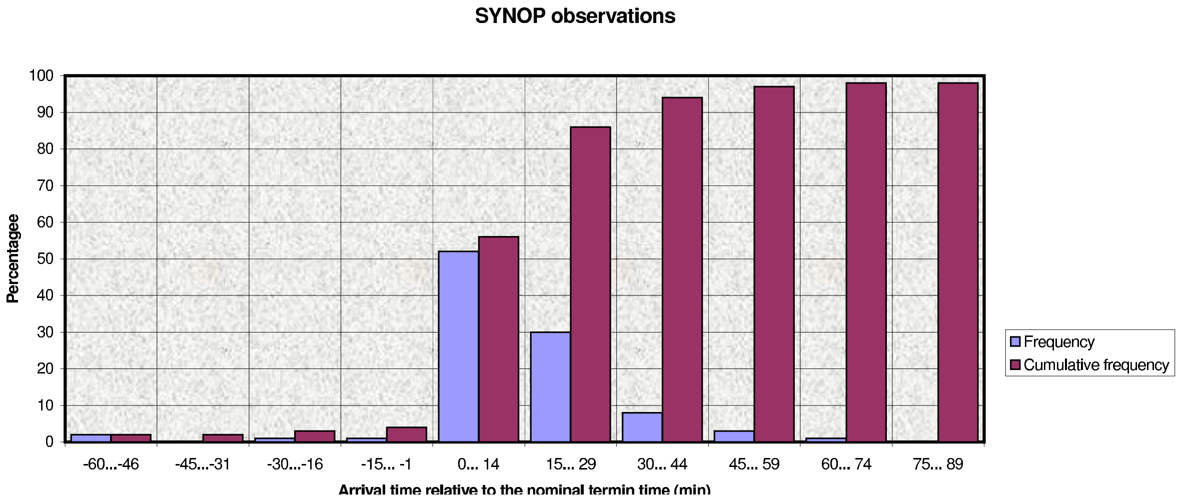

Figure 2: Statistics of arrival times of SYNOP observations: frequency and cumulative frequency in 15 min time slots relative to the nominal assimilation hour.

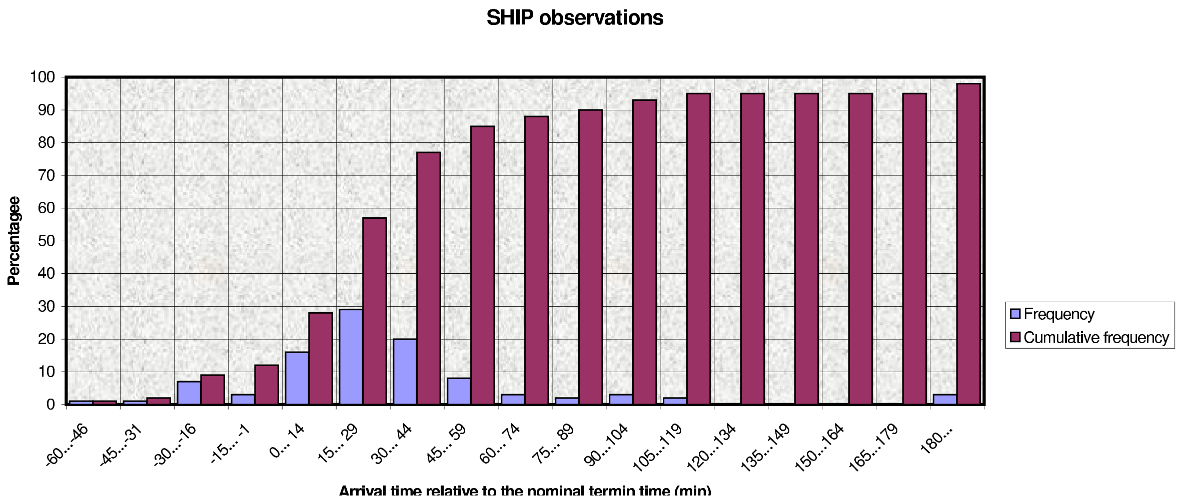

Figure 3: As Figure 2, but for SHIP observations.

Figure 4 shows the arrival times statistics for AIREP observations. Of them 76% arrive in 90 min, 84% in 120 min and 94% in 150 min The corresponding numbers for AMDAR observations

(Figure 5) are 87%, 98% and 99%. So AMDAR observations arrive very quickly and a two hour cut-off time would be sufficient, but about 10% of AIREP observations arrive between two and two and half hours. So if a two hour cut-off time is applied, a re-assimilation, using a longer cut-off might be beneficial, especially because these observations come mainly from the Atlantic Ocean, where every single observation may be valuable.

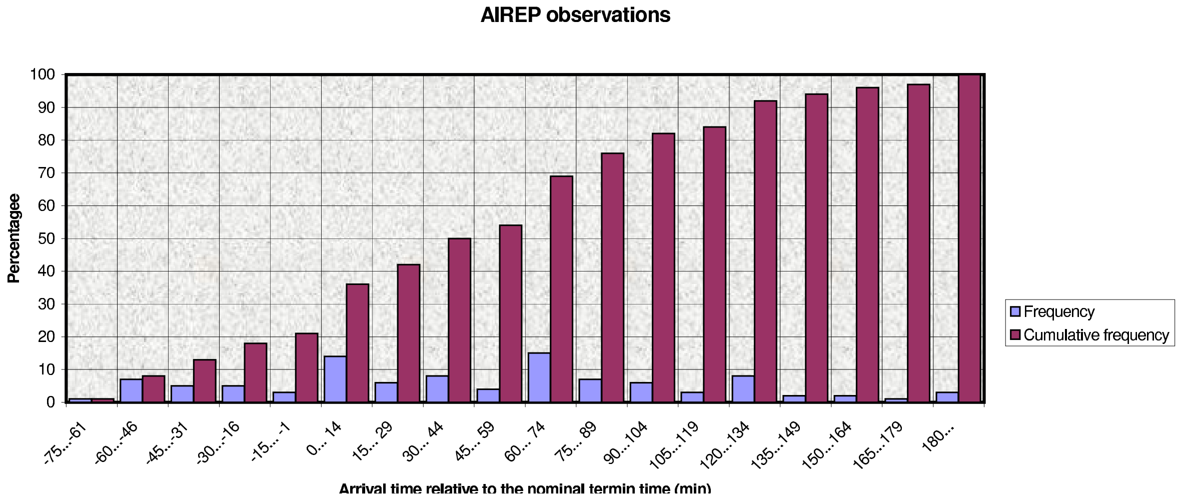

Figure 4: As Figure 2, but for AIREP observations.

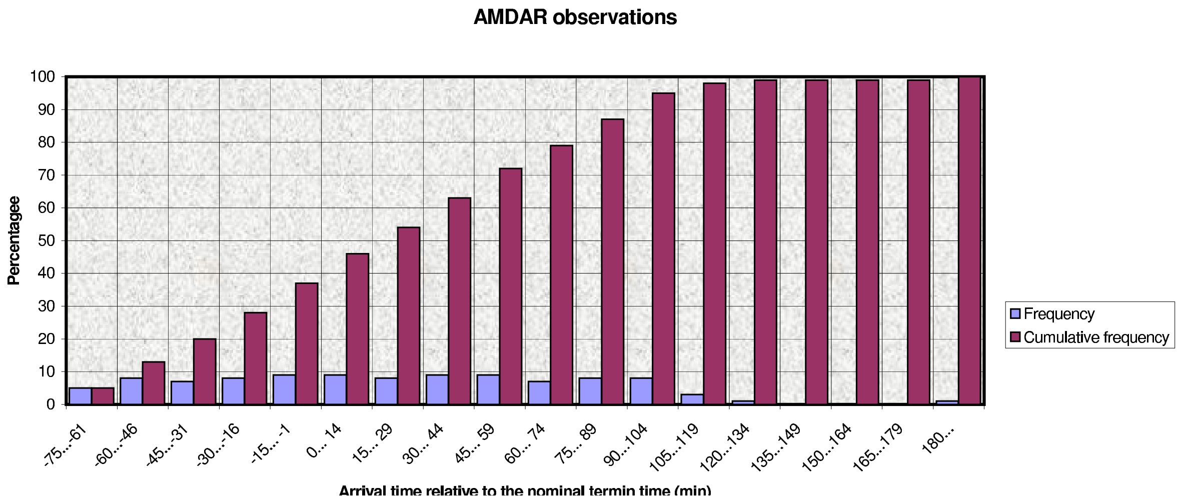

Figure 5: As Figure 2, but for AMDAR observations.

Looking at the BUOY observations (Figure 6) we see that they arrive rather slowly and the number of observations increases all the time when time goes on. The percentages of all observations for 90, 120 and 150 min are 43%, 57% and 67%. Note that 23% arrive later than three hours after the nominal observation hour. So the slow arrivals are not totally due to the asynoptic nature of BUOY observations, because the data window is supposed to be ±90 min. The statistics show that if all BUOY data are required, the needed cut-off time would be too long for operational purposes. This calls for considering a re-assimilation cycle, because the BUOY data come from the oceans, where the data coverage is not too dense.

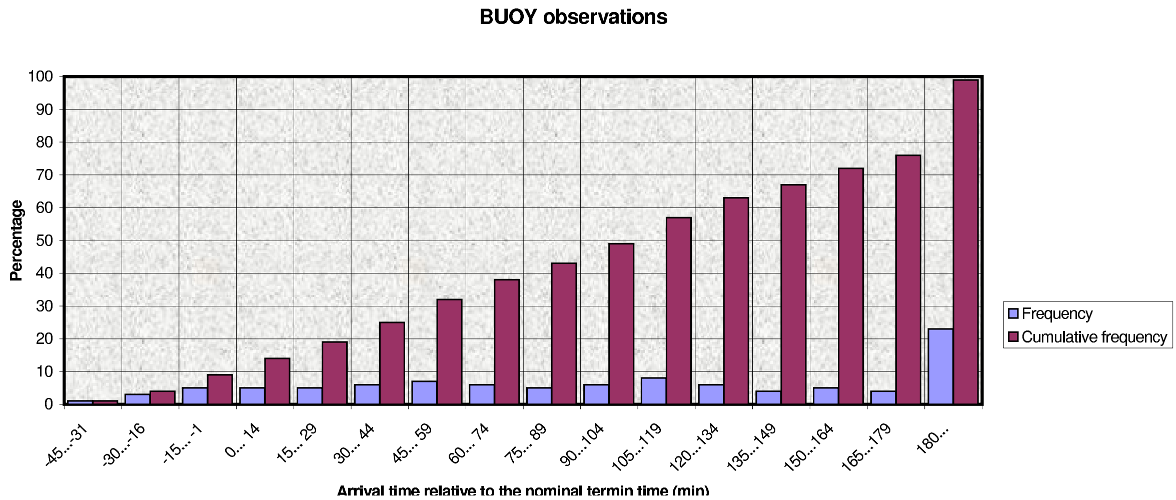

Figure 6: As Figure 2, but for BUOY observations.

The SATOB data arrive very quickly (not shown) and the cut-off time of 90 min would be enough to assimilate all SATOB data received at FMI at the moment.

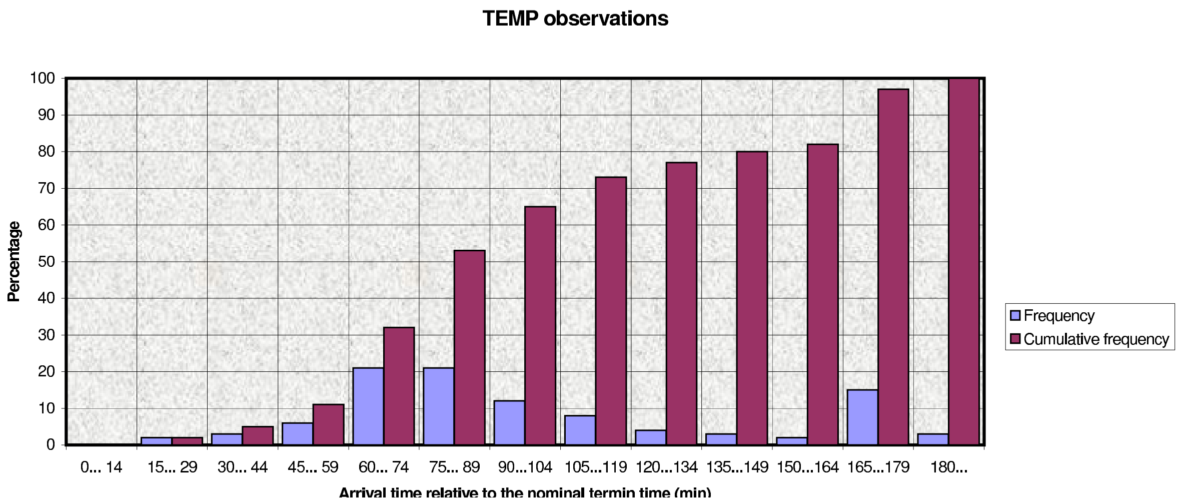

Figure 7: As Figure 2, but for TEMP observations.

3.2 Sounding data

Next the TEMP (Figure 7) and PILOT (Figure 8) soundings are discussed. They are the most important conventional observation types for HIRLAM, because they give vertical profiles of the atmosphere. From Figure 7 we see that a large part of soundings have arrived in two hours (71%), but there is a clear second maximum in receiving rates between 165 and 179 min after the nominal assimilation hour. Manual inspection revealed that many Russian soundings regularly arrive at this time. Because the arrival times of PILOT data are (Figure 8) recorded with an interval of one hour, we don't get as exact a picture for the PILOT soundings, but it seems that most PILOT soundings have arrived in 150 minutes.

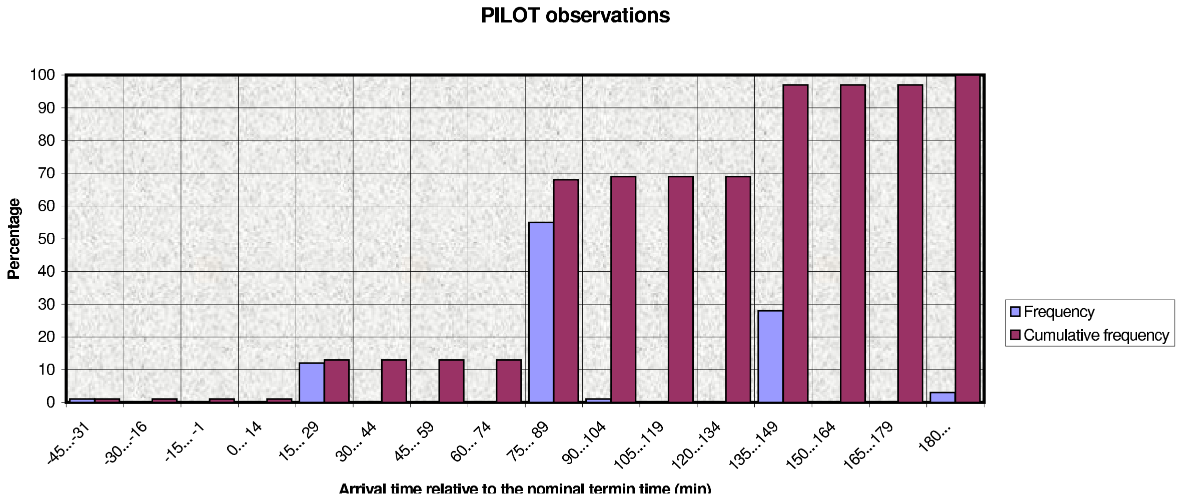

Figure 8: As Figure 2, but for PILOT observations.

Making a complete PILOT or TEMP sounding takes quite a long time, and therefore it is a common practice that the observing stations send the observation in parts. This means that the one, which comes last, is normally the most complete one. In the previous figures (Figure 7 and Figure 8) the most complete (maximum number of levels) is recorded. To get an idea, what information is available earlier, the arrival statistics for TEMP and PILOT are computed in two different ways. We record the arrival time of the most complete sounding (marked as "best"), but in addition we record the arrival time of the first observation for every station (marked as "first").

The cumulative frequency of the arrival times of the "first" and "best" TEMP sounding is shown in Figure 9. There is clear difference in the two statistics. For instance, if we select the cut-off time of two hours, we get at least a part of a sounding from 88% of the sounding stations, while the corresponding number is 71% for the "best" observation.

Because the arrival times of PILOT data are recorded with an interval of one hour, we don't get an exact picture for the PILOT soundings. However, Figure 10 suggests that discussion concerning the TEMP data is valid also for the PILOT data.

The previous discussion raises two questions. First, what is the geographical distribution of the observations arriving late, and secondly, how much information does the "first" observation contain.

First the geographical distribution of the soundings is discussed. For reference, Figure 11 shows the total number of TEMP observations for each station during the examined time period. Supposing two soundings a day, the maximum number of observations would be 31. Figure 12 show shows the number of "best" observations, arriving later than 120 minutes on different stations. It can be seen that typically the "best" observation from most Russian stations come later than 2 hours after nominal the observation time. On the other hand, it is encouraging to note that soundings from North America arrive very quickly. Looking at the same statistics for

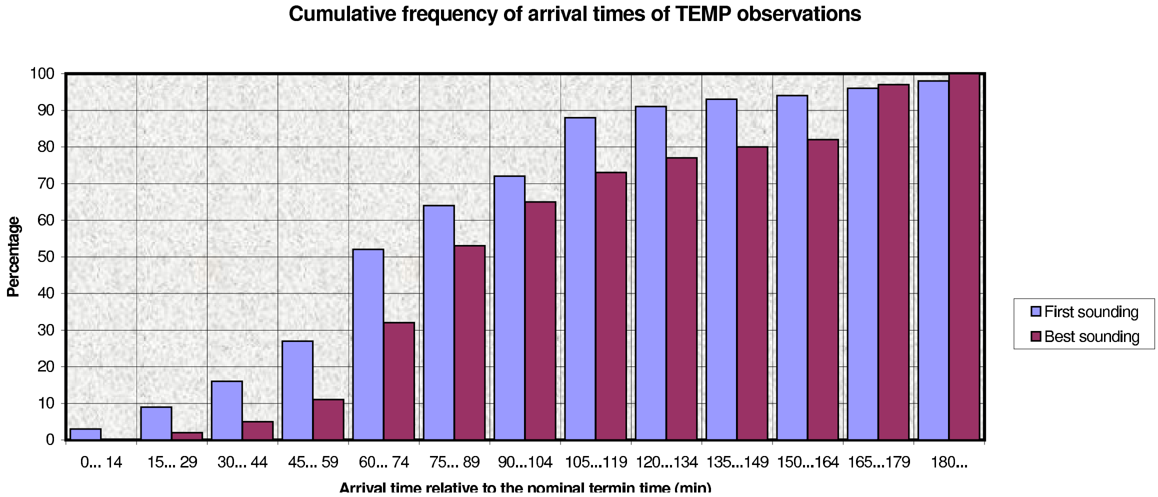

Figure 9: Cumulative frequency, in 15 min time slots relative to the nominal assimilation hour, of arrival times of the "first" and "best" TEMP sounding. For meaning of "first" and "best", see the text.

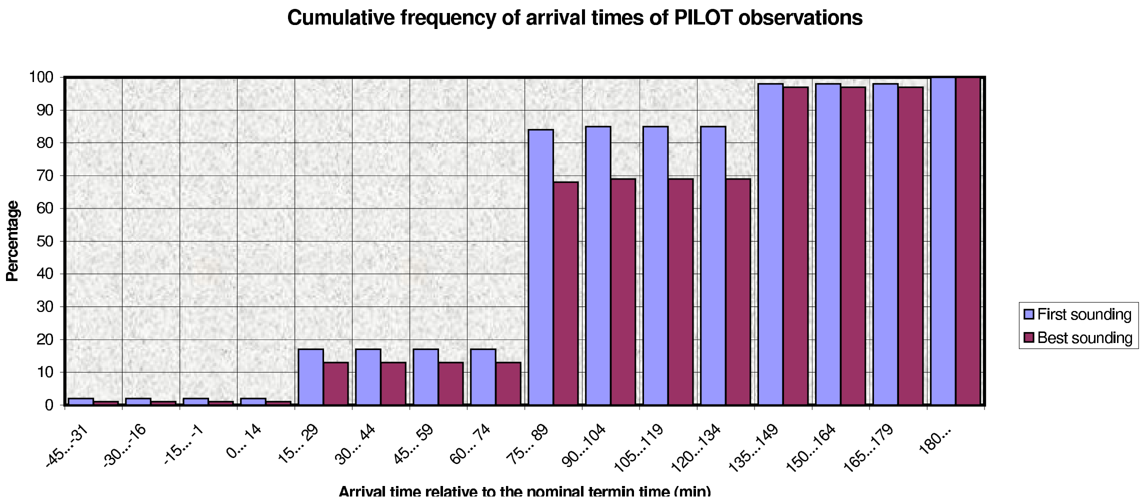

Figure 10: As Figure 9, but for PILOT observations.

the "first" observation (Figure 13), the situation is much better: in the mean, we get quite a good data coverage when using a two-hour cut-off and accepting the "first" TEMP observation. It is worth mentioning that there are several stations in Europe, from which observations arrive rather slowly. However, the observing network in this area is rather dense.

The next question is, how much information does the "first" observation typically contain. A quick manual inspection of the "first" soundings of the Russian stations shows that they normally contain parts A, B and D of the TEMP sounding. This means that for HIRLAM purposes we get a lot of information from these "first" observations. Of course a re-assimilation with a longer cut-off guarantees the best possible background field for the next cycle. The "best" sounding typically contains all parts (A, B, C, D) of the TEMP message.

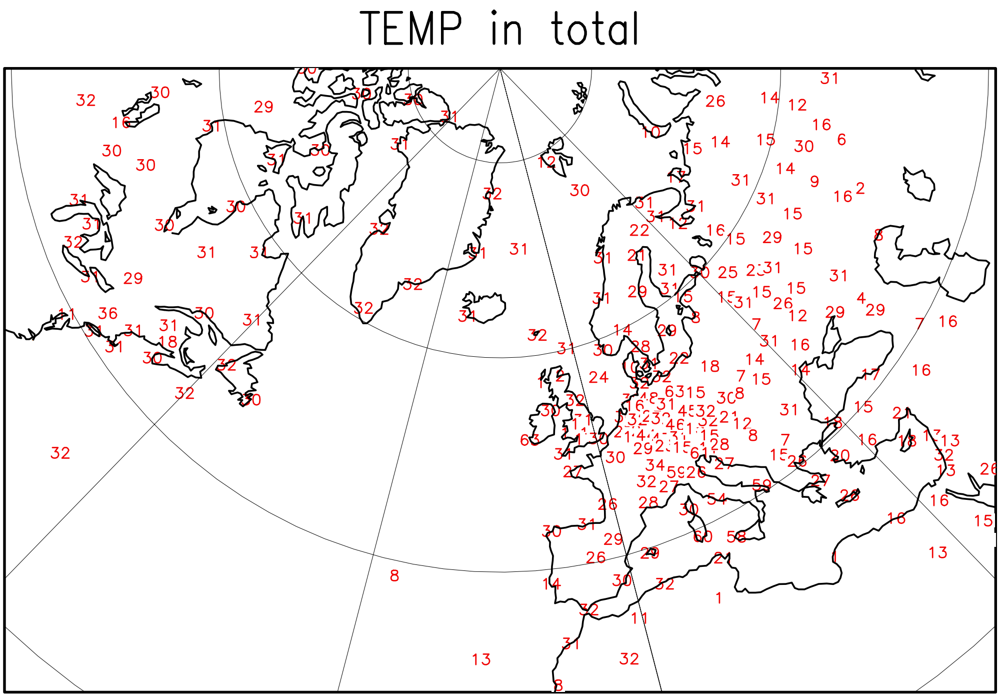

Figure 11: The total number of TEMP soundings on different stations.

In addition to statistical values, the number of observations in a long and short cut-off data assimilation run were checked in four randomly chosen cases, one for every main synoptic hour. The dates of these cases were 24 Oct. 00, 29 Oct. 06, 29 Oct. 12 and 29 Oct. 18 UTC runs. The cut-off times were two and six hours. Table 2 shows the number of active observations in the short and long cut-off data assimilation runs in these cases. It can be seen that especially the number of soundings and aircraft reports increases when applying the long cut-off data. This supports the statistical results shown earlier. On the other hand, the number of BUOY data does not increase with time, in contrast to the Figure 6. This may be due to the screening of observations in data assimilation, which is not taken into account in Figure 6. The effect of extra observations on the quality of analysis and forecasts is outside the scope of this study.

4 Discussion

The statistics of the arrival times for all observation types are summarized in Table 3. It shows the percentage of data received at cut-off times of 60, 90, 120, 150 and 180 min. As mentioned earlier, a three hour assimilation cycle is supposed and all the eight assimilation hours of the day are taken into account.

Table 3 and the results in the previous chapter show that SYNOP, SHIP, AMDAR and SATOB observations arrive very quickly. For them a 120 min cut-off time is sufficient, and even a shorter one could be considered. AIREP and especially BUOY data come slowly, but even if they contain only one level data, they are important, because they come from oceans, where the observation network is sparse.

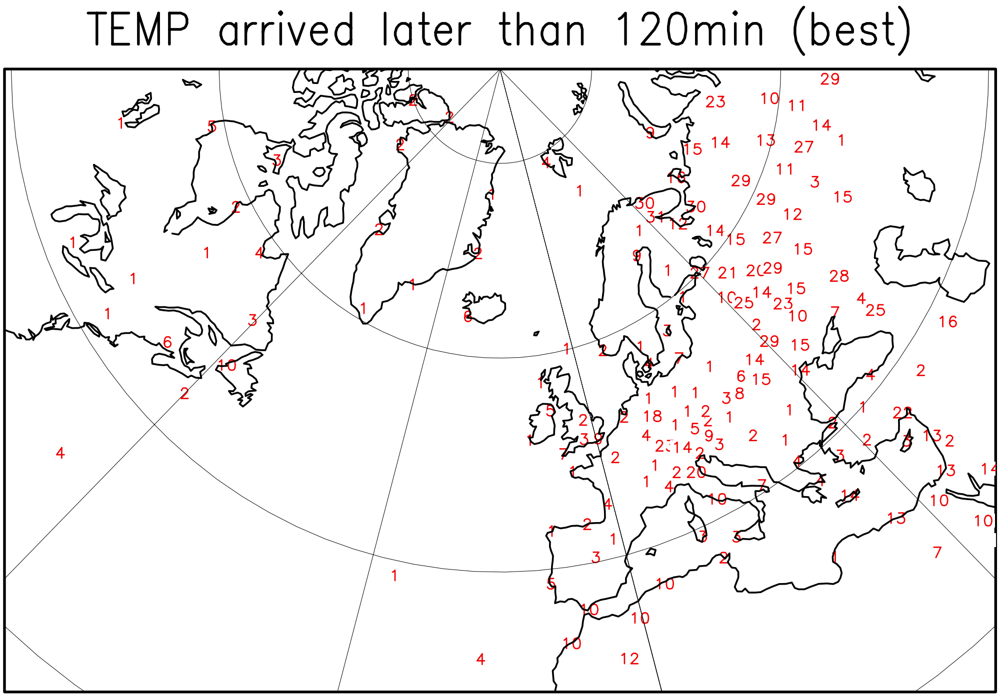

Figure 12: Number of "best" TEMP soundings at different stations, arriving later than 120 minutes.

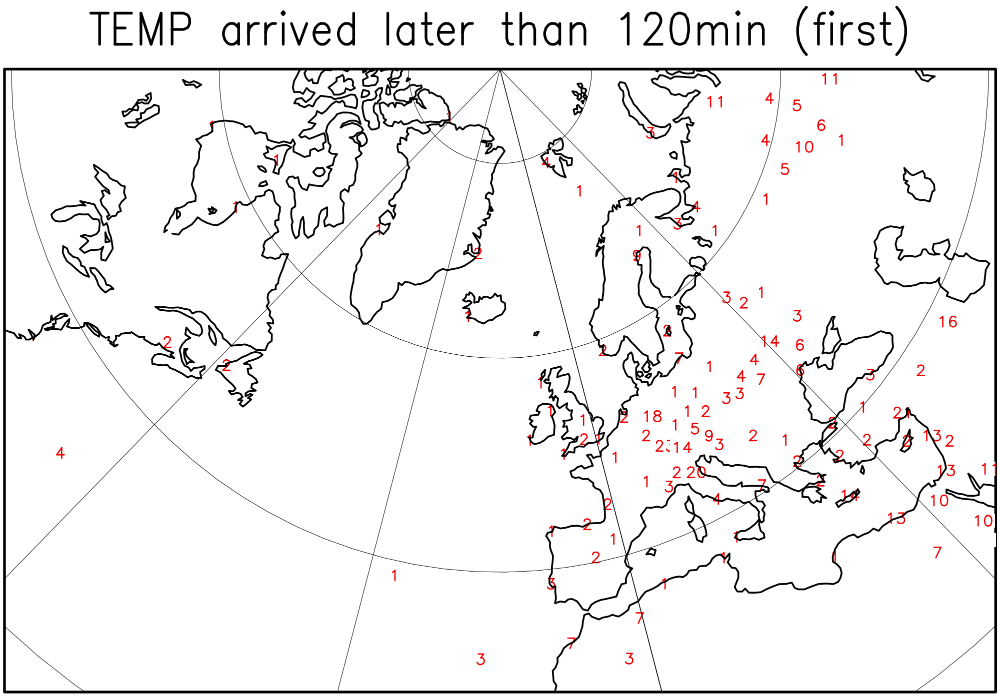

Figure 13: As Figure 12, but for "first" observations at different stations.

Table 2: The number of active observations in data assimilation in four situations (24 Oct 00 UTC, 29 Oct 06 UTC, 29 Oct. 12 UTC and 29 Oct. 18 UTC) when using 2 hour cut-off and 6 hour cut-off.

	19 Oct	**00 Z**	**29 Oct**	**06 Z**	**29 Oct**	**12 Z**	**29 Oct**	**18 Z**
	short	**long**	**short**	**long**	**short**	**long**	**short**	**long**
temp	151	182	26	33	133	144	18	19
pilot	2	3	4	15	3	7	3	16
synop	2055	2058	2387	2430	2363	2402	2376	2397
ship	330	341	328	374	298	350	305	331
aircraft	452	582	904	1330	1216	1709	1327	1653
buoy	108	112	117	122	110	114	100	101

Concerning sounding data (TEMP and PILOT), quite a good data coverage is achieved with a cut-off time of two hours, although the most complete observations especially from Russian stations arrive later. Almost a complete data set is achieved from North America within two hours.

Taking into account the fact that the HIRLAM forecast must be ready at the latest three hours after the observation time, cut-off time cannot be much longer than two hours. The results here show that two hours is sufficient to gather observations, but re-assimilation should be considered to guarantee the best possible background field for the next cycle on this very large area.

Table 3: Percentage of all observations arrived by given time for different observation types. For "first" and "best" see the explanation in the text.

	60 min	**90 min**	**120 min**	**150 min**	**180 min**
SYNOP	97	98	98	98	98
SHIP	85	90	95	95	95
AIREP	54	76	84	94	97
AMDAR	72	87	98	99	99
BUOY	32	43	57	67	76
SATOB	0	99	99	100	100
TEMP("first")	27	64	88	93	96
TEMP("best")	11	52	71	78	95
PILOT("first")	17	84	85	98	98
PILOT("best")	13	67	68	96	96

From the operational point of view there are two additional aspects, which are in favour of the re-assimilation cycle.

First, if the re-assimilation cycle of the previous six hour old run and the assimilation of the previous three hour old run are both executed just before the next cycle, new frame boundaries have probably arrived from ECMWF and all the three runs can use the same lateral boundaries. For example, the 00 UTC re-assimilation cycle and 03 UTC cycle would use the 00 UTC ECMWF boundaries, as well as the 06 UTC long forecast.

Second, technically it is easy to organize the "cold start" possibility, if the 00 UTC family, for example, contains 18 UTC re-assimilation, 21 UTC assimilation and 00 UTC assimilation and long forecast. For example, if 18 UTC run fails for some reason, it is enough to start the 00 UTC family. It will first run 18 UTC re-assimilation ("cold start" is possible, because 18 UTC ECMWF frames with full fields at 18 UTC, have already arrived), then 21 UTC assimilation and after that the actual 00 UTC assimilation and long forecast.

In the "re-assimilation model" the logical structure of one RCR run would schematically be the following, taking 12 UTC hour run as an example:

- 12 UTC Cycle
  - Start (check libraries, constant files etc.)
  - Check the latest boundaries
  - 06 UTC re-assimilation
    - Get 06 UTC observations
    - Analysis
    - 6 hour forecast
  - 09 UTC assimilation
    - Get 09 UTC observations
    - Analysis
    - 6 hour forecast
  - Check the latest boundaries
  - 12 UTC assimilation
    - Get 12 UTC observations
    - Analysis
    - 54 hour forecast

5 Conclusions

The arrival times of different conventional observation types at FMI for a two week period have been studied.

The following conclusion concerning the cut-off time can be made.

- The operational Hirlam runs can be safely be run using a cut-off time of two hours.
- If a longer cut-off time is used, a better coverage of TEMP, PILOT, BUOY and AIREP observation is achieved. The question, does this radically improve the preliminary field for the next cycle, remains open in this study.
- To get the best possible background field for the next cycle, the re-assimilation using a longer cut-off time should be considered.
- Every time before a new observation type will be implemented, the arrival time statistics should be studied.

From the meteorological point of view it seems that a RCR system with a three hour cycling, a two-hour cut-off time and a re-assimilation of the main synoptic hours is an appealing solution. Of course, the available computer resources and the simplicity of the running system must be taken into account.