

Experiences with Quality Evaluation of AMDAR Observations

Jitze van der Meulen (KNMI, the Netherlands)

abstract

Quality Evaluation of Eumetnet AMDAR data are performed starting in 1999. Experiences with quality evaluation practices and results are presented and discussed. This includes the methods used and the results. Evaluation reports inform on short and long term trends in quality and performances of these aircraft based observations. These reports advise AMDAR data operators on possible improvements. Moreover the archived reports will support data analyses of the upper-air for climate applications.

In this paper an historic background on the quality evaluation process is given, as well as the specific issues experienced to be highly relevant, but not foreseen as being trivial.

General

In-situ upper air observations are one of the basic requirements to assess the dynamical behavior and state of the atmosphere, used for forecasting and climate applications. Although remote sensing techniques provide promising results with increasing impact, in-situ observations are still essential. A number of techniques to measure the various physical quantities in the atmosphere, like with radiosondes are well known and also quite well understood. Also, automatic observations on board aircraft have a long historic background. Apart from the challenges to determine the variables accurately enough, data transmission to a ground station is critical for timely observations.

To reduce timeliness in data communications, a sophisticated data relay system was already operational in the past, and based on communications by satellites (ASDAR). Although successful, the high costs and therefore a limited number of observations prevented services to co-operate in this ASDAR programme. As a result, the already close cooperation with airlines was extended by using the aircraft to ground data communication facilities, reducing the costs significantly. Moreover, this aircraft meteorological data relay system (AMDAR) provides the ability to use any type of aircraft, provided that dedicated software is installed to produce AMDAR data, as required. In practice each aircraft type requires its own dedicated AMDAR software package¹, so implementation and maintenance requires quite complex data management and quality assessment apart from the certification procedures.

Observational data, or "AMDAR" data are typical air temperature, wind speed and wind direction, and also air pressure. Another relevant variable, which may be derived, is turbulence, today defined as the

¹ see <http://www.wmo.int/amdar>

eddy dissipation rate (EDR). Although many aircraft are able to determine its horizontal and vertical position using GNSS, the altitude is traditionally derived from the determined static pressure. It's usually expressed as flight level, but a better expression is pressure altitude. So in fact, a barometric quantity is expressed in the unit of length using the International Standard Atmosphere (ISA). The last decade autonomous sensors to determine humidity and icing are installed on a number of aircraft too. It is foreseen that data from these sensors will demonstrate increased impact and further improvement of weather forecasts, in particular for aviation.

Relevant to note here is that all the data are in fact third party data. Although the functional specifications of the sensors, the data processing and AMDAR software is well documented and maintained, the quality of the received data must be assessed by monitoring and evaluation. Evaluation is in fact a critical process, because procedures for data analyses are not trivial and requires talented approaches.

Short history of the quality evaluation project of the Eumetnet AMDAR programme.

Eumetnet, the European Meteorological Services Network is formed in 1996 by European national meteorological services. This cooperation made it possible to organize observational programmes of an international nature. Because many aircraft fly international route, the implementation of a European AMDAR programme was a logical action. Moreover, organizing profile observations in areas, where radiosonde data is sparse improved the homogeneity of upper air data over Europe, *i.e.* the areas for which the limited areas NWP models are sensitive. Starting in 1999 with a daily amount of about 15000 single level observations, reported by about 150 aircraft, nowadays 50000 observations are reported daily by more than 1200 aircraft. Depending on the phase of flight (ascending, descending or level flight), sets of data are used to generate profiles or horizontal data tracks. The system is configured to plan observations on forehand, based the daily flight charts of the aircraft, provided by the airlines. Using such a configuration planning service a significant improvement of the homogeneity of upper air data is established. Note that homogenization is important in terms of both space and time.

Data quality evaluation practices.

The process of quality evaluation consists of a continuous cycle mechanism based on quality monitoring and quality analyses research resulting in quality improvement proposals. These proposals are reported to the Eumetnet AMDAR Technical Coordinator and the national AMDAR operators. Such a feedback mechanism is in fact a crucial success factor to improve the performance of the system in general. It also mitigates impacts of serious outliers. Moreover, quality evaluation will result in improving the system on the whole, including software and applications.

Key issues for AMDAR quality evaluation practices.

In practice quality monitoring focuses on three specific areas:

1. Quality of the metadata, like location and time of observation
2. Quality the reported variables (derived from measurants, provided by the sensors), like air temperature, mixing ratio (humidity), wind vector (speed and direction)
3. Performance, like availability and timeliness.

Note that the measured static air pressure is used to derive pressure altitude, so a vertical position parameter. The quality of positional information is experienced to be extremely critical for NWP. Especially in data sparse areas, incorrect position information (*e.g.* temperature and wind associated with an incorrect position) may affect NWP dramatically and enforce unwanted instabilities.

The quality of the reported variables can be best monitored by using NWP output as background reference. It is relevant to be aware that such a reference is not representing true values with proven traceability to SI. Moreover these model output have uncertainties which are larger than the expected measurement uncertainties. Nevertheless, model output is proven to be very useful in determining continuous biases and incidental significant outliers. Typically, trends in biases per aircraft help to filter out data from such aircraft before its uncertainty reaches a critical limit. For operational practices first guess fields may be used, but also analysis results. Both types of references will deliver data sets demonstrating bias trends to initiate the alert state. In principle limited area models (like HIRLAM or HARMONIE in Europe) with a high resolution in space and time have preference, but it is experienced that global models (like from ECMWF) provide comparable results.

Meta data evaluation

In practice the quality control of reported metadata is extremely challenging. Biases in pressure altitude (*e.g.* due to incorrect pressure readings or altitude settings) are hard to detect. An appropriate test is to control if the aircraft is at the same altitude as the runway just before takeoff or after landing. For this check an accurate air pressure value reported by the airfield is essential. Figure 1 gives a typical example of such a check. Horizontal and vertical checks, which may help to find suspect positions is to evaluate the time dependent behavior of a series of reported positions. Aircraft flying with a horizontal or vertical speed higher or lower than expected should be flagged.

To report errors in reported time of observations (or any trends) are even harder to detect. Errors in date stamps are registered, typically just after 0 UTC. Some aircraft report timestamps in future, which can easily be flagged of course.

Evaluation of derived variables

As stated, the derived variables are compared to a background reference based on the output of NWP models. Such a reference is constructed in GRID fields and in various layers, each with their own resolution. Each gridpoint represents a specific horizontal and vertical position. The forecasted value of that point is calculated for specific timestamp ahead or with a series of such values (*e.g.* for each of the following whole hour). These series are frequently updated, *e.g.* every six hours. To be able to intercompare with an observation (at random position and time) it is necessary to implement an interpolation scheme to have a data point in space and time representative for the location of the

aircraft. If such an interpolation scheme is not well defined, unexpected biases may appear and which are not due to the observations.

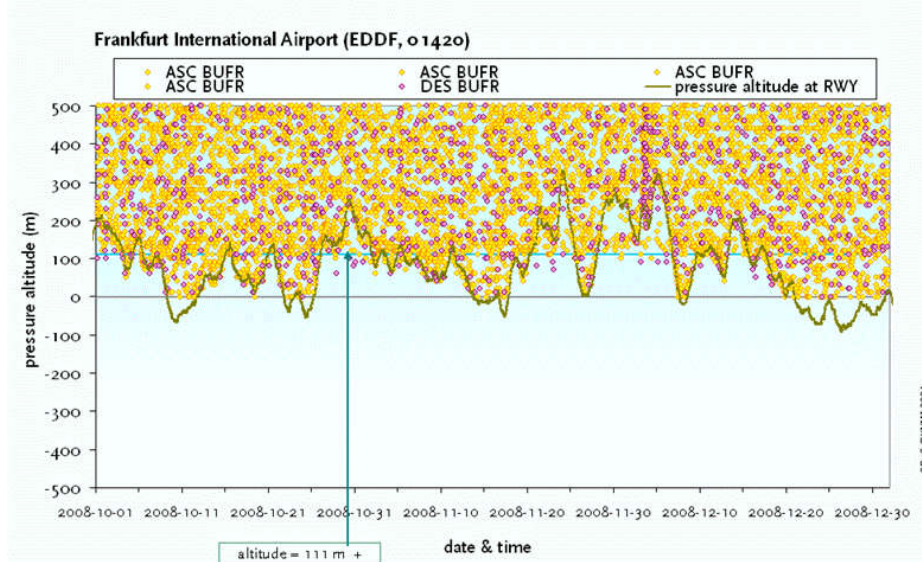


Fig. 1 Example of reported pressure altitudes presented together with the pressure altitude of the runway. Aircraft with reported pressure altitudes below the green line are incorrect.

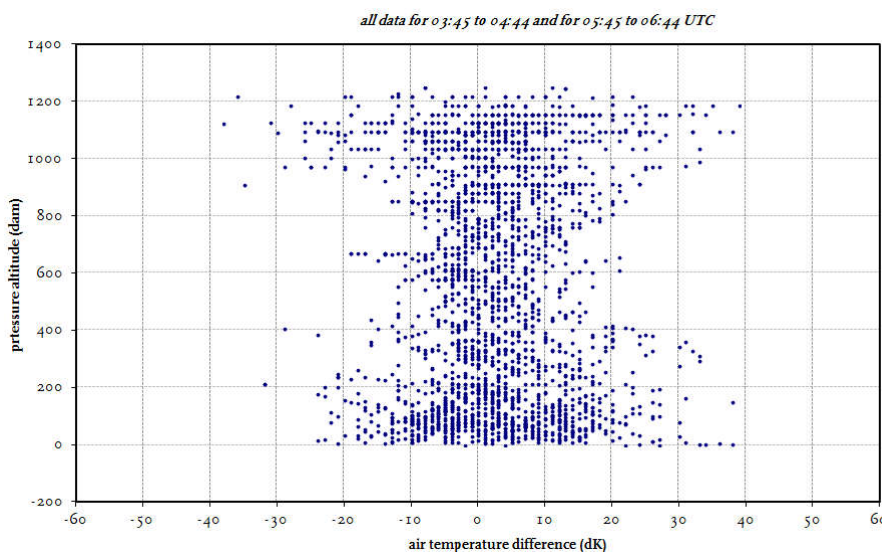


Fig. 2 Example of derived air temperature differences as a function of pressure altitude. Note its noisy behavior.

(note: Pressure altitude values are presented in dam, temperature in dK)

Comparing single level observations with such interpolated model results for each observation apart will give too noisy results (see fig. 2). The best method is to analyse for each aircraft datasets based on one or more days of observation. This method, and based on several model runs as well, will deliver medians or average differences, to be regarded as representative for the present bias of that aircraft. Due to the noisiness of these differences, largely due to the uncertainty of the model and the interpolation method used, the derived standard deviation has no meaning and should not be considered to determine the uncertainty of a single observation itself. Daily analyses, but also longer term evaluation provides details in e.g. trends in biases, and also differences between the various aircraft types and subtypes. Such long term analyses also provide details of the statistical distribution of the differences and for each variable. Because the phase of flight is reported, data analyses of differences associated with a particular phase of flight (e.g. ascending versus descending) can provide typical details which may help to modify data before

ingestion into NWP. Such long term evaluation is often more useful because also the daily results demonstrate a quite noisy behavior (see fig. 3). Nevertheless the set of daily data results (e.g. the daily medians), analyzed over a longer period (e.g. a month) will provide useful statistics.

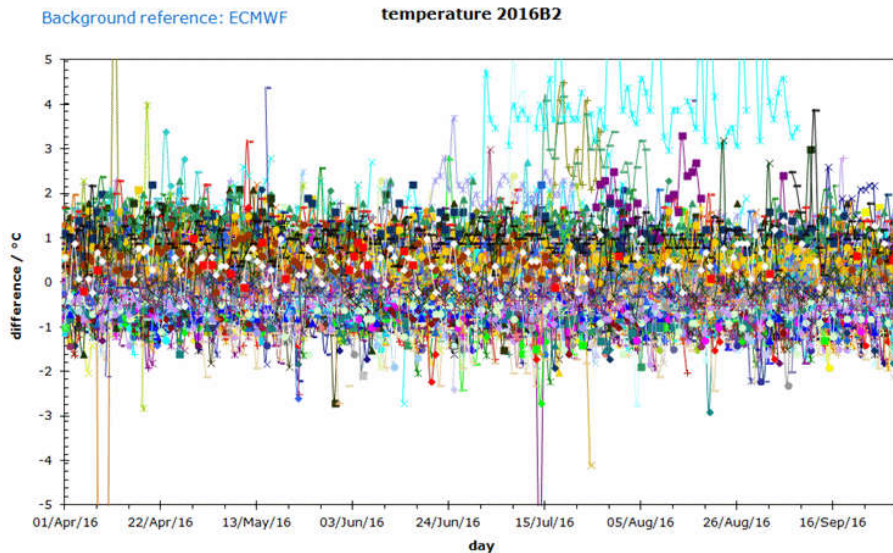


Fig. 3 Example of derived daily air temperature differences (medians) as a function of time. Note its noisy behavior and the typical outliers.

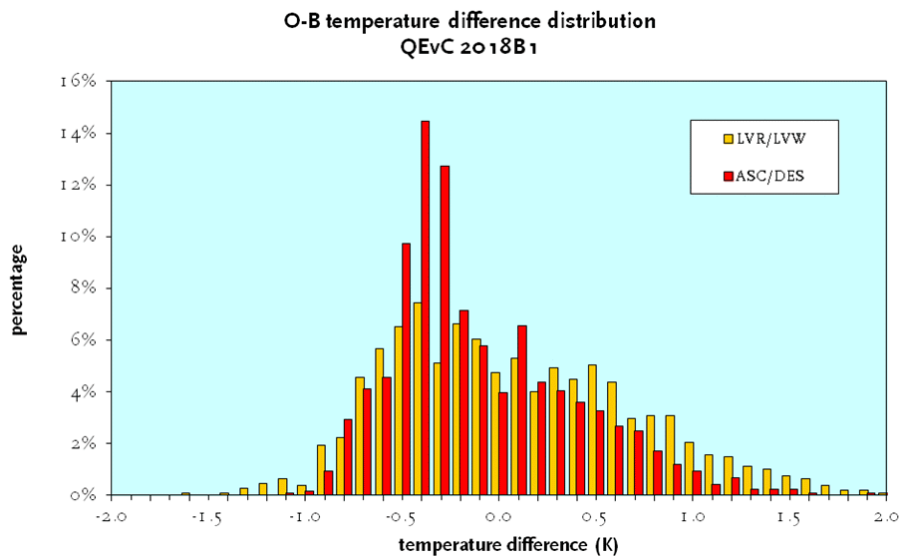


Fig. 4 Statistical distribution of daily air temperature differences (medians) as a function of temperature. In the figure results for ascending and descending data (ASC/DES) and for Level Flight data (LVR/LVW) are presented separately. Note the right warm bias wing of the distribution. (period Oct. 2017 - March 2018).

Such results will show distributions of these differences associated to each aircraft (see fig. 4). On the other hand the single level observations can be used too to show a total statistical distribution for all aircraft (e.g. for a quarter). Temperature and wind speed differences show normal distributions. Wind direction differences however show a typical behavior as shown in fig. 5.

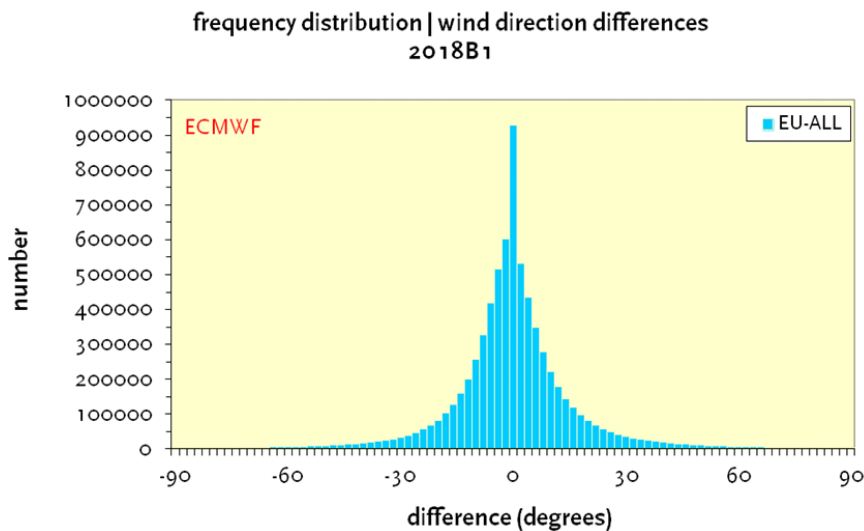


Fig. 5 Example of the distribution for the wind direction differences based on the total amount of single level observations (period Oct. 2017 - March 2018).

These evaluations are found to be very useful in finding typical biases associated with subtypes of aircraft. Figure 6 shows an example for Boeing 737, subtype 705, demonstrating a significant bias of about +1 K. It is found that more than 50% of all investigated aircraft have a negative bias (too cold), but there is a small set of aircraft demonstrating a serious warm bias (too warm), which is almost out of specification (about +1 K), see fig. 4.

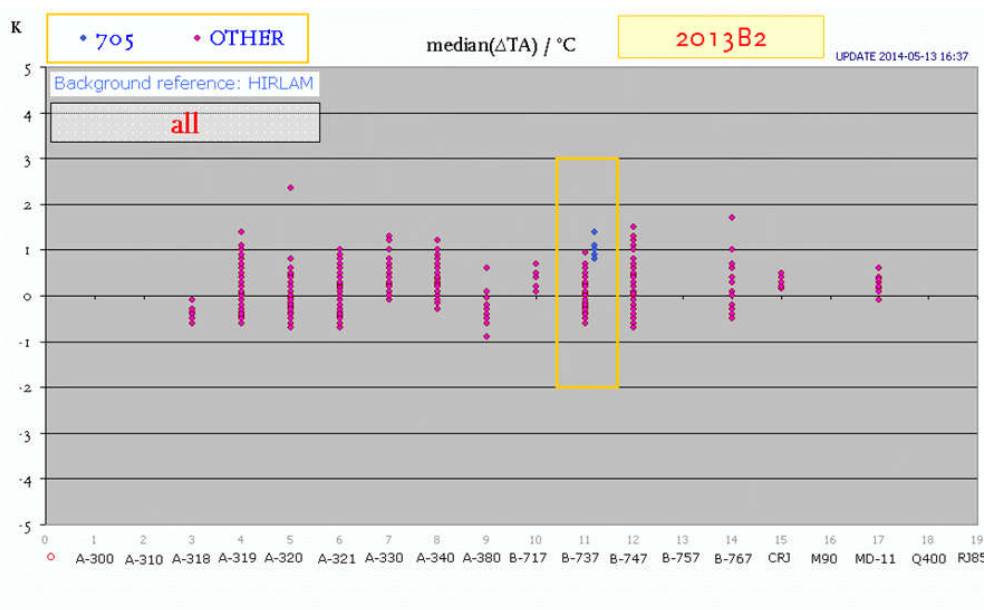


Fig. 6 Example of the distribution for the air temperature differences based separated over sets of aircraft types. Note the very warm biases of B737, subtype 507 (period Apr. 2013 - Sept. 2013).

Timeliness

Apart from the quality of data itself, its usefulness for daily use is important. For NWP and nowcasting applications, timeliness (the delay between the time of observation and the reception at the users' desk) is critical. Buffering data on board and the delay in waiting until a ground station is available to receive data may cause significant delays (the timeliness due to communication between a ground station and the MSS hubs are found to be negligible, *i.e.* < 1 min). A typical example of how timeliness is distributed is

show in fig. 7. Clearly, level flight, ascending and descending data demonstrate quite different figures, which can be understood quite well. An overview of the daily timeliness figures for the Eumetnet AMDAR data is shown in figure 8.

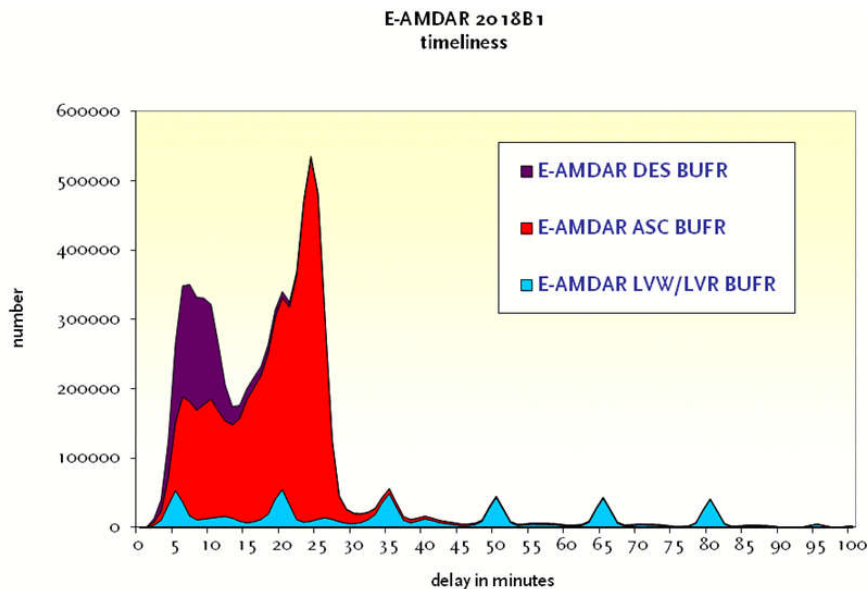


Fig. 7 Example of the distribution for timeliness as a function of the delay and split for the different phases of flight. The typical 'spikes' are due to buffering of data for a fixed period (20 or 15 min) to reduce transmission costs (period Oct. 2017 -

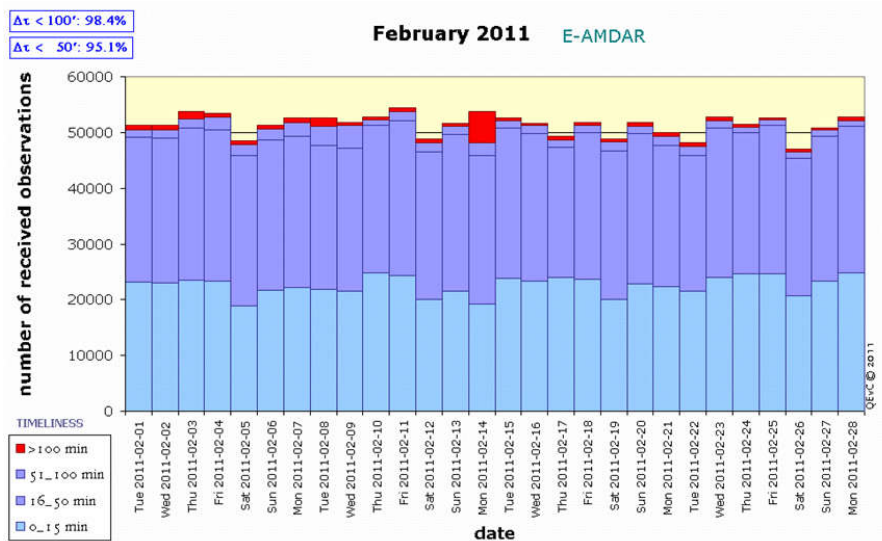


Fig. 8 Example of the daily behavior of the timeliness distribution. The red area represents data with a timeliness > 100 minutes; the light blue area is for data received within 15 minutes. (period Febr. 2011).

Humidity

The last couple of years a number of aircraft are equipped with instruments measuring the mixing ratio (of vapor). The technology is based on measuring the optical behavior of humid air. The Eumetnet AMDAR fleet is provided with 9 of such sensors². A typical example of observed mixing ratio data as a function of altitude is shown in fig. 9. Considering the %RH values obtained from the mixing ratio and by using the derived air temperatures, we found very stable results with uncertainties less than 5 %RH (see fig. 10).

² WVSS type II, by SpectraSensors Inc.

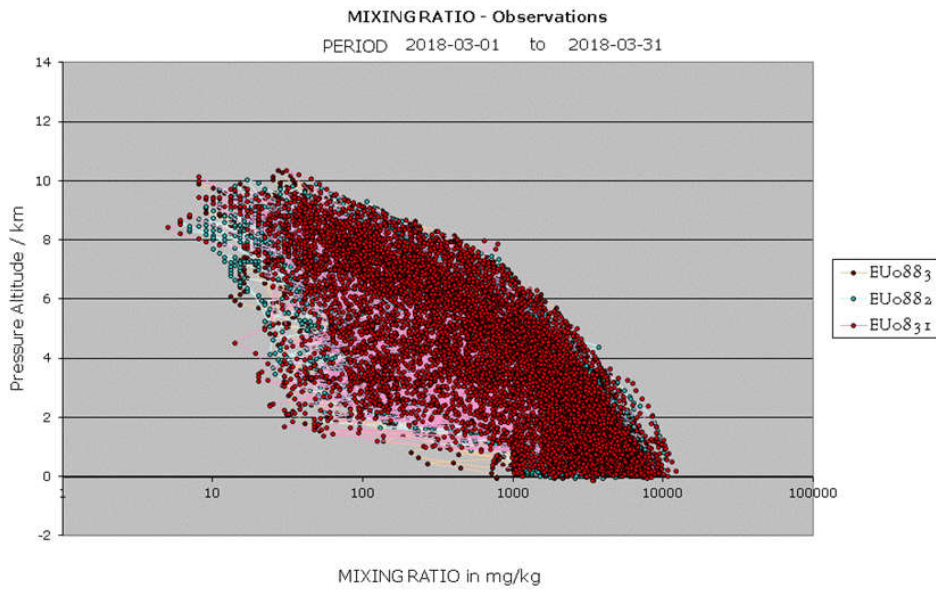


Fig. 9 Example of observed mixing ratio value as a function of pressure altitude (three aircraft shown; period March 2018).

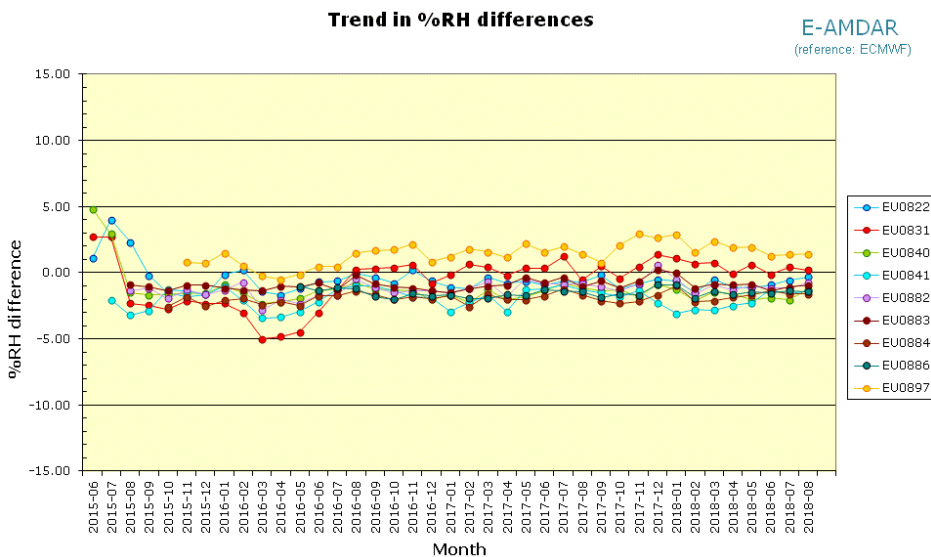


Fig. 10 Results for the monthly differences in %RH for the 9 aircraft equipped with the humidity sensor (period June 2015 - Aug. 2018).

Conclusion

Quality evaluation of AMDAR data has demonstrated its ability to monitor data using NWP as background references. Experiences with further data analyses and research are used to improve the quality of the data in general and to support further data management. In practice NWP forecast data can be used to develop a monitoring system that is able to flag data to indicate its usefulness (a quality status). Such a service may result in further improvements of the assimilation of data in NWP itself and the processing of data to generate near real-time profiles of the atmosphere. Moreover quality evaluation results will be of particular interest in case of (future) climate research of the upper air, based on aircraft based observations.