RMetS
Royal Meteorological Society

On the use of a Huber norm for observation quality control in the ECMWF 4D-Var

Christina Tavolato[a,b] and Lars Isaksen[a]*

[a]*European Centre for Medium-Range Weather Forecasts, Reading, UK*
[b]*Department of Meteorology and Geophysics, University of Vienna, Austria*

*Correspondence to: L. Isaksen, ECMWF, Shinfield Park, Reading RG2 9AX, UK.
E-mail: lars.isaksen@ecmwf.int

This article describes a number of important aspects that need to be considered when designing and implementing an observation quality control scheme in an NWP data assimilation system. It is shown how careful evaluation of innovation statistics provides valuable knowledge about the observation errors and help in the selection of a suitable observation error model. The focus of the article is on the statistical specification of the typical fat tails of the innovation distributions. In observation error specifications, like the one used previously at ECMWF, it is common to assume outliers to represent gross errors that are independent of the atmospheric state. The investigations in this article show that this is not a good assumption for almost all observing systems used in today's data assimilation systems. It is found that a Huber norm distribution is a very suitable distribution to describe most innovation statistics, after discarding systematically erroneous observations. The Huber norm is a robust method, making it safer to include outlier observations in the analysis step. Therefore the background quality control can safely be relaxed. The Huber norm has been implemented in the ECMWF assimilation system for *in situ* observations. The design, implementation and results from this implementation are described in this article. The general impact of using the Huber norm distribution is positive, compared to the previously used variational quality control method which gave virtually no weight to outliers. Case-studies show how the method improves the use of observations, especially for intense cyclones and other extreme events. It is also discussed how the Huber norm distribution can be used to identify systematic problems with observing systems.

Key Words: observation quality control; Huber norm; robust estimation; variational data assimilation

Received 24 March 2014; Revised 16 July 2014; Accepted 19 August 2014; Published online in Wiley Online Library

1. Introduction

Quality control (QC) of observations is an important component of any data assimilation system (Lorenc and Hammon, 1988). Observations have measurement errors and sometimes gross errors due to technical errors, human errors or transmission problems. The goal is to ensure that correct observations are used and erroneous observations are discarded from the analysis process. It has long been recognised that a good QC process is required because adding erroneous observations to the assimilation can lead to spurious features in the analysis (Lorenc, 1984).

In data assimilation, the use of departures of observations (o) from the short-range (background, b) forecast is an integral part of the QC. If observations, evaluated over a long period, systematically or erratically deviate from the background forecast they should be blacklisted, i.e. not taken into account at all in the analysis (Hollingsworth *et al.*, 1986). The remaining observations are, for each analysis cycle, also compared against the background and rejected if the background departures are large. Often departures, normalised by the expected observation error, are assumed to follow a Gaussian distribution. This means outliers are statistically very unlikely and will unjustly get the same full weight in the analysis as correct observations, increasing the risk of producing an erroneous analysis by using incorrect observations. This is usually resolved by applying fairly tight background departure limits which reject outliers. The background QC limits depend on the specified observation error and background error. For accurate observations and modern high-quality assimilation systems, these are both small, e.g. of the order of 0.5 hPa for automated surface pressure observations. So surface pressure observations will typically be rejected if they differ by more than about 4 hPa from background fields, corresponding to six standard deviations of normalised departures. In most cases this is reasonable, but for extreme events it may well happen that the short-range forecast is wrong by more than 4 hPa near the centre of cyclones. The QC decisions can be improved to some degree by introducing flow-dependent, more accurate, background errors, like the ones recently implemented at the European Centre for Medium-range Weather Forecasts (ECMWF; Isaksen *et al.*, 2010; Bonavita *et al.*, 2012). These errors would typically be larger near the centre of cyclones.

Section 2 of the article investigates the innovation statistics for some of the most important *in situ* observations. This leads to a discussion and description in section 3 of the gross error QC aspects which need to be considered for observations. The special problems that may occur for biases and bias correction of isolated stations is covered in section 4. After this general study of innovation statistics and gross error characteristics, section 5 describes a range of proposed probability distributions with fat tails which are candidates for innovation statistics and observation error specification. Based on the information presented in sections 2–5, it is found that a Huber norm (Huber, 1964, 1972) is the most suitable distribution to use. The method allows the inclusion of outliers in the analysis with reduced weight, because it is a robust estimation method. This is in contrast with a pure Gaussian approach where the analysis can be ruined by a few erroneous outliers. Section 6 covers the aspects which need to be considered when implementing a Huber norm QC in a numerical weather prediction (NWP) system. We describe how this is done in the Integrated Forecast System (IFS) at ECMWF, where it has been used operationally since September 2009. It is also explained how the background QC has been relaxed, and how observation error values have been reduced at the centre of the distribution to consistently reflect the Huber norm distribution. Section 7 presents general impact results and a number of case-studies.

2. Distribution of departure statistics for some important *in situ* observations

The main weakness of using background departure statistics for investigations of observation error distributions is that they are a convolution of observation and background information. Further information is required to uniquely determine the observation-related distribution, which is what we really are trying to estimate, as it is needed in the definition of the observation cost function. Despite this weakness innovation statistics are the most common observation-related diagnostics used in data assimilation. Additional research to identify if background errors are non-Gaussian is recommended, but it is outside the scope of this paper. Assuming the background error follows a Gaussian distribution, all non-Gaussian aspects of the innovation distribution can be assigned to the observation error distribution. Evaluation of the tails of innovation distributions is also likely to provide valuable information about the tails of the observation distributions.

The QC aspects are primarily related to small numbers of observations in the tails of the distribution. So to get a sufficiently large sample of relevant departure statistics, 18 months of data assimilation system departure statistics (February 2006 to September 2007) were used for these estimates. This was done for a large number of observation types, to determine the distributions that best represented the normalised departures for each of these sets. The model background fields are from the operational incremental four-dimensional variational (4D-Var) assimilation system (Courtier *et al.*, 1994) at ECMWF, taken at appropriate time ($\pm$15 min) and at T799L91 (25 km horizontal grid and 91 levels outer loop) resolution.

Figure 1 shows the departure distributions, normalised by the prescribed observation error, for a number of these observation types. The grey crosses represent the data counts for bins of width 0.1 in the range $\pm$10 of normalised departures. Traditionally this distribution would be plotted as a histogram, but crosses were easier to see on the figures. To put the focus on the tails of the distribution, the data are plotted on a semi-logarithmic scale. This means a Gaussian distribution shows up as a quadratic function, and an exponential distribution as a linear function. On the figure the best-fit Gaussian distributions (dashed-dotted line) are included. Figure 1(a) shows temperature data in the 150–250 hPa range for all Vaisala RS92 radiosonde measurements in the Northern Hemisphere (NH) Extratropics. A similar plot for the used data is shown in Figure 1(b) (used data are quality controlled data with more than 25% weight after applying the previously used 'Gaussian plus flat' distribution QC). The 'Gaussian plus flat' distribution QC and its implementation at ECMWF is described in Andersson and Järvinen (1999) (abbreviated below as AJ99). Vaisala RS92 radiosondes are known to be of very high quality with very low bias, very few gross errors, and with low random errors. Figure 1(c–f) show normalised departure statistics for other conventional observation types and their data distributions for the extratropical regions. Figure 1(c, d) show two different surface pressure observing systems (land surface pressure in the Southern Hemisphere (SH) Extratropics and ship surface pressure in the NH Extratropics), and Figure 1(e, f) show upper air and surface wind observations (aircraft winds from all levels in the NH Extratropics and winds observed by drifting buoys in the NH Extratropics).

The solid black curves on Figure 1 show the best-fit Huber norm distribution. The Huber norm, a Gaussian distribution with exponential distribution tails, is defined in section 5 (Eq. (1)). Because f in Eq. (1) is first-order continuous, the Huber norm distribution shows up as a quadratic function which smoothly transforms into a linear function in the tails of the distribution. It is seen that the background departure statistics are well described by a Huber norm distribution, because the data in the tails are in good agreement with the solid black curves. Indeed, these results indicate that the Huber norm distribution fits the data much better than a pure Gaussian distribution (dash-dotted curves). The 'Gaussian plus flat' distribution previously used in the operational assimilation system at ECMWF is included on Figure 1(a) as a thick grey curve. It is evident that the 'Gaussian plus flat' represents the tails of the normalised departure distribution very poorly for radiosonde observations. This is the case for all the variables shown in Figure 1, and for almost all other observation types that have been investigated (not shown). It is worth mentioning that a sum of Gaussian distributions does not produce a Huber distribution, so this is not the explanation for the fat tails.

It is noted that there is a factor of more than 1000 between the data counts in the tails (at 8–9 normalised departures). At the centre of the distribution (up to 2 normalised departures), departures are close to a Gaussian distribution for most observations. There are no indications of flat-tailed distributions, i.e. no indication of standard gross errors where the observed value is unrelated to the background field. There is rather an indication of an exponential distribution for many observations in the range 2–9 normalised departures. For the used data (Figure 1(b)), the departures to a large extent follow a Gaussian distribution. This is because the departures are from the pre-2009 operational assimilation system at ECMWF which applied the 'Gaussian plus flat' QC distribution, resulting in a sharp transition from full Gaussian weight to zero weight, as shown schematically in Figure 5(b) below.

3. Gross error QC aspects for observations

Extensive investigation of the normalised background departure statistics for many different observation types and parameters gave a useful insight into gross error aspects. Most distributions have fatter than Gaussian distributions beyond 1–2 normalised departures. The reason why there are only few examples of flat distributions in the tails may well be due to most observing systems now being automated. Automated systems reduce human-related gross errors like swapped latitude/longitude, east/west sign error and swapped digits. If innovation statistics from a station or platform show flat-tail gross error characteristics, it will often be due to a systematic malfunctioning that results in all observations being wrong. It is fairly easy to detect and eliminate ('blacklist') these observations via a pre-analysis monitoring procedure. They will then not be part of the observations presented to the analysis. We will now give some examples that highlight these issues.

3.1. *Chinese aircraft temperature observations*

Chinese AMDAR aircraft measurements were reported wrongly from March to May in 2007. Positive (greater than 0 °C)

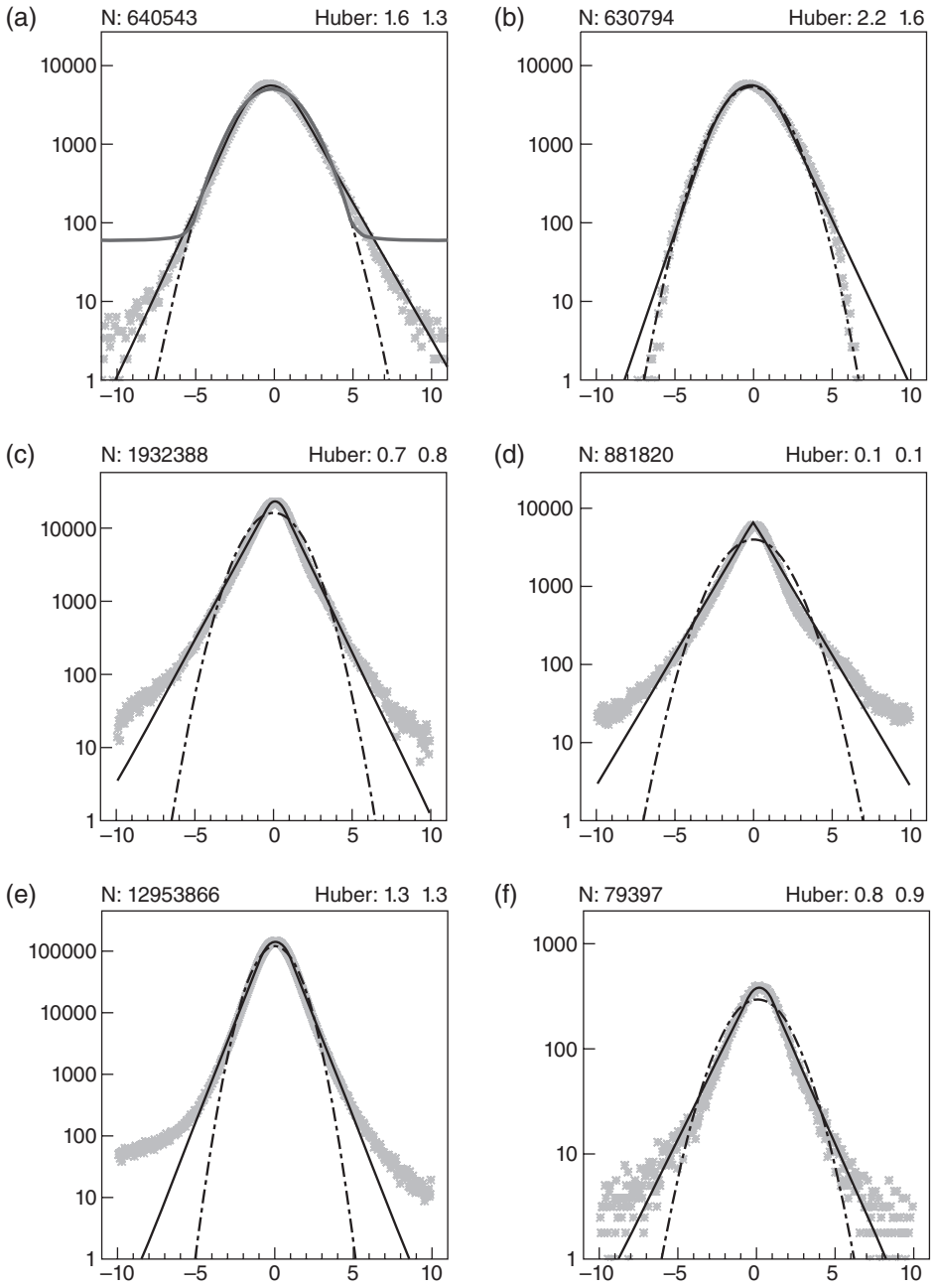

Figure 1. Innovation statistics, normalised by the prescribed observation error, for (a) all and (b) used Vaisala RS92 radiosonde temperature observations from 150 to 250 hPa in the NH Extratropics. (c) SYNOP surface pressure observations in the SH Extratropics, (d) SHIP surface pressure observations in the NH Extratropics, (e) aircraft wind observations in the NH Extratropics and (f) DRIBU wind observations in the NH Extratropics. The best fit Gaussian distribution (dash-dotted line) and Huber norm distribution (solid line) are included. The best fit Gaussian plus flat distribution is included on (a) as a fat solid grey line. N is the sample size. Huber numbers are the optimal left and right transition point, defined in Eq. (2).

temperatures were reported with the wrong sign as negative Celsius temperatures. This led to a negative tail of gross errors in the innovation statistics. Figure 2(a, b) show the distribution for all AMDAR and ACARS temperature departures for descending aircraft over the NH Extratropics from March to May 2007. The AMDAR data clearly show a large deviation from a Huber distribution for large negative departures which is due to the wrongly reported Chinese measurements. This is one of the few examples of a normalised innovation distribution with an almost flat tail. Over the same period, the ACARS temperature observation departures nicely follow a Huber norm distribution with only a slight misfit for very negative values.

3.2. *Surface pressure observations*

Figure 2(c) shows the distribution of surface pressure departures for NH extratropical synoptic land stations. A hump is clearly identifiable on the positive side of the background departure distribution. Detailed investigations revealed that this is related to the difference in model orography and station height for some observations. A high percentage of observations with positive background departures between 5 and 10 standard deviations are from stations located in Alpine valleys. The height of these stations tends to be lower than the height according to the forecast model orography since small valleys are not well resolved in the model. Specific QC, like orography difference-related blacklisting, ensures that those observations get rejected so this hump disappears in the distribution of the 'used' data.

Figure 2(d) shows the importance of not including blacklisted data in the estimation of the most suitable observation departure distribution. This example shows how the tropical METAR surface pressure data fit a Huber distribution well after excluding blacklisted data. It should be noted that the blacklisting is performed as a completely independent task to identify stations of consistently poor quality. This underlines the necessity of a good blacklisting procedure. It also shows the power of Huber norm distribution plots as a diagnostic tool to identify such outliers.

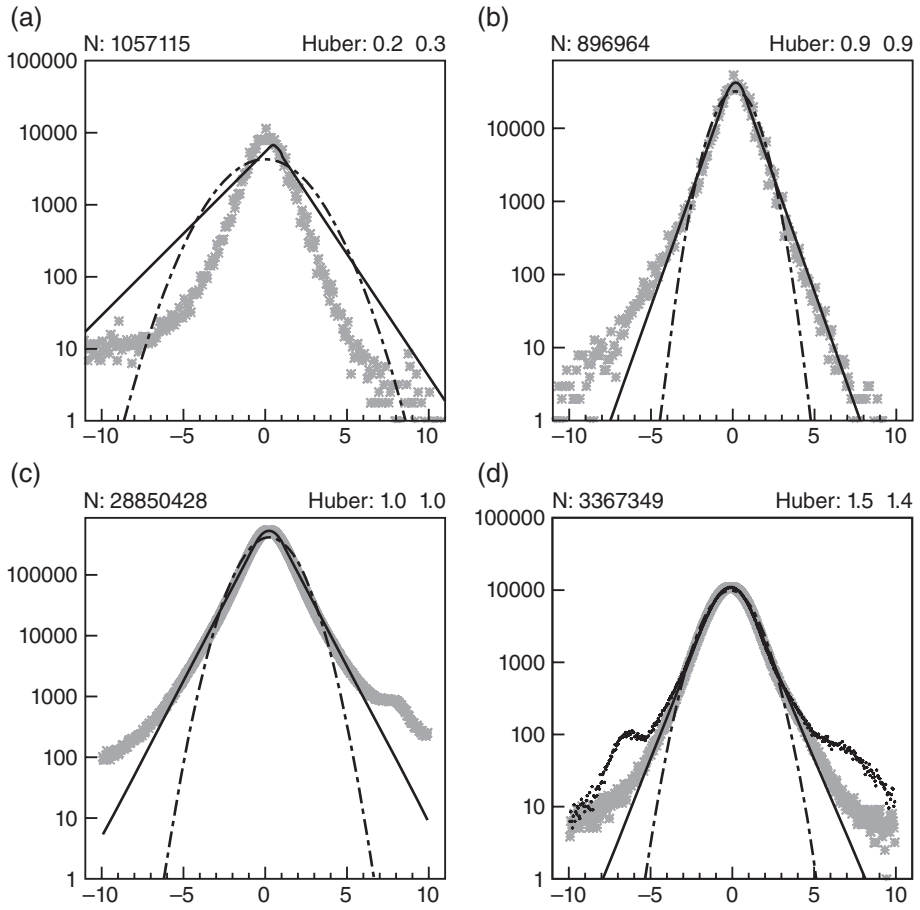

Figure 2. Departure statistics for (a) temperature data for all AMDAR (primarily data from European, Chinese and Japanese aircraft) descending over the NH Extratropics; (b) is as (a), but for ACARS (primarily data from North American aircraft); (c) SYNOP surface pressure; and (d) METAR surface pressure. In (d) the black dots show the data before blacklisting.

3.3. Humidity

Statistical distributions of humidity departures depend a lot on the selected variable. The innovation statistics for specific or relative humidity are far from Gaussian or Huber distributed, even after normalising by the specified observation error. A variable transformation, as the one used operationally at ECMWF (Hólm *et al.*, 2002; Andersson *et al.*, 2005), ensured a better fit. Figure 3(a) shows the distribution of radiosonde relative humidity departures, whereas Figure 3(b) shows statistics for humidity data normalised by the average of analysis and background data values, mimicking the variable transform method used at ECMWF. It is clear that relative humidity departures are poorly fitted by a Huber distribution, whereas the normalised data provide a reasonable fit.

3.4. Satellite data

In general it is more difficult to find regular distributions that fit satellite data departures well. Therefore we have not implemented a more relaxed QC for satellite data. We will discuss three reasons for this here. Firstly, most satellite data provide less detailed information than conventional data. The satellite data usually describe the broad features well for the whole swath area. The data seldom pinpoint small-scale weather events, for which a relaxed QC will make the biggest difference. Secondly, even though satellite data departures, for e.g. channels that peak in the stratosphere, typically follow a Huber norm distribution, they are more in accordance with a Gaussian distribution than conventional data. An example is given in Figure 3(c) (all data) and Figure 3(d) (used data) for AMSU-A channel 14, showing that there is a smaller benefit of switching to a Huber distribution. Thirdly, some satellite channels are contaminated by cloud and rain, leading to distributions with large humps, as shown in Figure 3(e), where all data for mid-troposphere peaking AMSU-A channel 7 are shown. This channel's atmospheric signal is contaminated by cloud and surface returns. Strict QC is applied to eliminate the contaminated tails of the distribution. Figure 3(f) shows the departure statistics for the used data for this channel, with the best Huber and Gaussian distribution included. Note that these plots, with their log-scaling and optimal Gaussian and Huber norm curves, also provide valuable diagnostic information. For example, the two plots for the AMSU-A channel 7 case identify that the cloud clearing is done very well, but it is not perfect for warm departures. Likewise, from comparing Figure 3(c and d), it is evident that the first-guess QC is too strict on AMSU-A channel 14 data. The normalised departure QC limits are 2–3, whereas without problem these could be increased to 6. Further investigations of relaxing background QC and using a Huber norm QC for satellite data will be done in the near future.

4. Bias correction problems for isolated biased observations

It is always difficult for an assimilation system to identify problematic isolated observations with large biases. This is a bigger problem when applying an observation QC that allows the use of outliers, because this effectively means relaxing the background QC considerably. This problem was identified at ECMWF in 1998 when hourly SYNOP surface pressure data were assimilated in the first 12 h 4D-Var implementation. Biased isolated stations influenced the analysis negatively. The solution was to identify time-correlated observation errors (Järvinen *et al.*, 1999) to reduce the likelihood of giving frequently reporting biased observations too high a weight. In 2005, ECMWF implemented a station-based bias correction scheme that dynamically corrected surface pressure observation biases (Vasiljevic *et al.*, 2006). This scheme only bias corrected observations in the range ± 15 hPa, which was

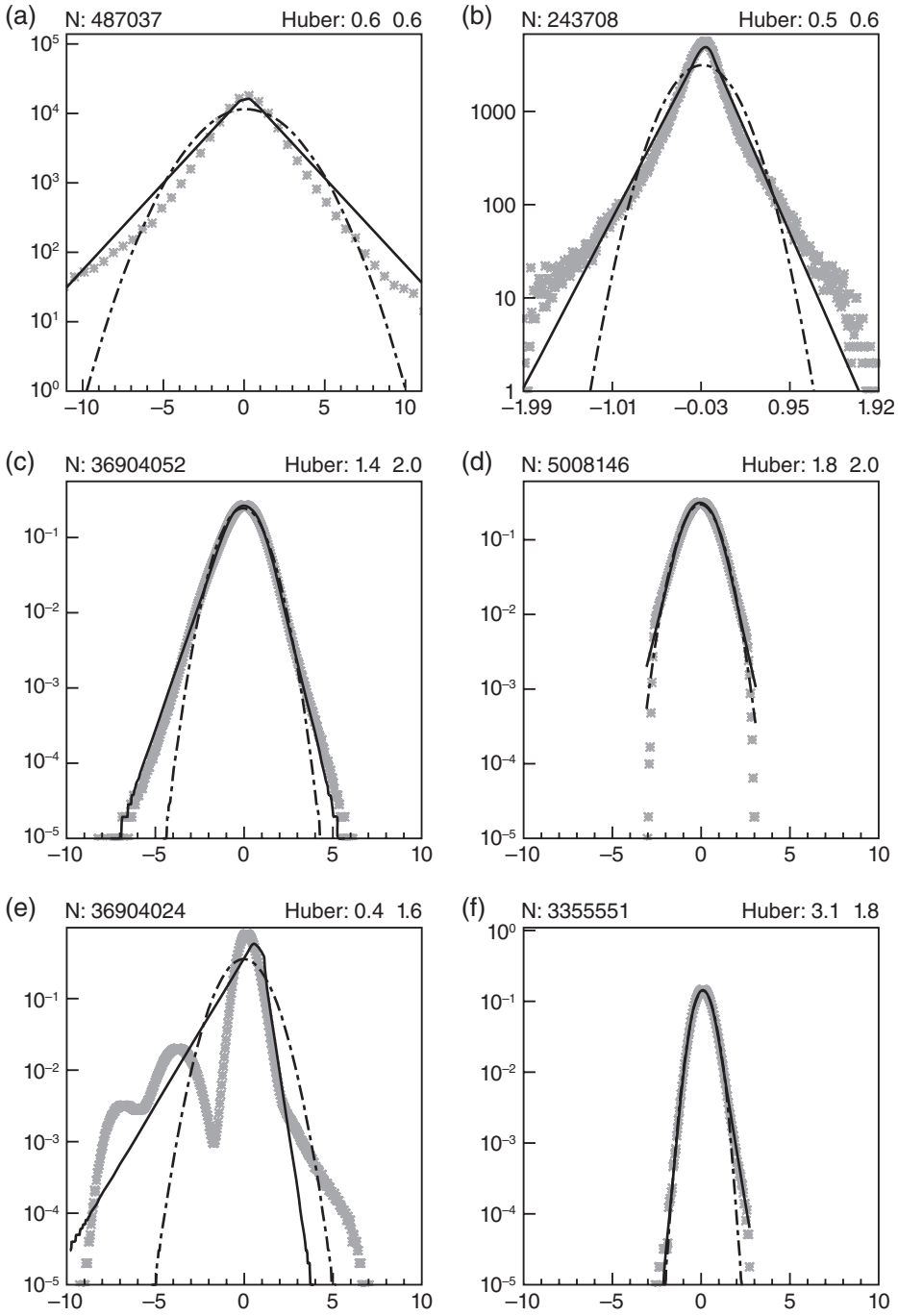

Figure 3. Departure statistics for (a) radiosonde relative humidity innovation distributions in the Tropics around 850 hPa. (b) Normalised humidity for radiosonde observations. (c) All brightness temperature departures from METOP-A AMSU-A channel 14 (stratospheric, peaking at 1 hPa) for the SH Extratropics. (d) is as (c) but for used data. (e) is as (c), all data, but for channel 7 (tropospheric, peaking at 250 hPa). (f) is as (e) but for used data.

safe because the then operational background check had much tighter limits than that.

With the introduction of the surface bias correction scheme, there was no longer a need to account for time-correlated observation errors in the assimilation system, so the scheme was abandoned. Relaxing the background check limits with the Huber norm implementation brought the problem back, because data with very large departures were allowed into the assimilation system again, and by mistake the limits for the surface pressure bias correction scheme were not extended accordingly. This provided a very useful reminder that relates to a similar problem to that found by Järvinen *et al.* (1999) related to remote surface stations with very large departures due to observation biases. A similar problem was also recently identified (H. Hersbach, 2014; personal communication) during the testing of the ECMWF ERA-20C surface pressure only reanalysis, where biased measurements from an isolated Pacific island station were spuriously assimilated. Therefore handling of biases is an important aspect to consider when developing an observation QC scheme. One example encountered during the development of the Huber norm QC scheme was the Antarctic station with WMO station identifier 89622. This station reported surface pressure hourly with an almost constant bias of 18 hPa in this data-sparse area, very likely due to a mis-specified station altitude. Figure 4(b) shows that until 26 December 1999 (observation counts 1–83, marked with the first vertical grey line) almost all the observations from this station were background QC rejected, leading to almost identical background and analysis departures for the first week. At this point the departures are reduced slightly, so some of the data just pass the relaxed background QC and become active observations in the subsequent 12 h analysis (observation counts 84–92 on Figure 4(b)). Each observation initially only got a small weight, but because the biased observations and the observation errors are strongly correlated, the sum of these small weights managed to draw the analysis somewhat towards the biased observations. The background forecast tried to correct the spurious analysis,

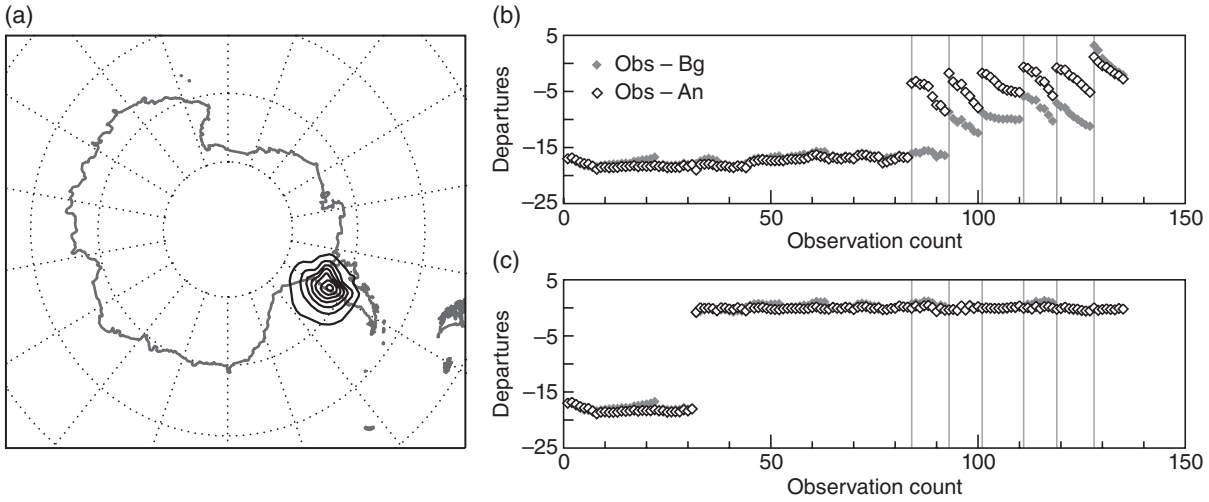

Figure 4. (a) Surface pressure difference of the Huber norm experiment to ERA-Interim; each black contour is 1 hPa, and solid lines indicate negative differences. Time series of obs minus background (grey diamonds) and obs minus analysis (open diamonds) departures for WMO station 89266 for the Huber norm experiment (b) without and (c) with relaxed surface pressure bias correction limits.

but each subsequent analysis was drawn more and more towards the biased observations. After four additional analyses cycles, identified by the grey vertical lines on Figure 4(b), the background has moved approximately 8 hPa towards the biased observations. The surface pressure bias correction was then activated in the next analysis cycle, after a spin-up period of five cycles, removing the bias between background and observation values (symbols after the last vertical grey line on Figure 4(b). Because the analysis within those five cycles had already been moved to a biased state by these uncorrected observations, the bias correction applied was 9 hPa. This introduced the spurious (9 hPa too deep) analysis difference seen on Figure 4(a). The difference was maintained for the subsequent period (not shown).

To avoid this problem, the limit for the surface pressure bias correction was extended beyond the values of the relaxed background check. Figure 4(c) shows the departure time series for an assimilation experiment with this extended limit of ± 25 hPa for the surface pressure bias correction scheme. After a spin-up period (at 30 observation counts on Figure 4(c)), the observations were bias corrected and used in the assimilation system. A bias correction of 17 hPa was performed and resulted in background departures very close to 0 for subsequent analyses, due to the simple almost constant bias pattern for this station. This example shows that careful bias correction is required, especially for remote frequently reporting stations, when very relaxed background limits are used. The problem was identified only because a control analysis was available.

5. Potential candidates for distributions with outliers and fat tails

The treatment of outliers in datasets has been discussed for centuries. The simplest method is to assume outliers are gross errors that are then discarded from the analysis of the data. In the 1960s Tukey (1960) and others investigated statistical methods that reduce the problems associated with the large sensitivity to outliers for the estimation of mean and standard deviation of a data sample assumed to follow, e.g. a Gaussian distribution. Tukey (1960) proposed to use a mixture of a one-sigma Gaussian plus a three-sigma Gaussian that represented the effect of fat tails in the distribution. Huber (1964) developed the concept of robust estimation where outliers could be accepted without ruining the estimation process. This method will be discussed further in section 6. Huber (2002) discussed important aspects of robust estimation methods. There are several methods for handling outliers, as the best method depends on the structure of the data and the users' view on the value or risk of outliers. In NWP, outliers occur due to erroneous observations (gross errors), valuable observations that can help to correct a poor background forecast, or observations than cannot be represented by the forecast model (representativeness errors). It is not always easy to distinguish between these groups of outliers. Robust methods are powerful because they allow the inclusion of outliers, but with some inbuilt safety so that the estimation of mean and standard deviation is less sensitive to the outliers. The most drastic robust method is to eliminate outliers completely. The background error QC described in section 6.6 is an example of such a method. The simplest is then to assume the remaining data are correct and follow e.g. a Gaussian distribution. The 'Gaussian plus flat' distribution is a refinement with a grey zone between correct data and gross error data. A certain small percentage of the data is assumed to be gross errors, without information, which follow a flat distribution. The remaining data are assumed to follow a Gaussian distribution. The variational QC which was used at ECMWF from September 1996 to September 2009 was based on such a formulation. The method and implementation was described in AJ99. The implementation in the variational data assimilation system at ECMWF is technically simple, resulting in only very minor modifications of the nonlinear, tangent linear and adjoint model code. The implementation was based on Dharssi *et al.* (1992) and Ingleby and Lorenc (1993), who argued for the use of a 'Gaussian plus flat' distribution, which assumes all outliers represent gross errors that are completely independent of the background field and therefore provide no useful information for the analysis. The Gaussian distribution and the flat distribution, describing the fraction of outliers, are estimated from large samples of innovation statistics and depend on the quality of each observing system and variable. For small innovations, a Gaussian distribution is typically a good assumption, but for outliers this is often not a correct or safe assumption. It should be noted that robust estimation is used in several areas outside NWP, e.g. finance, noise reduction of images and seismic data analysis (Guitton and Symes, 2003).

6. Aspects to consider when implementing a Huber norm QC

6.1. *Definition and formulation of the Huber norm*

Gross errors that are well represented by a flat distribution do exist for some observations, as discussed in section 3.1, but it is evident from Figures 1–3 (and many similar figures not shown) that the flat distribution is usually a poor representation of outliers. There is evidence that the majority of outliers cannot be considered as gross errors, but rather providers of some relevant information. This leads to fat tails in the distributions. In this article, it is

identified that these fat tails are well represented by a Huber norm.

The Huber norm distribution is defined as a Gaussian distribution in the centre of the distribution and an exponential distribution in the tails. Equations (1) and (2) define the Huber norm distribution as it was introduced by Huber (1972):

$$f(x) = \frac{1}{\sigma_o \sqrt{2\pi}} \exp\left\{-\frac{\rho(x)}{2}\right\} \tag{1}$$

with

$$\rho(x) = \begin{cases} \dfrac{x^2}{\sigma_o^2} & \text{for} \quad |x| \le c\,, \\ \dfrac{2\,c|x| - c^2}{\sigma_o^2} & \text{for} \quad |x| > c\,, \end{cases} \tag{2}$$

where c is the transition point, which is the point where the Gaussian part of the distribution ends and the exponential part starts. The definition ensures that f is continuous and the gradient of f is continuous. In our implementation, we allow the transition points to differ on the left (c_{L}) and the right (c_{R}) side of the distribution, enabling a better fit to the departure data.

The observation cost function (Lorenc, 1986) for one datum is defined as

$$J_o^{\mathrm{QC}} = -\frac{1}{2}\ln(p^{\mathrm{QC}}) = -\frac{1}{2}\ln\{f(x)\} = \rho(x) + \text{const.} \tag{3}$$

Note that J_o^{QC}, with the Huber norm distribution applied, is an L^2 norm in the centre of the distribution and an L^1 norm in the tails. This is the reason why the Huber norm QC is a robust method which allows the use of observations with large departures. Huber (1972) showed that the Huber norm distribution is the robust estimation that gives most weight to outliers; a higher weight on outliers makes the estimation of statistical moments theoretically unsafe.

Figure 5(a) shows the cost function, J_o^{QC}, for the Huber norm distribution, the pure Gaussian distribution, and the 'Gaussian plus flat' distribution. It is clearly seen that the pure Gaussian distribution has large values and large gradients for large normalised departures. The 'Gaussian plus flat' distribution has gradients close to zero for large departures. The Huber norm distribution is a compromise between the two.

Following AJ99, we define the weight applied to an observation as the ratio between the applied J_o^{QC} and the pure Gaussian J_o:

$$W = \frac{J_o^{\mathrm{QC}}}{J_o^{\mathrm{Gaussian}}}. \tag{4}$$

This defines how much the influence of the observation is reduced compared to the influence based on a pure Gaussian assumption. The definition of f in Eq. (1) ensures that the same weight factor is applicable to the gradient of the cost function, which controls the influence of the observations in the analysis. Figure 5(b) shows W for the three distributions discussed, as a function of departures normalised by the observation error standard deviation, σ_o. Near the centre of the distribution, both the Huber norm distribution and the 'Gaussian plus flat' distribution follow a Gaussian distribution, i.e. $W = 1$.

It can be seen that the 'Gaussian plus flat' distribution has a narrow transition zone of weights from 1 to 0, whereas the Huber norm has a broad transition zone. For medium-sized departures, the Huber norm reduces the weight of the observations and for large departures the weight is significantly higher.

A major benefit of the Huber norm approach is that it enables a significant relaxation of the background QC. With the previous QC implementation, rather strict limits were applied for the background QC, with rejection threshold values of the order of 5 standard deviations of the normalised departure values. For the implementation of the Huber norm, this has been relaxed considerably, as discussed in section 6.6. This is especially beneficial for extreme events, e.g. where an intensity difference or a small displacement of the background fields can lead to very large departures. Examples of this will be shown in section 7.

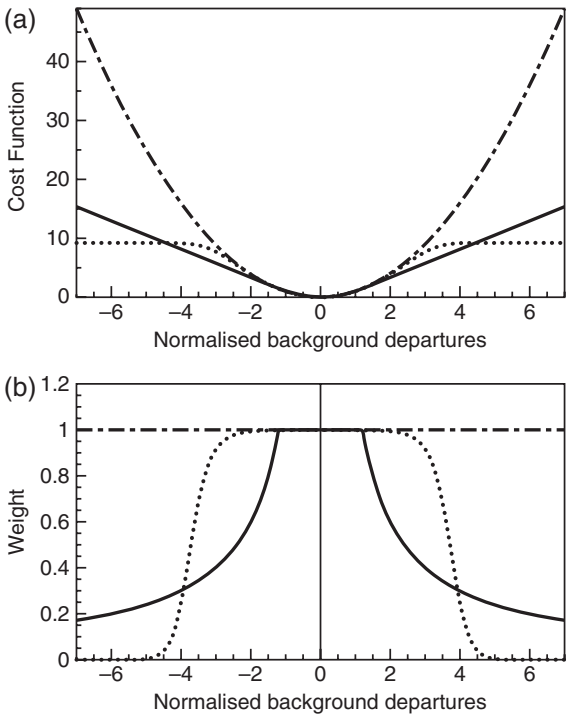

Figure 5. (a) Observation cost functions and (b) the corresponding weights after applying the variational QC; Huber norm distribution (solid), 'Gaussian plus flat' distribution (dashed), and Gaussian distribution (dash-dot).

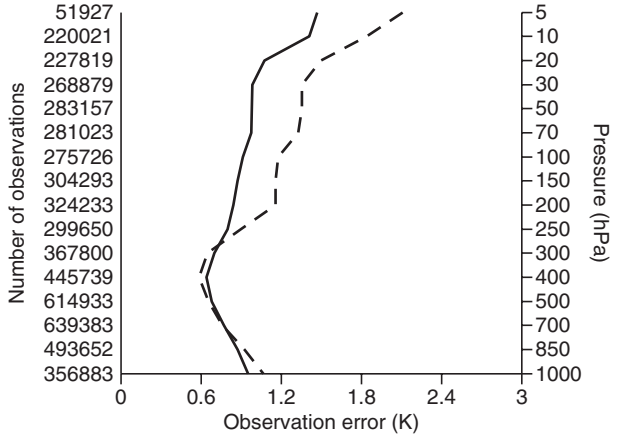

Figure 6. Profile of estimated observation errors (dashed) for Vaisala RS92 radiosonde temperature data over the NH Extratropics compared to the used observation errors (solid). The left axis shows the data count, and the right axis the pressure (hPa).

6.2. *Retuning of observation error*

The quality of each observation type is quantified by σ_o, the observation error standard deviation. While implementing the new variational QC scheme, a retuning of σ_o was done with the guidance from estimated observation errors (Desroziers *et al.*, 2005). This led to changes in the observation errors for radiosonde temperature measurements at high altitudes (above 200 hPa; Figure 6) and a retuning of the observation errors used for automatic and manual surface pressure measurements from ships. At the same time, airport surface pressure observation errors were adjusted to be similar to the observation errors

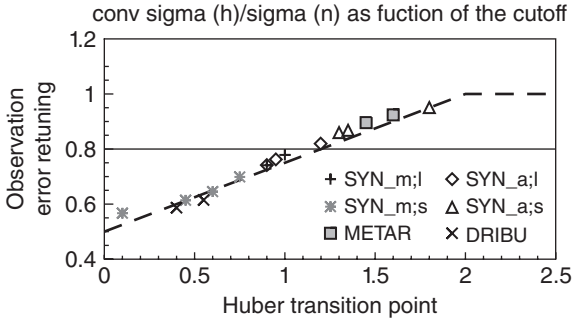

Figure 7. Used observation error tuning function (dashed line). The symbols indicate the ratio between the Gaussian and the Huber standard deviation for different kinds of surface pressure observations. SYNOP observations are split into manual (m) and automatic (a) as well as land (l) and ship (s). Every observation type is evaluated for three regions: Tropics and SH and NH Extratropics.

applied to automatic surface stations. The evaluation of the 18 months of departure data clearly supported these adjustments.

A retuning of the observation error was implemented for all data types for which the Huber norm VarQC was applied. This is highly recommendable because the specified observation error in the Huber norm implementation represents the good data in the central Gaussian part of the distribution, whereas it had to represent the whole active dataset in the old method. So theoretically the observation error should be reduced, especially for datasets with a small Gaussian range, i.e. with small Huber transition points.

We examined this for all the observing systems for which a Huber norm distribution was applicable. The symbols on Figure 7 show the ratio of the estimated σ_o for the optimal Huber norm distribution and the optimal value for a Gaussian distribution for a range of surface pressure observing systems. Values are plotted as function of the average Huber left and right transition points (c_L and c_R) for three different areas: NH Extratropics, Tropics and SH Extratropics. The selected observing systems cover a wide range of Huber transition points. It was found that on average the observation error is reduced to 80% of the previously used value. There is an approximately linear relationship between the observation error retuning factor and the Huber norm transition point.

The retuning factor can be estimated well with the simple function

$$T_{\sigma_o} = \min\left\{1.0\,,\ 0.5 + 0.25\left(\frac{c_L + c_R}{2}\right)\right\}. \qquad (5)$$

Here T_{σ_o} is the retuning value for a certain observation. We choose this simple linear function as it described the relation very accurately (dashed curve on Figure 7). There was no justification for implementing a more complex or statistically based tuning function.

6.3. Determination of the optimal Huber distribution and evaluation of the Huber norm transition points

The Huber distributions had to be computed for a large number of observing systems, their associated variables, for various layers for profiling data, and for each channel for satellite data. It was therefore beneficial to develop an objective method to determine the 'optimal' Huber distribution.

The algorithm is described here. First the bias of each data sample is removed, as it is considered an independent problem to address systematic errors. Therefore, as described in section 6.1, the Huber distribution is uniquely defined by σ_o, c_L and c_R. Note that the σ_o described the standard deviation of the central sample data of normalised departures between c_L and c_R. So if c_L and c_R are very large, the σ_o becomes identical to the value for the 'optimal' Gaussian distribution (shown with dashed-dotted curves on Figures 1–3).

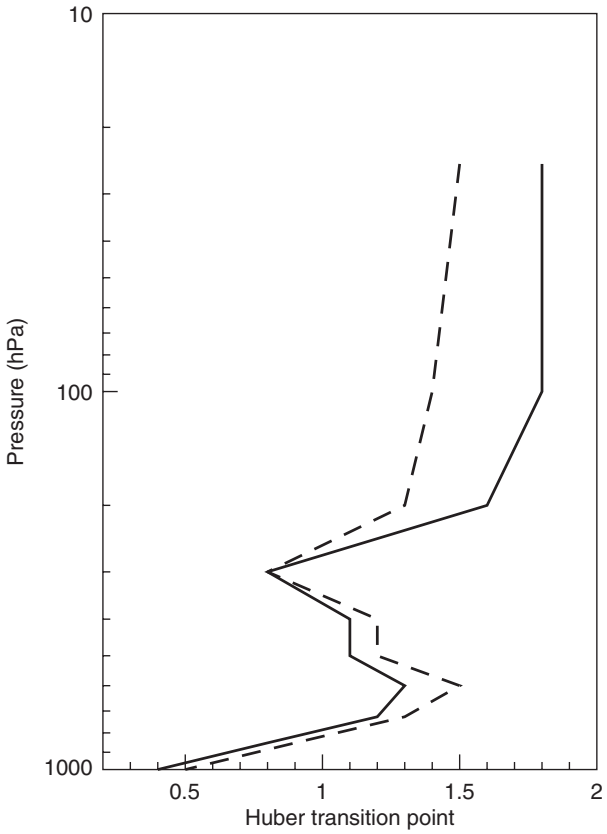

Figure 8. Profile of the optimal left (solid) and right (dashed) transition points for Vaisala RS92 radiosonde temperature data for each 100 hPa layer.

The optimal c_L and c_R for the Huber norm distribution were determined for each observation type and variable by searching among values in the range 0.0–5.0 in steps of 0.1. The best Huber norm fit was established by least-square-like curve fitting of normalised departures. This was done by computing a cost function, for each (c_L,c_R) pair, that describes the misfit between Huber distribution and the data sample. The misfit is defined as:

$$\sum_{i=1}^{n}\left[\,p(x_i)\ln\{p(x_i)\} - H(x_i)\ln\{H(x_i)\}\right]^2, \qquad (6)$$

where $p(x_i)$ is the population in range bin i and $H(x_i)$ is the population expected for the specific Huber distribution in the range bin i.

Typically the selected c_L and c_R values are identical or close to each other. It should be noted that different c_L and c_R introduce a bias due to the heavier tail to either left or right. This has a very small impact on the bias, because the asymmetry due to this typically represents much less than 1% of the data sample.

The evaluation of Huber transition points in general also confirmed the wide range of values for different observing systems and variables. It would be suboptimal to use a fixed value of (say) 1.0.

For profiling data, the vertical distributions of the Huber left and right transition points were computed for each 100 hPa vertical level. Figure 8 shows an example of this for Vaisala RS92 radiosonde temperatures. Investigations showed that the Huber norm transition points tended to be distinct for three layers in the atmosphere: the stratosphere (observations above 100 hPa), the free troposphere (observations between 100 and 900 hPa), and the boundary layer (observations below 900 hPa). So Huber norm distributions were computed and applied for these three layers for radiosonde, pilot, aircraft and wind profiler data. Notice that the transition points shown in Figure 8 differ in the stratosphere for the left and right transition points. This flexibility in the formulation gives us the opportunity to account for differences in the behaviour of the negative/positive temperature departures. Because we use departure distributions for the

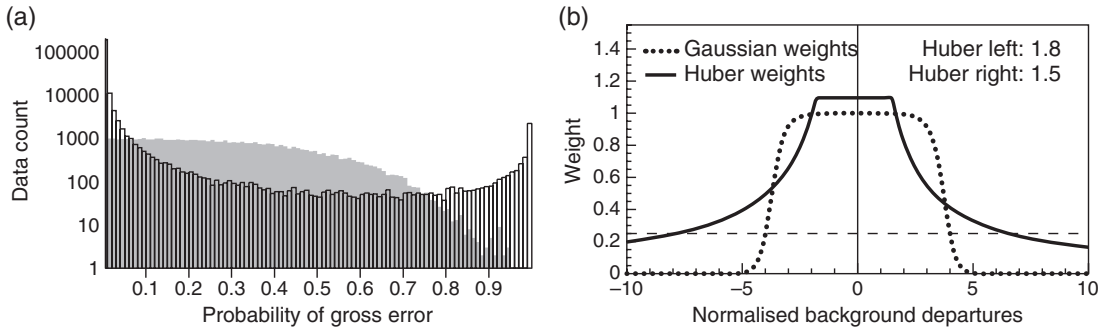

Figure 9. Illustration of the VarQC weights for a Huber norm compared to a Gaussian and flat distribution. (a) Data count as a function of the probability of gross errors, with the previously used 'Gaussian plus flat' distribution (black open bars) and the Huber norm distribution (grey solid bars), and (b) the corresponding weight functions for the two distributions. Huber values were taken from the radiosonde temperature observations in the stratosphere ($\leq$ 100 hPa).

evaluation, it is not clear if the observations or the background fields are responsible for asymmetric behaviour in the tails of the distribution. It could be questioned why the left and right transition points for radiosondes should change with height. This could possibly be linked to representativeness errors which in the ECMWF system (and most other assimilation systems) are treated as part of the observation error. But in several cases we are able to link asymmetries of temperature departures to issues with the observing system. Of course it is preferable to correct for systematic observation errors and model errors closer to the source.

6.4. *Huber norm VarQC implementation at ECMWF*

Further aspects that need to be considered when implementing the Huber norm VarQC in an operational NWP system are discussed here using the ECMWF operational implementation as an example. Because Huber norm VarQC is a robust method, it allows the relaxation of the background QC. This is a very important side benefit of the Huber norm method, because it makes observations with large departures active, so the data get a chance to influence the analysis. The observation errors were also adjusted as discussed in section 6.2.

The weights, W, are computed based on the high-resolution departures in the nonlinear outer loop of the incremental 4D-Var (Courtier *et al.*, 1994). The weights are kept constant during the minimisation (inner loop), because the Lanczos minimisation algorithm (Fisher *et al.*, 2009) used at ECMWF does not allow the function that is minimised to be modified during the minimisation process. Some minimisation methods are more lenient and would allow the weights to be adjusted slightly for each iteration of the minimisation process. But the benefit of the much faster, but strictly quadratic, Lanzcos algorithm outweighs the benefit of a more dynamic QC. The weights are updated at each of the three relinearisation outer loops applied at ECMWF; this makes it possible for the analysis to change the weights during the assimilation cycle.

In this article we concentrate our investigation on conventional observations. As mentioned in section 3.4, it is expected that the Huber norm QC will be most beneficial for conventional data. Of the conventional observing systems used in ECMWF's assimilation system, it was found that the distributions for the following observation types and variables were very well represented by a Huber norm distribution:

- Radiosonde observations: temperature and wind upper-air data (with special Huber norm distributions fitted to dropsondes);
- Aircraft observations: temperature and wind upper-air data;
- Pilot balloon observations: wind upper-air data;
- Wind profiler observations: wind upper-air data from American, European and Japanese wind profilers;
- Land surface observations: surface pressure data from automatic and manual SYNOP reports;
- Ship observations: surface pressure and wind data from automatic and manual SHIP reports;
- Airport observations: surface pressure data from METAR reports;
- Drifting and moored buoy observations: surface pressure and wind data from drifting and moored buoys.

So these observation types were all included in the operational analysis system update of the variational QC. The remaining observation types and variables kept the 'Gaussian plus flat' distribution.

The Huber norm QC is not implemented for humidity in the present implementation at ECMWF due to the difficulties discussed in section 3.3. It is planned to implement this in a forthcoming update.

6.5. *Weights for Huber norm VarQC*

Following the definition from AJ99, we define the probability of gross error, scaled to the range 0.0–1.0, to be $1 - W$. Figure 9(a) shows the distribution of gross error probabilities for the 18 months sample of stratospheric radiosonde temperature data. The transparent black bars are for the previously used 'Gaussian plus flat' distribution and the grey shaded bars are for the Huber norm distribution. Note the vertical scale is logarithmic and bars have a width of 0.01. It is seen that more than 99% (100 000) of the observations have gross error probabilities below 0.01. This is the case for both distributions. In the gross error probability range from 0.01 to 0.5 the Huber norm has similar data counts in each bin. For higher values the data counts fall off, because there are so few data values in the extremes of the departure distribution. For the 'Gaussian plus flat' distribution bin data counts are reduced between 0.01 and 0.5, and reach a level that is an order of magnitude lower than for the Huber norm distribution at 0.5. For higher gross error probabilities the data counts are increased for the 'Gaussian plus flat' distribution – a result of the sharp transition zone for gross error probabilities closer to the centre of the distribution, resulting in more observations with large probability of gross error values. Figure 9(b), similar to Figure 5(b), shows the corresponding weights for the optimal Huber norm distribution and the previously used 'Gaussian plus flat' distribution. It gives a qualitative understanding of the different shape of data counts for the gross error probabilities shown in Figure 9(a).

6.6. *Relaxation of the background QC*

Before the introduction of the Huber norm VarQC, the background QC had rather strict limits. Typical standard deviation values (α) would be around 5 (Järvinen and Undén, 1997) for the normalised departures, $(o - b)^2 < \alpha^2(\sigma_o^2 + \sigma_b^2)$, where σ_o and σ_b are the observation and background error standard deviations, respectively. For the Huber norm VarQC, this has been relaxed

Table 1. Data usage table showing the background QC and the VarQC rejections for 15 November 2008 to 31 December 2008 of operational data (Old) and the Huber norm assimilation experiment (Huber). VarQC rejected is defined by a weight smaller than 25%. The count for all observations is in thousands and is the same for both datasets. More detail is given in the text.

Obstype	Value	Level	All obs	% FG rej		% VarQC rej		BG QC limits		VarQC rej limits		
			*1000	Old	Huber	Old	Huber	Old	Huber	Old	Huber	
SYNOP	Ps	surf	5373	0.58	0.19	0.11	0.42	260.0	780.0	200.0	140.0	Pa
SHIP	Ps	surf	360	0.94	0.17	0.97	2.56	280.0	1100.0	200.0	180.0	Pa
SHIP	U/V	surf	350	0.77	0.02	0.43	5.44	11.2	12.7	10.8	5.4	$m\,s^{-1}$
DRIBU	Ps	surf	1156	1.17	0.55	0.47	0.97	360.0	800.0	200.0	200.0	Pa
DRIBU	U/V	surf	111	4.05	0.77	1.56	6.63	10.7	26.3	7.4	4.3	$m\,s^{-1}$
METAR	Ps	surf	2070	0.05	0.00	0.09	0.07	1000.0	>1600	340.0	80.0	Pa
TEMP	T	0-100	693	0.96	0.04	2.03	0.15	5.2	29.0	3.6	6.6	K
TEMP	T	100-900	1614	0.54	0.02	0.66	0.70	3.3	15.8	2.5	2.5	K
TEMP	T	1000-900	188	0.88	0.03	1.45	4.33	5.1	21.8	3.6	2.6	K
TEMP	U/V	0-100	716	0.49	0.09	0.78	0.61	13.9	22.5	10.2	11.5	$m\,s^{-1}$
TEMP	U/V	100-900	1237	0.35	0.08	0.39	0.79	11.2	23.5	9.1	6.5	$m\,s^{-1}$
TEMP	U/V	1000-900	189	0.44	0.06	0.48	2.67	11.1	30.9	9.2	6.5	$m\,s^{-1}$
AIREP	T	100-900	8477	0.08	0.00	0.05	0.19	4.4	15.9	3.8	1.4	K
AIREP	T	1000-900	1529	0.40	0.03	0.09	1.77	6.2	23.9	5.0	1.5	K
AIREP	U/V	100-900	8483	0.09	0.02	0.08	0.28	∼15.0	∼21.5	12.7	9.1	$m\,s^{-1}$
AIREP	U/V	1000-900	1483	0.63	0.17	0.11	0.62	∼15.0	no data	12.5	8.9	$m\,s^{-1}$
PILOT	U/V	0-100	238	0.53	0.04	0.81	0.71	14.3	24.6	10.3	11.6	$m\,s^{-1}$
PILOT	U/V	100-900	536	0.39	0.04	0.61	1.27	11.6	23.4	9.2	6.5	$m\,s^{-1}$
PILOT	U/V	1000-900	100	0.32	0.03	0.32	2.20	11.5	51.4	9.2	6.5	$m\,s^{-1}$
profiler	U/V	0-100	73	0.90	0.15	0.52	0.65	15.9	22.0	10.7	12.2	$m\,s^{-1}$
profiler	U/V	100-900	4061	0.15	0.03	0.10	0.25	12.7	22.2	9.2	6.5	$m\,s^{-1}$
profiler	U/V	1000-900	346	0.01	0.00	0.02	0.06	13.2	no data	9.2	6.5	$m\,s^{-1}$
EU-profiler	U/V	0-100	8	0.41	0.00	0.71	0.52	17.3	no data	10.7	12.4	$m\,s^{-1}$
EU-profiler	U/V	100-900	2036	0.08	0.02	0.06	0.13	12.7	24.2	9.2	6.5	$m\,s^{-1}$
EU-profiler	U/V	1000-900	246	0.01	0.00	0.02	0.08	13.2	no data	9.2	6.5	$m\,s^{-1}$
JP-profiler	U/V	100-900	303	0.18	0.01	0.49	0.85	13.2	22.2	9.2	8.4	$m\,s^{-1}$
US-profiler	U/V	0-100	46	1.36	0.24	0.70	0.93	15.9	22.0	10.7	12.2	$m\,s^{-1}$
US-profiler	U/V	100-900	1181	0.34	0.07	0.13	0.40	13.4	24.1	9.7	8.9	$m\,s^{-1}$

to around 15 standard deviations of the normalised innovation departure values. The 'BG QC limits' column in Table 1 shows the background QC values for the 'Gaussian plus flat' distribution (labelled Old) and for the Huber norm QC in details for all the involved observation types and variables. The values shown in Table 1 are absolute values in SI units. It could be argued that a background QC is not necessary any more when a robust estimation in the variational QC is applied, but the relaxed limits are still helpful in rejecting clearly erroneous gross errors, like 0 K temperatures.

7. General impact and case-studies

7.1. A general summary of QC decisions for the Huber norm implementation

The overall impact of the Huber norm implementation was evaluated over a three month data assimilation period in 2008, and for a number of intense weather events where the Huber norm implementation would be expected to make the biggest difference. For all the experiments presented in this Section the only difference between the control assimilations and Huber norm assimilations are the quality control and error distribution differences described in this article.

Table 1 shows the QC statistics for control (old) and Huber norm assimilations for all conventional observations that use the Huber norm QC. Upper air observation statistics are split up into three vertical bins, as described in section 6.3. The data were evaluated for the period from 15 November 2008 to 31 December 2008. The change in percentage of background rejected (FG rej) data is clear for all observation types, with significantly fewer rejections for the Huber norm assimilation experiment. The percentages of data with very low variational QC weight (less than 25%) are called VarQC rejected data, even though the data are not fully rejected. The data are still active data and influence the analysis according to their reduced weight. As discussed in section 6.5, the percentage of VarQC rejected data are generally larger for the Huber norm because this is the percentage of a much larger sample that pass the background QC. It is also related to the shape of the probability of gross error distributions, as shown in Figure 9. The final four columns show the approximate limits used by the different QC decisions. The term 'no data' means that no datawere background rejected for this data type during the six-week period evaluated. The VarQC limits show the range for which the weights get below 25%.

This change in variational QC was implemented into the operational forecasting system at ECMWF in September 2009 (Tavolato and Isaksen, 2010) and has proved to have a positive impact on the use of conventional observations within the assimilation system.

A number of impact studies and general investigations have been performed to evaluate the impact of the Huber norm QC. Assimilation experiments over a period of three months in 2008 showed a small positive impact over Europe and the NH Extratropics in general, and neutral scores for the SH Extratropics.

During the last week of December 1999, two small-scale lows affected Europe with intense gusts and storm damage. These storms are ideal case-studies due to the high-density, high-quality synoptic land station surface pressure network over France and Germany. These surface pressure observations captured the intensity and location of the storms very well, and neighbouring stations consistently supported each other. However, the strength of these storms was poorly represented in both the operational ECMWF analysis and the ERA-Interim (Dee *et al.*, 2011). Both assimilation systems used the old ('Gaussian plus flat' distribution) QC method.

Several case-studies were performed to investigate the assimilation impact of applying the Huber norm VarQC in the analysis system. The Huber norm experiments were run with the same model version as ERA-Interim, and for most experiments at the same resolution.

7.2. *Lothar; 26 December 1999*

The first of the December 1999 storms that hit Europe on 26 December 1999 is known as *Lothar* (Ulbrich *et al.*, 2001). It followed a path from the Atlantic to France, moving eastwards into Germany. The position of this storm was well predicted in both analyses (ERA-Interim as well as the Huber norm experiment), but the intensity was not captured well in ERA-Interim. Indeed, the SYNOP observations reporting the lowest surface pressure were background rejected in ERA-Interim. The Huber norm experiment showed a reduced central pressure compared to the reanalysis because many more observations were assimilated. However, the analysis was still significantly above the lowest observed surface pressure. One of the reasons is that the analysis is not able to capture the small scale of this event well enough at the reanalysis resolution.

To evaluate the influence of the resolution, several Huber norm experiments with different resolutions (inner and outer loop) were carried out and the results are shown in Table 2. It shows that increased resolution is beneficial.

Table 2. The lowest observed and analysed surface pressures on 26 December 1999, 0600 UTC. The Huber norm assimilation experiment deepens the low compared to ERA-Interim. A further improvement is found when the resolution (outer/inner loop) is increased.

System	Resolution	Lowest P_s (hPa)
ERA-Interim	T255 / T95,T159	978.0
Huber expt.	T255 / T95,T159	976.9
Huber expt.	T319 / T95,T159	976.4
Huber expt.	T319 / T95,T255	975.6
Huber expt.	T511 / T95,T255	974.3
Observation	–	962.4

7.3. *Martin, 27 December 1999*

The second storm was the very intense *Martin* which reached the French coast on 27 December 1999 (Ulbrich *et al.*, 2001). It was poorly predicted, being too weak and misplaced in the operational ECMWF analysis; ERA-Interim produced similarly poor results. Most surface pressure observations near the cyclone centre were rejected by the background QC (shown as filled triangles on Figure 10(a)) even though a hand analysis showed that all the observations from France were correct. This led to an analysis with the storm centre further to the east than surface pressure observations would suggest. The lowest surface pressure observation at 1800 UTC on 27 December 1999 reported 963.5 hPa. It was one of the background QC rejected observations in ERA-Interim.

Figure 10(b) shows rejections and observation weights from the Huber norm assimilation experiment. The numbers show the QC weight associated with each surface pressure observation: they are 16% or higher for all stations. More observations get higher QC weights than in the reanalysis due to the Huber norm. The centre of the low has correctly moved further to the west in good agreement with the observations. Furthermore, the minimum surface pressure is reduced significantly.

The analysis and the observation rejections for the December 1999 storm cases have also been discussed by Dee *et al.* (2001). They use an adaptive buddy check QC approach with the same effect as the Huber norm method to analyse this case.

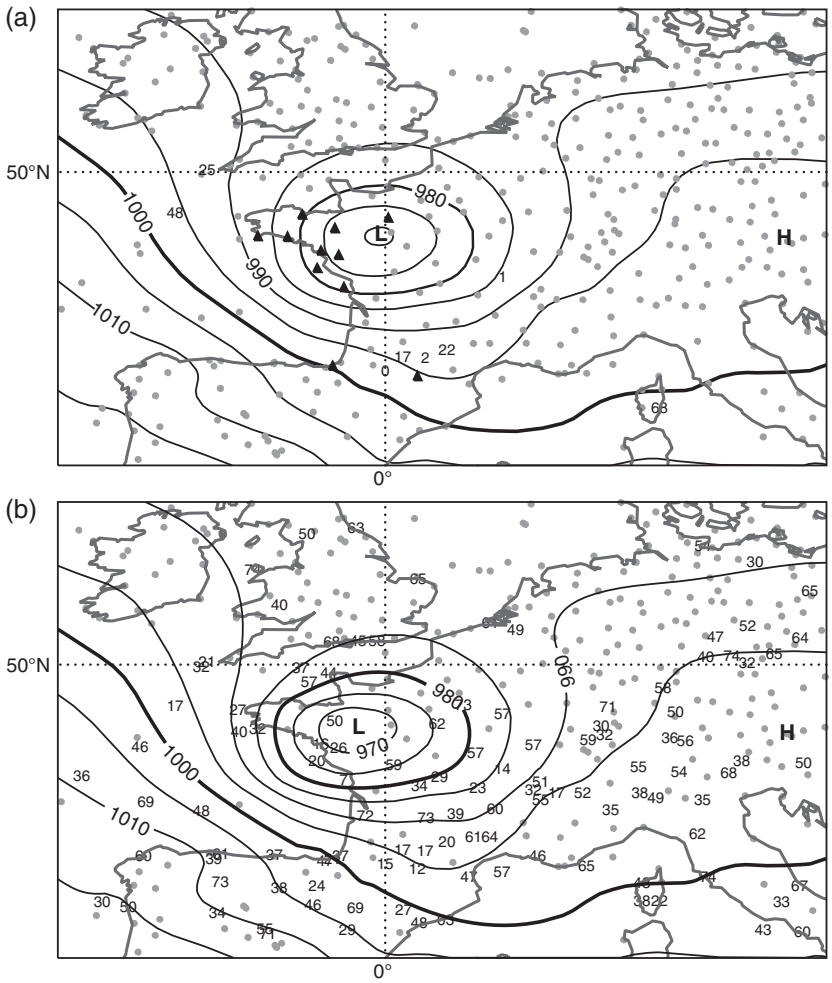

Figure 10. Rejections on 27 December 1999, 1800 UTC, for (a) ERA-Interim, (b) Huber norm experiment. The contours show the analysed surface pressure field for each experiment. Black triangles indicate background rejected observations, and numbers the effective VarQC weights for quality controlled stations. Grey circles indicate observations with weights higher than 75%.

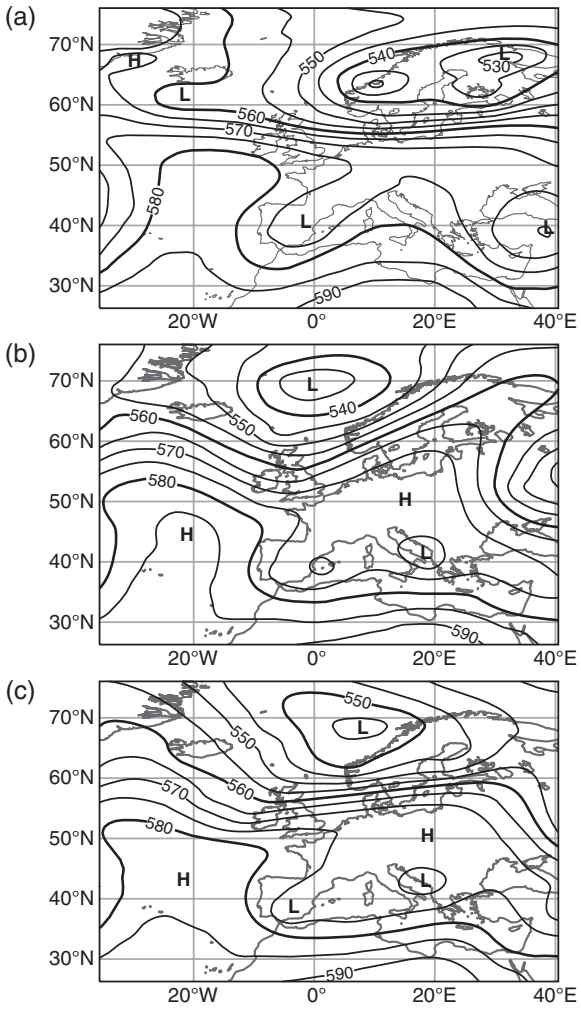

Figure 11. Analysis of 500 hPa geopotential height for (a) 11 June 2008, and the 5-day forecast valid at the same time from (b) the operational ECMWF system, and (c) the Huber norm experiment.

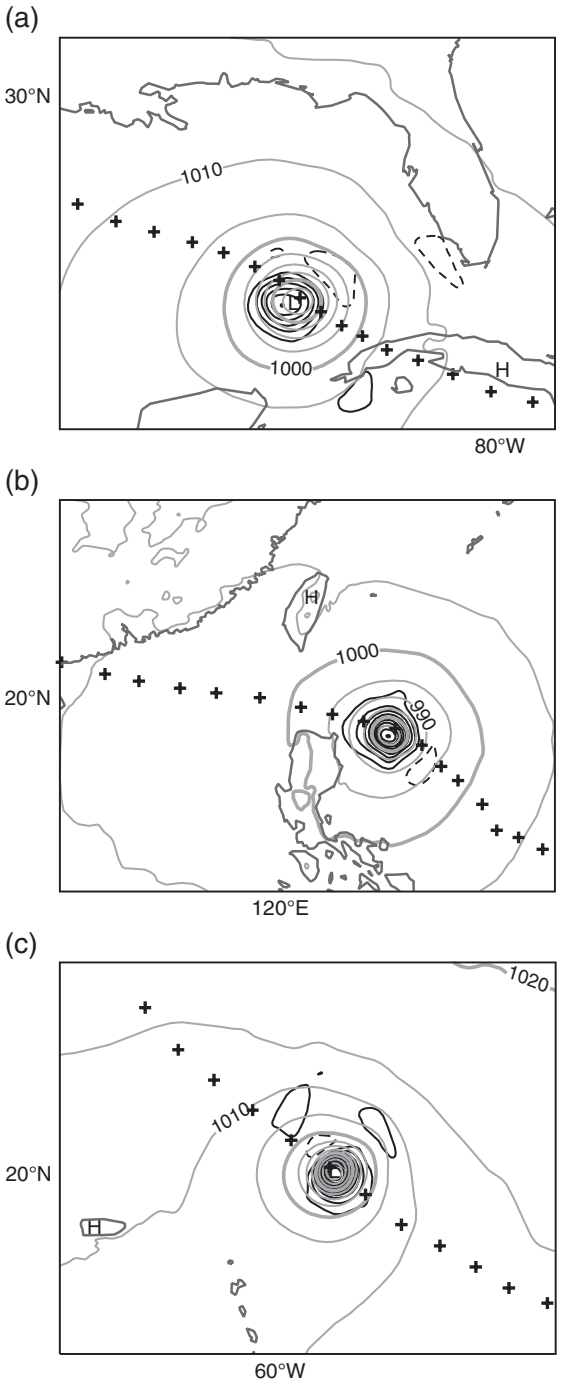

Figure 12. Improvement in tropical cyclone analyses due to Huber norm QC. (a) Hurricane *Ike* on the 10 September 2008 1800 UTC in the Gulf of Mexico approaching Texas. (b) Typhoon 0814 *Hagupit* on the 21 September 2008 1800 UTC in the Pacific approaching the Chinese coast. (c) Hurricane *Bill* on the 20 September 2009 0000 UTC in the Caribbean Sea. The grey contours (at 5 hPa intervals) show the MSL pressure analysis, and black crosses indicate the cyclone centre's position every 6 h. Solid/dashed black contours indicate reduction/increase in the surface pressure analysis compared with the control; the contour interval is 1 hPa in (a) and (b) and 2 hPa in (c).

However, the Huber norm method is simpler to implement in the IFS.

7.4. June 2008 extratropical event

At the beginning of June 2008, exceptionally low forecast scores were seen for the 5-day 500 hPa geopotential height forecast over Europe (anomaly correlation errors for 500 hPa geopotential height were below 0) in several NWP models (not shown).

In the operational ECMWF system, this drop in performance was linked to the rejection (mainly background rejection) of radiosonde and aircraft observations around 200 Pa over North America. Most of the background rejected data had relatively small background departures, just outside the QC limits. Applying the Huber norm VarQC had the effect that all these observations were used and the 5-day forecast improved drastically. Figure 11 shows the verifying analysis over Europe on 11 June 2008 and the two 5-day forecasts. The westerly flow over Europe is predicted much better in the Huber norm VarQC experiment.

7.5. Tropical cyclones

The Huber norm QC and relaxation of rejection limits are also applied for dropsonde wind and temperature observations, resulting typically in more correctly analysed tropical cyclones.

Results for hurricane *Ike*, hurricane *Bill* and typhoon *Hagupit* from September 2008 (*Ike*, *Hagupit*) and August 2009 (*Bill*) are discussed here. The two Atlantic hurricanes were well observed by dropsondes. Usage statistics for this period confirms that more dropsonde wind and temperature data were used in the Huber norm experiment than in the operational system.

Figure 12 shows the observed cyclone track, marked with crosses for every 6 h, and the analysis of surface pressure for the three tropical cyclones at one selected analysis time during the most intense cyclone phase. It is evident that all three tropical cyclones have been intensified very significantly by using the revised observation QC. Figure 13 shows the time series of core surface pressure every 6 h. These results indicate that the use of the Huber norm intensified the core pressure compared with the analysis using the 'Gaussian plus flat' distribution in the QC

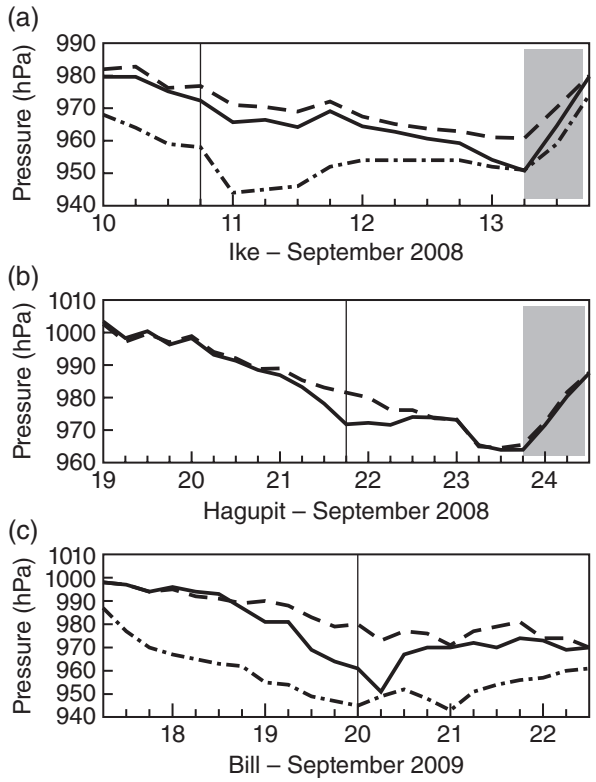

Figure 13. Time series of tropical cyclone centre surface pressure (hPa) for the storms (a) *Ike*, (b) *Hagupit* and (c) *Bill*. The solid line shows the surface pressure analysis from the Huber norm assimilation experiment, and the dashed curve is the control experiment. The dash-dotted line shows the observed surface pressure if available. The grey shaded area indicates the time after the landfall of the cyclone and the vertical line marks the date and time used in Figure 12.

for many analysis cycles during the intense phase of the tropical cyclones.

For the Atlantic hurricanes *Ike* and *Bill*, measurements of the core MSL pressure are available. For storm *Hagupit* no core MSL pressure observations are available, but the intensity estimates indicates it developed into a typhoon from a tropical storm on 20 September 2008. This means it was too weak in both analyses. So all three time series show that the Huber norm experiment improved the surface pressure analysis of the tropical cyclones.

It is clear that, when extensive dropsonde data are available, as with hurricane *Bill*, deeper and more accurate analyses were obtained when the Huber norm QC was applied.

8. Conclusions

The article describes a number of aspects that are important to consider for QC of observations used in data assimilation systems. Observations have measurement errors, representativeness errors and sometimes gross errors. In data assimilation, innovations are used extensively for observation QC and provide generally very valuable information (Hollingsworth *et al.*, 1986). But the background forecasts also have errors, sometimes very large ones, so it may be difficult to determine if an observation or the equivalent model value is the outlier. Monitoring time series for individual *in situ* stations and satellite channels provide a powerful method for detecting poorly performing or erratic platforms. It is shown how semi-logarithmic plots of normalised departures also are able to identify groups of outliers. Studying these outlier samples often makes it possible to identify problems with observations used in the data assimilation system, e.g. due to representativeness errors in mountainous areas. For humidity data, normalisation is required to obtain Gaussian-like innovations. For satellite data, clouds, rain and surface emissivity may contaminate the atmospheric signal. The semi-logarithmic plots of normalised departures also provide useful guidance in detecting this. A comprehensive evaluation was performed of the departure statistics for every observing system and every variable used in the ECMWF assimilation system over an 18-month period. After filtering out systematic outliers ('blacklisting' of stations and satellite channels) there is very little evidence of gross errors for the observations used in the departure distributions. This is likely because most observations are now automated and therefore either are of nominal quality or all display gross errors which are easy to detect.

The difficulties related to assimilation of isolated frequently reporting stations with biases is discussed. It is shown how important it is to do bias correction of isolated stations, especially when observation QC limits are relaxed. It is important to consider this when an observation QC scheme is developed.

Various QC methods for handling outliers are presented. The main issue is how much weight to assign to outliers. The 'Gaussian plus flat' distribution method (Andersson and Järvinen, 1999), used at ECMWF from 1999 to 2009, had fairly strict background QC limits and so there was a sharp transition from observations being active to being given virtually no weight in the analysis. The article describes the introduction of a Huber norm QC which makes it safe to use observations with large departures in the analysis. This is because it is a robust method where the moments of the distribution are affected very little by a few erroneous outliers.

Evaluating the 18-month sample of innovation data from ECMWF showed that almost all departure distributions were well described by the Huber norm distribution, after removing systematically erroneous data. The fit was much better than for the pure Gaussian distribution or a 'Gaussian plus flat' distribution. It was also shown to be beneficial to introduce the flexibility of allowing different left and right transition points from the Gaussian to exponential part of the Huber distribution. It is acknowledged that a Huber distribution fit for the normalised innovations does not prove that the observation-error distribution follows a Huber distribution. Innovation distributions are a convolution of observation- and background-error distributions. For the background QC, it is theoretically correct to use the innovation statistics, but for the observation cost function term it is not. But it is shown that it is more beneficial to relax the observation QC and allow outliers to influence the analysis under the assumption that observation errors follow the robust Huber distribution. Several case-studies show how the Huber norm QC deserves the credit for improved analyses and forecasts of extreme events such as extratropical storms and tropical cyclones. The examples show the strength of the robust Huber norm approach which enables the analysis to benefit from observation outliers in the situations when several observations deviate significantly and consistently from the model background. The previously used QC method would reject such observations.

The Huber norm QC has been implemented successfully at ECMWF in September 2009 for wind, temperature and surface pressure measurements from all conventional observations available. In the future this will be extended to humidity and some satellite data.

This work has also shown that refined QC and observation error tuning can be an important method to help extract more information from observations. It is an area of data assimilation where there is potential for further improvements.

Acknowledgement

The authors want to thank Hans Hersbach (ECMWF) for providing the software to compute the best fit Huber distribution. Elias Hólm and Erik Andersson, both ECMWF staff, provided helpful comments. We thank Rob Hine and Anabel Bowen, ECMWF staff, for improving our figures significantly. Part of this work was funded by project P21772-N22 of the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF).

References

Andersson E, Järvinen H. 1999. Variational quality control. *Q. J. R. Meteorol. Soc.* **125**: 697–722.

Andersson E, Bauer P, Beljaars ACM, Chevallier F, Hólm EV, Janisková M, Kållberg P, Kelly G, Lopez P, McNally AP, Moreau E, Simmons AJ, Thépaut J-N, Tompkins AM. 2005. Assimilation and modeling of the atmospheric hydrological cycle in the ECMWF forecasting system. *Bull. Am. Meteorol. Soc.* **86**: 387–403.

Bonavita M, Isaksen L, Hólm EV. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.* **138**: 1540–1559, doi: 10.1002/qj.1899.

Courtier P, Thépaut J-N, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1388.

Courtier P, Andersson E, Heckley W, Vasiljevic D, Hamrud M, Hollingsworth A, Rabier F, Fisher M, Pailleux J. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.* **124**: 1783–1807.

Dee DP, Rukhovets L, Todling R, da Silva AM, Larson JW. 2001. An adaptive buddy check for observational quality control. *Q. J. R. Meteorol. Soc.* **127**: 2451–2471.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597.

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation-, background- and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 3385–3396.

Dharssi I, Lorenc AC, Ingleby NB. 1992. Treatment of gross errors using probability theory. *Q. J. R. Meteorol. Soc.* **118**: 1017–1036.

Fisher M, Nocedal J, Trémolet Y, Wright SJ. 2009. Data assimilation in weather forecasting: a case study in PDE-constrained optimization. *Optim. Eng.* **10**: 409–426.

Guitton A, Symes WW. 2003. Robust inversion of siesmic data using the Huber norm. *Geophysics* **68**: 1310–1319.

Hollingsworth A, Shaw DB, Lönnberg P, Illari L, Arpe K, Simmons AJ. 1986 Monitoring of observations and analysis quality by a data assimilation system. *Mon. Weather Rev.* **114**: 861–879.

Hólm EV, Andersson E, Beljaars ACM, Lopez P, Mahfouf J-F, Simmons AJ, Thépaut J-N. 2002. *Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans*, *ECMWF Technical Memorandum* 383. ECMWF: Reading, UK.

Huber PJ. 1964. Robust estimates of a location parameter. *Ann. Math. Stat.* **35**: 73–101.

Huber PJ. 1972. Robust statistics: A review. *Ann. Math. Stat.* **43**: 1041–1067.

Huber PJ. 2002. John W. Tukey's contribution to robust statistics. *Ann. Math. Stat.* **30**: 1640–1648.

Ingleby NB, Lorenc AC. 1993. Bayesian quality control using multivariate normal distributions. *Q. J. R. Meteorol. Soc.* **119**: 1195–1225.

Isaksen L, Bonavita M, Buizza R, Fisher M, Haseler J, Leutbecher M, Raynaud L. 2010. *Ensemble of Data Assimilations at ECMWF*, *ECMWF Technical Memorandum* 636. ECMWF: Reading, UK.

Järvinen H, Undén P. 1997. *Observation Screening and Background Quality Control in the ECMWF 3D-Var Data Assimilation System*. *ECMWF Technical Memorandum* 236. ECMWF: Reading, UK.

Järvinen H, Andersson E, Bouttier F. 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus* **51A**: 469–488.

Lorenc AC. 1984 'Analysis methods for the quality control of observations'. In Proceedings of ECMWF Workshop on the Use and Quality Control of Meteorological Observations for Numerical Weather Prediction, Reading, UK, 6–9 November 1984, 397–428.

Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.

Lorenc AC, Hammon O. 1988. Objective quality control of observations using Bayesian methods – Theory, and practical implementation. *Q. J. R. Meteorol. Soc.* **114**: 515–543.

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

Tukey JW. 1960. A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, Olkin I, Ghurye S, Hoeffding W, Madow W, Mann H. (eds.): 448–485. Stanford University Press: Stanford, CA.

Ulbrich U, Fink AH, Klawa M, Pinto JG. 2001. Three extreme storms over Europe in December 1999. *Weather* **56**: 70–80.

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

