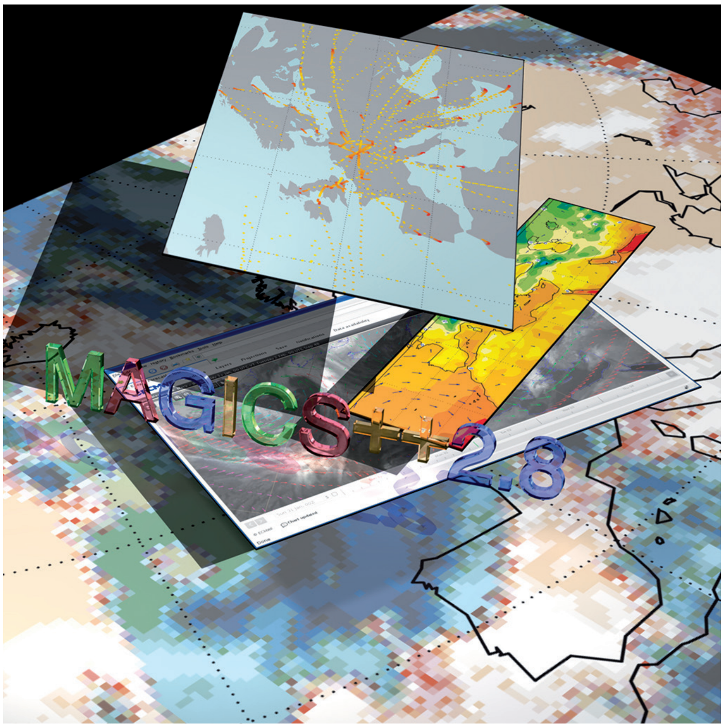
ECMWF Feature article

from Newsletter Number 122 – Winter 2009/10

METEOROLOGY

Huber norm quality control in the IFS



www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/4k12rmv6

This article appeared in the Meteorology section of ECMWF Newsletter No. 122 – Winter 2009/10, pp. 27–31.

Huber norm quality control in the IFS

Christina Tavolato, Lars Isaksen

Quality control (QC) of observations is a very important part of any data assimilation system. Observations contain measurement errors and sometimes gross errors due to technical errors, human errors or transmitting problems. The aim is to ensure that correct observations are used and erroneous observations are discarded from the analysis process. In this quality control process it is best to be cautious because accepted erroneous observations can lead to spurious features in the analysis.

Observations are compared against a short-range (6–12 hours) forecast from the previous analysis and they are discarded by automated QC procedures if they differ significantly from the forecast value. In the QC that was used operationally at ECMWF until recently, the threshold limits for exclusion of data was fairly tight to avoid the risk of using incorrect observations. This meant that, for example, surface pressure observations were rejected if they differed by more than about 6 hPa from the model field. In most cases this may be reasonable, but for extreme events it may well happen that the short-range forecast is wrong by more than 6 hPa near the centre of lows. To overcome this problem a new approach based on the Huber norm was implemented in cycle 35r3 of the Integrated Forecast System (IFS) on 8 September 2009.

The Huber norm

The new Huber norm quality control (QC) is a robust method that allows the use of observations with larger departures with a low risk of ruining the analysis locally. It has been introduced for conventional data in the ECMWF variational data assimilation system. Figure 1 shows schematically how the quality control method has changed.

Figure 1a shows how the QC-weight assigned to observations with large departures from the model are reduced, compared to the full weight given by the Gaussian distribution. The QC-weight, a value between 0 and 1, defines how much the impact of a suspect observation is reduced in the analysis. The Huber method consists of a Gaussian distribution near the centre of the distribution (full weight of the data) combined with an exponential distribution towards the tails of the distribution which leads to gradually decreasing weights. The previously used QC method had a Gaussian distribution in the centre plus a flat distribution in the tails (see *Andersson & Järvinen*, 1999). It can be seen that the old method has a narrow transition zone of weights from one to zero, 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 QC-weight is significantly higher.

Figure 1b shows the associated cost functions, where the Gaussian corresponds to a quadratic function and the Huber norm to a quadratic function for small departures and a linear function for large departures; the old QC cost function is flat for large departures.

The Huber norm is a so-called robust estimation method. The presence of a few incorrect outliers is less likely to ruin the analysis because their weights have been reduced compared to that of a purely Gaussian norm. On the other hand, if several outlier observations support each other, they will influence the analysis and their QC-weight will increase as the analysis manages to get closer to the observed values. For a purely Gaussian approach (with a QC-weight of 1) this would be potentially damaging, so for such an approach outliers have to be removed before the analysis. Consequently a major benefit for the Huber norm approach is that it enables a significant relaxation of the pre-analysis QC.

With the previous QC implementation, rather strict limits were applied for the first-guess (pre-analysis) QC, with rejection threshold values of the order of 5 standard deviations of the normalised departure value. For the implementation of the Huber norm this was typically relaxed to 15 standard deviations. This is beneficial for extreme events where the first-guess feature is more likely to be misplaced or too weak. The Huber norm also warranted a retuning of the observation error; this is discussed in Box A.

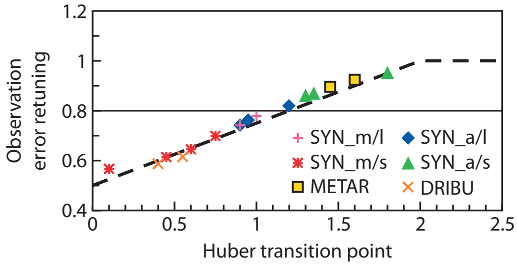
 doi:10.21957/4k12rmv6

Retuning of observation error

A

The quality of each observing system is quantified by the observation error. As part of the quality control revisions we took the opportunity to check whether the specified observation error was reasonable, using the *Desroziers et al.* (2005) method and background departure statistics. This led to an increase in the specified observation error for radiosonde temperatures above 200 hPa, and retuning of the observation error for automatic and manual surface pressure measurements from ships. At the same time METAR surface pressure observation errors were adjusted to be similar to the observation error applied to automatic SYNOP data.

An overall retuning of the observation error was implemented for all data types for which a Huber norm was applied. This is justified because the standard deviation represents the good data in the central Gaussian part of the distribution, whereas it has to represent the whole active data set in the old method. The figure shows the tuning function and the ratio of standard deviations for a range of surface pressure observations as function of the Huber transition point. The different symbols signify different observation types over three different areas: northern hemisphere, tropics and southern hemisphere. This chosen function fits well for all observation types and areas. The retuning factor describes the ratio of those two standard deviations and has been estimated with the dashed curve shown in the figure. So the observation error is on average reduced to 80% of the previously used value.



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 in manual and automatic (m or a) as well as land or ship (l or s). Also shown are results for METAR and DRIBU (buoy data). Every observation type is evaluated in three regions: tropics, southern hemisphere and northern hemisphere.

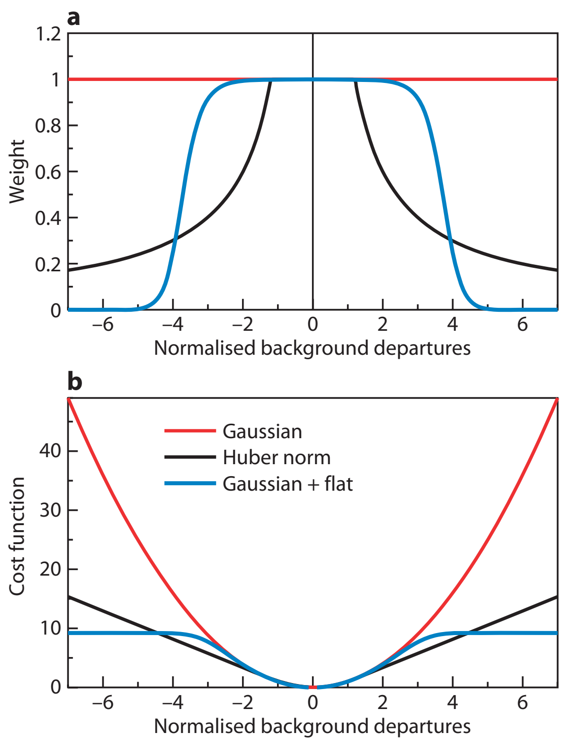


Figure 1 (a) Relative weight of an observation relative to a Gaussian distribution and (b) associated cost function values for the Gaussian 'normal' distribution (red line), Huber norm distribution (black line), and Gaussian plus flat distribution (blue line).

doi:10.21957/4k12rmv6

The Huber norm describes the data well

It is important to assess whether the Huber norm is suitable for describing actual observation distributions. Background departure statistics (observed–background) are the only easily available measure to evaluate observation-related distributions. Their main weakness is that the background departure statistics contain both observation and background information. It is difficult to isolate the observation-related part, which is what we are really trying to estimate.

The QC affects only a small number of observations in the tails of the distributions. So, to get a sufficiently large sample of relevant statistics, 18 months worth of data (February 2006 to September 2007) was examined. This was done for a large number of observation types to determine the Huber distribution that best represented the departures. Figure 2 shows the distributions for a number of these observation types.

Figure 2a shows ‘all temperature data’ at 150–250 hPa for Vaisala RS92 radiosondes in the northern hemisphere. A similar plot for the ‘used’ data is shown in Figure 2b. Also shown are the corresponding statistics for ‘not blacklisted’ data for southern hemisphere land surface pressure (Figure 2c), northern hemisphere surface pressure (Figure 2d), tropical METAR surface pressure (Figure 2e) and northern hemisphere buoy wind (Figure 2f). The results are plotted on a semi-logarithmic scale, so a Gaussian distribution appears quadratic and an exponential appears linear. In this diagram, a Huber distribution appears quadratic near the centre and linear in the tails.

Figure 2 shows the best fit to the background departure statistics by the Huber norm distribution (black line) and the Gaussian distribution (red line). It is clear that actual background departure statistics are best described by a Huber norm distribution. Indeed, these results indicate that the Huber norm distribution is also much better than using Gaussian plus flat distribution that until recently has been used in operations. This is the case for all the variables shown in Figure 2 and for almost all other variables that have been investigated.

For the ‘used data’ (Figure 2b) in the old operational QC implementation it is clear that to a large extent the data is either assumed Gaussian or rejected. In Figure 2e ‘all data’ values (blue dots) have been included in addition to the ‘not blacklisted data’ (green stars). This shows the importance of removing blacklisted data from the data sample, for less reliable observing systems, because it may eliminate strange humps due to biases and gross errors.

Data types that use Huber norm quality control

We have concentrated on conventional data distributions because they are the most important for the analysis of extreme events. Small-scale, fast-developing weather systems are mainly analysed by conventional observations, whereas satellite data benefits the broader temperature and humidity analyses. The first operational implementation has introduced Huber norm for the majority of conventional observation types.

- Temperature and wind data from radiosondes, dropsondes, pilots, wind profilers, aircraft, SYNOP stations, ships, moored buoys and drifters.
- Surface pressure data from SYNOP stations, ships, aviation weather reports (METARs), moored buoys and drifters.

Humidity data is more difficult to represent and requires the use of a normalized variable. Satellite data is affected by cloud or surface contamination which makes the QC work more difficult. So both humidity and satellite data have been left out in the first Huber norm implementation.

Investigations showed that the Huber norm distributions tended to be distinct for three layers in the atmosphere: the stratosphere (observations above 100 hPa), the troposphere (observations between 100 hPa 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.

 doi:10.21957/4k12rmv6

Some issues with surface pressure observations and satellite data

Investigating the background departure statistics for different observation types and parameters highlighted some unexpected behaviour. In cases where a Huber norm was difficult to fit to the data, it was usually due to erroneous data or gross errors. A few examples will be presented here.

Figure 2d shows the distribution of surface pressure departures for northern hemisphere SYNOPs. A hump is clearly identifiable on the positive side of the background departure distribution. 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 model orography as small valleys are not well resolved in the model. Specific QC ensures that those observations get rejected so this hump disappears in the distribution of the 'used' data.

Figure 2e shows the importance of not including blacklisted data in the estimation of the Huber norm distribution: without blacklisted data the Huber norm fits the distribution well. The blacklisted data add spurious humps for both positive and negative departures. This underlines the necessity of a good blacklisting procedure. It is also important to perform bias correction of surface pressure data to avoid spurious analyses for isolated stations when the first-guess quality control check is relaxed.

Satellite data has not been included in the Huber norm so far for three reasons. Firstly, most satellite data provides less detailed information than conventional data so the satellite data usually describes the broad features of small-scale weather events where the Huber norm is most beneficial. Secondly, satellite data seems to have a distribution more nearly Gaussian than conventional data. Thirdly, some satellite channels are contaminated by cloud and rain leading to distributions with large humps.

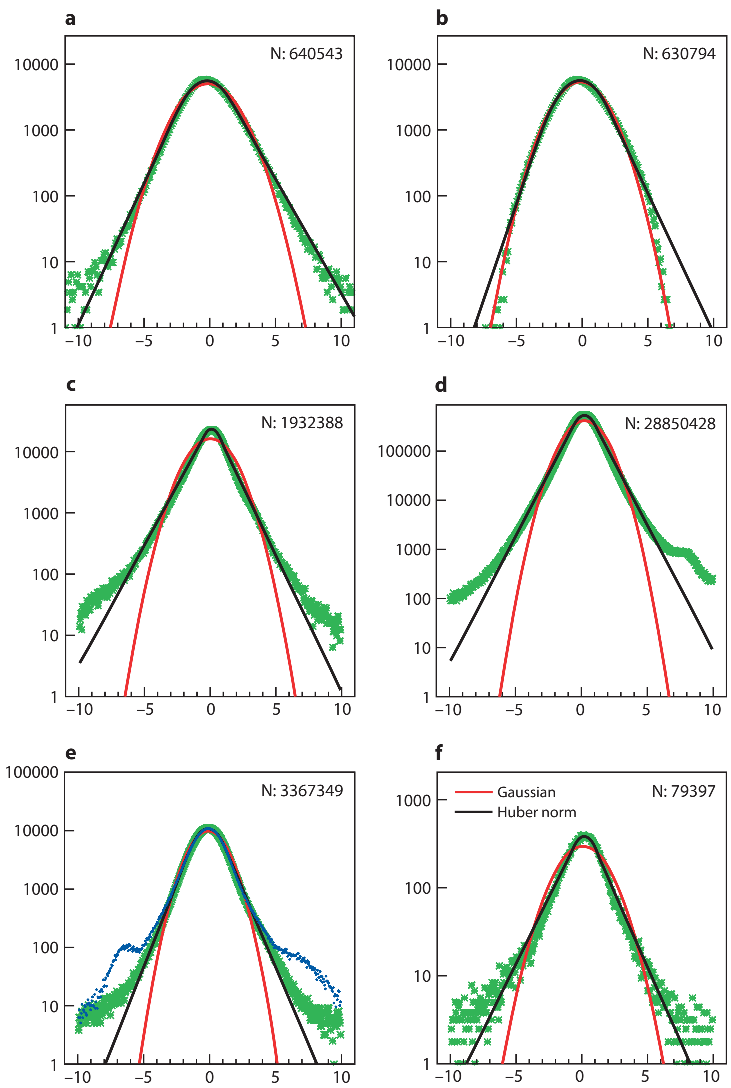


Figure 2 Best fit to background departure statistics by the Huber norm distribution (black line) and Gaussian distribution (red line) for a number of observing systems. (a) All Vaisala RS92 radiosonde temperature data at 150–250 hPa. (b) As (a) but for all used data. Results are also shown for (c) 'not blacklisted' southern hemisphere SYNOP surface pressure data, (d) As (c) but for northern hemisphere, (e) 'all' and 'not blacklisted' tropical METAR surface pressure data, and (f) 'not blacklisted' buoy wind speed data. The green stars are the 'not blacklisted data'. In (e) 'all data' values (blue dots) have been included. The number N each panel is the number of observations.

Extratropical storm impact studies

A number of impact studies and general investigations have been performed to evaluate the impact of the Huber norm quality control. Long runs over a period of three months in 2008 showed a small positive impact over Europe and the northern hemisphere in general, and neutral scores on the southern hemisphere.

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 SYNOP station network over France and Germany. These surface pressure observations captured the intensity and location of the storms, and neighbouring stations consistently support each other. However, the strength of these storms is poorly represented in both the operational analysis and the ECMWF climate reanalyses runs (ERA) that all used the old quality control method.

Two case studies investigated the difference in data rejection of the Huber norm assimilation experiment and the most recent ECMWF reanalysis, ERA-Interim (http://www.ecmwf.int/research/era/do/get/era-interim). The Huber norm experiment is run at the same resolution and with the same model version as ERA-Interim.

Storm Lothar

The first storm that hit Europe on the 26 December 1999 is known as Lothar. 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 is not captured well in ERA-Interim. Indeed, the SYNOP observations reporting the lowest surface pressure were first-guess rejected in the ERA-Interim analysis. The Huber norm experiment showed a reduced central pressure 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.

Storm Martin

The second storm was the very intense Martin that reached the French coast on 27 December 1999. It was poorly predicted being too weak and misplaced in the operational analysis; the ERA-Interim reanalysis produced similar results. Most surface pressure observations near the cyclone centre were rejected by the first-guess quality control (shown as filled triangles on Figure 3a) 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 and was first-guess rejected in ERA-Interim.

Figure 3b shows rejections and observation weights from the Huber norm assimilation experiment. The numbers show the effective percentage 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. However, the Huber norm method is simpler to implement in the IFS.

doi:10.21957/4k12rmv6

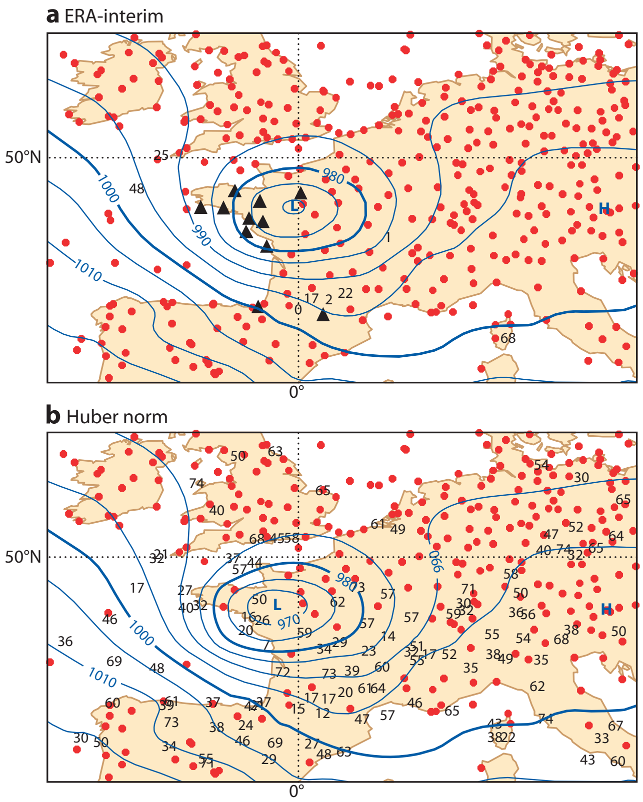


Figure 3 Mean sea level pressure chart (5 hPa contour level) valid at 1800 UTC on 27 December 1999 for (a) ERA-Interim and (b) Huber norm experiment showing the location and usage of surface pressure observations. The contours show the analysed surface pressure field for each experiment. Black triangles indicate first-guess rejected observations. The numbers indicate the effective percentage weight for observations with a partial weight, as defined by the quality control. Red dots indicate observations with weights higher than 75%.

Tropical cyclones

Another benefit from the use of the Huber norm method is that it provides the opportunity to relax the parameters defining rejection limits even further for special observation types. This is done for dropsonde wind and temperature observations. Dropsondes provide highly accurate measurements of tropical cyclones. With our relaxation of dropsonde QC thresholds the analysed surface pressure of tropical cyclones is typically deeper and the centres are more correctly positioned.

We will now consider results for Hurricane Ike and Typhoon Hagupit that occurred during September 2008. Both tropical cyclones were observed by dropsondes. Usage statistics for this period showed that more dropsonde wind and temperature data from low levels was used in the Huber norm experiment compared to the operational system.

Significantly deeper and more accurate analyses (not shown) were also obtained for Hurricane Bill in August 2009 when the Huber norm quality control was applied.

Figure 4 shows (a) the analysis of surface pressure at a specific time and (b) the time series of core surface pressure for Hurricane Ike. These results indicate that use of the Huber norm intensified the core pressure compared with the analysis that used the Gaussian plus flat distribution in the quality control. Also comparison with observations shows that it improved for the surface pressure analysis.

The results for Typhoon Hagupit shown in Figure 5 are similar to those for Hurricane Ike, but in this case there are no surface observations against which the analysis can be assessed.

doi:10.21957/4k12rmv6

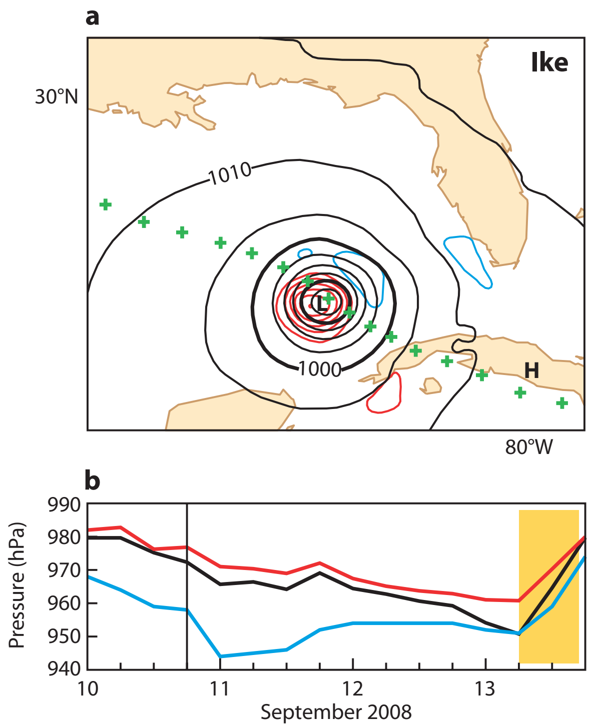


Figure 4 Impact of the Huber norm quality control impact on the analysis of tropical cyclone Ike2008 in the Gulf of Mexico approaching Texas. (a) Surface pressure analysis of the Huber norm assimilation experiments at 1800 UTC on 10 September (black line, pressure interval 5 hPa). The red (blue) contours indicate how the surface pressure analysis has intensified (weakened) compared to the control analysis that used the Gaussian plus flat distribution in the quality control. The green crosses show the observed cyclone track. (b) Time series of core surface pressure (in hPa) from the Huber norm assimilation experiment (black line) and the control experiment (red line), along with the observed surface pressure (blue line). The shaded area indicates the time after the land fall of the cyclone and the grey line marks the date and time used in (a).

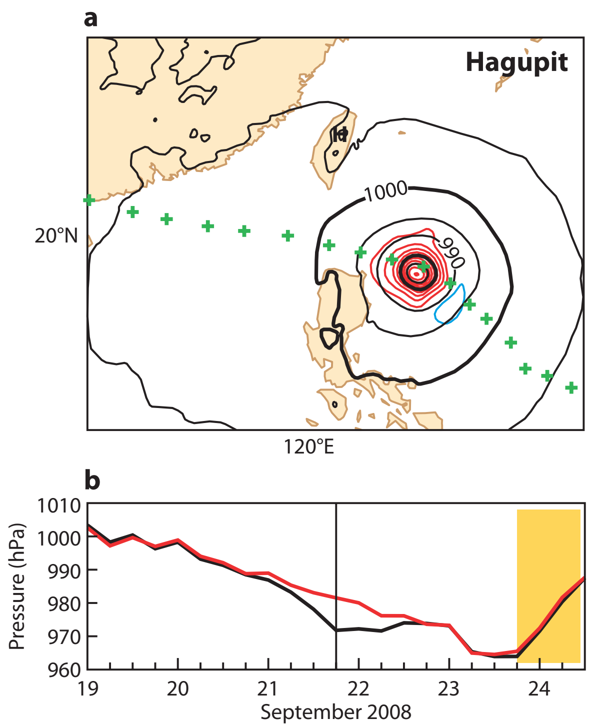


Figure 5 As Figure 4 but for Typhoon Hagupit in the Pacific approaching the Chinese coast. Note that (a) shows results for 1800 UTC on 21 September 2008 and in (b) there are no surface pressure measurements as usually none are available for tropical cyclones in the West Pacific.

doi:10.21957/4k12rmv6

Concluding remarks

The introduction of the Huber norm quality control has allowed the use of more observations with large departures in the analysis. This has resulted in more correct analyses of extreme events such as extratropical storms and tropical cyclones. If several observations deviate significantly and consistently from the model background the Huber norm method ensures that they influence the analysis. The previously used quality control method would reject the observations.

The Huber norm quality control has been implemented successfully for wind, temperature and surface pressure measurements for most conventional data. In the future this will be extended to humidity and some satellite data.

This work has shown that refined quality control and observation error tuning can be an important method to help extract more information from observations.

Further Reading

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

Dee, D.P., L. Rukhovets, R. Todling, A.M. da Silva & **J.W. Larson,** 2001: An adaptive buddy check for observational quality control. *Q.J.R. Meteorol. Soc.*, **127,** 2451–2471.

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

doi:10.21957/4k12rmv6