Observations and diagnostic tools for data assimilation:
October 1998

By **Heikki Järvinen**

Abstract

The purpose of the observation preprocessing and screening is to produce a clean array of observations in an easily accessible format to be used in the data assimilation. At the preprocessing stage an array in a suitable format is created for the data assimilation. Observation screening then selects a subset of observations to be presented for the assimilation itself. After the assimilation step a feedback file is created using the preprocessing software. This file contains all the relevant information regarding the use and impact of observations in the assimilation. This enables detailed diagnostic studies to be carried out afterwards on the performance of the assimilation and observing systems.

Table of contents

1. OBSERVATION PREPROCESSING

1.1 The incoming observations

The observations arrive at ECMWF through GTS (Global Telecommunications System) and are stored in a decoded format in the RDB (Report Data Base). Prior to the data assimilation the observations are extracted from the data base. These data have already undergone some rudimentary quality control, e.g. a check for the observation

format and position, for the climatological and hydrostatic limits as well as for the internal and temporal consistency, respectively. Then an observation file suitable for assimilation is created in an observation preprocessing module. This entails format conversions, change of some observed variables, like calculation of relative humidity from dry and wet bulb temperatures, as well as assignment of observation error statistics. The resulting file contains all the observational information from the data window (currently six hours) and is an input for the IFS (Integrated Forecast System). The observation screening then selects the best quality and unique observations. In 3D-Var closeness to the middle of the data window is preferred as the background is not interpolated to the exact time of the observation whereas in 4D-Var the screening can be performed hourly. Unlike the OI, the 3D/4D-Var data assimilation is global and therefore no separate data selection for analysis boxes is needed (OI analysis involves a matrix inversion of the size of the number of observations and therefore the analysis equation is solved separately for smaller areas where the number of observations is sufficiently small).

1.2 Bias correction

The feedback files are extensively used for monitoring the performance of the observing and assimilation systems. One use is to determine the bias corrections for some observing systems, currently for TEMP temperature observations, TOVS radiances and scatterometer (SCATT) winds.

Bias correction, in general, is a very difficult task as there is no fixed reference point with respect to which the bias should be corrected. If one removes, for instance, all the bias between the model background field and the TOVS radiances, there is a risk that part of the removed bias actually originates from the forecast model rather than from the observing system. In this case, the true effect of the bias removal is that the observations will actually enforce the model bias in the subsequent assimilations. Due to the risks involved, often a policy of "conservative bias correction" has been adopted, i.e. removing for instance only a half of the bias appearing in the observations.

The biases change in time due to changes in observing and assimilation systems and therefore the bias correction has to be updated from time to time. An update to the bias correction coefficients for TOVS radiances is performed once a month on the past 2 to 4 weeks of radiance background departure statistics. The bias correction is calculated with an off-line code using feedback files as input. The coefficients are substituted to input observations at the pre-processing stage.

2. THE OBSERVATION SCREENING

The ECMWF 3D/4D-Var data assimilation system makes use of an incremental minimization scheme to reduce the computational cost. The variational data assimilation starts with the first (high resolution) trajectory run. During this run the model counterparts for all the observations are calculated through the non-linear observation operators. As soon as these background departures are available for observations, the screening can be performed. Options for 3D- and 4D-screening are available. 3D-screening time window extends over the whole assimilation time window (currently six hours), whereas in 4D-screening the assimilation time window is partitioned into one hour time slots where the screening decisions are taken independently of the other time slots.

2.1 Screening of conventional observations

2.1 (a) Preliminary checks of observations. The observation screening begins with a preliminary check of the completeness of the reports. For instance, the observation and background errors should not be missing, as otherwise the background quality control cannot be performed. Also the reporting practice for SYNOP and TEMP mass observations (surface pressure and geopotential height) is checked.

Next the observations are scanned through for blacklisting. The blacklist consist formally of two parts. First, the selection of variables for assimilation is done using the data selection part of the blacklist file. This controls which observation types, variables, vertical ranges etc. will be selected for the assimilation. Some more complicated decisions are also performed through the data selection file. For instance, an orographic rejection limit is applied in the case of the observation being too deep inside the model orography. This part of the blacklist also provides a handy tool for experimentation. Second, a monthly monitoring blacklist is applied for discarding the stations that have recently been reporting in an excessively noisy or biased manner as compared with the ECMWF background field.

2.1 (b) Background quality control. The background quality control is performed for all the variables that are intended to be used in the assimilation. The procedure is as follows. The variance of the background departure $y - H(x_b)$ can be estimated as a sum of observation and background error variances $\sigma_o^2 + \sigma_b^2$, assuming that the observation and the background errors are uncorrelated. After normalizing with σ_b, the estimate of variance for the normalized departure is given by $1 + \sigma_o^2/\sigma_b^2$. In the background quality control, the square of normalized background departure is considered as suspect when it exceeds its expected variance more than by a predefined multiple. For the wind observations, the background quality control is performed simultaneously for both wind components. There is also a background quality control for the observed wind direction. For the SCATT winds, a test for high wind speeds and cold SST (possible sea-ice) is applied. An example of the background quality control rejections is given in Fig. 1 . It shows that the background quality control effectively cuts off the tails of observation minus background departure distribution.
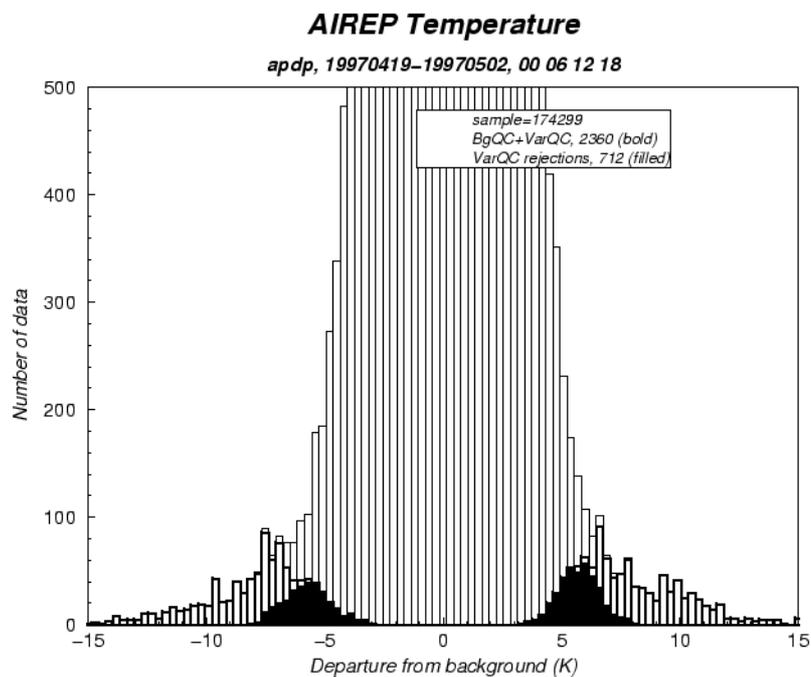
Figure 1. An example of a histogram of background departures for AIREP temperature observations. Variational and background quality control rejections are denoted by filled and outlined columns, respectively.

2.1 (c) Vertical consistency of multi-level reports. The multi-level reports are checked for the vertical consistency and the duplicated levels are removed from the reports. The vertical consistency check of multi-level reports

is applied in such a way that if four consecutive layers are found to be of suspicious quality, then these layers are rejected, and in the case of geopotential observations also all the layers above these four are rejected.

2.1 (d) Removal of duplicated reports.   The removal of duplicated reports is performed by searching pairs of co-located reports of the same observation types and then checking the content of these reports. It may, for instance, happen that an AIREP report is duplicated having only a slightly different station identifier but the observed variables inside these reports are exactly the same ones, or partially duplicated. The pair-wise checking of duplicates results in a rejection of some or all of the content of one of the reports.

2.1 (e) Redundancy check.   The redundancy check of the reports, together with the level selection of multi-level reports, is performed next for the active reports that are co-located and originate from the same station. For land SYNOP and PAOB reports, the report closest to the centre of the screening time window with most active data is retained whereas the other reports from that station are considered as redundant and are therefore rejected from the assimilation. For ship SYNOP and DRIBU observations the redundancy check is done in a slightly modified fashion. These observations are considered as potentially redundant if the moving platforms are within a circle with a radius of one degree latitude. Also in this case only the report closest to the centre of the screening time window with most active data is retained. All the data from the multi-level TEMP and PILOT reports from same station are considered at the same time in the redundancy check. The principle is to retain the best quality data at the significant levels (i.e. the turning points of the sounding) and closest to the centre of the screening time window. One such datum will however only be retained in one of the reports. A wind observation, for instance, from a sounding station may therefore be retained either in a TEMP or in a PILOT report, depending on which one happens to be of a better quality. A SYNOP mass observation, if made at the same time and at the same station as the TEMP report, is redundant if there are any TEMP geopotential height observations that are no more than 50hPa above the SYNOP mass observation.

2.1 (f) Thinning.   Finally, a horizontal thinning is performed for the AIREP and TOVS reports. The horizontal thinning of reports means that a predefined minimum horizontal distance between the nearby reports from the same platform is enforced. For AIREP reports the free distance between reports is currently enforced to about 125 km. The thinning of the AIREP data is performed with respect to one airliner at a time. Reports from different airliners may however be very close to each other. In this removal of redundant reports the best quality data is retained as the preceding quality control is taken into account. In vertical, the thinning is performed for layers around standard pressure levels thus allowing more reports for ascending and descending flight paths. Thinning of TOVS reports is done at two stages. First a minimum distance of about 70 km is enforced, and thereafter a repeated scan is performed to achieve the final separation of roughly 250 km between reports from one platform. The thinning algorithm is the same as used for AIREPs but in case of TOVS reports a different preference order is applied: a sea sounding is preferred over a land one, a clear sounding is preferred over a cloudy one and finally, the closeness of observation time to centre of the screening time window is preferred. Fig. 2 gives an example of the over-all usage of TOVS reports. There is also an option for further thinning of SSM/I and SATOB observations within the IFS.

TOVS reports 24/12/1997 09-15UTC
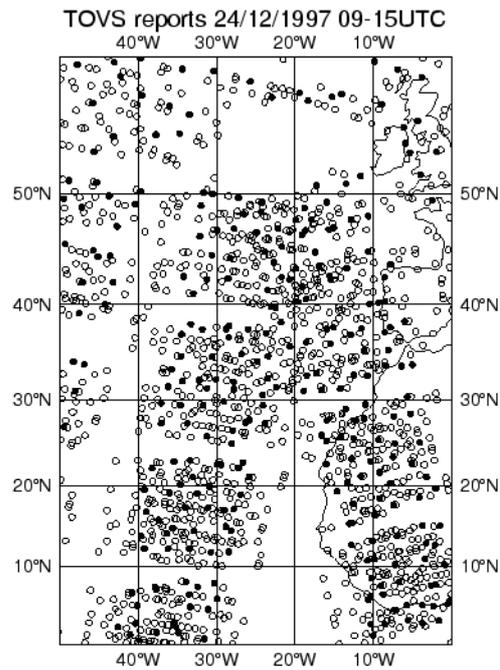
Figure 2. The usage of TOVS reports in the assimilation on the North Eastern Atlantic. Filled rings mark reports contain one or more channels used in the assimilation, whereas the empty rings denote rejected reports. Most of the rejections are due to the horizontal thinning and much less due to the quality reasons. Note that both edges of the swath are rejected.

The effect of observation screening on SYNOP surface pressure observations is summarized in Fig. 3 in the case of 3D-Var and 4D-Var, demonstrating the potential of 4D-Var in using observations from frequently reporting stations.

(a) Synop (4D–Var)

Surface pressure (5. – 19. 12. 1996)

(b) Synop (3D/4D–Var)

Surface pressure (5. – 19. 12. 1996)
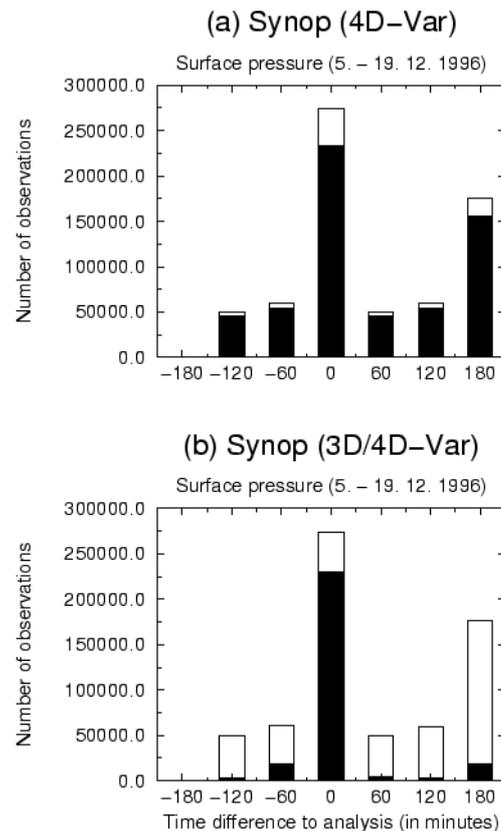
Figure 3. The effect of the observation screening on SYNOP surface pressure observations. Column height gives the number of observations available, while the shaded part displays those actually used in the assimilation. (a) 4D-screening for 4D-Var, and (b) 3D-screening for 3D/4D-Var

2.2 Screening of satellite radiances

The TOVS radiances (currently 120 km resolution) are preprocessed in a dedicated module which performs several functions to allow the assimilation of TOVS radiances in 4D-Var (the NESDIS retrievals are not used in 4D-Var but only monitored with the background profiles). This module is called ADVAR and it is called for each TOVS observation with the model background temperature, specific humidity and ozone profiles and surface parameters interpolated to the location of the observations. For each analysis cycle there are typically 20,000 TOVS observations in total, for a dual polar orbiter system. In the screening run, ADVAR is called twice.

2.2 (a) Input. The fast radiative transfer model for TOVS radiances requires an input profile from 1000 to 0.1 hPa. For the current 31 level model the background profiles are only available up to 10 hPa and so an extrapolation has to be performed up to 0.1 hPa for temperature using the NESDIS retrievals to 1 hPa and then a simple extrapolation based on model atmospheres above this level. Climatological mean profiles are assumed for water vapour and ozone. For the next version of the ECMWF forecast model with levels in the stratosphere this extrapolation is not necessary any more. Once the full profile from 1000 to 0.1 hPa is defined and checked radiative transfer model is called to compute the background radiances from the background profiles.

2.2 (b) Quality control. Several quality checks are applied to the measured and background radiances. The gross checks applied are:

     (i)   Check that the background profile is within realistic limits (e.g. temperature in range 150 to 350 K, specific humidity positive and not supersaturated, ozone within climatological extremes).

     (ii)   Check that the measured and background brightness temperatures are present for all required channels and within the range 150 to 350 K.

A series of more critical tests are then applied:

     (i)   Gross background check (i.e. measured radiance departures from the background are less than 20 K).

     (ii)   The background temperature, specific humidity and ozone profiles are checked to make sure they are close to or within the range encompassed by the diverse 32 (or 35 for ozone) profile dataset for which the radiative transfer model is valid.

     (iii)   A fine background check where the square of the radiance departures are flagged if they are greater than $16 \times [\mathbf{KBK}^T + \mathbf{O} + \mathbf{F}]$.

     (iv)   A check for cloud contamination for the HIRS channels is included by checking the radiance departure for HIRS channel 10 is inside the range –4 to +8 K.

     (v)   Radiances at the two extreme edge positions of the swath are flagged at present and not used in 4D-Var.

     (vi)   Checks are also made that the bias correction coefficients, satellite id, and scan position are all valid before proceeding.

2.2 (c) Retrieval.   The main task for ADVAR is to perform a 1D-Var retrieval of temperature, water vapour and ozone profiles. Each radiance profile is assigned to be clear, partly cloudy or cloudy by NESDIS and different TOVS channels and observation errors are used for each type. The background error covariances **B** are also specified in a file and for temperature are close to the global mean background errors assumed in 4D-Var. For specific humidity the background errors assumed in 1D-Var follow the same formulation as in 4D-Var and the correlations are the same as in 4D-Var.

The minimisation of the cost function is performed using the method of Newtonian iteration and up to 5 iterations are allowed before the minimisation fails. If the cost function of the observed radiance in any of the channels exceeds a predefined threshold then the set of radiances is indicated as inconsistent. The output of 1D-Var includes background and retrieved temperature, water vapour and ozone profiles together with several retrieved surface parameters also included in the 1D-Var control vector.

A final check on the stability of the retrieved profile is provided in the code but not implemented as the profiles are not used in 4D-Var.

2.2 (d) SSM/I radiances.   SSM/I radiances are also screened in a similar module which performs a similar set of functions to ADVAR retrieving total column water vapour, surface wind speed and cloud liquid water path. At the time of writing the SSM/I radiances are used operationally only in a passive mode enabling a full scale performance monitoring.

2.2 (e) Scatterometer processing.   A horizontal thinning is performed for the ERS scatterometer reports with respect to the particular measurement geometry of the instrument. The backscatter data are acquired within individual cells related to a 450 km wide grid with a mesh of 25 km in the across and along track directions. 19 measurement nodes are thus defined across the scatterometer's swath, while 19 rows are also considered in the along track direction to gather the data in squares of 19 by 19 points. The thinning is then achieved by keeping only every

fourth point within these squares. The data are thus used at a resolution of 100 km instead of the original 25 km sampling distance.

Apart from the thinning, the other observation dependent decisions involved by the screening of the SCATT data come essentially from the application of a sea-ice contamination test from the model sea surface temperature analysis, using a minimum threshold of 273 K, and a high wind rejection test with an upper wind speed limit set to 25 m/s for the higher of the SCATT and background winds.

An extra quality control is done on the wind retrieval residual or so-called "normalized distance to the cone". This quantity is tested in global average over the six hours of the analysis cycle for each of the 19 measurement nodes across the swath. All the data are then rejected in bulk if an excessive value is found for any node (more than 1.3 times the expected average) whereas the number of data taken into account is judged significant (more than 500). While the first check performed locally aims at avoiding geophysical effects not explained by the transfer function (CMOD4), for example rain or sea-state effects in the vicinity of deep lows, this global quality control on distance to the cone allows to detect technical anomalies not reported in real time by ESA and likely to affect the measurements in a correlated way and at larger scales. Such anomalies occur typically in the case of orbital manoeuvres.

2.3 A summary of the current use of observations

A summary of the current status of use of observations in the 4D-Var data assimilation is given in Table 1 below.

TABLE 1 . A SUMMARY OF THE CURRENT USE OF OBSERVATIONS IN THE 4D-VAR DATA ASSIMILATION AT THE ECMWF. ps STANDS FOR SURFACE PRESSURE, 2 m rh FOR RELATIVE HUMIDITY AT 2 m LEVEL, AND T_b FOR BRIGHTNESS TEMPERATURE, RESPECTIVELY.

Observation type	Variables used	Remarks
SYNOP	u, v, ps, (or z), rh	u and v used only over sea, in the tropics also over low terrain (< 150 m). Orographic rejection limit 6hPa for rh , 100 hPa for z and 800 m for ps
AIREP	u, v, T	Not used in full resolution. Used only below 50 hPa
SATOB	u, v	Selected areas and levels
DRIBU	u, v, ps	Orographic rejection limit 800 m for ps
TEMP	u, v, T, q, 2 m rh	Used at significant levels. q only below 300 hPa. 10 m u and v used over land only in tropics over low terrain (< 150 m). Orographic rejection limit 10 hPa for u and v , 100 hPa for z/T , 6 hPa for rh and –4 hPa for q
PILOT	u, v	Used at significant levels. 10 m u and v used over land only in tropics over low terrain (< 150 m). Orographic rejection limit 10 hPa for u and v
SATEM	T_b	Selected channels and areas. NESDIS retrievals are not used any more

TABLE 1 . A SUMMARY OF THE CURRENT USE OF OBSERVATIONS IN THE 4D-VAR DATA ASSIMILATION AT THE ECMWF. ps STANDS FOR SURFACE PRESSURE, 2 m rh FOR RELATIVE HUMIDITY AT 2 m LEVEL, AND T_b FOR BRIGHTNESS TEMPERATURE, RESPECTIVELY.

Observation type	Variables used	Remarks
PAOB	ps	Used south of 19°S. Orographic rejection limit 800 m for ps
SCATT	u, v	Not used in full resolution. Used if SST warmer than 273 K or if both observed and background wind less than 25 m/s

2.4 Compression of the CMA-file

After the observation screening roughly 15% of all the observed data are active and the compressed observation array for the minimization run only contains those data. That large compression rate is mainly driven by the number of TOVS data as after the screening there are only 10–20% of the TOVS reports left, whereas for the conventional observations the figure is around 40%. As a part of the compression, the observations are resorted among the processors for the minimization job in order to achieve a more optimal load balancing of the parallel computer.

2.5 A massively parallel computing environment

The migration of operational codes at the ECMWF in 1996 to support a massively parallel computing environment set a requirement for reproducibility. The observation screening should result in exactly the same selection of observations when different number of processors are used for the computations. In the observation screening there are the two basic types of decisions to be made. Independent decisions, on one hand, are those where no information of any other observations or decisions is needed. In a parallel computing environment these decisions can be happily made at different processors fully in parallel. For dependent decisions, on the other hand, a global view of the observations is needed which implies that some communication between the processors is required. The observation array is however far too large to be copied for each individual processor. Therefore, the implementation of observation screening at the ECMWF is such that only a minimum necessary information of the reports is globally communicated in order to provide the global view to the observations needed for the dependent decisions.

The global view of the observations is provided in the form of a global "time-location" array for selected observation types. This array contains compact information of the reports that are still active at this stage. For instance, the observation time, location and station identifier as well as the owner processor of that report are included. The time-location array is composed at each processor locally and then collected for merging and redistributed for each processor. After the redistribution the array is sorted locally at the processors according to the unique sequence number. Every processor has thus exactly the same information to start with and the dependent decisions can be performed in a reproducible manner independently of the computer configuration.

The time-location array is just enough for all the dependent decisions, except for the redundancy checking of the multi-level TEMP and PILOT reports. This is a special case in the sense that the information of each and every observed variable and from each level is needed. This actually means that the whole multi-level report has to be communicated. The other way out of this would be to force the observation clusters of the multi-level reports always into one processor without splitting them. In that case codes responsible for creation of the observation arrays for assimilation should ensure that geographical integrity of the observation arrays distributed for processors. This is, however, not possible in all the cases, and the observation screening has to be able to cope with this. Currently, it

is coded in such a way that only a limited number of multi-level TEMP and PILOT reports, based on the time-location array, are communicated between the appropriate processors as copies of these common stations.

3. USE OF FEEDBACK INFORMATION

The feedback files are extensively used for monitoring the performance of the observing and assimilation systems and some of the use is listed below (and some is discussed further in the chapter "Diagnostic tools for an assimilation system").

- observation statistic generation
- station-by-station monitoring
- observation plotting
- bias correction (or bias tuning)
- observation and background error estimation

4. DIAGNOSTIC TOOLS FOR AN ASSIMILATION SYSTEM

An operational assimilation system is a (ever increasingly) complex machinery comparable with any large-scale industrial application: the scheduling is tight, an effective but robust functioning is required and a quick trouble-shooting is needed in case something goes wrong in an operational run. The complexity of the system dictates that several aspects of the system have to be monitored and diagnosed to make sure the output is reliable. A number of diagnostic tools are presented in this chapter. They are collected under headings according to their most obvious use.

4.1 Code development and trouble-shooting

4.1 (a) Test the correctness of tangent linear and adjoint codes. In the IFS there are tangent linear and adjoint codes associated with the forecast model and the observation operators. A test for the correctness of the tangent linear code can be derived from a Taylor expansion for the perturbed non-linear model state

$$H(\mathbf{x} + h) = H(\mathbf{x}) + \mathbf{H}h + O(\|h\|^2)$$

by dividing by $\mathbf{H}h$ and reorganizing to a formula which behaves asymptotically according to

$$\lim_{h \to 0} \left(\frac{H(\mathbf{x} + h) - H(\mathbf{x})}{\mathbf{H}h} \right) = 1$$

It is best to do the test for an individual routine at the time of writing, but the test can also be applied to the whole tangent linear model.

The adjoint and tangent linear codes have to form an adjoint pair which can be tested using the definition of the adjoint operator

$$\langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle_F = \langle \mathbf{x}, \mathbf{A}^*\mathbf{y} \rangle_E$$

where the inner products are defined in their respective spaces *E* and *F*. In practise, $\mathbf{x}$ and $\mathbf{y}$ are (randomly generated) input for tangent linear and adjoint codes (subroutines), respectively, and the inner products have to result in

the same value within the computing accuracy.

IFS contains a large number of tangent linear and adjoint routines which are tested at the time of writing. It is best to do the testing individually for each routine and also for the model as a whole. In the IFS there is a built-in facility to test the tangent linear and adjoint of the forecast model but not observation operators. From the maintenance point of view, there are frequent changes to the non-linear code, the observation operators for example, and each such change has to be incorporated in the corresponding tangent linear and adjoint routines. Also changes in the internal data structures or subroutine arguments need to be done consistently in the tangent linear or adjoint codes. Currently at ECMWF, the tangent linear and adjoint coding is finished, however adding new features, like a new observation type which requires a new observation operator, brings along a need for development of the linear codes.

4.1 (b) Gradient test. Testing the gradient of the cost function is similar to that of testing the tangent linear code: the gradient of the cost function must asymptotically point to the same direction as is the difference between two realizations of the cost function which are separated by a small perturbation in model state. A Taylor expansion for the cost function is given by

$$J(\mathbf{x}_0 + \delta\mathbf{x}) = J(\mathbf{x}_0) + (\delta\mathbf{x})^T \nabla J(\mathbf{x}_0) + O(\delta\mathbf{x}^2)$$

The perturbation of cost function is given by $\delta\mathbf{x} = -\alpha \nabla J(\mathbf{x}_0)$

and therefore the quantity

$$[J(\mathbf{x}_0 + \delta\mathbf{x}) - J(\mathbf{x}_0)] / (\alpha \|\nabla J(\mathbf{x}_0)\|^2) = 1 - O(\alpha)$$

approaches unity from below. There is a range of orders of magnitude of α for which this is true. Outside the range it is not true because of the computing accuracy for too small values of α, or because of the gradient of being non-quadratic for too large values of α. In practise, the value if α is repeatedly decreased by one order of magnitude resulting in a printout with more and more of 9's appearing until the computing accuracy is been reached.

A failure in the gradient test is a definite signature of an error somewhere in the variational assimilation system and not necessarily just in the tangent linear or adjoint coding. There are many ways of trouble-shooting, one of which is to reduce the dimension of the problem, for instance limiting oneself to a single observation case. The gradient may pass the test if a coding error in the adjoint code creates only a relatively small error in the gradient, so it is important to keep testing the tangent linear and adjoint codes as explained above.

4.1 (c) Convergence checks. The minimization of the cost function faces convergence checks. A trivial test of convergence is to check that the value of cost function decreases in every iteration. This is actually a built-it feature of the decent algorithm used in the IFS. For quadratic minimization problems, the norm of the gradient of the cost function should decrease in every iteration, apart from the rounding errors. The cost function at ECMWF assimilation system is non-quadratic and therefore the norm of the gradient can locally be larger than in the previous iterations when entering a new "valley" in the cost function topology. The gradient test is performed in every minimization at the first and the last minimization steps, as described above. The user also receives a note from the minimization algorithm if the norm of the gradient has not been reduced by more that a predefined factor which is dependent on the number of iterations.

4.1 (d) J_o break-down and screening statistics. The observation term of the cost function describes the misfit of the model state to the observations scaled with their relative accuracy, which is for an individual datum

$$J_o = \left(\frac{y - H(x_b)}{\sigma_o} \right)^2$$

The expectation for the term before the minimization is given by

$$\langle J_o \rangle = \frac{\sigma_o^2 + \sigma_b^2}{\sigma_o^2} = 1 + \left(\frac{\sigma_b}{\sigma_o} \right)^2$$

and should always be greater than one. If the quality of the background and the observations is similar then the value should be around two. The observation term can be broken down to contributions from different observation types, areas and observed variables and an average J_o contribution for those can be computed by dividing by the cost function by the number of observations. A troublesome subset of observations will show up in this way.

The printout of screening statistics comprises tables of the number of observations rejected (and for which reason) and the number used in the assimilation, and reveals for instance if an observation type is missing. This diagnostic printout as well as the J_o break-down are produced by default in IFS and together they tell reliably

- if two assimilation experiments use the same observations as input (identical printout of the screening statistics)
- if two assimilation experiments have been started from the same initial state (for the same observations as input, the initial value of the cost function should be identical)
- if the version of the IFS is the same for two experiments (for the same observations as input, also the final value of the cost function should be identical)

In research experimentation at ECMWF, a common wish for new experiments is that there is a comparison available, either an operational products or another experiment.

4.2 Experimentation

4.2 (a) Forecast scores. Modifications to the operational assimilation system are usually justified with positive or neutral forecast scores (defined by anomaly correlation) as compared with the operational scores. A common practice is to perform one or several two-week assimilation experiments in order to objectively see the effect of the changes in assimilation or forecast model. Often the experiments are run for different seasons, as well. For major changes in the operational suite also a separate e-suite parallel to the operations is run to ensure the quality of the products and a smooth transition to the revised system.

Figure 4 gives an example of the forecast scores in a typical two-week pre-implementation experiment. In this case an hourly observation screening is tested in 4D-Var, i.e. allowing more observations from frequently reporting stations into assimilation (dotted line). The forecast scores for Northern Hemisphere are comparable with 4D-Var experiment using six-hourly observation screening (dashed) and better than 3D-Var (full) but for the Southern Hemisphere the hourly screening is clearly a bad option for 4D-Var. Based on these experiments it was decided to continue 4D-Var experimentation using the six-hourly screening of observations (or 3D-screening), and to investigate the reasons behind the bad performance on the Southern Hemisphere.
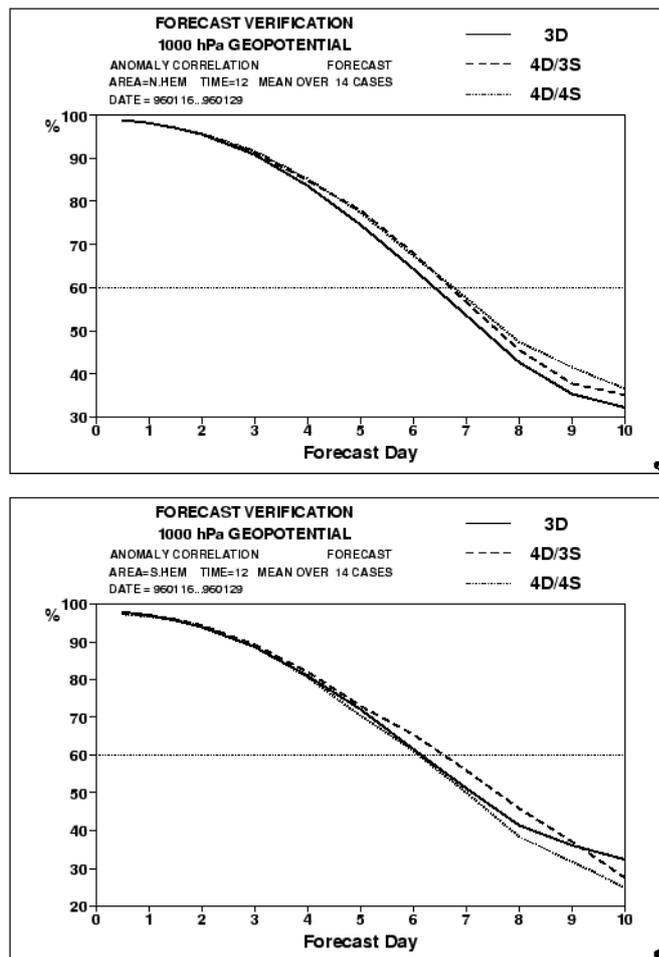
Figure 4. An example of the forecast scores in a two-week assimilation experiment for Northern Hemisphere (top panel) and Southern Hemisphere (bottom panel) for geopotential height at 1000hPa. Solid line is for 3D-Var, dashed line for 4D-Var using same observations as 3D-Var (3D-screening) and dotted line for 4D-Var using extra surface observations from frequently reporting stations (4D-screening).

4.2 (b) Observation r.m.s. fit and histograms.   The fit of the observations to the background and analysis can be conveniently examined by r.m.s. plots and histograms which are automatically generated for each assimilation experiment. An example of the r.m.s. plot for AIREP wind and temperature observations used in an assimilation experiment is given in Fig. 5 . One can see that the r.m.s. difference is smaller for the analysis departures (dotted lines) than for the background departures (solid lines) - the analysis is said "to have drawn to the data". The biases are also displayed and they have generally been reduced in the assimilation. Note that in these plots a desirable feature is a small r.m.s. of the background departures. This value is generally smaller, for instance, in 4D-Var than in 3D-Var indicating improved accuracy of the 4D-Var assimilation compared to 3D-Var. A small r.m.s. of the analysis departures is however not a design criterion as such. One could, for instance, specify too small observation errors which would result in unrealistically small r.m.s. of the analysis departures which might deteriorate the subsequent short range forecast, i.e. r.m.s. of the background departures would increase.

A similar diagnostic plot is the histogram of departures which is usually plotted for single level observations, like SYNOP or DRIBU reports. Figure 6 gives an example of histogram for SATOB (or cloud track) wind observations. Both the background and analysis departures are displayed. One can note that the mean and standard deviation of

the departure distribution is smaller after the assimilation which means that information has been extracted from the observations. The distribution of background departures should be approximately Gaussian with mean near zero.
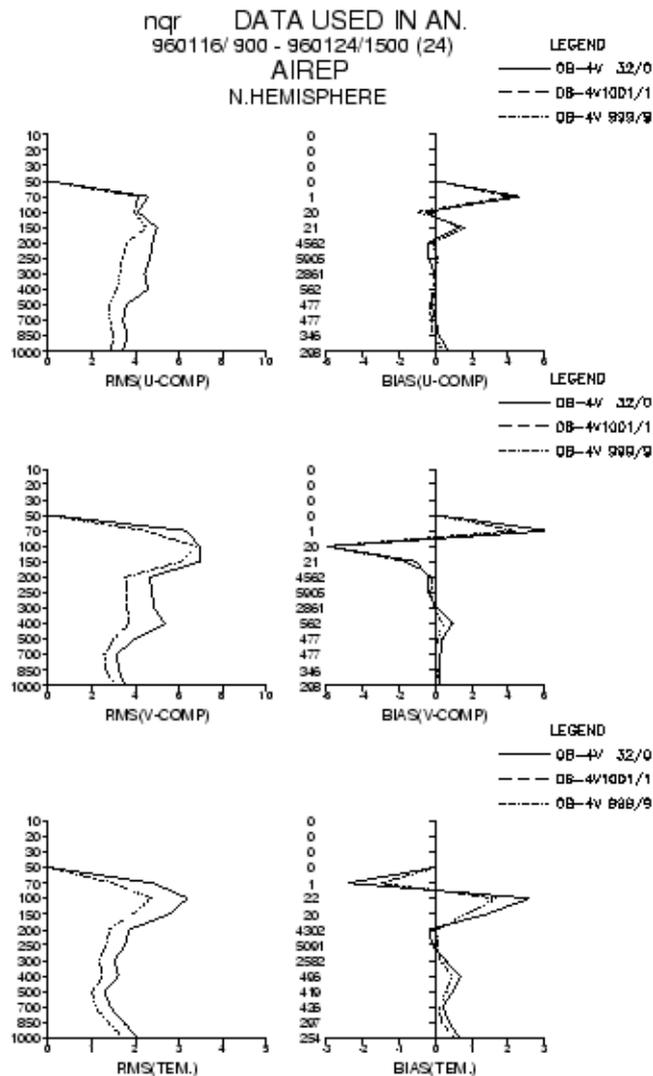
Figure 5. An example of an r.m.s. plot for AIREP wind and temperature observations. r.m.s. on the left and bias on the right, and number of observations used in the assimilation in the middle. Solid line is for background departures and dotted for analysis departures.

4.2 (c)   Mean and r.m.s. of analysis increments.   The analysis increments can be reconstructed after the assimilation by subtracting the background from the analysis. The mean and r.m.s. of these increment fields can reveal a lot of the performance of the assimilation system. First, large mean increments may result from using biased observations which may be for instance due to incorrect bias correction. It may also be a sign of an unsuccessful model change which has introduced a model bias which may appear only locally. For instance an albedo change over snow covered areas may cause a bias to appear in the background which the unbiased observations try to correct. Second, the r.m.s. of the analysis increments should be small which is a sign of consistency of short range forecast and observations.
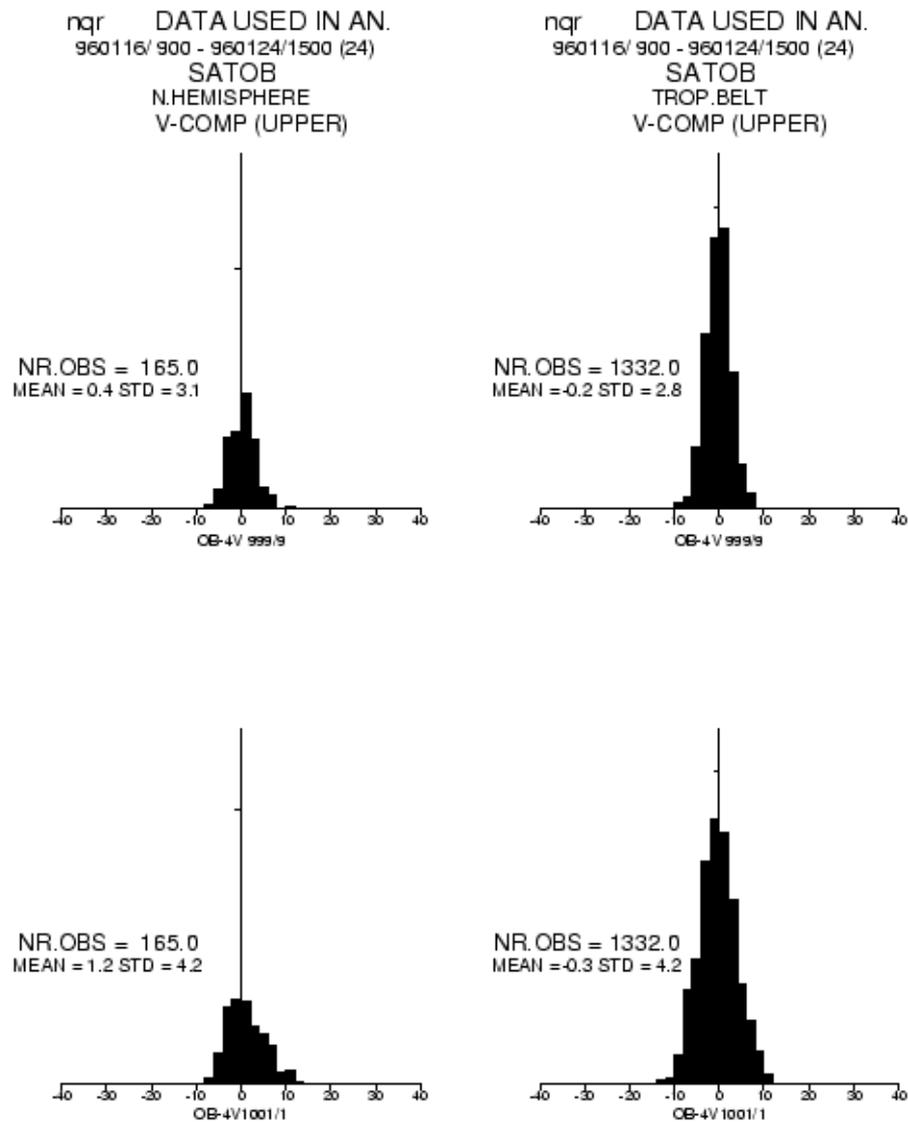
Figure 6. An example of the histogram of the SATOB wind (v-component) fit to the analysis (top panel) and background (bottom panel).

When 4D-Var was about to be implemented at ECMWF, one of the strong points for the implementation was the smaller analysis increments in 4D-Var compared with 3D-Var. Later when a modification of 4D-Var to use more observations from frequently reporting stations by applying serial correlation of observations errors was discussed, one aspect for the implementation was the further reduced analysis increments (Fig. 7), for instance over the Northern Atlantic. The impact due to the addition of more observations can be revealed simply by comparing the difference between the analyses from the two assimilation systems in the r.m.s. sense (Fig. 8). The largest impact is, as expected, over the areas where the conventional observational coverage is not a very dense one, and in areas where the atmospheric flow tends to be more unstable, like the storm track areas.
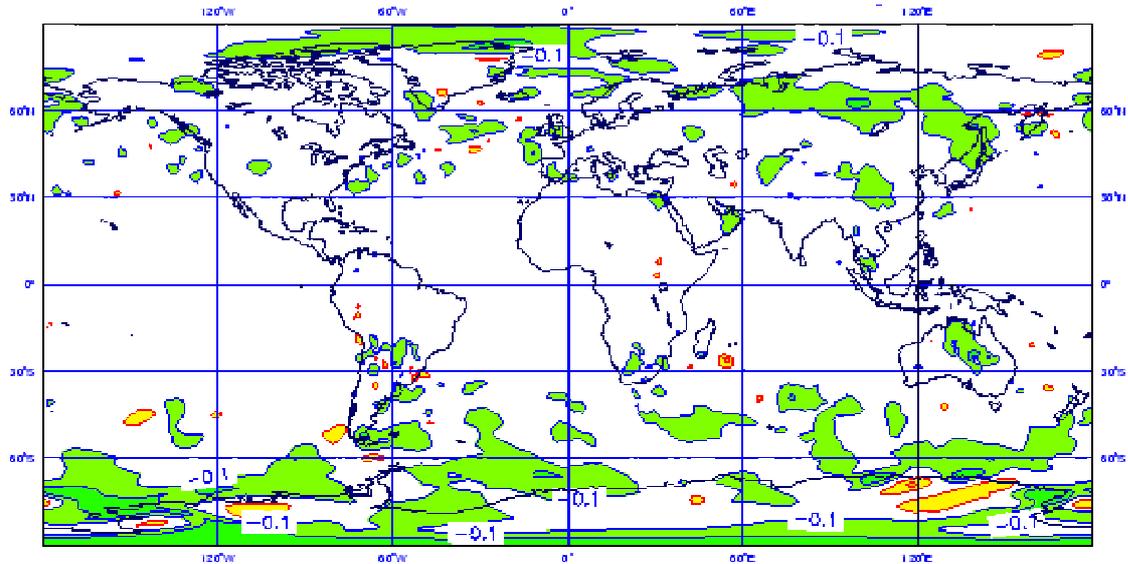
Figure 7. The improvement of the consistency of the background field with observations when using 4D-screening (plus serial observation error correlation plus joint variational quality control). The quantity is the 1000hPa geopotential difference between r.m.s. of analysis increments in the experiment and its control, for period 11 to 24 December 1997. Contours are +/−0.1, +/−0.25 and +/-0.50 decametres. Green (orange) areas denote smaller (larger) analysis increments in the experiment than in its control.
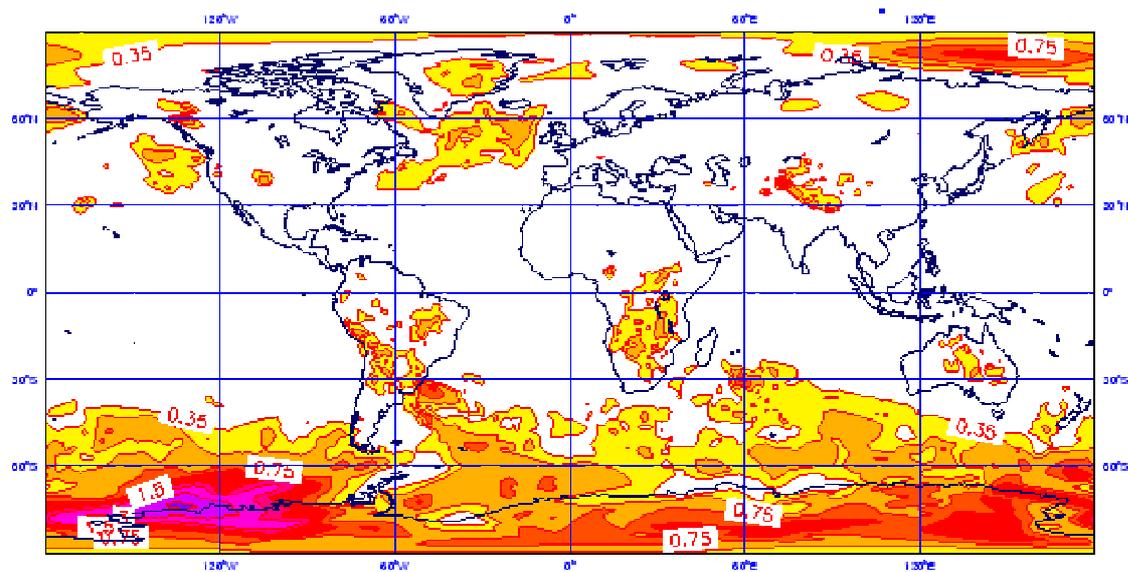
Figure 8. The impact on analyses of applying 4D-screening (plus serial observation error correlation plus joint variational quality control). The quantity is the 1000hPa geopotential r.m.s. of analysis differences between the experiment and its control., for period of 11 to 24 December 1997. The contours are 0.35, 0.50, 0.75,.1.00,.1.50,.2.00 and 3.00 decametres. The largest impact is over the areas of sparse conventional observational coverage.

4.3 Operational monitoring

4.3 (a) Cross-validation with satellite products. The operational department at ECMWF is constantly monitoring the quality of the operational production, e.g. use and quality of observations, their availability, character of the analysis increments etc. Many of the suggestions for improving the assimilation system actually come from the results of this intense monitoring. More details of their activities are given in the appropriate Training course module. One method which is used both by the operations and the research is the cross-validation with satellite products. There are some parameters for which direct (in situ) observations are scarce, like clouds or position of a tropical storm, and for those a visual comparison with satellite products may be very useful.

4.3 (b) Back-tracking problems with sensitivity products. An often occurring situation in weather forecasting is an unpredicted small scale flow pattern, followed by a question why it was not predicted. In these cases error back-tracking has long been used (even with subjective forecasts). The adjoint model provides one extra tool for doing the back-tracking. Sensitivity to analysis "errors" can be calculated using the adjoint model in the following way. Two day forecast error is fed to the adjoint model as a forcing and the adjoint calculations result in a gradient, or sensitivity pattern, with respect to the initial condition. This sensitivity pattern tells where and in which direction the initial condition should be perturbed in order to achieve a smaller two day forecast error. Of course, the two day forecast error is not entirely due to an inaccurate initial condition but also due to the model error over the two day integration time. Nevertheless, this sensitivity pattern can give a useful clue for the analyst about where the reason for the forecast failure may be found. This method has been successfully used at ECMWF.

4.4 Estimation and tuning

4.4 (a) Observation and background errors. The specification of observation and background error covariances for the assimilation system is an essential step which determines the relative weight of the observations and the background, respectively. These statistics are not known exactly but are estimated for each assimilation system. Therefore, as the observing network or the assimilation system changes, the statistics may require tuning for optimal performance.

There is a reliable method (Hollingsworth-Lönnberg method) for observation and background error estimation over data rich areas (as explained elsewhere in Lecture Notes). An example of the behaviour of background error covariances is given in Fig. 9 for AIREP temperature observations over North America at 200hPa. The background departures are correlated at short distances and the correlation rapidly decreases with increasing distance. With distances over about 500km there is hardly any correlation left. In the estimation method it is assumed that the observation errors are not correlated between the stations. This enables partitioning the perceived short-range forecast error variance into contributions from the observation and background errors. A curve is fitted (dashed line in Fig. 9) to the histogram of covariance values (filled circles in Fig. 9) and the intersect of the fitted curve with the ordinate gives an estimate of the background error variance, the rest of perceived short-range forecast error variance being due to the observation error.

4.4 (b) Verification of structure functions. The structure functions are specified from a sample of short-range forecast differences (24-hour minus 48-hour forecast differences in the NMC method). The Hollingsworth-Lönnberg method is not for re-tuning or changing them, but the method can be used for verifying how well the shape of specified structure functions is supported by the covariance of background departures. An example of the specified structure function is given in Fig. 10 for temperature at model level 10 (about 200hPa) at mid latitudes. Comparing Figs. 9 and 10 reveals the sharper horizontal structure of the short range forecast error as estimated from AIREP observations departures (calculated at resolution T213) than the modelled structure function at truncation TL159. The difference is partly explained by the resolution. More importantly, the modelled structure function is

a global one dominated by Southern Hemisphere mid latitudes, whereas the estimated one is from Northern America with a very dense data coverage which tends to shorten the horizontal scale of short-range forecast error.
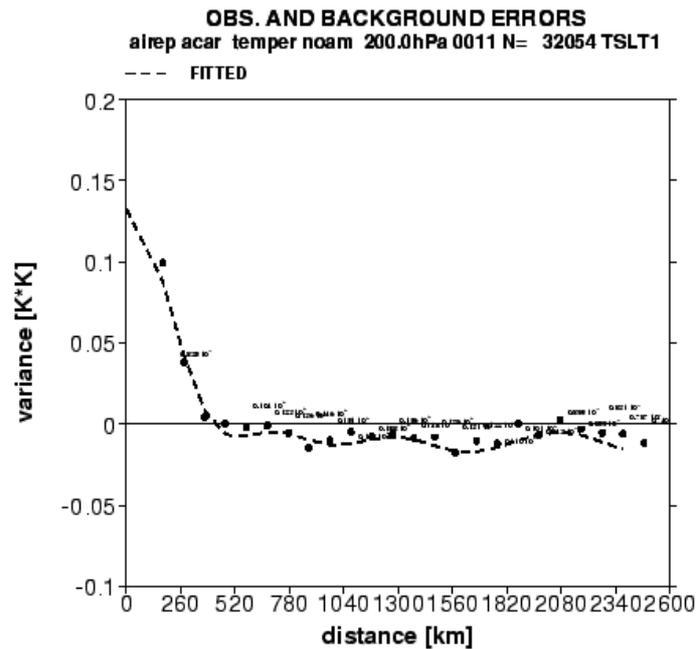
Figure 9. An example of background error covariance for AIREP (ACAR) temperature observations in 4D-Var over the period of 1 September 97 - 14 October 97 over North America at 200hPa. In this case, the estimated background error variance at zero distance is about $0.13K^2$ which would indicate a background error of about $0.36K$. As the total perceived error variance is $1.03K^2$ (not shown), the estimated observation error is therefore $0.95K$.
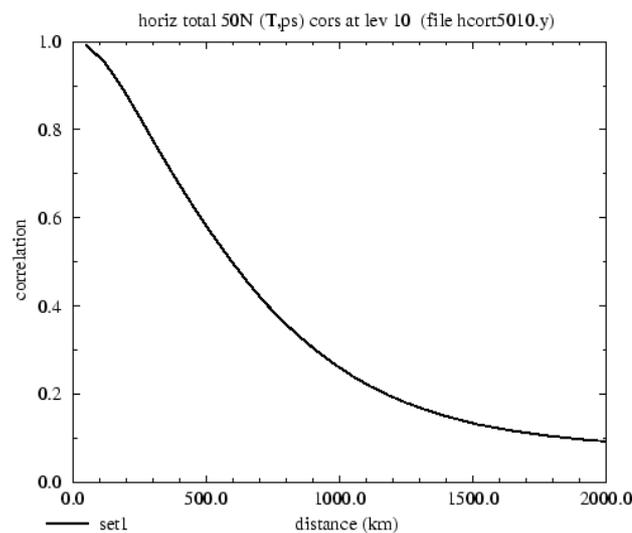
Figure 10. The specified structure function for temperature at latitude 50°N. Note that the horizontal scale of the absissa is different from Fig. 9 .

REFERENCES

Further reading:

Daley, R., 1991: Atmospheric data analysis. Cambridge University Press. 457pp.

Eyre, J.R. 1992: A bias correction scheme for simulated TOVS brightness temperatures. ECMWF Research Dept. Tech. Memo. 186.

Harris, B. 1997: A revised bias correction scheme for TOVS radiances. ECMWF Technical Memorandum No. NN.

Hollingsworth, A. and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111-136.

Hollingsworth, A. and P. Lönnberg, 1989: The verification of objective analyses: Diagnostics of analysis system performance. *Meteor. Atmos. Phys.*, **40**, 3-27.

Järvinen, H and P Undén, 1997: Observation screening and background quality control in the ECMWF 3D-Var data assimilation system. ECMWF Technical Memorandum No. 236, 26pp

Klinker, E., Rabier, F. and Gelaro, R., 1998: Estimation of key analysis errors using the adjoint technique. *Q. J. R. Meteorol. Soc*. **124**,1909-1933

Lönnberg, P, 1989: Developments in the ECMWF analysis system. ECMWF Seminar on Data assimilation and the use of satellite data. 5-9 September 1988, 75-119.

Lönnberg, P and D Shaw, 1985: Data selection and quality control in the ECMWF analysis system. ECMWF Workshop on The Use And Quality Control of Meteorological Observations, 6-9 November 1984, 225-254.

Lönnberg, P. and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A**, 137-161.

Lönnberg, P and D Shaw (Eds.), 1987: ECMWF Data Assimilation Scientific Documentation. Research Manual 1.

Norris, B, 1990: Pre-processing - General data checking and validation. ECMWF Meteorological bulletin, 28pp. Available in request from ECMWF, Shinfield Park, RG2 9AX, Reading, Berkshire, England.

Pailleux, J., 1993: Organisation of 3D variational analysis within the "IFS/ARPEGE" project. Future plan at Meteo France. ECMWF Workshop Proceedings on Variational assimilation, with special emphasis on three-dimensional aspects. 9-12 November 1992. 37- 47.

Rabier, F., Klinker,E., Courtier, P and Hollingsworth, A., 1996: Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc*. **122**,121-150