

Observation screening and
background quality control in the
ECMWF 3D Var data
assimilation system

H. Järvinen and P. Undén

Research Department

September 1997

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.





Observation screening and background quality control in the ECMWF 3D-Var data assimilation system

by Heikki Järvinen and Per Undén

ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England

Abstract

A system for observation screening and background quality control has been developed and implemented for the variational data assimilation at ECMWF. The first operational implementation of the 3-dimensional variational data assimilation at ECMWF in January 1996 relied on the Optimum Interpolation system to provide a clean quality-controlled array of observations for the assimilation. The new module for observation screening and background quality control removes, together with two other recent developments, the dependence of the variational data assimilation scheme on the Optimum Interpolation. These other developments are the conjugate gradient method for providing an estimate of the analysis error and the variational quality control of observations. They were implemented operationally together with the new observation screening and background quality control system at ECMWF in September 1996.

The first implementation of the new system for observation screening and background quality control is designed to reproduce the data selection of the Optimum Interpolation for two reasons. First, it enables a smooth continuation of the current operational production. Second, a close reproduction is needed for verifying the correctness of the new system against the previous one. The newly developed system is however much more flexible and enables modifications for more optimal use of observations in the 3/4-dimensional variational data assimilation. It also fully supports a massively parallel computing environment with a distributed memory.

The performance of the new system for observation screening and background quality control is assessed both in the pre-operational parallel tests as well as in the operational production with the ECMWF 3-dimensional variational data assimilation system. The impact of the first implementation is, as expected, neutral in terms of the quality of analysis and forecast products of ECMWF.

1. INTRODUCTION

The variational formulation of the data assimilation problem essentially consists of minimizing a cost function which measures the closeness of fit of the model state (the analysis) to observations, and to an a priori estimate of the state of the atmosphere (the background). The current formulation of the ECMWF operational 3-dimensional variational (3D-Var) data assimilation system (*Courtier et al.*, 1993, 1997) has a cost function with quadratic terms for observation and background departures, except for ambiguous scatterometer winds (*Gaffard and Roquet*, 1995). That is a direct consequence of the underlying assumption that the observation and the background errors follow Gaussian distributions. This is true to a large extent for the observations (*Hollingsworth et al.*, 1986) but it imposes a strong constraint for the quality control of observations. In the real world,

there are always observations grossly in error due to a multitude of reasons and these errors do not follow a Gaussian distribution (*Hollingsworth et al.*, 1986; *Lorenc and Hammon*, 1988). The formulation of the assimilation problem is therefore not valid unless the outlying observations are removed from the analysis. In practice, in the variational data assimilation, as described in *Courtier et al.* (1997), the gradient of the cost function with respect to the control variable will get unrealistically large values for such observations and large analysis increments are produced at these locations. An observation background quality control is therefore a necessary step in data assimilation and it aims to recognize and reject the observations grossly in error.

The observation term in the cost function can be formulated to be non-quadratic by applying a non-Gaussian distribution for observation errors, accounting for a probability of gross observational error (*Lorenc and Hammon*, 1988). The functional form of the cost function can be regulated through the non-Gaussian errors in such a way that for large observation departures the cost function flattens out instead of increasing quadratically, as is the case for the Gaussian errors. Consequently, the gradient of the cost function with respect to the control variable will approach zero for large departures and such observations no longer affect the minimization. This mechanism is called the variational quality control (*Lorenc and Hammon*, 1988) and it is thus able to reject or re-activate outlying observations during the minimization by combining all information, not just that of the background. It seems therefore to be possible to abandon the background quality control once the assimilation system is adapted for non-Gaussian errors and can naturally deal with outlying observations through variational quality control. In practice, the background quality control is still necessary for various other reasons. Redundant observations, for instance a duplicated aircraft report, increase the initial value of the observation term of the cost function resulting in a spurious decrease of the cost function during the minimization as the model field is more closely fitted to that observation. More seriously, redundant observations increase the weight of that particular piece of information and as a result the analysis will be drawn more to that data than it should be. Furthermore, if that data happens to be biased, the resulting analysis may suffer as these observations support each other, and in the worst case may lead to rejection of other, possibly correct, observations through variational quality control. There are also computational cost reasons for rejecting the observations with obvious gross errors in the background quality control. The observation screening together with the background quality control and the variational quality control can therefore be seen as complementary measures for controlling the quality of the observations and of the analysis through the selection of the most representative data for the data assimilation.

This report concentrates on the observation screening and the background quality control in the ECMWF 3D-Var data assimilation system. Throughout the report the term observation screening is intended to cover not only the pure background quality control of observations but all the decisions that are made for the use of observations in the assimilation. The first implementation of the observation screening is designed to reproduce the selection of observations that the ECMWF implementation of the Optimum Interpolation (OI) (*Gandin*, 1963; *Lorenc*, 1981 and 1986; *Shaw et al.*, 1987) used to make. There are two reasons for this. The first reason is that over the years, a lot of experience has been gathered through the operational monitoring of the global observing system. This experience has been built into the ECMWF observation selection algorithm of the OI analysis code (*Lönnerberg and Shaw*, 1985 and 1987; *Lönnerberg*, 1989). There is a risk that this knowledge might be lost if it is not properly migrated to the new observation screening algorithm for the 3D-Var data assimilation system. The second reason is the need for a thorough verification of the new software. It must be verified that no observations are unintentionally rejected or activated for assimilation but really that the operational practice in the OI can be reproduced to a high degree



using the new observation screening and background quality control system of the IFS (Integrated Forecasting System).

The plan of this report is as follows. The logical structure of the selection algorithm is explained in detail in Chapter 2 while the results of the assimilation and forecast experiments are given in Chapter 3. Some special requirements for the observation screening in the 3D/4D-Var assimilation system are addressed in Chapter 4. Some remaining issues related to a massively parallel computing environment with distributed memory are briefly reviewed in Chapter 5.

2. THE STRUCTURE OF THE OBSERVATION SCREENING

2.1 The ECMWF operational 3D-Var data assimilation system

The 3D-Var data assimilation is an intermittent scheme, like the OI, where the analysis of meteorological fields and of their error structures is performed at one analysis time, followed by a forecast step to advance the analysed states over a short period of time. The implementation of the ECMWF 3D-Var data assimilation system is such that there are four analysis steps each day with data windows of six hours centred around the observing times at 00, 06, 12 and 18UTC. The forecast model is used to advance fields by six hours from each analysis to form the background fields for the next data window. A ten day forecast is launched once a day from the 12UTC analysis. Each analysis step begins with the observation preprocessing. The operational practice is to wait past the end of the data window until the so called cut-off-time is reached. The observations are then extracted from a data base of decoded data. 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 (Norris, 1990). The so called RDB (Report Data Base) flag is assigned at this stage. Then an observation file suitable for assimilation is created in a separate 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 six hour data window available at the cut-off-time and is an input for the IFS. The observation screening then selects the best quality and unique observations preferably close to the middle of the data window as the background is not interpolated to the exact time of the observation. Unlike the OI, the 3D-Var data assimilation is global and therefore no separate data selection for analysis boxes is needed.

The ECMWF 3D-Var data assimilation system makes use of an incremental minimization scheme (Courtier *et al.* 1994 and 1997) to reduce the computational cost. In that scheme the minimization is performed at a lower model resolution, and in the 4D-Var context currently with a near-adiabatic model version, around a constant model state. The model initial state at full resolution is then updated using the analysis increments of the lower resolution minimization. These incremental minimization steps are repeated until a necessary convergence is achieved. The variational data assimilation therefore starts with a 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. The observation screening is thus integrated with the first high resolution trajectory run of the incremental minimization scheme. For the observation screening, the background errors are interpolated to the observation locations for the observed variables. Therefore the observation screening logically depends also on the estimation of background errors.

Technically, the final result of the observation screening is a pair of observation arrays. The original observation array now contains observations complemented by the background departures together with quality control information for most of the observations. This array is stored for later feedback. The compressed observation array is a subset of the original array of observations and it is passed to the subsequent minimization job. The compressed file contains only the observations to be used in the variational assimilation. Memory-wise, the first trajectory run is therefore the most demanding one as all the observations are kept in memory together with the full resolution model fields.

2.2 General rationale of observation screening in the 3D-Var assimilation system

The general logic in the 3D-Var observation screening algorithm is to make the independent decisions first, i.e. the ones that do not depend on any other observations or decisions. One example is the background quality control for one observed variable. These decisions can be carried out in any order without affecting the result of any other independent decision. The rest of the decisions are considered as mutually dependent on other observations or decisions, and they are taken next following a certain logical order. For instance, the horizontal thinning of TOVS reports is only performed for the subset of reports that passed the background quality control. Finally, the observation array is compressed for the minimization in such a way that it only contains the data that will be used.

2.3 The independent observation screening decisions

The observation screening begins with a preliminary check of the completeness of the reports. None of the following values should be missing from a report: observed value, background departure, observation error and vertical coordinate of the observation. Also a check for a missing station altitude is performed for SYNOP, TEMP and PILOT reports. The reporting practice for SYNOP and TEMP mass observations (surface pressure and geopotential height) is checked, as explained in the Appendix. At this stage also, the observation error for SYNOP geopotential observations is inflated if the reported level is far from the true station level. The inflation is defined as a proportion of the difference between the reported level and the true station altitude by adding 2% of the height difference to the observation error.

Next the observations are scanned for blacklisting. This consists formally of two parts. First, the selection of variables for assimilation is performed through the blacklist using a special data selection file. This controls which observation types, variables, vertical ranges etc. will be selected for the assimilation. Some other 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 with the observing system, as well as with the assimilation system itself. Second, a normal monthly monitoring blacklist is applied for discarding the stations that have recently been reporting in an excessively noisy or biased manner compared with the ECMWF background field. A full documentation of the new blacklisting mechanism is found in *Järvinen et al.* (1996).

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 by more than a predefined multiple. For the wind observations, the background quality control is performed simultaneously for both wind components. In practice, there is an associated background quality control flag with four possible values, namely zero for a correct, one for a probably correct, two for a probably incorrect and three for an incorrect observation, respectively. Table 1 gives the predefined limits for the background quality control in terms of multiples of the expected variance of the normalized background departure. Note that for AIREP, SATOB, DRIBU and PILOT wind, for DRIBU and PAOB surface pressure and for AIREP temperature the background quality control rejection limits are modified as explained in the Appendix.

Table 1: The predefined limits for the background quality control, given in terms of multiples of the expected variance of the normalized background departure. Flag values are denoted by 1 for a probably correct, 2 for a probably incorrect and 3 for an incorrect observation, respectively. The variables are denoted by u and v for wind components, z for geopotential height, ps for surface pressure, dz for thickness, T for temperature, rh for relative humidity and q for specific humidity, respectively.

Variable	Flag 1	Flag 2	Flag 3
u, v	8.00	18.00	20.00
z, ps	12.25	25.00	36.00
dz	2.25	5.06	7.56
T	5.00	10.00	12.00
rh, q	9.00	16.00	25.00

There is also a background quality control for the observed wind direction. The predefined error limits of 60° , 90° and 120° apply for flag values 1, 2 and 3, respectively. The background quality control for the wind direction is applied only above 700hPa for upper air observations for wind speeds larger than 15m/s. If the wind direction background quality control flag has been set to a value that is greater or equal to 2, the background quality control flag for the wind observations is increased by one.

For TOVS reports a background quality control procedure is applied in a separate 1D-Var module. This module is called from the appropriate observation operator for each TOVS profile with the model background humidity and temperature profiles that are interpolated to the observation location. The module makes 1-dimensional variational retrieval (Eyre *et al.*, 1993) for the measured radiances. The background quality control is done on the radiance departures on selected channels and also on the convergence of the variational retrieval (Kelly *et al.*, 1991). The module also delivers some upper and lower boundary conditions used for 3-dimensional TOVS retrieval in 3D/4D-Var (Andersson *et al.*, 1994). One should note that currently the ECMWF forecast model top level is at 10hPa whereas the TOVS instrument measures radiance upwelling partly from above the 10hPa level in the stratosphere (Eyre, 1989). The NESDIS retrievals are not used any longer in the ECMWF assimilation system.



2.4 The dependent observation screening decisions

Just before performing the dependent screening decisions, the flag information gathered so far is converted into a status of the reports, namely active, passive, rejected or blacklisted, and also into a status of the data in the reports. The reports with an RDB report flag value 2 (probably incorrect) or higher for latitude, longitude, date and time are rejected. For the observed data there are RDB datum flags for the variable and for the pressure, i.e. the pressure level of the observation. The rejection limits for these are as follows: all data is rejected for the maximum RDB datum flag value 3 (incorrect), non-standard level data is rejected for the maximum RDB datum flag value 2, and for the pressure RDB datum flag the rejection limit is 1 (probably correct). The background quality control rejection limits are flag value 3 for all the data and flag value 2 for the non-standard level data.

The first dependent decisions are the vertical consistency check of multilevel reports and the removal of duplicated levels from the reports. The vertical consistency check of multilevel reports is applied in such a way that if four consecutive layers are found to be of suspicious quality, even having a flag value one, then these layers are rejected, and in the case of geopotential observations also all the layers above these four are rejected. These decisions clearly require the quality control information and are therefore "dependent" on the preceding decisions.

The duplicated reports will be removed next. This is performed by searching pairs of co-located reports of the same observation type and then checking the content of these reports. It may, for instance, happen that an AIREP report is formally duplicated having 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.

The redundancy check of the reports, together with the level selection of multilevel 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 analysis time 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 analysis time with most active data is retained. All the data from the multilevel 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 in the vicinity of standard levels and closest to the analysis time. 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.

Finally, a horizontal thinning is performed for the AIREP and SATEM (which consists mainly of 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 125km. Reports from different airliners may however be very close to each other thus ensuring as many independent pieces of information as possible for the assimilation. This is done by treating the AIREP reports from one airliner separately



from the other platforms. In this removal of redundant reports the best quality data is retained as the preceding quality control is taken into account. In the vertical, the thinning is performed for layers around standard pressure levels thus allowing more reports for ascending and descending flight paths. Thinning of SATEM reports is done at two stages. First a minimum distance of about 70km is enforced, and thereafter a repeated scan is performed to achieve the final separation of roughly 250km between reports from one platform. The thinning algorithm is the same as used for AIREPS but in the case of SATEM 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 analysis time is preferred. A summary of the current status of use of observations in the 3D-Var data assimilation is given in Table 2.

Table 2: A summary of the current use of observations in the 3D-Var data assimilation at ECMWF. The variables are as in Table 1, with the addition that *2m rh* stand for 2 metre relative humidity and *Tb* for brightness temperature. The observation types are shortened by *SYNOP* for synoptic surface observations, *AIREP* for aircraft reports, *SATOB* for satellite cloud track winds, *DRIBU* for drifting buoy reports, *TEMP* for radiosonde soundings, *PILOT* for wind soundings, *SATEM* for satellite temperature soundings, *PAOB* for pseudo observations of surface pressure and *SCATT* for scatterometer reports

Observation type	Variables used	Remarks
SYNOP	<i>u, v, ps (or z), rh</i>	<i>u</i> and <i>v</i> used only over sea, in the tropics also over low terrain (<150m). Orographic rejection limit 6hPa for <i>rh</i> , 100hPa for <i>z</i> and 800m for <i>ps</i>
AIREP	<i>u, v, T</i>	Not used in full resolution. Used only below 50hPa
SATOB	<i>u, v</i>	Selected regions and levels
DRIBU	<i>u, v, ps</i>	Orographic rejection limit 800m for <i>ps</i>
TEMP	<i>u, v, z, q</i> and <i>2m rh</i>	Used on or closest to standard pressure levels. <i>q</i> only below 300hPa. 10 metre <i>u</i> and <i>v</i> used over land only in tropics over low terrain (<150m). Orographic rejection limit 10hPa for <i>u</i> and <i>v</i> , 100hPa for <i>z</i> , 6hPa for <i>rh</i> and -4hPa for <i>q</i>
PILOT	<i>u, v</i>	Used on or closest to standard pressure levels. 10 metre <i>u</i> and <i>v</i> used over land only in tropics over low terrain (<150m). Orographic rejection limit 10hPa for <i>u</i> and <i>v</i>
SATEM	<i>Tb</i>	<i>Tb</i> used everywhere with a varying selection of channels over different regions and over land/sea. NESDIS retrievals not used.
PAOB	<i>ps</i>	Used south of 19°S. Orographic rejection limit 800m for <i>ps</i>
SCATT	<i>u, v</i>	Used if SST warmer than 273K and both observed and background wind less than 25m/s

After the observation screening roughly a fraction of 1/10 of all the observed data are active and so the compressed observation array for the minimization run only contains those data. That large compression rate is mainly driven by the number of SATEM data as after the screening there are only roughly 5% of the SATEM 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.

3. CHARACTERISTIC PERFORMANCE OF THE OBSERVATION SCREENING

3.1 Performance in the 3D-Var data assimilation experiments

Several weeks of assimilation experiments were performed at various stages of the development and migration effort of the operational data assimilation codes at ECMWF. The experimentation was followed by a two and a half month parallel run with operations, from the beginning of July 1996 until the operational implementation on the 18th of September 1996. Separating the effect of the new observation screening on the forecast scores from the effects of all the other developments is not possible, especially as also a change in model physical parametrization was included in the last part of the parallel run. The similarity in performance can be shown by studying the model background and analysis fit to the observations in terms of root mean square (RMS) differences and bias and also by studying the number of active observation used in the analyses. Both the model fit to observations (not shown) and the number of observations are very similar in the operational and experimental runs. There are some minor differences in the use of observations and they are believed to result mainly from the use of different background errors in the background quality control. Different algorithmic choices in the observation screening code also necessarily result in minor differences. Some of the differences are due to coding errors that were identified in the earlier operational implementation and were removed. These changes result in a more correct use of observations in the new system.

Tables 3-9 show examples of the performance of the new screening system compared with the previous operational one. As mentioned before, the differences between the two systems are not only due to the new screening code. There were also very different background errors and the OI quality control check was replaced by variational quality control. The impact in terms of the number of data used are displayed for one week, i.e. 28 assimilation cycles, of the pre-operational parallel run on the Fujitsu VPP700 with the complete new data assimilation system. The result is typical of the parallel run and of research experiments being done prior to that, except for correction of a few mistakes identified in the earlier experiments.

Table 3 displays the number of TEMP wind observations used in the one week period. There are slightly more TEMP winds (Table 3) in the new system at low levels. At 1000hPa level the global difference is +11% for the new system. At medium and high levels the numbers are very similar or slightly less for the new system. The reason for the difference at the lowest level is a coding error in the old system where the 925hPa level observation was not used when the surface observation was rejected. In the new system the error is corrected and thus the numbers are higher. The same error affected also the PILOT winds (not shown) and in the new system there is an increase of PILOT winds at the lowest levels in the tropics by a factor of two. Another small difference in the use of PILOT winds in the old and in the new system arises from the pressure assignment for the observation levels for which only the height is reported. In both systems, a pressure for the PILOT wind observation is assigned by integrating the hydrostatic equation from the model surface using the



model temperature and humidity profiles at the observation location. The old system had however a deficiency in making this assignment for observations close to the model orography.

Table 3: Number of TEMP winds used in one week of experimental assimilation (new) 1-7 August 1996 for all four cycles of the day compared with operations (old)

Level [hPa]	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
1000	4308	4729	618	720	218	281
850	5350	5390	739	742	372	371
700	5815	5806	784	783	392	393
500	5858	5843	750	748	391	390
400	5816	5804	731	730	393	393
300	5763	5748	711	711	390	390
250	5707	5687	708	708	382	378
200	5671	5655	700	698	374	373
150	5588	5574	684	682	371	366
100	5441	5430	655	646	368	368
70	5013	5006	499	490	342	344
50	4816	4811	446	437	317	316
30	4319	4310	353	349	283	283
20	3383	3375	250	248	204	205
10	1789	1784	118	118	52	61

Geopotentials from TEMP reports (Table 4) are slightly fewer at the lowest level but a few more at higher levels outside the tropics. This is due to a slightly different (better) level selection algorithm of multi-level reports (TEMP and PILOT reports).

The amount of humidity data used from TEMP reports (Table 5) is similar in the two systems except for the lowest level where substantially less data is used in the new system. The difference at the lowest level is mainly due to a changed orographic rejection criterion. In the new system this depends on the difference between the model surface pressure and reported pressure, rather than on the station level pressure as in the old system. This change also affected the SYNOP humidities

(Table 9) and has since been corrected for SYNOPS but not for TEMPs since the TEMPs have many additional data and since almost all TEMPs have a co-located SYNOP which is used in most cases.

Table 4: As Table 3 but for TEMP geopotentials

Level [hPa]	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
1000	5109	5022	709	708	598	592
850	5452	5469	729	737	665	665
700	5597	5577	755	748	670	671
500	5564	5570	737	733	668	674
400	5496	5504	734	729	664	669
300	5432	5446	722	729	659	665
250	5363	5382	711	705	652	658
200	5300	5317	701	695	652	659
150	5217	5234	694	684	629	635
100	5032	5065	639	625	617	621
70	4662	4726	475	471	569	576
50	4492	4561	435	430	526	536
30	3996	4046	367	362	444	451
20	3263	3300	285	280	339	343
10	1749	1765	150	147	92	92

Table 5: As Table 3 but for TEMP humidities

Level [hPa]	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
1000	5824	3422	1213	676	570	405
850	7608	7243	1305	1215	800	768
700	6305	6257	1039	1036	718	705
500	6350	6271	1010	1006	704	700
400	6239	6166	981	967	662	658
300	5758	5684	940	933	570	564



Tables 6 and 7 display the use of AIREP data in the assimilation. For AIREP reports the thinning of dense observations is done in the new system with a quite different algorithm compared with the old one. In the new system the same thinning algorithm is used both for the AIREP and SATEM reports. The horizontal resolution of the thinned AIREP data is tuned to be approximately 125km which is the same as in the old system. The choice of the thinning resolution is arbitrary anyway and will be subject to a relaxation in the future. As can be seen in Tables 6 and 7, for AIREP winds and temperatures respectively, the number of data used in the new system is generally a bit higher. The exception is around 850hPa in the extra-tropics where very dense data around certain airports have been thinned more in the new system. Some, obviously erroneous, reports from above 50hPa are blacklisted in the new system.

Table 6: As Table 3 but for AIREP winds

Level [hPa]	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
1000	1781	1898	171	260	1867	1611
850	2170	2223	124	175	1292	1066
700	3028	3228	247	300	2205	2211
500	2848	3359	222	255	1695	1768
400	2800	3151	156	160	1565	1575
300	10562	11631	789	866	1979	2148
250	26748	29731	1960	2332	1985	2400
200	25167	29326	1092	1341	2050	2438
150	55	58	5	4	6	8
100	55	54	0	0	0	0
70	3	2	0	0	0	0
50	6	0	0	0	0	0
30	0	0	0	0	0	0
20	1	0	0	0	0	0
10	1	0	2	0	3	0



Table 7: As Table 3 but for AIREP temperatures

Level [hPa]	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
1000	1795	1751	157	213	2056	1609
850	2114	2089	116	148	1533	1080
700	2919	3025	237	268	2399	2211
500	2755	3130	205	214	1726	1759
400	2704	2951	149	149	1603	1577
300	10201	10548	763	787	1970	2112
250	25974	26792	1804	2068	1959	2377
200	25794	28717	1109	1317	2019	2401
150	49	50	5	5	6	8
100	50	50	0	0	0	0
70	0	0	0	0	0	0
50	0	0	0	0	0	0
30	0	0	0	0	0	0
20	0	0	0	0	0	0
10	0	0	0	0	5	0

The number of SATOB winds used (Table 8) is generally fairly similar in the old and new systems. At the upper levels however, especially in the tropics, there is a certain loss of data in the new system compared with the old one. This is an effect of the lower background errors for wind in the new system. The background quality control rejection limits for SATOB winds are set tightly in the new system so that they cut inside the range where the background departures are normally distributed and thus any change of background errors will have a noticeable effect. The loss of data is not considered to be serious since it is only a slight increase of the rejections that already take place. Re-turning of the limits will have to be done regularly anyway, in light of the performance of the observing systems.



Table 8: As Table 3 but for SATOB winds

Level [hPa]	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
1000	38	37	241	239	476	478
850	5601	5561	10615	10497	10834	10844
700	1727	1708	3362	2983	3529	3323
500	321	330	597	577	389	375
400	525	537	551	528	1170	1106
300	1140	1227	2158	2053	1629	1459
250	1539	1594	3205	3065	543	466
200	1239	1219	4373	4044	87	62
150	227	218	1025	936	13	10
100	2	0	8	6	0	0
70	0	0	0	0	0	0
50 >	0	0	0	0	0	0

Table 9 displays the use of SYNOP observations. There are somewhat fewer SYNOP winds in the new system than in the old one, probably due to different horizontal interpolation of orography and land sea mask in the two systems. The number of SYNOP winds used in 3D-Var is low anyway compared with the number of SYNOP pressure data used. Winds from SYNOPs are only used over sea or low tropical islands. The number of SYNOP pressure observations used is very similar in the two systems, with slightly more used in the new system in the extra-tropics. For the SYNOP humidities a substantial drop in numbers can be seen for the new system. This is happening mainly at 06 and 18UTC, not very much at 00 or 12UTC. At 06 and 18UTC the observing practice is different for a large group of SYNOP stations, so that they report Mean Sea Level pressure rather than station level pressure. Due to a poor formulation of the orographic rejection criterion, the difference between the reported pressure and model surface pressure were used, rather than the difference between the station height and model orography. This error has since been corrected (implemented operationally 8th of January 1997), and the numbers are now very similar.

Table 9: As Table 3 but for SYNOP winds, pressures and humidities

	N.HEMISPHERE		TROP.BELT		S.HEMISPHERE	
	old	new	old	new	old	new
<i>u</i>	11093	9954	3781	3699	1224	1213
<i>ps</i>	89947	90828	20573	20460	14259	14323
<i>rh</i>	49954	36326	13176	10783	8259	8756



DRIBUS and PAOBs are used in very similar numbers in the old and new systems (not shown). For SATEM reports there is a larger difference. As mentioned earlier, the thinning algorithm is different in the new system. Also the thinning is now done after the background quality control, not before as in the old system, and that adds to the number of data being let through the system. For SATEM reports there are about 15- 20% more data used in the new system.

3.2 Performance in the operational data assimilation

The observation screening was implemented operationally at ECMWF on 18th of September 1996 along with other developments prepared for the Centre's new massively parallel vector computer, a Fujitsu VPP700. Some aspects of the current operational use of observations are studied here with aid of the operational feedback files that are stored for each analysis cycle. These files contain all the observations together with information of the background and analysis departures, quality control flags, etc. In this section, the operational feedback files are studied for October 1996, in total for 124 assimilation cycles.

Figure 1 displays the time distribution for all the surface pressure observations from SYNOP and DRIBU reports during October 1996, as seen through the 6 hour data window. A histogram of the time difference between observation time and analysis time is shown using a binning of 30 minutes. Most of the SYNOP stations report every 6 hours (Fig. 1a) but there is also a considerable number of observations available on an hourly and a 3 hourly basis. The observation screening used with the 3D-Var assimilation selects predominantly observations made within the 30 minute interval closest to the analysis time. Roughly 85% of the observations made within this interval are passed for the assimilation whereas only a small fraction of the off-time observations, i.e. those with a larger time difference to the analysis time, are retained. For the reports from drifting buoys (Fig. 1b) the situation is similar with the exception that there are relatively more off-time observations available in every 30 minute interval. A rather small fraction of those are really used in assimilation, except from the interval closest to the analysis time where around 75% are passed for the assimilation. As a summary of Fig. 1 one may note that less than a half of the surface pressure observations are used in the assimilation. It is however not easy to make good use of the remaining observations within the 3D-Var context.

The time distribution for AIREP reports is displayed in Fig. 2 for October 1996. One can note that AIREP reports are fairly evenly distributed in time, and they are passed to the assimilation without preference for any particular observing time. Currently roughly 40% of the AIREP temperature observations (Fig. 2a) and 55% of the AIREP wind observations (Fig. 2b) are used in the 3D-Var assimilation, while the rest are rejected, mainly as redundant reports in the observation screening.

For the TEMP soundings (Figs. 3-5) the time distribution is such that the observations are mostly available at the main observing times, and also one hour before those times. Note that the reported times are launch times, not the actual observing times. The current use of standard pressure level TEMP geopotential (Fig. 3a) and wind observations (Fig. 4a) is at the level of 90% which is very satisfactory. There is however a large number of non-standard level observations still available (Figs. 3b and 4b) that are not used in the assimilation. Only about a half of the specific humidity observations (Fig. 5) are used, predominantly at the standard pressure levels (Fig. 5a). Note that above the 300hPa pressure level, the specific humidities are not used due to quality reasons. Without any doubt, the current use of TEMP observations is below the capabilities of the 3D-Var data assimilation system. The high rejection rate of TEMP observations is because they are only used on or around the 15 standard pressure levels, whereas the operational assimilation is performed for



the 31 levels currently in use in the forecast model. The TEMP wind and humidity observations could be used with a higher vertical resolution without almost any code changes, whereas the use of geopotentials at levels other than the standard ones requires some software development for calculation of significant level geopotentials in a consistent way from the reported significant level temperatures. Currently at ECMWF, the TEMP geopotentials are calculated for significant levels from the temperature and humidity observations but they are only used if the closest standard level geopotential observation is missing.

The majority of the standard pressure level wind observations from the PILOT soundings (Fig. 6a) are used in the assimilation, whereas only a small minority is used from other levels (Fig. 6b). The low overall usage of PILOT winds is mainly due to two factors. First, in the screening, the PILOT and TEMP winds are considered together and a co-located PILOT report is redundant to the TEMP report. Second, PILOT winds are used only on 15 standard pressure levels, whereas they are reported with much higher vertical resolution, i.e. also on the significant levels. Satellite cloud track winds (Fig. 7) are used at a level of 50% which is probably in a fair accordance with their respective impact.

A summary is given next of the use of observations in operational data assimilation at ECMWF. The number of different screening and data assimilation events is given in Table 10 for selected observation types for October 1996.

Table 10: Number of events for selected observation types due different reasons during the operational observation screening and data assimilation for October 1996 at ECMWF. Symbols are as in Table 1 and 2. with the addition that *Act* stands for the percentage of active reports, *Fg 1*, *Fg 2* and *Fg 3* stand for the percentages for the background quality control flag values 1, 2 and 3, respectively, and *Vqc* stand for the percentage of variational quality control failures. *A/C-part* (*B/D-part*) stands for the standard (non-standard) pressure levels of TEMP and PILOT reports.

Observation type and variable	Total no.	Act	Fg 1	Fg 2	Fg 3	Vqc
SYNOP <i>ps</i>	1 381 803	41.89%	0.47%	0.18%	0.63%	0.37%
AIREP <i>T</i>	894 922	41.11%	1.13%	0.13%	1.45%	0.02%
AIREP <i>u (v)</i>	876 987	57.15%	3.36%	0.16%	0.89%	0.01%
SATOB <i>u (v)</i>	596 535	40.88%	11.17%	0.92%	6.53%	0.28%
DRIBU <i>ps</i>	112 835	25.90%	0.74%	0.30%	1.44%	0.57%
TEMP <i>z</i> A/C-part	441 815	86.18%	1.19%	0.29%	0.64%	0.42%
TEMP <i>z</i> B/D-part	616 610	1.98%	0.41%	0.13%	0.24%	0.07%
TEMP <i>u (v)</i> A/C-part	386 261	94.20%	0.41%	0.04%	0.17%	0.11%
TEMP <i>u (v)</i> B/D-part	419 577	6.21%	0.97%	0.09%	0.40%	0.04%
TEMP <i>q</i> A/C-part	311 252	53.79%	2.08%	0.26%	0.36%	0.04%
TEMP <i>q</i> B/D-part	643 347	4.39%	3.06%	0.48%	0.59%	0.01%
PILOT <i>u (v)</i> A/C-part	118 038	78.23%	0.65%	0.06%	0.19%	0.13%
PILOT <i>u (v)</i> B/D-part	191 003	9.46%	0.66%	0.05%	0.25%	0.03%

The main reasons for the number of active observations in Table 10 are as follows. A lot of SYNOP and DRIBU reports from frequently reporting stations are rejected as redundant, not because of quality problems. AIREP reports are often received with a high horizontal resolution but a horizontal thinning is applied to limit their number. For SATOB reports, a stringent background quality control is applied and all the cloud track wind products are not fully used due to quality problems. For TEMP and PILOT reports, the standard pressure level observations are predominantly used and if missing, the closest level is used instead. No TEMP humidities are used above 300hPa level thus explaining the low percentage. For PILOTS, the percentage is low because preference is given to TEMP winds in case of close-in-time or simultaneous reports.

The number of observations with different flag values in Table 10 is in a qualitative agreement with those reported for the ECMWF OI analysis system (*Lönnberg and Shaw, 1985*). The background quality control of observations set the flag values 1, 2 and 3. The observations with flag value 1 (probably correct) are usually not rejected but accepted for the assimilation - in the redundancy checking, of course, a datum with flag value 1 is rejected before one with a flag value 0. It seems in light of Table 10 that the current error limit for flag value 1 is rather tight if compared with the number of the flag values 2 and 3. The number of flag values 2 or 3 is currently well balanced with the number of rejections due to variational quality control. It should be pointed out that the background quality control rejection limits, as given in Table 1, are the same ones that were used in the ECMWF OI analysis system. Along with the implementation of the 3D-Var on Fujitsu VPP700, the method of estimation of the background errors was also changed (*Fisher and Courtier, 1995*). This would indicate a need for retuning the background rejection limits which has indeed been done for later IFS cycles. The rather high proportion of cases with flag value 1 in Table 10 may just reflect this change.

In Table 10, the large number of background quality control flags for AIREP and SATOB winds is a result of tight error limits that have been applied, as explained in the Appendix. For wind observations from TEMP reports, for instance, the error limits are not modified from those in Table 1.

In terms of rejection of observations, different data selection decisions are responsible for the majority of rejections. Next comes the background quality control but with an order of magnitude fewer rejections. As a final cleanup, the variational quality control makes few well targeted rejections during the minimization. As a whole, the impact of the new observation screening and background quality control system can be regarded as neutral in terms of the ECMWF analysis and forecast products.

4. SOME FURTHER ASPECTS OF THE OBSERVATION SCREENING IN 3D/4D-Var

4.1 Considerations for the 3D-Var data assimilation

There are some aspects in the current observation screening that can be modified for more optimal use of observations in the 3D-Var data assimilation. Probably the most important of them is a need to improve the use of TEMP and PILOT observations which are the backbone observing system for operational NWP, especially in the Northern Hemisphere. As discussed earlier, the geopotential height information for TEMP reports is currently received only for the 15 standard pressure levels whereas the temperatures are available also for significant levels, i.e. at selected local extreme



values of the sounding profile. With a software development, the geopotentials could also be recovered for significant levels in a consistent way (*Chouinard and Staniforth, 1993*). Wind and humidity information, on the other hand, is immediately available at full resolution and could easily be incorporated into the 3D-Var assimilation. This is true for the temperature as well, but at ECMWF the geopotential is the chosen mass variable. This development will be a subject for further impact studies.

Another aspect of possible improvement in the current observation screening is the horizontal thinning of the aircraft reports. They are currently thinned down to a rather coarse horizontal resolution of 125km which can be easily modified. This choice should be assessed through an impact study.

4.2 Particular aspects for the 4D-Var data assimilation

The general feature of observation screening for the 3D-Var data assimilation is the tendency to accept only observations that are close to the analysis time (see e.g. Fig. 1), i.e. to the middle of the 6 hour data window. This is a genuine feature which is in accordance with the static nature of the 3D-Var assimilation scheme. For the 4D-Var data assimilation, however, that restriction can be relaxed to a large extent and the observation screening can be carried out essentially on a hourly basis. In principle, one could think of providing the model counterparts for the observations exactly at the observed time from the closest model time steps. The implementation is, however, such that the model counterparts for the observations are calculated from the model time step closest to a full hour and thus the observation screening is restricted for the hourly observation slots. That is a technical restriction, though, not a scientific one.

A good example of the potential benefit of the hourly screening is given in Fig. 1b for the drifting buoys reporting surface pressure. Applying the hourly screening would approximately double the number of accepted surface pressure observations from drifting buoys in 4D-Var compared with the current use in the 3D-Var data assimilation. The impact of this improved use of observations is a subject for further impact studies.

The background quality control will also be affected by the introduction of the 4D-Var. Currently the observation errors are inflated if the observation time differs from the analysis time, i.e. from the middle of the 6 hour data window of the 3D-Var. In the 4D-Var context, the model background field is close to the observation time and the observation error need not to be inflated. Instead, an error growth law can be assumed for the background errors thus accounting for the relative accuracy of the observations and the background fields, respectively.

5. A MASSIVELY PARALLEL COMPUTING ENVIRONMENT

The migration of operational codes at ECMWF to support a massively parallel computing environment set some very special requirements. One of them is a requirement for reproducibility of the results in different computer configurations. When performing data assimilation or forecast experiments on a parallel computer, the executable should be such that the result of computations is independent of the actual number of processors used for the computations. In the assimilation code there may be a risk, for instance, for the value of the observation term of the cost function to be different when using a different computer configuration. The difference may arise, for instance, because the observations appear in a different order in the summations which in turn lead to



different rounding errors. One way of meeting this requirement is to incorporate a truly unique observation sequence number that is invariant from run to run and which can be used for sorting when calculating sensitive global sums. The problem is in this case transferred partly to the codes that initially decode the report data base files and create an observation array suitable for assimilation.

The requirement for reproducibility of results also implies that 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, as mentioned earlier. 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 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. That 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 multilevel 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 multilevel report has to be communicated. The other way out of this would be to force the observation clusters of the multilevel reports always into one processor without splitting them. In that case codes responsible for creation of the observation arrays for assimilation should ensure the 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 multilevel TEMP and PILOT reports, based on the time-location array, are communicated between the appropriate processors as copies of these common stations.

6. SUMMARY

The newly developed observation screening and background quality control system performs well when verified against the old ECMWF operational system that was based on the screening and background quality control of the OI system. The use of observations in the new system is very similar, but not identical, when compared with the old system. The fit of the observations to the background forecast and to the analysis is almost identical. The code is however much more modular and flexible for modifications required for more optimal use of observations in the 3D/4D-Var data assimilation. The observation screening code is interfaced with an external blacklisting system. This allows fairly complex criteria to be used for selecting observations without IFS code



modifications. The new system is also fully parallelized and performs well on a truly parallel computer with a distributed memory. The impact of the newly developed observation screening and background quality control system on the forecast scores is neutral.

ACKNOWLEDGEMENTS

The authors are grateful to Erik Andersson, Mats Hamrud, Sami Saarinen and Drasko Vasiljevic for their help in the development of the observation screening and background quality control system at ECMWF. Many discussions as well as their technical assistance is warmly acknowledged.



The quality control of the observations is performed in two steps. In the first step, the observations are screened for gross errors. In the second step, the observations are screened for background errors.

Background quality control

The background quality control is performed by comparing the observations with the background fields. The background fields are the fields of the ECMWF 3D-Var system.

APPENDIX A1

Bad reporting practice of SYNOP and TEMP reports

The way the synoptic surface stations report mass observations (pressure or geopotential height) is considered as bad if the

- station altitude is above 800m and station reports mean sea level pressure
- station altitude is above 800m and station reports 1000hpa level
- station altitude is above 1700m and station reports 900hpa level
- station altitude is below 300m and station reports 900hpa level
- station altitude is above 2300m and station reports 850hpa level
- station altitude is below 800m and station reports 850hpa level
- station altitude is above 3700m and station reports 700hpa level
- station altitude is below 2300m and station reports 700hpa level
- station altitude is below 3700m and station reports 500hpa level

The reporting practice is also considered as bad if the station reports 500gpm, 1000gpm, 2000gpm, 3000gpm or 4000gpm level pressure, respectively, and station altitude is more than 800m different from the reported level.

For TEMP geopotentials the reporting practice is considered as bad if the

- station altitude is above 800m and station reports 1000hpa level
- station altitude is above 2300m and station reports 850hpa level
- station altitude is above 3700m and station reports 700hpa level

APPENDIX A2

Revised background quality control rejection limits for selected observations

The background quality control rejection limits are applied more strictly for some observation types than stated in Table 1. The special cases are the following ones

- AIREP wind observations with zero wind speed are rejected if the background wind exceeds 5m/s
- for AIREP and DRIBU wind observations the rejection limit is multiplied by 0.5, and for PILOT wind by 0.8
- for SATOB wind observations the rejection limit is multiplied by 0.1, except below 700hPa level where it is multiplied by 0.2
- no background quality control is applied for SCATT winds
- for DRIBU surface pressure observations the rejection limit is multiplied by 0.9, and for PAOB surface pressure by 0.7
- for AIREP temperature observations the rejection limit is multiplied by 1.6

For SATOB wind observations also an asymmetric background check is performed. If the observed wind is more than 4m/s slower than the background wind, the background error limits are revised in the following manner

- if pressure level is below 700hPa, error limits are multiplied by 0.15
- in the tropics (20°), the rejection limit is multiplied by 0.07
- otherwise a speed dependent error limit $0.075 - 0.00125V$, where V is the background wind speed

THE UNIVERSITY

... ..

... ..

... ..

... ..

... ..

... ..

... ..

... ..

... ..

... ..

REFERENCES

- Andersson, E, Pailleux, J, Thépaut, J-N, Eyre, J R, McNally, A P, Kelly, G A and P Courtier, 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Q J R Meteor Soc*, 120, 627-653.
- Chouinard, C and A Staniforth, 1995: Deriving significant level geopotentials from radiosonde reports. *Mon Wea Rew*, 123, 222-229.
- Courtier, P, E Andersson, W Heckley, G Kelly, J Pailleux, F Rabier, J-N Thépaut, P Undén, D Vasiljevic, C Cardinali, J Eyre, M Hamrud, J Haseler, A Hollingsworth, A McNally and A Stoffelen, 1993: Variational assimilation at ECMWF. ECMWF Technical Memorandum No. 194, 84pp.
- Courtier, P, Andersson, E, Heckley, W, Pailleux, J, Vasiljevic, D, Hamrud, M, Hollingsworth, A, Rabier, F and Fisher, M, 1997: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. Submitted to *Q. J. R. Meteorol. Soc.*
- Courtier, P, J-N Thépaut and A Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q J R Meteor Soc*, 120, 1367-1387.
- Eyre, J R, 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. *Q J R Meteorol Soc*, 115, 1001-1037.
- Eyre, J R, Kelly, G A, McNally, A P, Andersson, E and A Persson 1993: Assimilation of TOVS radiance information through one-dimensional variational analysis. *Q J R Meteor Soc*, 119, 1427-1463.
- Fisher, M and P Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Technical Memorandum No. 220, 28pp.
- Gaffard, C and H Roquet, 1995: Impact of ERS-1 scatterometer wind data on the ECMWF 3D-Var assimilation system. ECMWF Technical Memorandum No. 217, 21pp.
- Gandin, L S, 1963: Objective analysis of meteorological fields. Translated from the Russian by the Israeli Program for Scientific Translations (1965).
- Hollingsworth, A, Shaw, D B, Lönnberg, P, Illari, L, Arpe, K, and A Simmons, 1986: Monitoring of observations and analysis quality by a data assimilation system. *Mon Wea Rew*, 114, 861-879.
- Järvinen, H, S Saarinen and P Undén, 1996: User's guide for blacklisting. 51pp. Available on request from ECMWF, Shinfield Park, RG2 9AX, Reading, Berkshire, England.
- Kelly, G, Andersson, E, Hollingsworth, A, Lönnberg, P, Pailleux, J and Z Zhang, 1991: Quality control of operational physical retrievals of satellite sounding data. *Mon Wea Rew*, 119, 1866-1880.
- 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.

Lönnerberg, 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.

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

Lorenç, A C, 1981: A global three-dimensional multivariate statistical interpolation scheme. Mon Wea Rev, 109, 701-721.

Lorenç, A C, 1986: Analysis methods for numerical weather prediction. Q J R Meteorol Soc, 112, 1177-1194.

Lorenç, A C and O Hammon, 1988: Objective quality control of observations using Bayesian methods. Theory and practical implementation. Q J R Meteorol Soc, 114, 515-543.

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.

Shaw, D B, P Lönnerberg, A Hollingsworth and P Undén, 1987: Data assimilation: The 1984/85 revisions of the ECMWF assimilation system. Q J R Meteorol Soc, 113, 533-566.

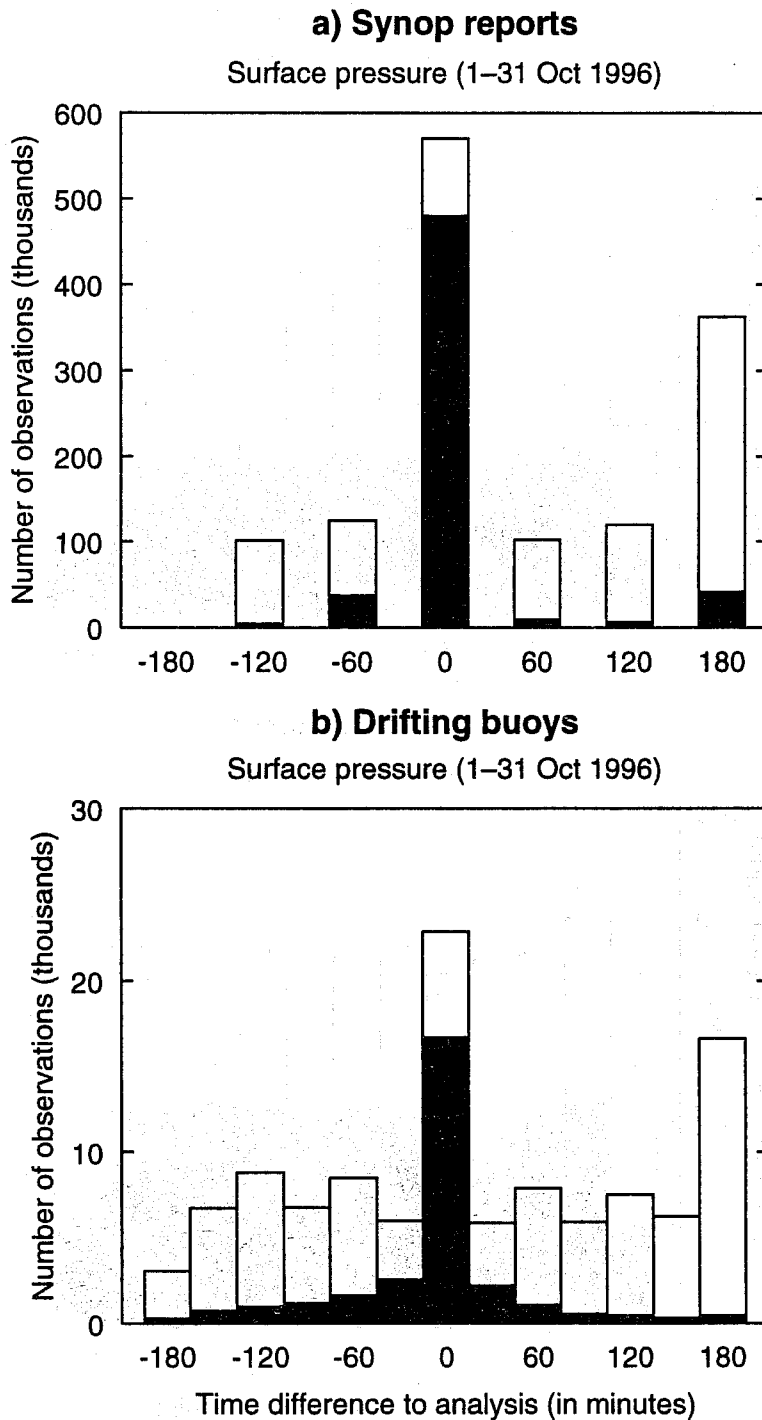


Fig. 1: A histogram of the time distribution of the observations within a 6 hour data window, using a binning of 30 minutes. Column height gives the total number of observations while the shaded part displays the number of observations used by the 3D-Var assimilation during the period of 1–31 October 1996. (a) Surface pressure from synop reports, and (b) surface pressure from drifting buoys. Note the different scale of ordinate in (a) and (b).

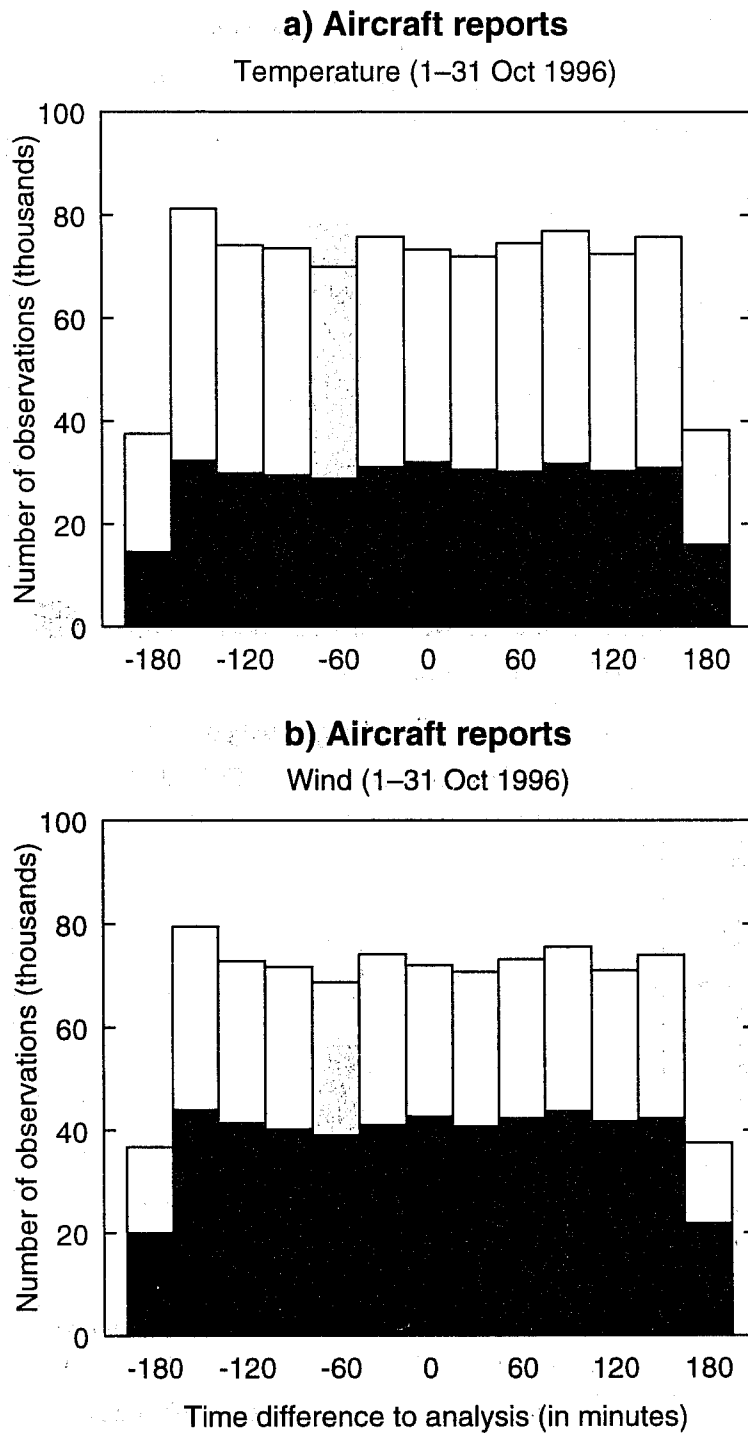


Fig. 2: As Fig. 1 but for (a) aircraft report temperatures, and (b) aircraft report winds, respectively.

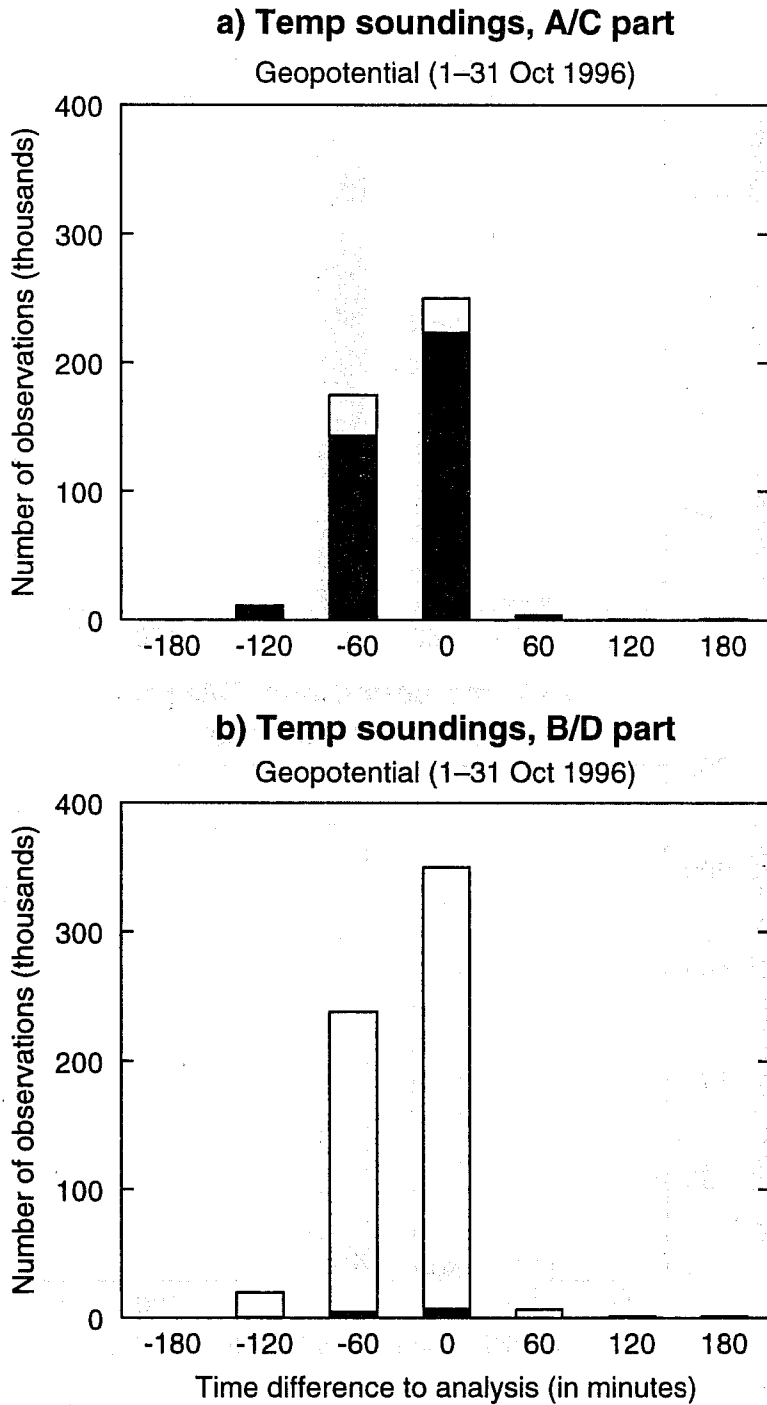


Fig. 3: As Fig. 1 but for geopotential observations from temp sounding at standard pressure levels (a), and at non standard pressure levels (b), respectively.

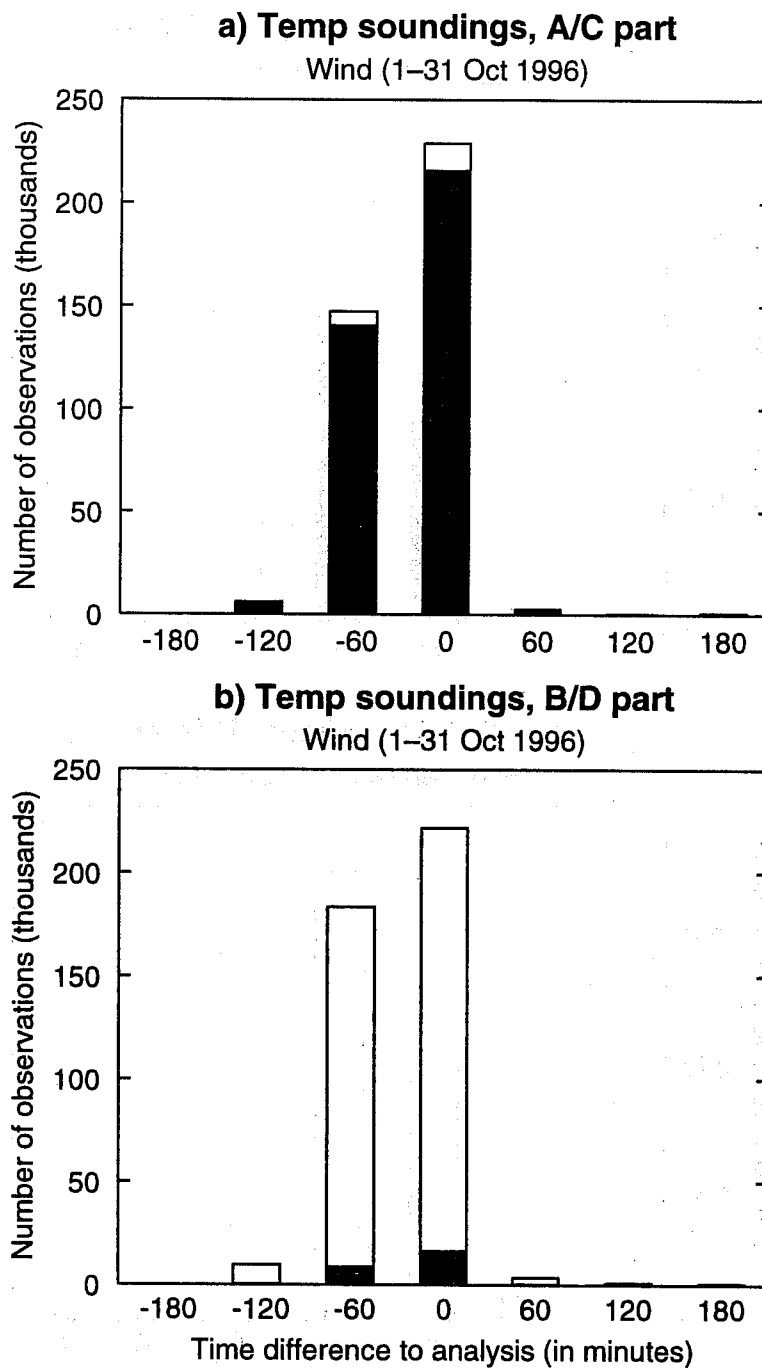
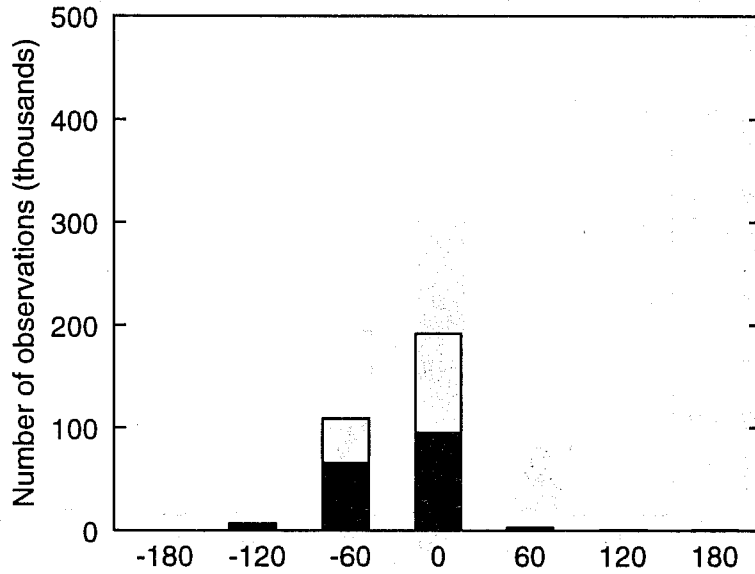


Fig. 4: As Fig. 3 but for wind observations from temp sounding.

a) Temp soundings, A/C part

Specific humidity (1-31 Oct 1996)



b) Temp soundings, B/D part

Specific humidity (1-31 Oct 1996)

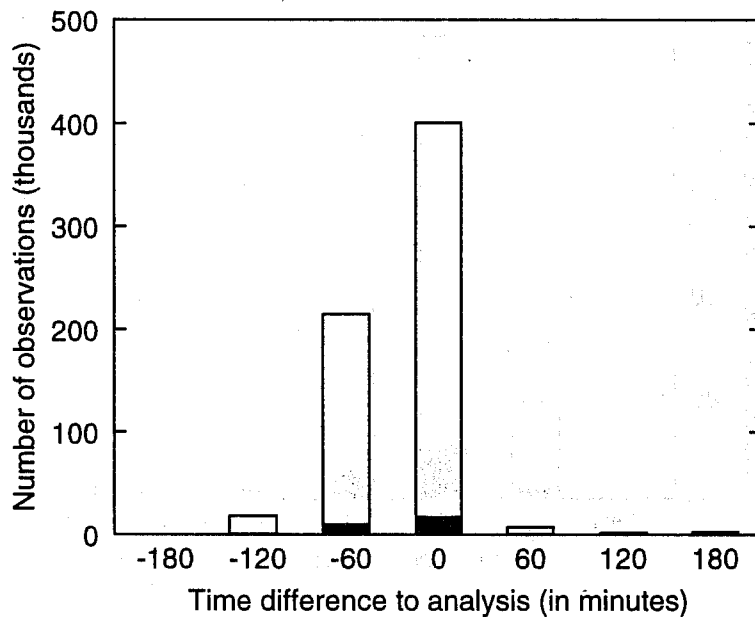


Fig. 5: As Fig. 3 but for specific humidity observations from temp sounding.

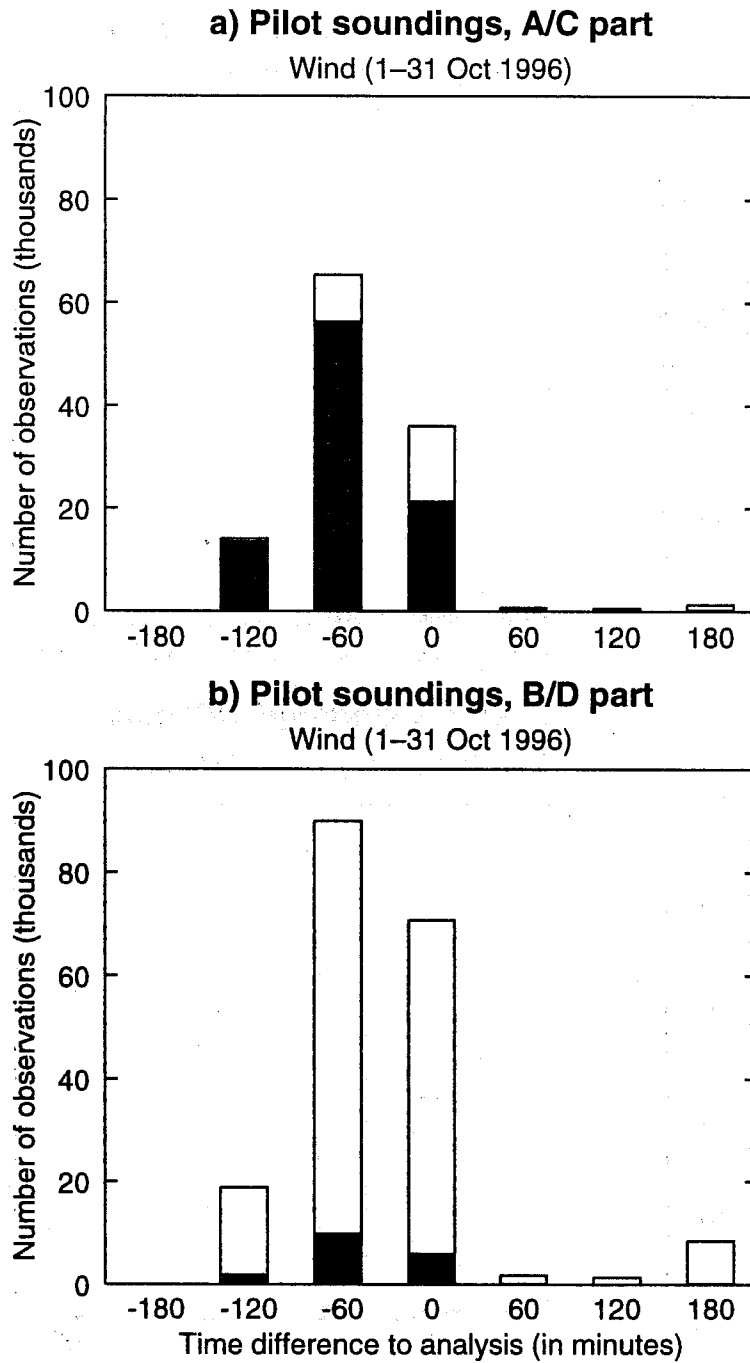


Fig. 6. As Fig. 3 but for wind observations from pilot sounding.

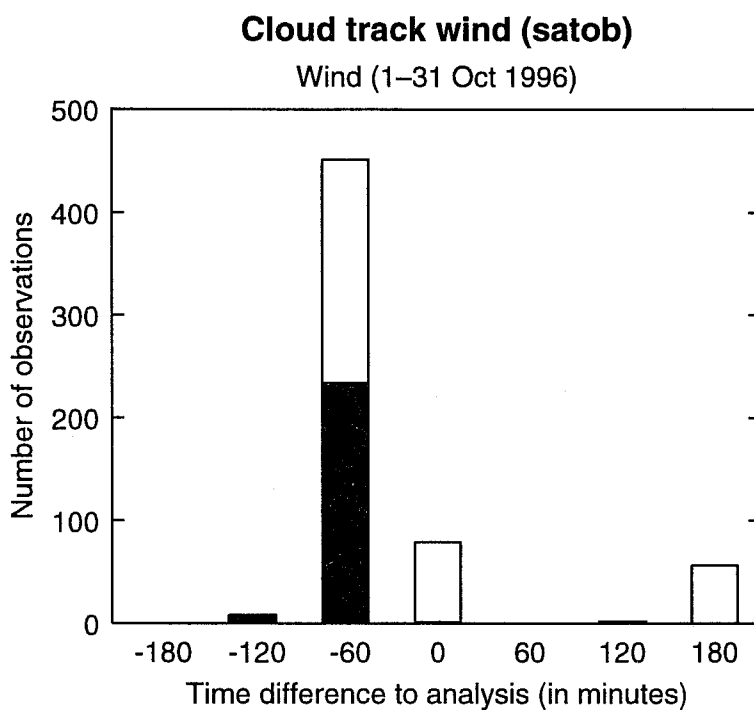


Fig. 7. As Fig. 1 but for satob cloud track winds.