TECHNICAL MEMORANDUM

313

Observing systems experiments in the ECMWF 4D-Var data assimilation system

F. Bouttier and G. Kelly

Research Department

February 2001

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Abstract

A set of global observing system experiments has been run by denying classes of observations from the ECMWF four-dimensional variational data assimilation and forecast system. This indicates how efficiently the observations are used in this particular system. The observing systems tested were mainly TOVS radiances, SATOB atmospheric motion winds, radiosondes, aircraft, drifting buoys and PAOB pseudo-observations. The impact has been assessed in both summer and winter 1999, on a total of six weeks, in terms of the average objective quality of the deterministic forecast of tropospheric wind and geopotential at medium (5-7 day) and short (1-3 day) ranges. It is shown that the choice of verification method matters in data-poor areas such as tropics and the Southern Hemisphere.

Some interesting conclusions can be drawn about the ECMWF 4D-Var system. Each observing system tested has a notable positive impact on the medium-range forecast performance. This is not always true at shorter ranges. The TOVS radiances have a large impact on all areas and ranges. Drifting buoys and PAOBs both have some impact in the Southern Hemisphere. In the Northern Hemisphere, the relative impact of TOVS, aircraft and radiosondes depends on the area considered. For Europe, the statistical significance of the results is limited, but the most important observing system among those tested appears to be the radiosondes.

1 Introduction

Observing System Experiments (OSEs) are regularly conducted at the European Centre for Medium-Range Weather Forecasts (ECMWF) in order to review the efficiency of the data assimilation system. This data assimilation system has been developed as an accumulation of incremental modifications to the data assimilation algorithm, to the forecast model, and to the data selection. Each modification is known to improve the forecast skill, but there are interactions between them, and the observing systems have themselves evolved a lot over the past few years, with an increasing coverage and quality of satellite-based observing systems, and a reducing coverage of many conventional networks.

Reports from OSEs in other numerical weather prediction centres can be found in relevant publications of the World Meteorological Organization, for instance see Pailleux et al (1997).

Previous OSEs have been carried out at ECMWF using the Optimal Interpolation system with emphasis on the TOVS radiances and SATOB winds (Uppala et al 1985, Andersson et al 1991, Kelly et al 1993), and a few years later using a three-dimensional variational assimilation system (Undén et al, 1997), with emphasis on the scatterometer winds. The set of OSEs presented here has been run using the four-dimensional variational data assimilation system (4D-Var) described in Rabier et al (2000), Mahfouf and Rabier (2000), Derber and Bouttier (1999), Järvinen and Undén (1997) and Klinker et al (2000) which contains a few OSE results as well. The data selection is summarized in the following section. Apart from the assimilation system itself, the most original feature of this set of OSEs is the relatively large sample of almost 6 weeks spread over winter and summer. Its making has represented over 328 days of 4D-Var data assimilation and 321 10-day forecasts.

2 Experimental framework

The data assimilation system, forecast model and data used are nearly the same as the ones used in operations at ECMWF in the first half of 2000. The assimilating and forecast model is a global spectral primitive equations model with T319 linear truncation and 60 hybrid levels in the vertical, the lowest at 10m above ground, the highest at 10Pa. A comprehensive physics package is used, including prognostic treatment of cloud variables, ozone and ocean surface waves to couple surface wind and sea roughness. The atmospheric data assimilation algorithm is an incremental 4D-Var run in 6-hour cycles, at a reduced resolution of T63 to compute the corrections to the previous forecast (Courtier et al, 1994). Asynoptic observations are processed at the correct time, frequently reporting platforms can be used at up to hourly frequency. The observation operators are non-linear and, for most observations, use the calibrated measured physical quantity from each observing instrument.

The reference atmospheric data assimilation system processes about 200,000 pieces of information per 6-hour period, comprising: SYNOP-land reports (station pressure and screen-level relative humidity), SYNOP-ship (surface pressure and anemometer wind), drifting and moored buoys (surface pressure and anemometer wind), TEMP and PILOT profiles (wind, temperature and humidity where available, potentially at all reported levels), US profilers (wind), dropsondes (wind and temperature), aircraft at cruise level and in ascents/descents (wind and temperature), PAOB pseudo-observations (surface pressure), SATOB retrievals of atmospheric motion winds (wind), SCAT ambiguous wind retrievals from ERS scatterometer backscatter (10-m winds over sea), SSM/I microwave sensor retrievals (total water vapour and surface wind speed, over sea), TOVS level 1-C MSU/AMSU-A radiances (mainly sensitive to stratospheric and upper tropospheric temperature, some channels are sensitive to lower tropospheric temperatures as well). All data are subjected to a complex quality control (Järvinen and Undén 1997, Andersson and Järvinen 1999), and spatially dense datasets are thinned down to a resolution similar to that of 4D-Var in order to minimize correlated representativeness errors. Note that a few North Atlantic dropsondes were used in the summer cases, and that the winter cases did not use US profilers. Except for the data that are explicitly denied to the experiments, all data are used in each experiment as it would have been in operations, with the same cutoff time and real-time decisions for data selection.

In order to sample a variety of atmospheric situations most OSEs were broken down into two periods, a three-week summer period with assimilation from 20 Sept 1999, 00UTC to 10 October 1999, 12UTC, and a winter period from 16 December 1998, 00UTC to 6 January 1999, 12UTC. 10-day forecasts have been run from each 12UTC analysis, totalling respectively 22 and 21 forecasts from the summer and winter periods, with a 12-hour warm-up period before the first forecast in each period. A very similar setup was used in Cardinali (2000) in order to assess the impact of European radiosondes and regional aircraft data, so the results of these two studies can be compared directly. As in any OSE, the forecasts drawn from consecutive days are not statistically independent, and all the weather situations have not been sampled. Consequently, one needs to interpret the results with caution, since they might not be consistent with the effect of denying data for much longer periods.

3 Verification method

This section explains the methodology used in interpreting the forecast differences between the OSEs.

Experience gathered at ECMWF with this assimilation and forecast system shows that two three-week periods represent the minimum sample necessary to obtain a reliable indication on whether a typical change to the system is beneficial or detrimental to the average medium-range forecast quality. This may not be enough for changes that have a very small impact. For significant changes, the sign of the impact is reliably estimated up to about six-day forecast range over large areas such as the Northern and Southern Hemispheres. Even then, one needs to be careful not to choose spells for which the atmospheric circulation exhibits unusual and persistent anomalies, or emphasizes weaknesses in the assimilation and forecasting system. In order to obtain a fully reliable estimate, with useful indications on smaller areas, several months would have been necessary, but it was unfortunately not affordable for this work.

A meaningful measure of the impact a given modification on the forecast quality is normally obtained by averaging over time and areas the forecast scores verified against the ECMWF operational analysis. In the medium range, six weeks of experimentation provide a reliable estimate of the impact over the extratropical hemispheres (North of 20N and South of 20S) at about 5-day range and to a lesser extent on smaller, data-rich areas such as Europe, North America, Eastern Asia or Australia/New Zealand. A result is convincing only if the impact on the scores is consistent across several ranges, vertical levels (1000, 500, 200hPa were checked here) and parameters (wind vector rms error and geopotential anomaly correlation). The comparison of one experiment against another over a three-week period will usually contain a mixture of positive and negative impacts, so it is hazardous to look at individual cases: the assimilation system is only designed to minimize the analysis error on average. No single feature of the system ever provides a systematic improvement to the forecasts, which always have some random, chaotic character. A T-test for statistical significance can be applied in order to increase confidence on the forecast score differences, but even this will not guarantee a fully reliable interpretation of the results, because forecast errors can be correlated in time.

It is interesting to assess the forecast impact on shorter ranges as well. In general, the spread of forecast score differences is smaller at one- to three-day range, yielding more stable averages; unfortunately, the forecast verification is more difficult at these ranges, because the forecast errors are smaller than at medium range. In data-poor areas, and particularly over the tropics, the very nature of OSEs make this a noteworthy problem, even in the medium range: the atmosphere being chaotic and only partially observed, large errors can develop over time in the data assimilation. It means that the analysis itself may differ notably between one assimilation and the other. This is compounded by the fact that the spectral 4D-Var analysis system performs a global analysis which allows differences to propagate instantaneously over large distances (this has been observed on large-scale diurnal and semi-diurnal atmospheric waves). As a result, very short-range forecasts from any given experiment may differ a lot from the operational ECMWF analysis, simply because they started from different analyses, without one analysis being really better than the other. A solution is to verify each short-range forecast from the analyses of the assimilation it started from, which gives by construction a zero forecast error at zero range. Each assimilation experiment considered here compared well with a large number of observations, yielded forecasts of reasonable quality and is thus believed to provide realistic verifying analyses.

The verification of each experiment against its own set of analyses is satisfactory for most purposes, but not necessarily in OSEs, because the difference between two such experiments arises from many scattered observations. These observations perturb the assimilated state, making each forecast more different from the previous one; this is noticeable over data-poor areas such as the Tropics and the Southern Hemisphere. As a result, using more data sometimes results in forecasts looking poorer against their own analyses over data-poor areas. The answer is to use observations as a completely independent verification system. Obviously, this only works in areas where good-quality observations are regularly available; following WMO recommendations, quality-controlled radiosonde observations of wind and geopotential have been used here. This is useful over Australia/New Zealand and in average over the whole tropical area. Unfortunately, there is no fully satisfactory verification method

Northern Hemisphere (NH)	North of 20N
Southern Hemisphere (SH)	South of 20S
Tropics	Between 20N and 20S
Europe	75N–35N, 12.5W–42.5E
North America	60N–25N, 120W–75W
Asia	80N–25N, 45E–170E
East Asia	60N–25N, 102.5E–150E
Australia/New Zealand	12.5S–45S, 120E–175E
North Atlantic	65N–25N, 70W–10W
North Pacific	60N–25N, 130E-145W.

Table 1: Definition of the areas named in the text.

available over smaller areas in the tropics or elsewhere in the Southern Hemisphere. Perhaps high-quality aircraft, buoy or satellite data will provide a better reference for forecast verification in the future.

There is insufficient space available in this publication to display all the score plots that justify the statements made in the following sections. For the sake of clarity, only a few representative areas, variables and levels are presented, and the forecast range has been limited to 7 days (3 days for scores against observations) The area names follow WMO recommendations for verification; they are defined in table 1:

The statements expressed in the following sections have all been based on a comparison of all three verification methods (operational analysis, own analysis, observations), on each available period, for several levels and parameters; only the genuinely convincing features have been cited. Even then, one should remember that the periods considered are rather short and thus not representative of all weather conditions.

4 Nomenclature

In the plots, each experiment is identified by a name according to the following nomenclature:

ECMWF the reference summer and winter assimilation/forecast experiments, using exactly the same software and data as the other experiments on the corresponding dates, as mentioned. This is very close to, but not completely identical to ECMWF operations at the same times, mainly because the 'ECMWF' experiments contain some recent software improvements, that have only been implemented in ECMWF operations in either October 1999 or April 2000.

noTOVS the reference minus the TOVS level 1C radiances: MSU channels 2, 3, 4 from NOAA-14, AMSU-A channels 5 to 13 from NOAA-15.

noSATOB the reference minus the SATOB atmospheric motion winds: infrared and water vapour from GOES-8 and 10, infrared, water vapour and visible from Meteosat-5 and 7, infrared from GMS-5.

noUPPERSAT the reference minus all upper-level satellite data, i.e. TOVS, SATOB, and SSM/I total water vapour retrievals from DMSP (but SCAT winds and SSM/I low-level wind speed are still used).

noAIRCRAFT the reference minus all aircraft data (wind and temperature), including AIREP, ACARS, AMDARS.

noSONDE the reference minus all in-situ upper-level sounding data, including TEMP, PILOT, US profilers and dropsondes (the reference used no profilers and dropsondes on the winter period).

noDRIBU the reference minus all buoy wind and pressure data received as such over the Global Telecommunications System.

noPAOB the reference minus all PAOB pressure pseudo-observations. The reference only used them South of 20S.

noDRIBUnoPAOB the combination of noDRIBU and noPAOB impacts.

All experiments used ambiguous scatterometer winds over sea and surface pressure data from ships and land stations. These observing systems, along with the ones mentioned in section 2, guarantee that each experiment used a minimum amount of data to keep the data assimilation system in a realistic state in all areas.

The summer experiments can directly be compared with the first three weeks of the EUCOS/GADS impact experiments presented by Cardinali (2000). All these experiments used the 'ECMWF' experiment as reference, extended over a two-month period. Only a tiny average impact was observed on the regional forecast scores, which suggests that, in the medium range, on large areas and over a few weeks, observing systems experiments can only be used to assess the impact of large-scale changes of the observing network. No definite result can be obtained in this context for limited, regional changes.

5 The impact of the observing systems

5.1 Comparison between satellite observing systems

The impact of upper-level satellite data can be assessed by comparing the ECMWF, noTOVS, noSATOB, and noUPPERSAT experiments (Fig. 1 and 2). The noUPPERSAT experiment includes the denial of SSM/I total water retrievals, whose separate impact has been studied by Gérard and Saunders (1999): in terms of scores, it is usually weak, but not always so. Most notably, SSM/I data affect the model climate, so one has to assume that they play a non negligible part in the noUPPERSAT impact.

A common feature of TOVS, SATOB and SSM/I water retrievals is that they all affect the very large scales of the assimilation system, mainly over the tropical and southern oceans. They also are remote-sensed data with difficult representativeness problems, which means that the measurement quality can be low. Remote-sensed data can easily be misused in a way that will create local analysis errors. Such errors will limit the overall short-range forecast quality. On the other hand, it is generally accepted that medium-range forecasts are mainly sensitive to the analysis of the large scales of the

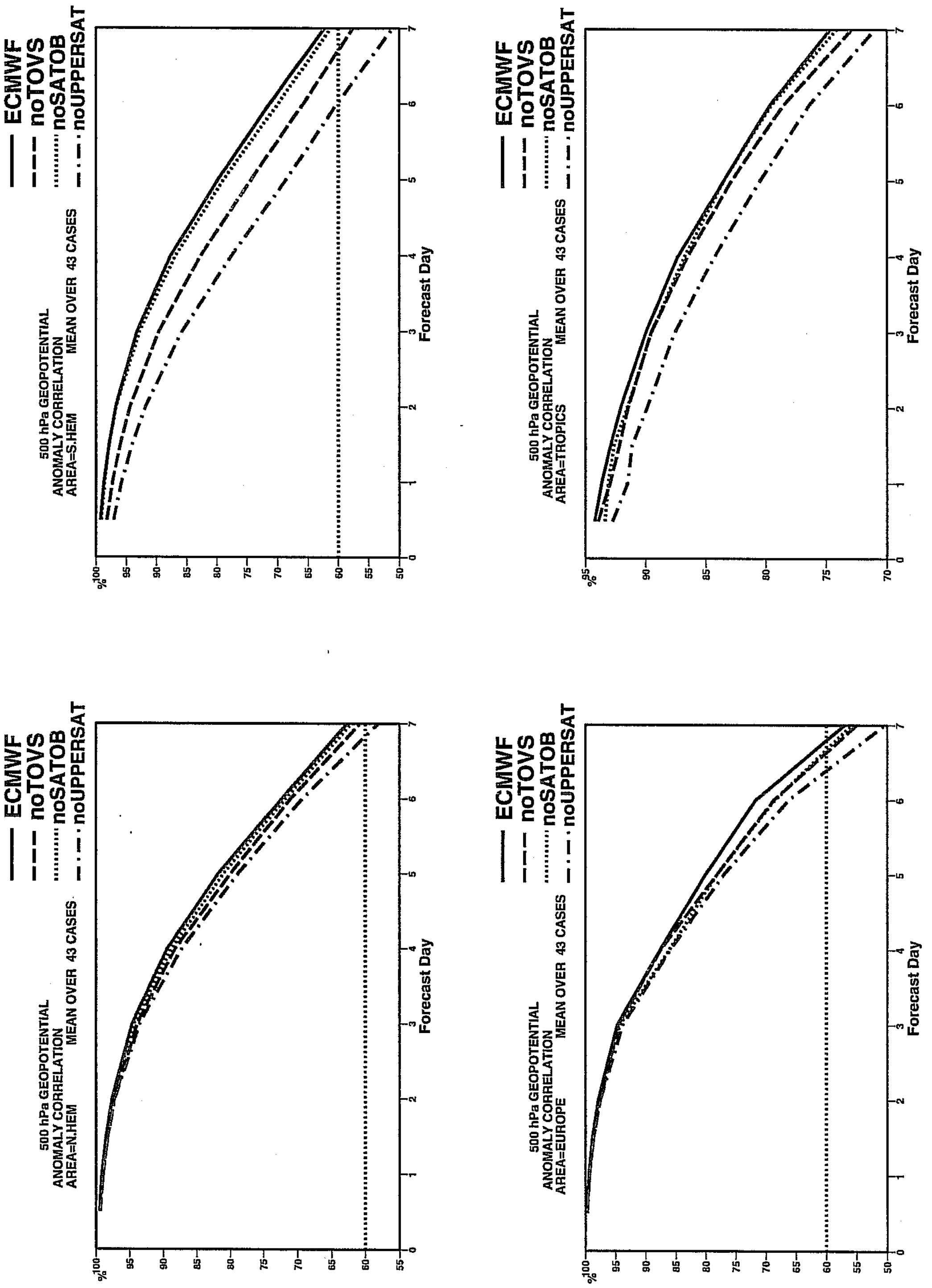

Figure 1: Average impact on the forecast of the z500 anomaly correlation, over 4 areas, of denying TOVS, SATOB and all upper-level satellite data in the ECMWF 4D-Var assimilation system (verification against the ECMWF operational analysis).

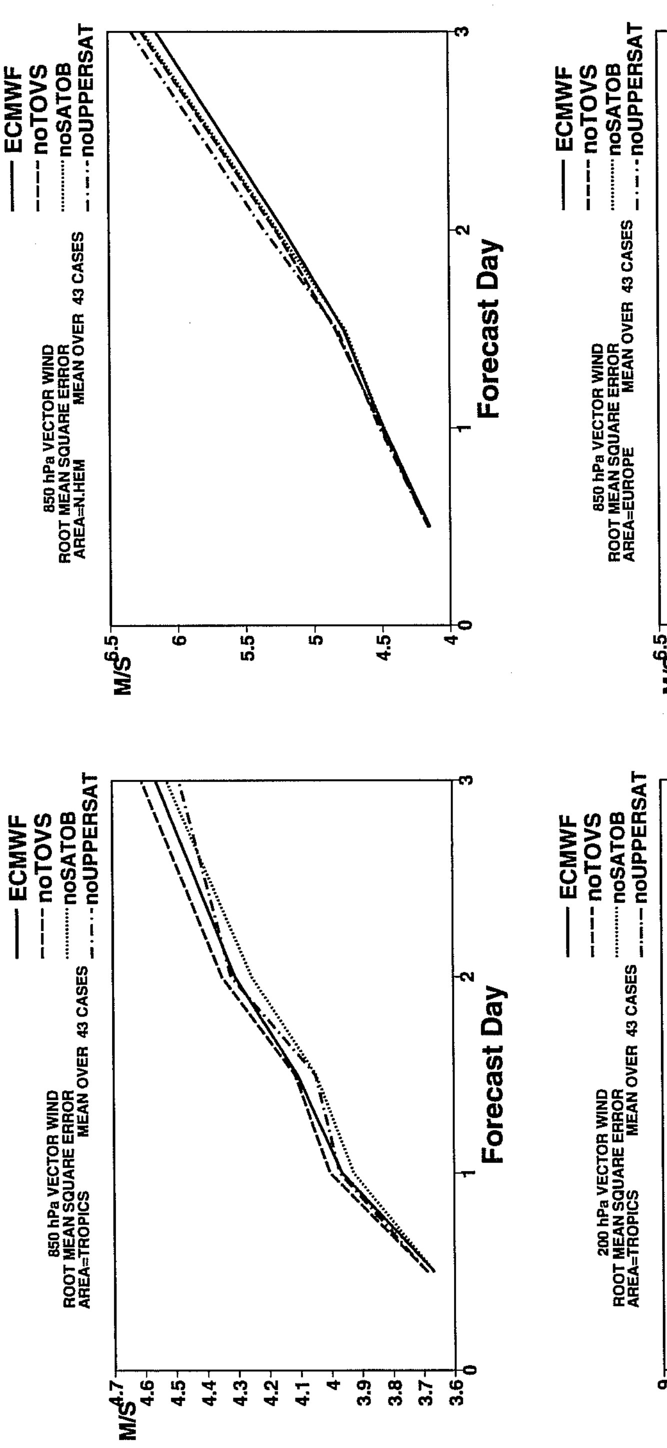

Figure 2: Average impact on the rms forecast error of the 850hPa wind over 3 areas (plus 200hPa wind over the Tropics), of denying TOVS, SATOB and all upper-level satellite data in the ECMWF 4D-Var assimilation system (verification against radiosonde observations).

atmospheric circulation. Hence, one expects TOVS, SATOB and SSM/I water retrievals to have a significant impact in the medium range. The main results shown in Fig. 1 and 2 are:

in the Northern Hemisphere, the average medium-range impact of TOVS is larger than SATOBs. The combined impact of upper-level satellite data is large and significantly beneficial, they contribute to about 12 hours of forecast skill at the 7-day range.

in the Southern Hemisphere and the Tropics, the medium-range impact of all satellite data is even larger than in the Northern Hemisphere, bringing a combined benefit of about 30 hours at the 6-day range, i.e. about 25% of the total forecast skill.

over Europe, the relative benefits of using TOVS and SATOBs are similar, and are mainly apparent in the medium range, beyond the 4-day range.

in the short range, TOVS are very beneficial but the impact of SATOBs is unclear until the 48-hour range. This could be caused by specific SATOB retrieval problems, such as speed biases (Tomassini et al, 1999), or an improper wind height assignment technique. A SATOB speed bias with respect to radiosondes would cause SATOBs to look particularly detrimental on scores verified against observations (Fig. 2).

the impacts are not additive in the short range over the Northern Hemisphere and Europe, because the noUPPERSAT impact is clearly different from the sum of noTOVS and noSATOB. This could be related to problems using the SSM/I data, and to interactions between TOVS and SATOB.

in the Tropics, the choice of verification method is crucial, notably in the short range: SATOBs do improve the upper tropospheric forecasts (around 200hPa, fig.2), but there are problems at low levels and in the short range (probably because of speed biases, as suggested above). Imbalances between the assimilation and the model's physics (such as the depiction of convective systems) are suspected in this area.

5.2 Comparison between satellite and conventional observing systems

The impact of conventional vs satellite data can be seen from the comparison between ECMWF, noUPPERSAT, noSONDE, and noAIRCRAFT experiments (Fig. 3 and 4). This is far from assessing the total weights of the whole conventional and satellite observing systems, because all experiments used a common dataset (SYNOP and SHIP pressure data, and SCAT winds), which is known to have a large impact on the assimilation system. Note that the noSONDE experiment measures the impact of conventional radiosondes, US wind profilers and dropsondes. The latter two are known to have some limited impact over North America, the North Atlantic and Europe. A comparison between TOVS/SATOBs, radiosondes and aircraft data in OSEs has already been already carried out in previous years (Undén et al 1997). The present work can thus be used to gauge the improvement in the use of satellite data at ECMWF. Unfortunately, the comparison with earlier OSEs is affected by the decline in radiosonde data coverage, which has been rather dramatic over the Siberian area. On the other hand, the ECMWF use of radiosondes has been improved since the early nineties by switching from using geopotential standard level data to temperature data (they are more informative

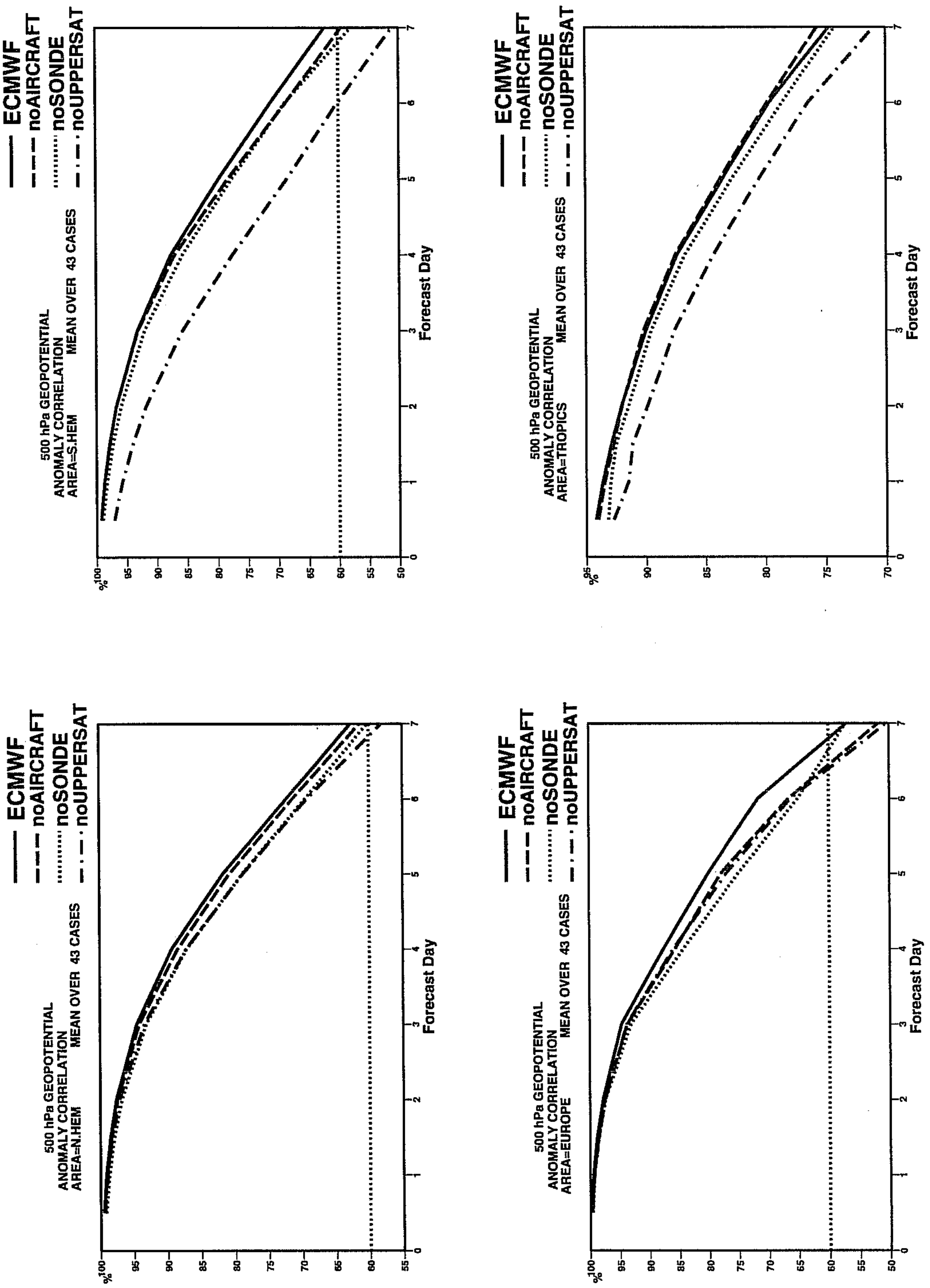

Figure 3: Average impact on the forecast of the z500 anomaly correlation, over 4 areas, of denying radiosonde, aircraft and upper-level satellite data in the ECMWF 4D-Var assimilation system (Verification against the ECMWF operational analysis).

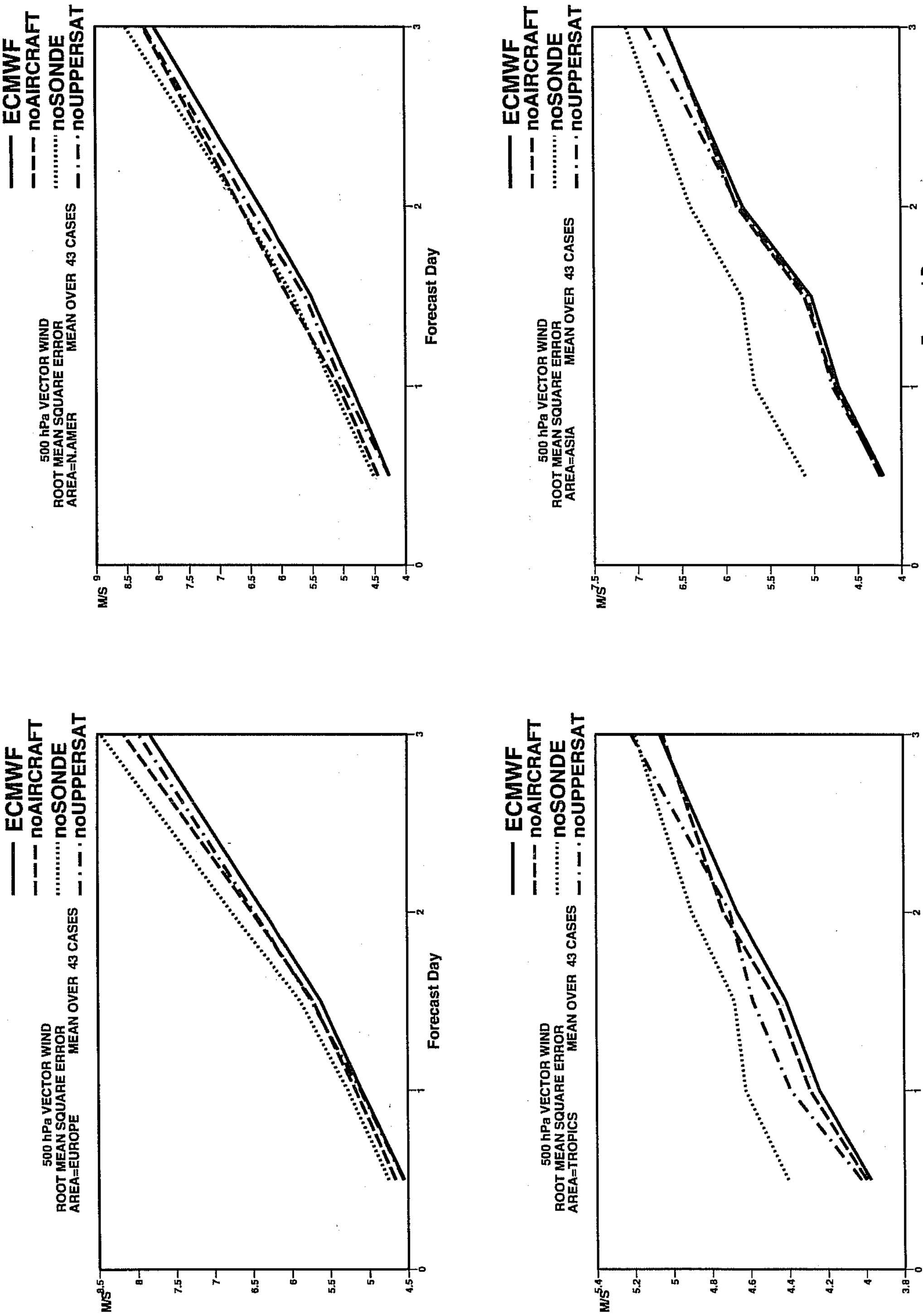

Figure 4: Average impact on the rms forecast error of the 500hPa wind, over 4 areas, of denying radiosonde, aircraft and upper-level satellite data in the ECMWF 4D-Var assimilation system (verification against radiosonde observations).

and less sensitive to deep bias structures). The use of aircraft data has improved, with more and more data being made available with increasing quality. The main results shown in Fig. 3 and 4 are:

in the Northern Hemisphere, the most-important medium-range forecast impact is obtained equally from satellite data and from sondes; aircraft have a smaller impact. This is at variance with previous experiments which showed a dominating impact from the sondes in earlier years. For instance, in Fig. 7 of Undén et al (1997), radiosonde data are found to provide 24h of forecast skill in the 5-day range.

in the Southern Hemisphere, the satellite data clearly dominate, with both sondes and aircraft impacts amounting to less than one third of the benefit of satellite data, and actually less than TOVS taken on their own.

over Europe in the medium range, the impact can only be appreciated up to the 5-day range; the sondes contribute more to the medium-range skill than satellite data. They provide about 16 hours of forecast skill at the 5-day range, which is less than in previous years.

in the Tropics, the aircraft have very little impact (few are available in this area; see fig.4 rather than fig.3 which is affected by the choice of verification; the sign of such a small impact cannot be reliably estimated in these experiments). The sondes do have a significant beneficial impact, but less than satellite data.

in the Northern Hemisphere in the short range, the impacts of aircraft and satellites are comparable, and about half of the sondes impact. There are wide variations from one sub-area to the other (fig 4): **in Europe,** the aircraft have more short-range impact than satellite data, probably because of the excellent aircraft coverage over North America and the North Atlantic. **Over North America,** the situation is different, with the impact from satellite similar to aircraft and radiosonde in the short (and medium) range. A possible explanation is that satellites are the major contributors to the analyses over the tropical and northern Pacific. **Over Asia,** the impact of sondes is extremely large, whereas aircraft and satellite data have little effect. The large relative impact of radiosondes seems to stem from the high density of radiosonde networks over Europe and Western Russia. The small impact of satellite data probably stems from the fact that few satellite data are used over land at ECMWF. Note that over all areas (including Australia/New Zealand) the impact of radiosondes dominates in the very short range (less than one day).

The impact of observations over the Northern Hemisphere depends highly on the range and area considered, because the observing systems are very irregular in space. The correspondence between the forecast impact in a given area and the location of the observations that caused it is not clear. Only conjectures can be made, since these are global observing systems experiments: the way perturbations will propagate from observations to features of the forecasted fields is rarely intuitive. Errors can develop during the assimilation process in very complex ways. In order to obtain detailed information, one must resort to more expensive OSEs as undertaken by institutions such as NAOS (North American Observing System) or EUCOS (EUmetnet Composite Observing System), or to using more specific methodologies such as adjoint-based techniques (Baker and Daley, 2000).

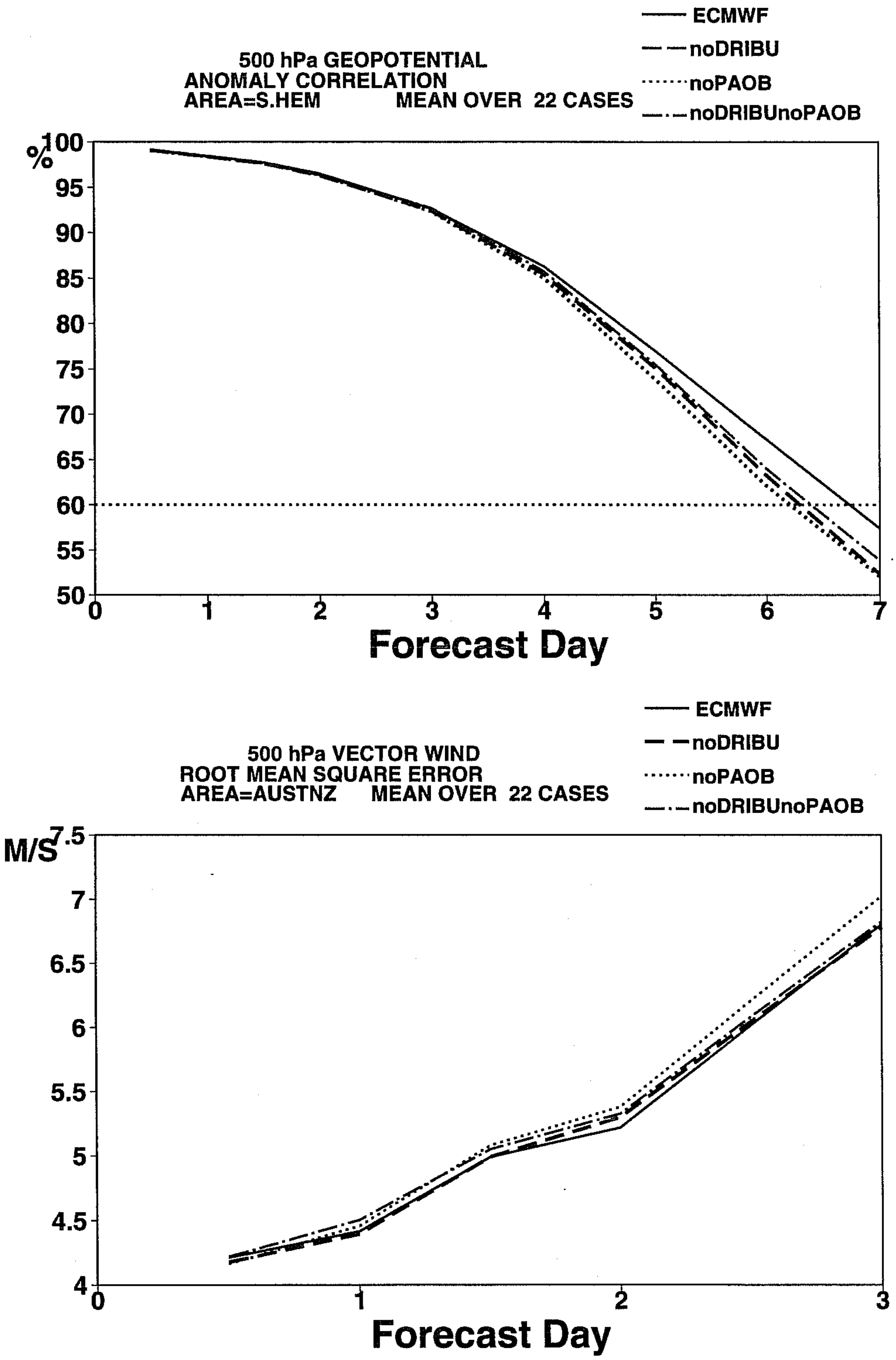

Figure 5: Average impact of denying DRIBU, PAOB and both observing systems in the ECMWF 4D-Var assimilation system. Top panel: medium-range impact on the 500hPa geopotential anomaly correlation over the Southern Hemisphere, verifying against the ECMWF operational analysis. Bottom panel: short-range impact on the rms error of the 500hPa wind over Australia/New Zealand, verifying against radiosonde observations.

5.3 Comparison between buoy and PAOB observing systems

The impact of buoys and PAOB data can be seen by comparison between ECMWF, noDRIBU, noPAOB and noDRIBUnoPAOB experiments shown in Fig. 5. This was only done on the winter period, i.e. the summer in the Southern Hemisphere, which may not provide a complete picture. Buoys and PAOBs are expected to have an impact mainly in the Southern Hemisphere, where they are the main source of surface pressure data over the vast oceans. Buoys winds and pressures are used, but low-level winds only have a very small weight in the 4D-Var system. The weight is small because the assumed wind observation error is set to a large value. The assumed observation error is large because of uncertainties on the surface roughness and possible incompatibilities with speed bias corrections in SCAT and SSM/I wind retrievals. The weight is small for another reason, too: 4D-Var will by construction set the equivalent background error variance to a low value. The background error covariance model was calibrated using model forecast differences that are damped near the ground (Derber and Bouttier, 1999), so the modelled background error is small. In 4D-Var the effective background error variance may be further reduced because the adjoint model in 4D-Var is diffusive in the planetary boundary layer and not forced at the bottom by a suitable model for roughness uncertainties (Mahfouf and Rabier, 2000). This is acknowledged as a weakness in the current 4DVar system; it implies that the so-called DRIBU impact is in effect a measure of the DRIBU surface pressure data. The main results shown in 5 are:

in the Northern Hemisphere and Tropics, the impact of DRIBU and PAOBs is very small and below the measurement accuracy of the OSE technique (plots not shown).

in the Southern Hemisphere in the medium range, both PAOBs and DRIBUs are very beneficial, and PAOBs have the largest impact. The impacts are far from additive, which suggests that there may be some incompatibilities between both observing systems.

in the Southern Hemisphere and in the short range, the forecasts can only be reliably verified against observations and in a data-rich area, here Australia/New Zealand. There, the DRIBUs and PAOBs are both beneficial, except in the very short range which seems to show an imbalance problem in the data assimilation system (perhaps a wind speed bias between the DRIBU and TEMP observation operators). Interestingly, the combined use of DRIBUs and PAOBs is beneficial at all ranges.

These impacts are very small, sensitive to the verification technique and probably to the choice of period as well. A likely cause for the smallness of the impacts is the fact that all experiments used scatterometer wind data.

6 Statistical significance tests

For guidance, an objective statistical significance test has been applied to the score plots that are presented in this paper. The test is the Student T-test applied to the forecast score differences between each impact experiment and the ECMWF control. The range is at forecast day 5 for the medium-range plots verified against the ECMWF operational analysis, at day 2 for the short-range plots verified against the radiosonde observations; the other characteristics of the scores are exactly as on the corresponding plots. In the list below, the percentage written between parentheses is the best significance level that was found: the smaller the percentage, the more significant the impact. The maximum significance tested was 0.1%; a difference that was not even significant at the 10% level is regarded as non-significant. The '(y%)' notation means that the results are significant at the y% level. The main conclusions can be summarized as follows:

noTOVS at day 5 degrades z500 on NH (0.2%), SH (0.1%), Europe (10%), Tropics (5%).

noTOVS at day 2 degrades the 850hPa wind on NH (0.2%); no significant impact on Europe.

noTOVS at day 2 in Tropics has no significant wind impact at 850hPa or 200hPa.

noSATOB at day 5 has no significant impact on z500 on NH, S, Europe or Tropics.

noSATOB at day 2 has no significant 850hPa wind impact on NH or Europe.

noSATOB at day 2 in Tropics *improves* the wind at 850hPa (10%), but degrades it at 200hPa (5%).

noUPPERSAT at day 5 degrades z500 on NH (0.1%), SH (0.1%), Europe (5%), Tropics (0.1%).

noUPPERSAT at day 2 degrades the 850hPa wind on NH (0.5%) and Europe (10%).

noUPPERSAT at day 2 in the Tropics degrades the wind at 200hPa (1%).

noUPPERSAT at day 2 degrades the 500hPa wind on N. America (10%), but has no significant impact on N. America, Tropics or Asia.

noSONDE at day 5 degrades z500 on NH (0.1%), SH (0.5%), Europe (10%), Tropics (5%).

noSONDE at day 2 degrades the 500hPa wind on Europe, N. America, Tropics, Asia (all at 0.1%).

noAIRCRAFT at day 5 degrades z500 on SH (0.5%); no significant impact on NH, Europe, Tropics.

noAIRCRAFT at day 2 degrades the 500hPa wind on Europe (10%), N. America (0.1%), Tropics (10%), Asia (10%).

on SH at day 5, z500 is degraded by noDRIBU (5%), noPAOB (0.5%), noDRIBUnoPAOB (10%).

on Australia/New Zealand at day 2, the 500hPa wind is degraded by noDRIBUnoPAOB (5%), the impacts of noDRIBU and noPAOB are not significant.

One should not exaggerate the value of these tests, because they are based on two assumptions: that the errors are uncorrelated between the forecasts, and that they are a representative sample of the distribution of possible forecast errors. These assumptions are not strictly true because the experiments only span a fraction of the variety of the possible atmospheric situations. Some specific comments made in the previous sections may seem to contradict the above list, because the author's opinion is based on many more forecast results than these T-tests, which are only shown as an help in the interpretation of the plots, and to highlight the most significant impacts. For instance, the degradation in noSATOB over NH and Europe is visible on most scores plots, and it is by confirmed by a study of the 4D-Var cost function as suggested in section 8 below. Unfortunately, it is too small to be deemed significant by the T-test on these areas.

7 Geographical aspects

The objective forecast scores suggest that some of the impact of denying observations is sensitive to the choice of verification area. There are several reasons why the impact should have spatial variations. One is that the observing systems are not uniform in space; for instance, most of the satellite data are used over sea. Another reason is that the error growth in the data assimilation and forecasting system is not uniform in space. It is sensitive to the local weather and in particular to the presence of dynamical instabilities such as baroclinically unstable jets or active convective areas. Of course, some of the error growth is due to flaws in the design of the model and data assimilation system itself. Such flaws may introduce extra errors in specific areas, in a way that can be a function of the weather conditions.

These considerations make it worthwhile to look at the maps of forecast error differences. Maps give an indication of the spatial variability of the errors inside each verification area, which is a way to assess the statistical significance of score differences: the same variation in average rms forecast error could be obtained either by a uniform, small improvement over the whole verification area, or by a mixture of large small-scale improvements and degradations. Small-scale features are likely to be caused by a few exceptional cases, which would not stand out as much in averages over a longer period.

Maps of average forecast error differences have been examined for each impact experiment, for the 500hPa geopotential at day 5. The maps are presented in terms of the relative change in rms error averaged in time, where 'relative change' is defined as follows: if $\varepsilon_{\mathrm{exp}}$ is the average rms error of an impact experiment named 'exp' and $\varepsilon_{\mathrm{ECMWF}}$ is the average rms error of the reference 'ECMWF' experiment, the relative change is

$$\frac{\varepsilon_{\mathrm{exp}} - \varepsilon_{\mathrm{ECMWF}}}{\max(\varepsilon_{\mathrm{exp}}, \varepsilon_{\mathrm{ECMWF}})}$$

This definition ensures that the relative change is between plus and minus 100%. The result is a set of maps that show where each observing system has the largest impact on the geopotential forecast; it does not tell what is the relative importance of each observing system in each area.

There is no room in this publication to display maps for all impact experiments; an example is displayed in Fig. 6. The results (for the 500hPa geopotential only, at a range of five days) can be summarized as follows:

- Maps of relative change of rms forecast error, for all impact experiments, show rather noisy structures, with many small-scale features. This suggests that the experimentation is too short for a detailed depiction of the impact of each observing system. The following tentative conclusions are likely to be specific to the weather conditions encountered in the experiments.

- All experiments show a mixture of large positive and negative impact along the Southern extratropical jet–stream; apparently the structures are entirely caused by the local weather conditions, meaning that the impact in the Southern Hemisphere can only be appreciated using averaged scores.

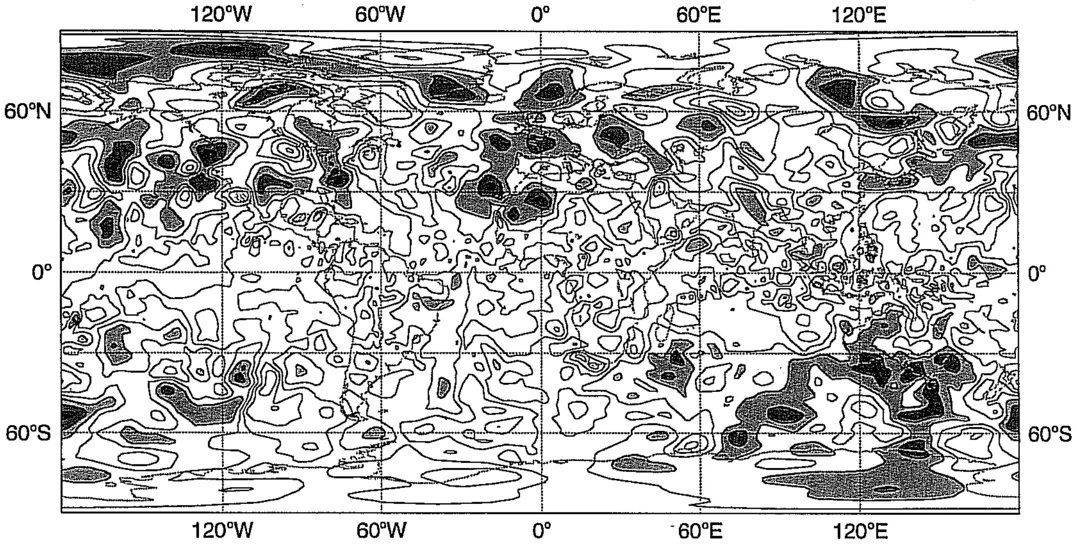

Figure 6: Impact of sounding data on the forecast skill of the 500hPa geopotential height at the five-day range. This is the relative change (see text) for the rms of forecast errors of the noSONDE experiment with respect to the ECMWF experiment. The isoline interval is 10%, with an isoline at 5%. Positive values (larger than 5%) are shaded, which indicates where sounding data reduce most the forecast errors.

- No experiment shows a large impact in the tropical region (between 20N and 20S), except the noSATOB and noUPPERSAT experiments which show that these data have a large beneficial impact. The interpretation may not be obvious because the 500hPa geopotential in the tropics is sensitive to diurnal and tidal oscillations.

- In the Northern Hemisphere, the use of TOVS data provides a well-defined improvement around the North Pole and over Siberia. Over Europe the impact is small. Over the other areas (notably North America) there is a mixture of improvement and degradation, with no obvious pattern.

- In the Northern Hemisphere, the impact of the noSATOB and noUPPERSAT experiments has a complex structure, with no obvious geographical pattern.

- The benefit of using radiosonde data are most pronounced over Europe and the Northeastern Atlantic, as well as over the Pacific area including Japan and the American West Coast. In the Southern Hemisphere most of the benefit is found in the Southern Indian Ocean and over Australia.

- Using aircraft data in the Northern Hemisphere provides the greatest benefit downstream of areas with many ascent and descent data (Europe and the USA), i.e. over the Atlantic, and over Asia.

- Using drifting buoy data and/or PAOB data is beneficial poleward of the Southern extratropical jet-stream, suggesting that the data are mainly helpful in correcting the location of low pressure systems.

8 Impact on the variational analysis

The previous sections have highlighted the value of several observing systems for improving the forecast performance. For numerical weather prediction centres, using data has a cost. Some of the cost is purely computational, because data processing requires computational power and may slow down the analysis algorithm. Another aspect of the cost is the effect of data on the statistical optimality of the analysis. Here we examine the impact of each experiment on the computer cost of the incremental 4D-Var analysis only (excluding data preparation, but including quality control checks). This impact needs to be considered in relation with possible changes in the way 4D-Var uses the observational data in the OSEs.

The impact on CPU time (measured on a Fujitsu VPP5000 machine) is a sensitive issue since most of the CPU cost of performing the ECMWF data assimilation is in the 4D-Var process. The impact of radiosondes, drifting buoys and PAOBs is negligible (below 1%). TOVS, SATOB and aircraft data are responsible, respectively, for 5%, 1.5% and 2% of the CPU cost. The main conclusion is that the benefit of radiance data is obtained at a non-negligible numerical cost, which suggests that special computer resources shall be allocated for the use of future high-volume radiances from systems such as AIRS, IASI or geostationary satellites.

The impact on the speed of convergence of 4D-Var can be measured by the rate of decrease of the cost-function gradient norm in the minimization procedure (Rabier et al, 2000). Currently the bulk of the cost of 4D-Var is proportional to the number of iterations performed in the minimization, which is set to 50 and 20 for each of the two minimization steps (Mahfouf and Rabier, 2000). Experience shows that performing more iterations does not improve the forecasts, but doing many less would degrade them. Thus it is important to know if any observing system is responsible for a loss of conditioning of 4D-Var, which would be tantamount to an increase in CPU cost. SInce 4D-Var is preconditioned by the background term, one might expect that, everything else being equal, using more observations would degrade the convergence.

In fact, the speed of convergence is generally improved when more data are used. This is probably because in the data assimilation, using more data improves the background fields, which in turn means that the observation departures are smaller at the start of the minimization. The speed of minimization appears to be more sensitive to the meteorological quality of the background fields than to the number of observations being processed. Given the speed of convergence of 4D-Var there is no reason to be restrictive about the number of data allowed in the assimilation. The only exception is SSM/I, which noticeably degrades the speed of convergence of 4D-Var when linearized physics are used (the final gradient norm increases by 30% when SSM/I data are used). A likely explanation is the poor conditioning of the humidity analysis in the ECMWF 4D-Var, which assimilates total precipitable water information in terms of the model's specific humidity variable. Convergence problems related to humidity observations have been documented by Andersson et al (2000).

Another important issue is how each observing system impacts on the use of the other observations. A first issue to check is how the quality control system behaves in the impact experiments. When withdrawing one observing system from the data assimilation, it is not obvious whether the remaining observations are still used in the same way, because some quality control decisions are sensitive to changes in the background fields. Indeed it is found that the number of used observations changes by less than 0.1% for each observing system and in each impact experiment, except of course for the observing systems that were intentionally withdrawn.

A final diagnosis is how well the observations are fitted by the analysis. This can be measured by the average value of the observation cost function J_o at the end of the 4D-Var analyses. The function J_o is to a large extent a sum on all the observed variables of the squared distance between model fields and observations, normalized by the expected observation error variance. A rigorous interpretation of the J_o values is given in Talagrand (1999). The average value of J_o (weighted by the numbers of observations used) provide an non dimensional measure of the quadratic distance between model and observations for each data assimilation experiment. In most observing system experiments, this quantity is found to increase by several tens of percent, meaning that in the full system each observing system is supported by the others to produce consistent analysis fields. The exception is aircraft data, perhaps because of incompatibilities with radiosonde or SATOB observations; it would be interesting to investigate this problem further.

9 Conclusions

The OSEs have confirmed that there is a clear and beneficial impact of using all the considered observing systems in the ECMWF global data assimilation system. This shows that, even if all the useful information may not be extracted from the observations, the 4D-Var system is certainly able to blend them in a sensible way. OSEs in previous years suggested that radiosondes were the main component of the observing system in the Northern Hemisphere (Undén et al 1997). The key property demonstrated here is the oustanding role played by satellite data, and notably the TOVS radiances. Over data-poor areas, satellite data seem to be the main contributor to the forecast performance. Over data-rich areas, satellite data contribute with the same order of magnitude as conventional observing systems such as radiosondes and aircraft. This means that the impact of satellite data has increased, because the products themselves have improved through better instruments and preprocessing systems, and because the ECMWF data assimilation system makes better use of these observations thanks to better quality control procedures, observation operators, forecast model and the 4D-Var algorithm.

Some of the decrease in the relative impact of radiosondes is obviously caused by the degradation in the observing network; at ECMWF the number of radiosonde reports received daily has decreased by about 15% over the past 8 years (F. Lalaurette, personal communication). It would be interesting to conduct such OSEs on older periods in order to demonstrate this effect. This will be facilitated by the reanalysis databases that are being prepared in several institutions.

Retrieved observations, such as SATOBs, often have quality problems in that they do not always clearly improve the short-range forecasts. Apparently, they create imbalances in the data assimilation system. A similar argument could be made for PAOBs and SSM/I retrievals on some areas. However, they do improve the medium-range forecasts, because they are useful in observing large-scale features.

Conversely, sparse observing systems such as drifting buoys clearly improve the forecasts, even in the short range, over data-poor areas such as the Southern Hemisphere. In this region one could have assumed that conventional in-situ measurements such as buoys and radiosondes are unimportant compared to satellite data. In reality, a good use of satellite data relies on some careful tuning; this tuning can only be set up and maintained if a decent network of in-situ observations is continuously available. This is crucial for the use of scatterometer and passive radiance sounding data.

Generally, the findings of this work are in agreement with similar studies carried out in other modelling centres such as NCEP and UKMO (see contributions in Pailleux et al 1997), notably the general impression about the usefulness of radiosondes and SATOBs.

There are relationships between the quantity of observations being assessed in an OSE and the sample length required to measure the impact on forecast quality. Here we have seen that two three-week periods seem sufficient to assess the large-scale forecast impact up to the 5-day range or so, for major components of the global observing system. A contrario, the impact of a smaller number of observations can barely be estimated in a two-month period such as for instance in the EUCOS study in Cardinali (2000). This suggests that OSEs cannot be used to assess the value of a regional observing system in the medium range, unless extremely long and expensive impact experiments are funded. Still, regional observing systems can probably be assessed in terms of the impact on short-range, regional forecasts, and on exceptional situations such as destructive storms, if enough independent cases can be selected for objective testing.

Acknowledgements

Thanks are due to A. Hollingsworth and A. Simmons for valuable comments on the manuscript, and to J. Haseler and N. Wedi who provided the facility to run assimilation and forecast experiments. Two anonymous reviewers and A. Lorenc provided detailed editorial comments that notably improved the quality of the paper.

references

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

Andersson, E., Hollingsworth, H, Kelly, G., Lönnberg, P., Pailleux, J. and Zang, Z., 1991: Global observing systems experiments on operational statistical retrievals of satellite sounding data. *Mon. Weather Rev.,* **199,** 1851–1864

Andersson, E., Fisher, M., Munro, R. and McNally, A., 2000: Diagnosis of background errors for observed quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Q. J. R. Meteorol. Soc.,* **126,** 1455-1472

Baker, N. and Daley, R., 2000: Observation and background adjoint sensitivity in the adaptive observation targeting problem. *Q. J. R. Meteorol. Soc.,* **126,** 1431–1454

Cardinali, C., 2000: Observing system experiments on the European conventional observing system. *Proceedings of the second CGC/WMO workshop on the impact of observing systems on NWP,* Toulouse, France, 6-8 March 2000. Available from WMO, Geneva, Switzerland

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

Derber, J. and Bouttier, F., 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus,* **51A,** 195-222

Gérard, É and Saunders, R., 1999: 4D-Var assimilation of SSM/I total column water vapour in the ECMWF model. *Q. J. R. Meteorol. Soc.,* **125,** 3077–3102

Järvinen, H. and Undén, P., 1997: Observation screening and background quality control in the ECMWF 3D-Var data assimilation system. *ECMWF Research Department Technical Memorandum no.236,* Available from ECMWF, Shinfield Park, Reading RG2 9AX, UK

Kelly, G., Pailleux, J., Rabier, F. and Thépaut, J.-N., 1993: Observing system experiments made with the ECMWF system. *World Weather Watch Tech Report no.16. WMO/TD no.594.* Available from WMO, Geneva, Switzerland

Klinker, E., Rabier, F., Kelly, G. and Mahfouf, J.-F., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.,* **126,** 1191–1216

Mahfouf, J.-F. and Rabier, ,F., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.,* **126,** 1171–1190.

Pailleux, J. (Editor) and contributors, 1997: Impact of various observing systems on numerical weather prediction. *WMO/TD No.868, proceedings of CGC/WMO workshop,* Available from WMO, Geneva, Switzerland

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: experimental results with simplified physics. *Q. J. R. Meteorol. Soc.,* **126,** 1143-1170.

Talagrand, O., 1999: A posteriori evaluation and verification of analysis and assimilation algorithms. *ECMWF workshop proceedings on diagnosis of data assimilation systems,* 17–28. Available from ECMWF, Shinfield Park, Reading RG2 9AX UK

Tomassini, M., Kelly, G. and Saunders, R., 1999: Use and impact of satellite atmospheric motion winds on ECMWF analyses and forecasts. *Mon. Wea. Rev.,* **127,** 971–986.

Undén, P., Kelly, P., Le Meur, D. and Isaksen, L., 1997: Observing system experiments with the 3D-Var assimilation system. *ECMWF Research Department Technical Memorandum no.244,* Available from ECMWF, Shinfield Park, Reading RG2 9AX, UK

Uppala, S., Hollingsworth, A., Tibaldi, S. and Kållberg, P, 1985: Results from two recent observing system experiments. *ECMWF seminar/workshop on "data assimilation systems and observing system experiments, with particular emphasis on FGGE",* Reading, 165–202, available from ECMWF, Shinfield Park, Reading RG2 9AX UK