Overview of Observing System Experiments

Richard Dumelow

Met Office, Fitzroy Road, Exeter, EX1 3PB
e-mail:richard.dumelow@metoffice.com

1. Introduction

The effective use of observations is essential for the production of NWP forecasts. It is thus important to carry out studies that assess the impact of the Global Observing System (GOS), or components of it. An effective way of doing this is by performing Observing System Experiments (OSEs). In an OSE, a continuous assimilation/forecast cycle is run under various observation use scenarios and the results from the runs compared. Using this method, the impact of extra data obtained through special field campaigns can be assessed (e.g. Cardinalli 2000) or the impact of groups of observations on downstream regions investigated (e.g. Cress & Wergen 2001; Hirshberg et al. 2001). The relative benefit of whole observing systems can also be estimated (e.g. Zapotocny et al. 2002). The results from OSEs can give valuable guidance to observation network designers within organisations such as EUCOS[1] and WMO.

Provided an operational NWP system can be adapted and sufficient computing resources are available, OSEs are relatively straightforward to run. They can be used to assess the impact of existing observing systems using existing assimilation techniques.

In order to assess the potential impact of future observing systems, Observing System Simulation Experiments (OSSEs) are sometimes carried out. ‘Pseudo-observations’, with characteristics similar to those expected from the future observing system, are generated by a numerical model then assimilated into another model and their impact assessed. To represent the fact that real observations contain more information than a numerical model, the model and assimilation scheme used to assess the impact are usually less sophisticated than the model that generates the observations (Graham & Bader, 1996).

An advantage of OSSEs over OSEs is that they attempt to assess the impact of ‘new’ observations. Disadvantages are that they are more complex and expensive to run than OSEs, and the ‘new’ observations generated are not independent of NWP and cannot fully replicate the characteristics of real observations. Consequently, OSSEs tend to over-estimate the benefits of a new observing system. As with OSEs, OSSEs can only assess impact using existing assimilation techniques.

OSEs can take up a large amount of computing resource, so they should only be run if relevant questions are being asked. Is data assimilation performance being investigated or an observation network question being asked? Observation use scenarios must then be designed to answer the questions. When carrying out the experiment, care is required to ensure that each run uses the same NWP system otherwise the results from each run may reflect differences in the NWP systems rather than differences in observation use. It is best to use an operational NWP system since such systems have been subject to rigorous testing and it can be assumed that they are producing good results.

---

[1] EUCOS - EUMETNET Composite Observing System (see http://www.eucos.net/)

2. Sensitivity of results

The results from OSEs are sensitive to a number of factors as discussed below.

2.1. Observation availability

Results from OSEs are influenced by the geographical distribution, frequency of reporting and quality of the observations used.

Maps showing typical distributions of observations of different types, that passed ‘background’ and ‘buddy’ quality control checks (Ingleby, 1992) and were assimilated by the Met Office model in a particular operational run, are given in Figure 1. Each dot on the maps represents the horizontal position of an observation that has passed quality control checks. Note that the dots may represent the position of a single level observation or profile of observations.

It can be seen that a large coverage of upper-level radiance observations is available [Figure 1(a)]. The only consistent gaps in the coverage are over certain land areas where the modelling of surface emissivity is difficult. Atmospheric Motion Vectors (AMVs) from geostationary satellites are used equatorward of 60º [Figure 1(b)]. Note that most observations are assigned to the two levels 850 and 250 hPa. Upper level profile reports from radiosondes, that provide direct measurements of wind, temperature and humidity, are widely but irregularly distributed over land areas [Figure 1(c)]. There are also a small number of observations taken from oceanic islands and ships. In-situ observations of surface parameters, taken from land stations, ships, oceanic drifting buoys and moored buoys, are widely distributed throughout the globe [Figure 1(d)]. Measurements from aircraft provide observations at flight levels. An increasing number of aircraft now take observations during ascent and descent thus providing profile information near airports [Figure 1(e)]. Currently, the density of aircraft reports over North America and Europe is considered too high to be effectively assimilated. Consequently aircraft reports are thinned before assimilation. In the Met Office scheme, one report in each 100 km box, within a 50 hPa pressure band, and within a two hour time period is used (Dalby & Berney, 1999).

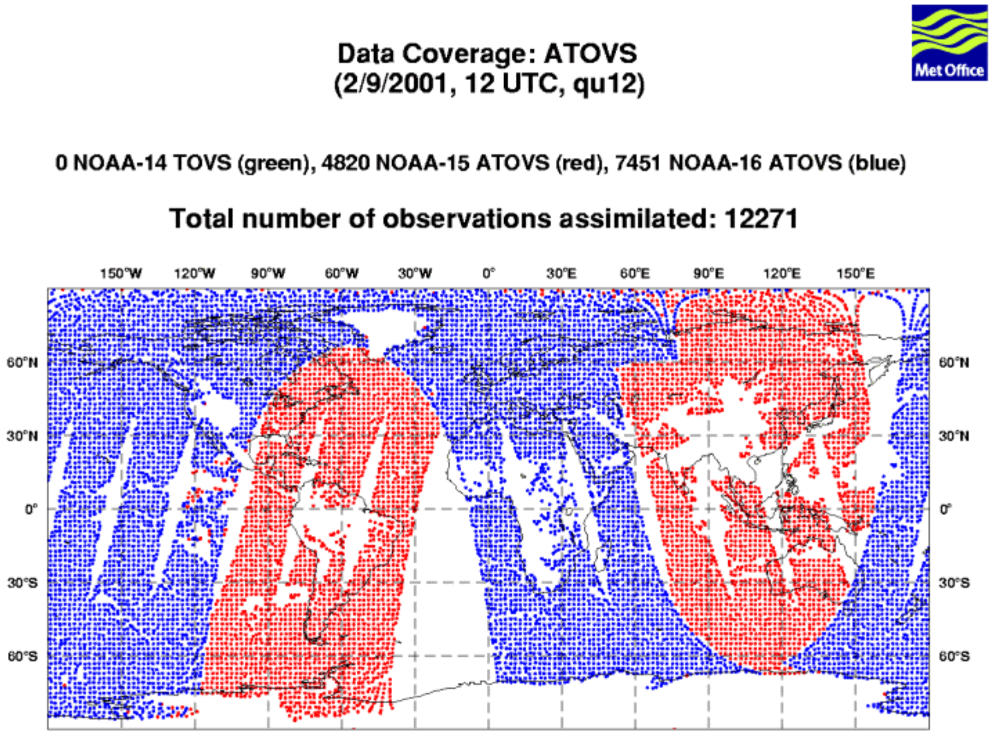


Figure 1. Positions of observations assimilated by the Met Office model at 12UTC 2/9/01. (a)

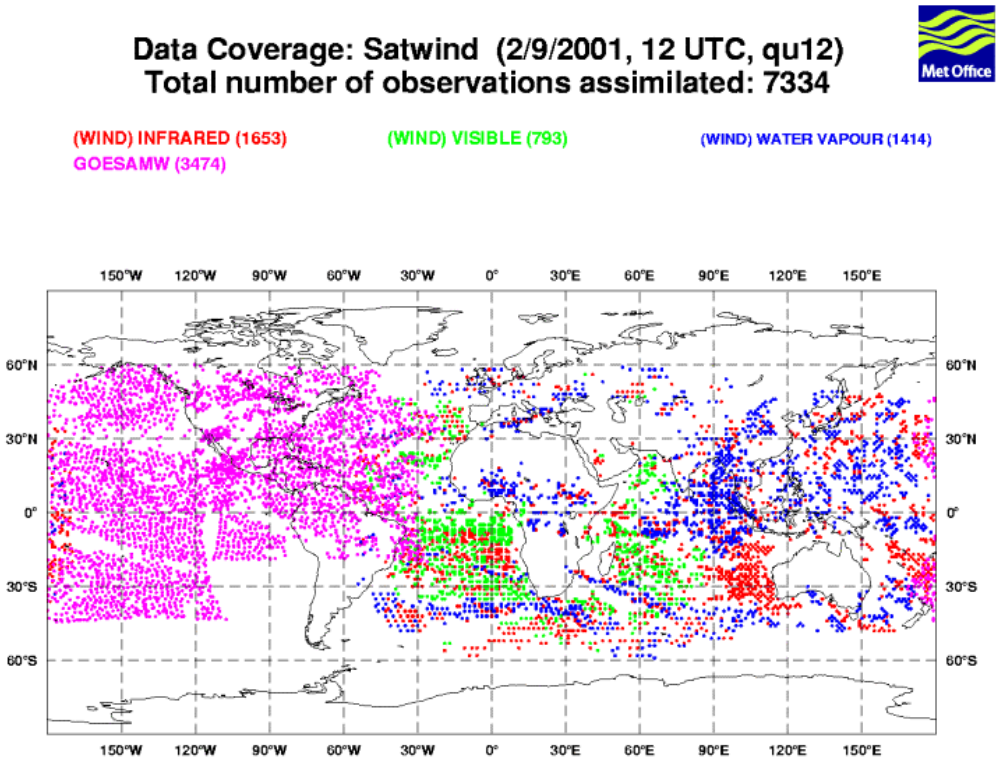


Figure 1(b)

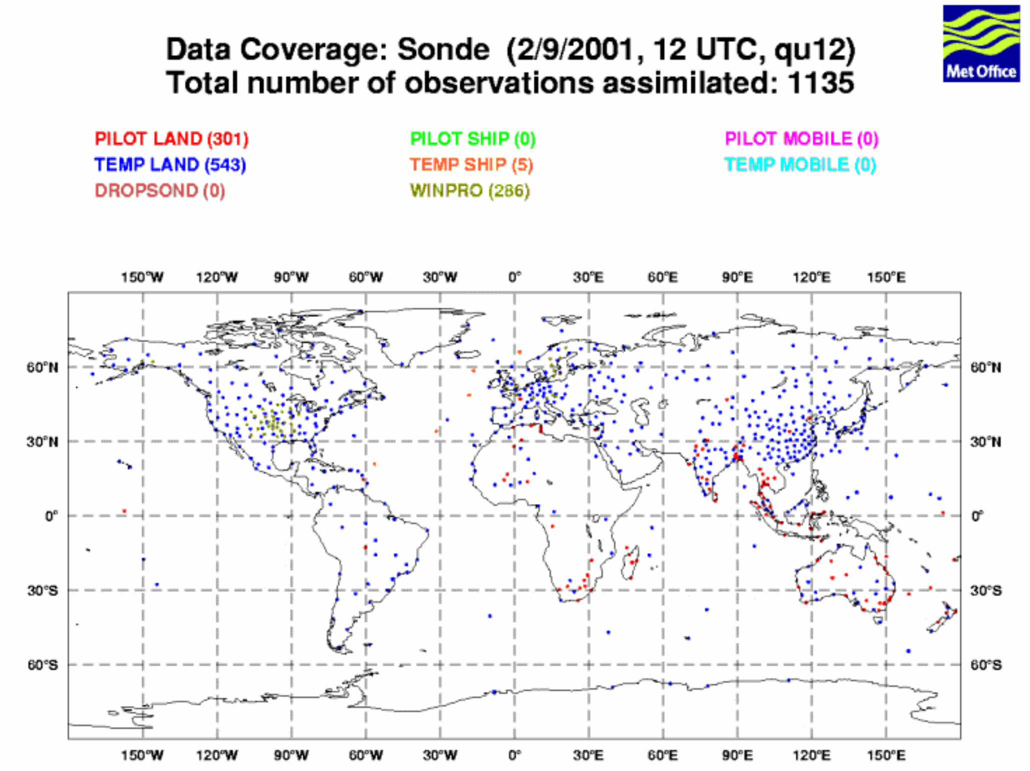


Figure 1(c)

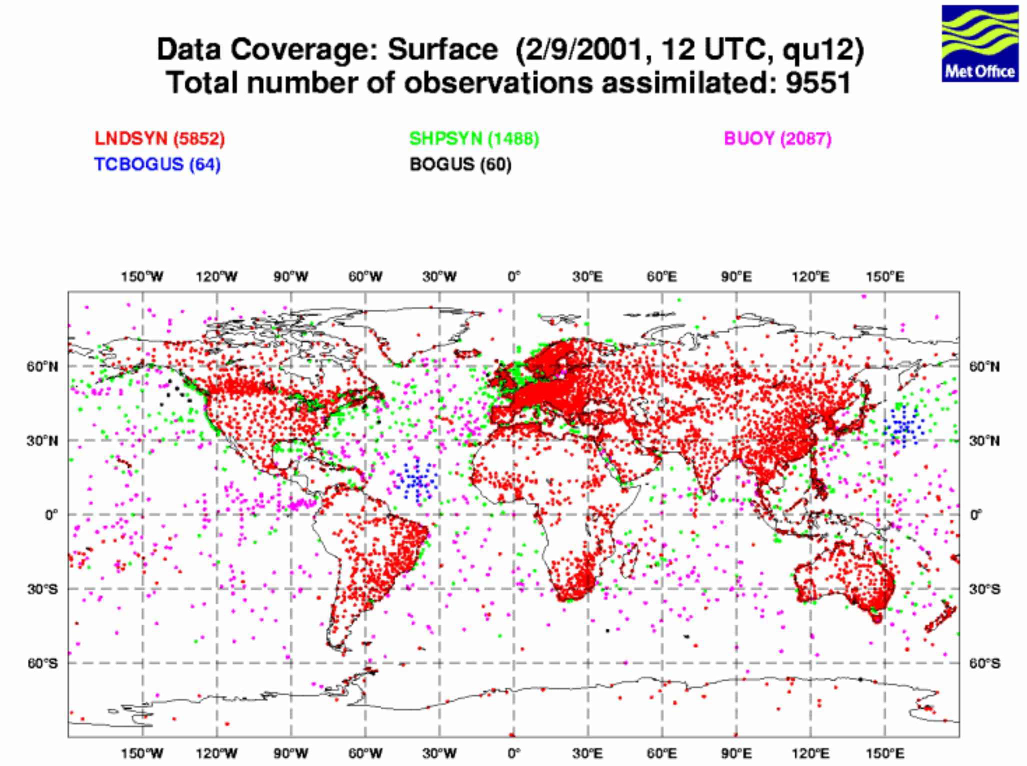


Figure 1(d)

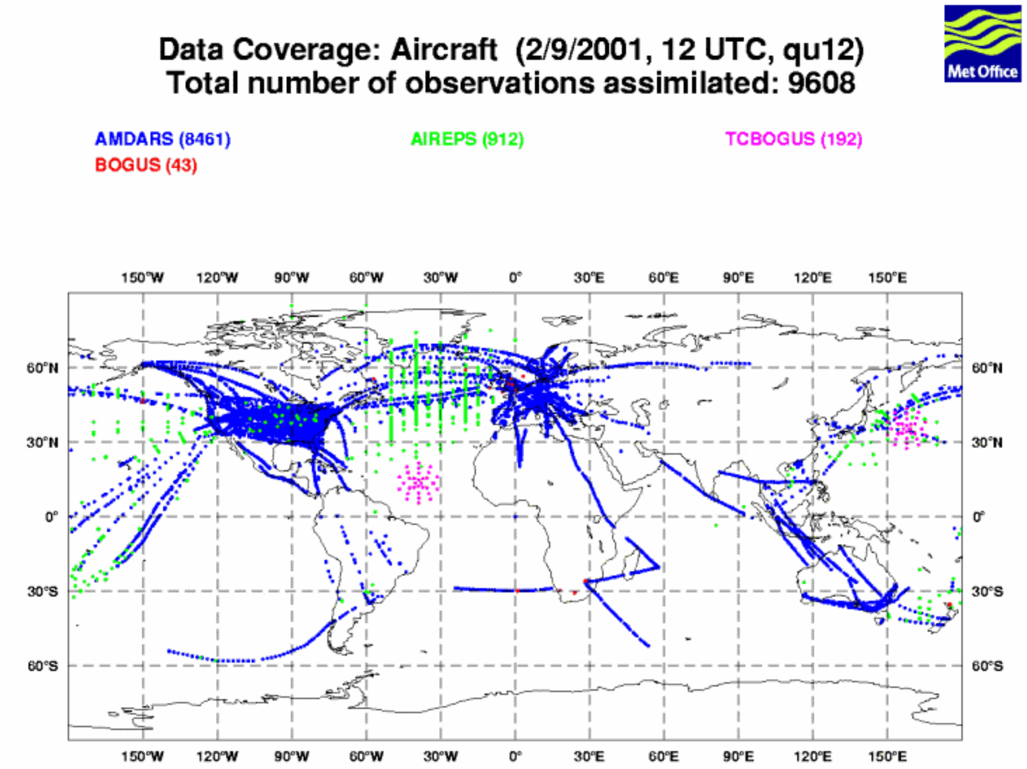


Figure 1(e)

In Figures 1(d) and 1(e) 'BOGUS' and 'TCBOGUS' observations are also plotted. 'BOGUS' observations are created manually by Met Office forecasters and are usually based upon a subjective analysis of satellite imagery. 'TCBOGUS' observations are generated objectively from tropical advisory messages (Heming et al, 1995).

2.2. Verification method

Verifying NWP forecasts against observations is desirable because observations are independent of NWP. However, verification against observations can give variable and biased results because observations are irregularly distributed in space and time and mainly occur over land areas (see Figures 1(c), 1(d)). Also, short-range NWP forecasts can be correlated with the observations (e.g. radiosondes) that are being used for verification. An alternative method of verification is to use NWP analyses. Statistics can then be calculated on a uniform grid covering all land and sea areas. In this case, the choice of analysis can influence the results that are obtained. For example, verifying each run of an OSE against its own analyses may give over-optimistic results (Simmons, 1999). In practice, both verification against analysis and observations is preferable.

Another choice concerns the type of statistics calculated. Root mean square errors give an easy to understand measure of forecast performance and can be used to verify forecasts against both observations and analysis. However it should be noted that for certain features, such as low pressure systems with strong gradients, a small displacement error will lead to a large RMS error even if the system is forecast with the correct central pressure and gradients. Consequently, RMS errors tend to rise during winter when there is a greater frequency of deep lows.

To obtain a ‘fairer’ verification of intense features with small displacement errors, the anomaly correlation coefficient (ACC) has been defined (see Simmons, 1995). The ACC is a useful score for verifying height forecasts but it cannot be used for verifying wind forecasts and is less intuitive than RMS.

2.3. Sampling method

The results from an OSE will partly depend on the number of forecasts that are assessed as well as the time of year for which they are run. An example of how the sampling period can influence the results is given below in Section 3.1. It is considered that at least 30 forecasts must be assessed to obtain meaningful results from an experiment.

2.4. NWP system

As data assimilation methods and forecast models improve, so does the ability to assimilate a wide range of observations and produce forecasts.

An example of how the NWP system can influence results is given in Section 3.1.

3. Examples of Observing System Experiments

3.1. Global data denial

A good way of assessing the overall effectiveness of a data assimilation method is to run an OSE as a global data denial trial. In such OSEs, whole observing systems are omitted and the impact on forecasts compared with a run that includes all observations. If the data assimilation system is working effectively, then it would be expected that the mean impact on forecasts of denying an observing system would be negative.

Such OSEs have been carried out at some NWP centres such as ECMWF (Bouttier & Kelly 2001; subsequently referred to as BK2001) who used a four-dimensional assimilation (4D-Var) scheme. A similar OSE was carried out at the Met Office. This global data denial experiment, which investigates the impact of all observing systems, is the first such one to be carried out within the Met Office using three-dimensional variational assimilation (3D-Var). Two one-month periods of observations, from July 2001 and January 2002, were used in this experiment and sixty forecasts assessed.

3.1.1. Description of the Experiment

The OSE was run using a version of the Met. Office unified forecast model (Cullen, 1993) and 3D-Var data assimilation scheme (Lorenc et al, 2000) using the configuration that was operational in December 2001. In order to reduce the computational expense, the forecast model was run at a reduced horizontal resolution (90 km rather than 60 km) but at the operational vertical resolution of 30 levels. Two one-month periods from different seasons - July 2001 and January 2002 - were chosen in order to sample a large variety of flow regimes. 6-day forecasts were run from the 12 UTC analysis on each day of both periods.

The reference or ALL DATA run used the following observations:

(i) ‘in-situ’ profile observations of temperature, wind and humidity from TEMP, PILOT, dropsonde and wind profiler reports;

(ii) satellite radiance data from the HIRS/3, AMSU-A, AMSU-B instruments on NOAA-15 and NOAA-16;

(iii) AMVs derived from infrared images produced by GOES 8/10; infrared, water vapour and visible images produced by METEOSAT 5/7; infrared, water vapour and visible images from GMS;

(iv) aircraft wind and temperature observations from AREP, AMDAR and ACAR reports;

(v) surface pressure from SYNOP and drifting buoy reports, surface pressure and wind from ship, moored buoy, rig and platform reports;

(vi) surface wind speed from the SSM/I satellite;

(vii) a small number of ‘BOGUS’ and ‘TCBOGUS’ data.

Six data denial scenarios were run in which the following observations were removed from the data assimilation system.

1. NO SONDE: ‘in-situ’ profile observations.
2. NO STRAD: satellite radiance data.
3. NO AMV: AMV data.
4. NO SAT: satellite data in the NO STRAD and NO AMV runs and SSM/I winds.
5. NO AIRCRAFT: aircraft data.
6. NO SURF: observations from the surface network.

Note also that a small number of ‘bogus’ data were eliminated in scenarios 1, 5, 6.

For each of these scenarios, sixty forecasts up to 6-days were assessed.

Additional runs investigated the impact of the sub-components of the surface network. These runs used July 2001 data only and thus 30 forecasts up to six-days were assessed. The following observations were eliminated from the data assimilation system.

1. NO SYNOP: SYNOP reports.
2. NO MARINE: observations from surface marine reports (buoys, ships, rigs and platforms).
3. NO SHIP: observations from ships, platforms and rigs.
4. NO BUOY: observations from moored and drifting buoys.

3.1.2. Verification

All the forecasts were verified against both observations (radiosonde and surface) and the analyses from the ALL DATA run. The geographical regions over which the statistics are calculated are defined in Table 1. Since the statistics have been averaged over sixty forecasts from two different times of year, it is expected that the mean statistics estimate the impact of different observing systems on a large variety of different flow regimes. Ideally, OSEs should be run over even more flow regimes covering many seasons but such comprehensive testing was not possible in this experiment due to limited availability of computing resources.

Conclusions from the study were made only if they were based upon impacts that were consistent over forecast parameter, level and forecast range. Since a large number of statistics were examined, for brevity only a representative sample is presented.

In order to indicate how the results from OSEs can be influenced by the period chosen and the number of forecasts verified, a comparison of some mean scores from July and January is presented. Also, plots showing the daily fluctuation in scores are shown.

Northern hemisphere	North of 20°N
Southern hemisphere	South of 20°S
Tropics	Between 20°N and 20°S
Europe	70°N-25°N, 10°W-28°E (for height) 75°N-35°N, 12.5°W-42.5°E (for wind - CBS region)
North America	60°N-25°N, 120°W-75°W
Asia	80°N-25°N, 45°E-170°E

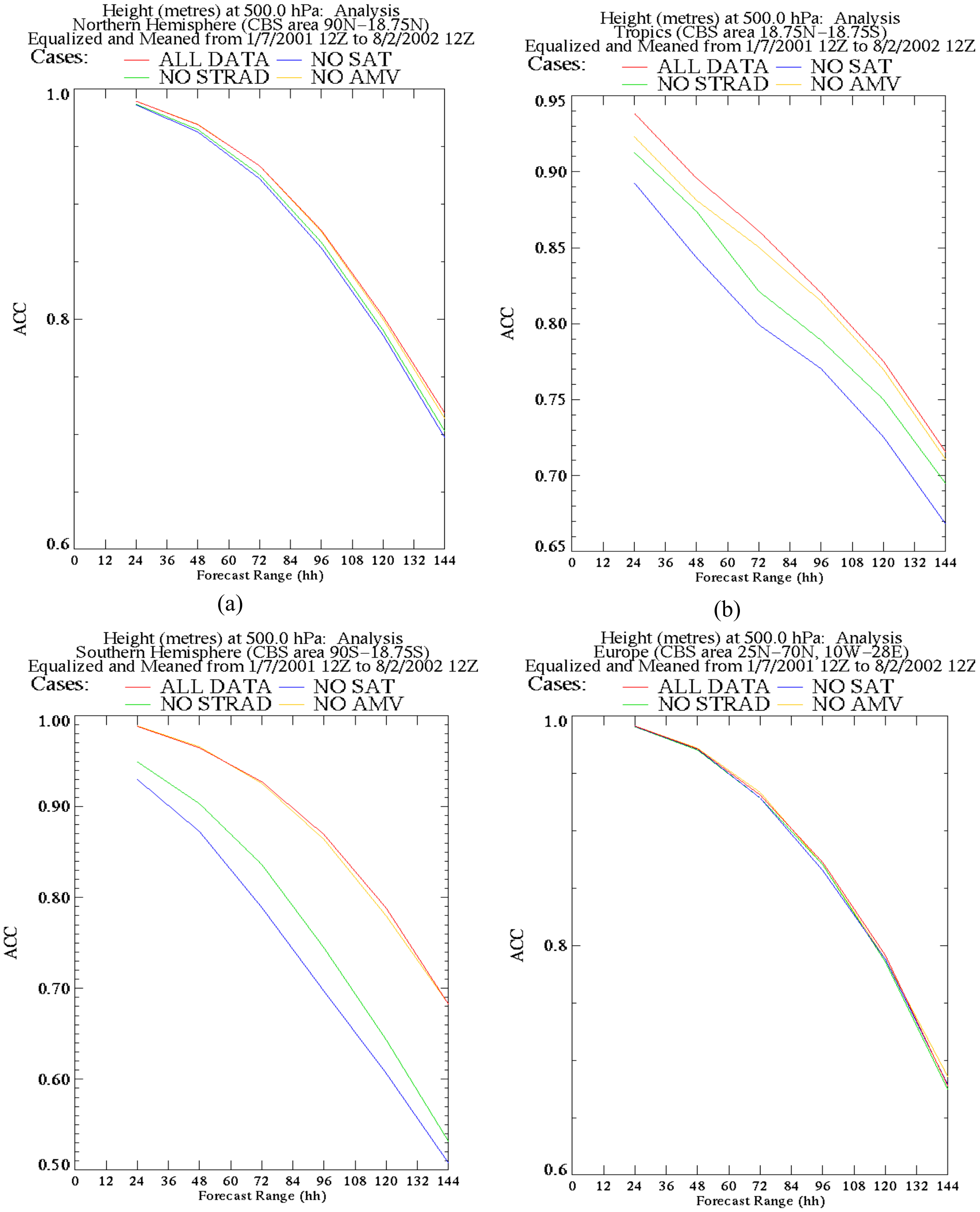
Table 1. Definition of verification areas.

3.1.3. Impact of satellite data

The impact of satellite data on height and wind forecasts is illustrated in Figures 2 & 3. The main points to note are as follows.

Satellite data have the largest positive impact on both geopotential height and wind forecasts in the southern hemisphere. In the southern hemisphere, satellite data improve the skill of geopotential height forecasts by 24-48 hours [Figures 2(c)]. In the tropics the impact is less but greatest on the geopotential height field (18-36 hours) compared with the wind field (up to 18 hours) [Figures 2(b), 3(a) & 3(c)]. In the northern hemisphere, the total impact is much smaller at less than 6 hours for both height and wind forecasts [Figures 2(a) & 3(b)].

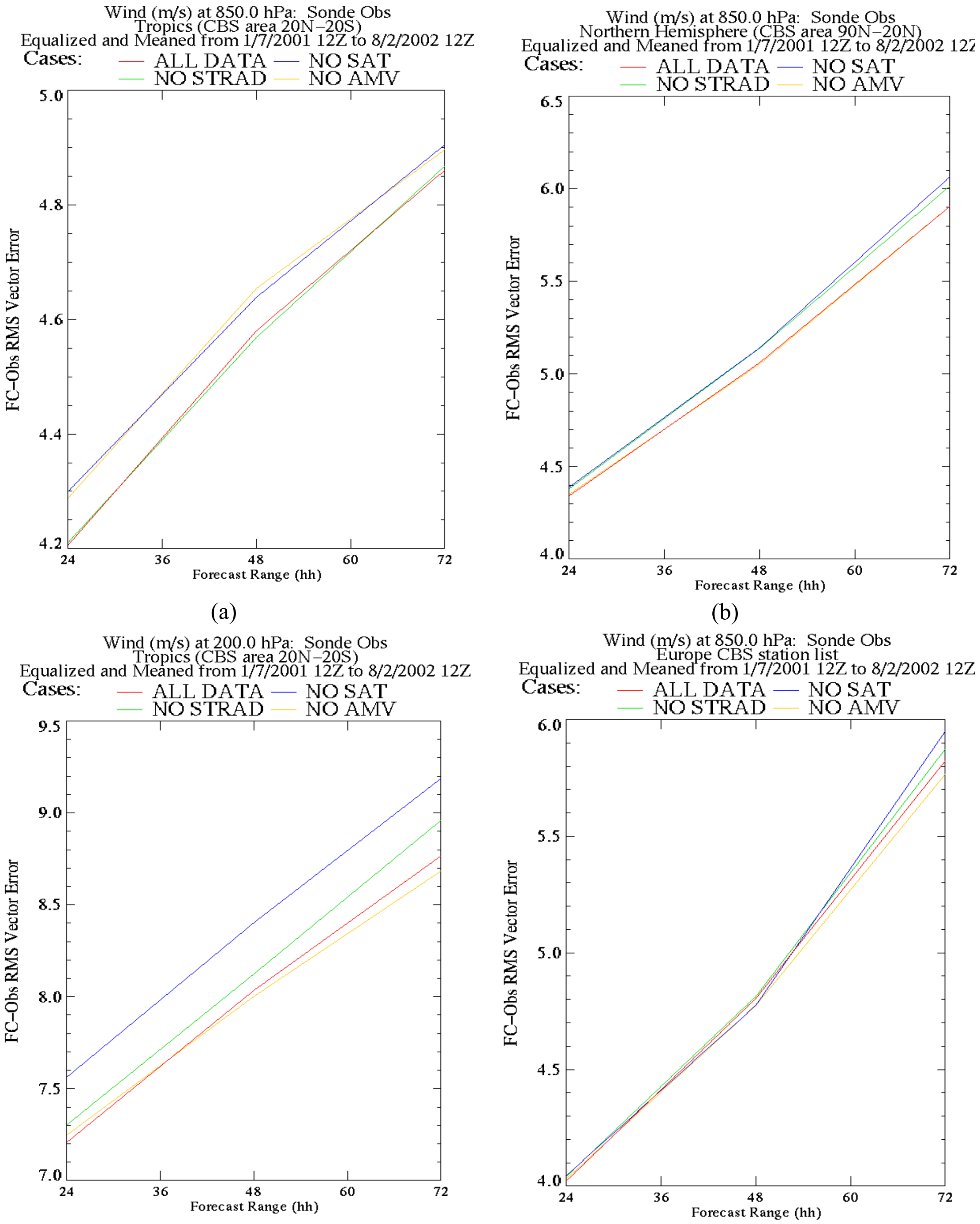
Satellite radiance data have a larger positive impact on forecasts in general than AMV data. The impact of AMV on 500 hPa height forecasts in both the northern and southern hemispheres is neutral, whereas satellite radiance data have a positive impact of up to 6 hours in the northern hemisphere and about 30 hours in the southern hemisphere [Figures 2(a) & 2(c)]. Satellite radiance data also have much bigger impact on wind forecasts than AMV data in both the northern hemisphere [Figure 3(b)] and southern hemisphere [Figure 3(e)].



(a) (b)

(c) (d)

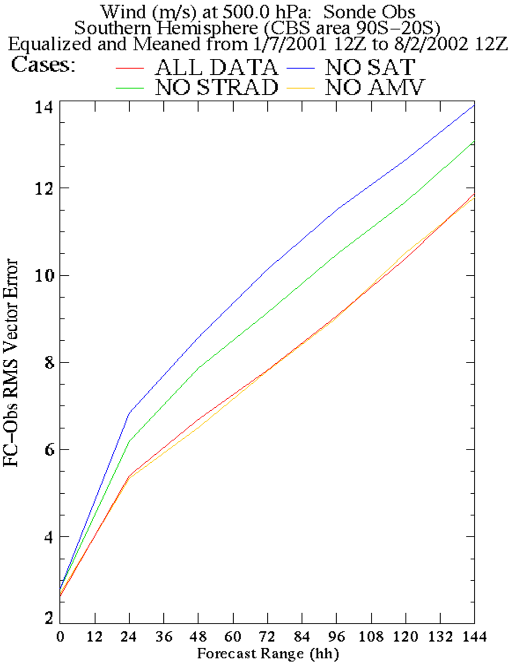
Figure 2. Comparison of satellite data impact. Anomaly correlation coefficient (versus 'All data' analysis). (a) Northern hemisphere (90N - 18.75N) 500 hPa height, (b) Tropics (18.75N - 18.75S) 500 hPa height, (c) Southern hemisphere (90S - 18.75S) 500 hPa height, (d) Europe (25N - 70N, 10W - 28E) 500 hPa height



(a) (b)

(c) (d)

Figure 3. Comparison of satellite data impact for wind. RMS vector wind errors versus radiosondes. (a) Tropics at 850 hPa, (b) Northern hemisphere at 850 hPa, (c) Tropics at 200 hPa, (d) Versus Europe CBS station list at 850 hPa,



(e)

Figure 3(e) Southern hemisphere at 500 hPa

The benefit of AMV can be most clearly seen in the tropics. For 850 hPa wind forecasts in the tropics, the most of the benefit of satellite data is due to AMV data [Figure 3(a)], although at 200 hPa the impact is neutral [Figure 3(c)]. For forecasts of 500 hPa geopotential height in the tropics, AMV data give a benefit of about 12 hours up to T+72 [Figure 2(b)]. Such positive impacts are not seen in either the northern hemisphere [Figures 2(a) & 3(b)] or southern hemisphere [Figures 2(c) & 3(e)].

The impacts of satellite radiance and AMV are not additive. By comparing the NO SAT and NO STRAD curves in Figure 2(c), it can be deduced that AMV data only have a measurable impact on southern hemisphere forecasts of 500 hPa height if satellite radiance data are not present. A similar result is seen for 500 hPa height forecasts in the tropics where the combined effect of removing all satellite data is greater than removing the observation types individually [Figure 2(b)]. For wind forecasts in the tropics, AMV data have a large impact of about 12 hours when satellite radiance data are not present [Figure 3(c)].

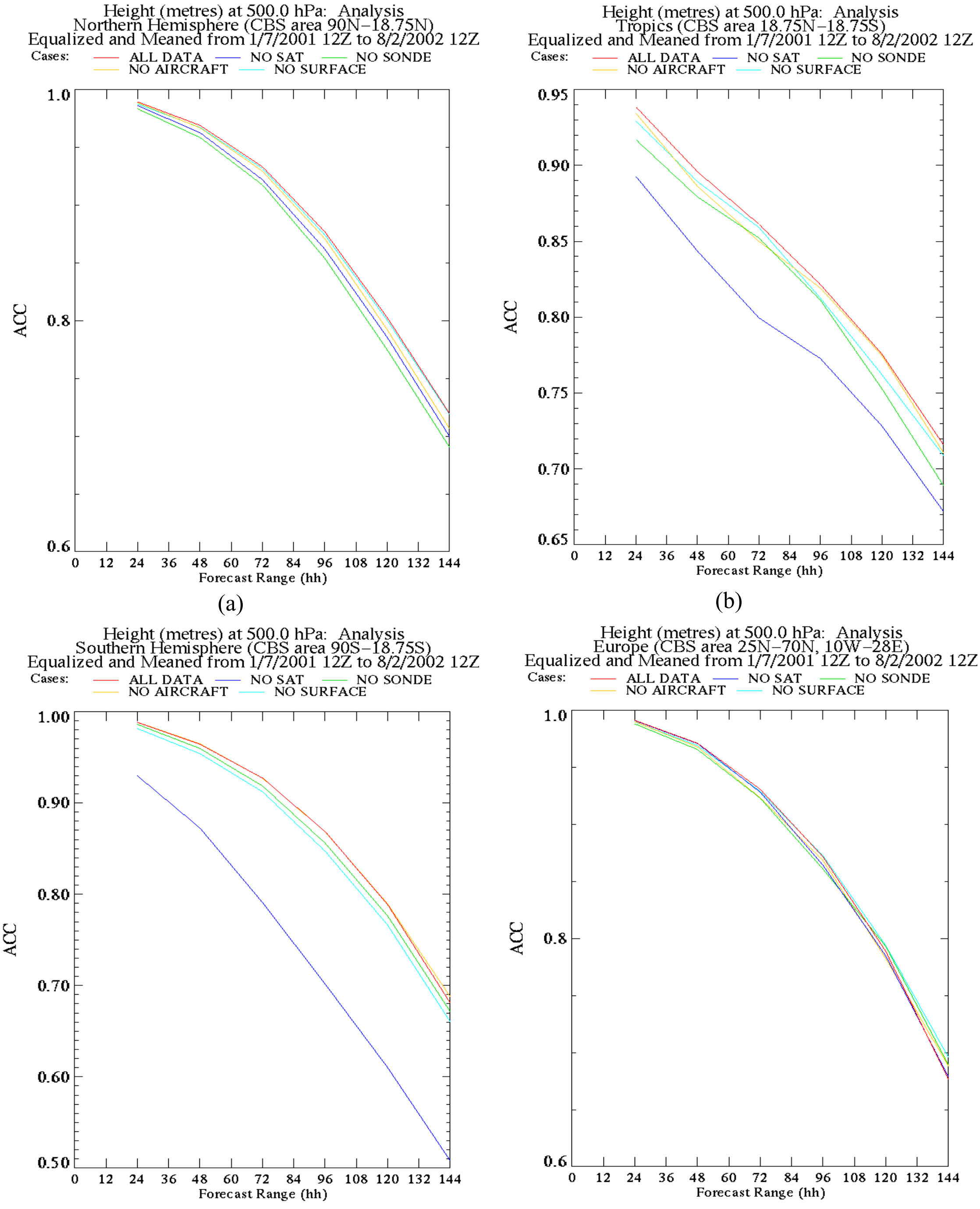
The impact of satellite data over Europe is neutral in these runs. A small impact is seem against both 500 hPa geopotential height [Figures 2(d)] and 850 hPa wind [Figure 3(d)].

The impacts discussed above are largely consistent in the vertical (not shown).

3.1.4. Impact of satellite data vs surface-based data

The impact of satellite observations is compared with the impact of surface-based observations in Figures 4 and 5. The plots shown are the same as those in BK2001, Figures 3 & 4. The main points to note are as follows.

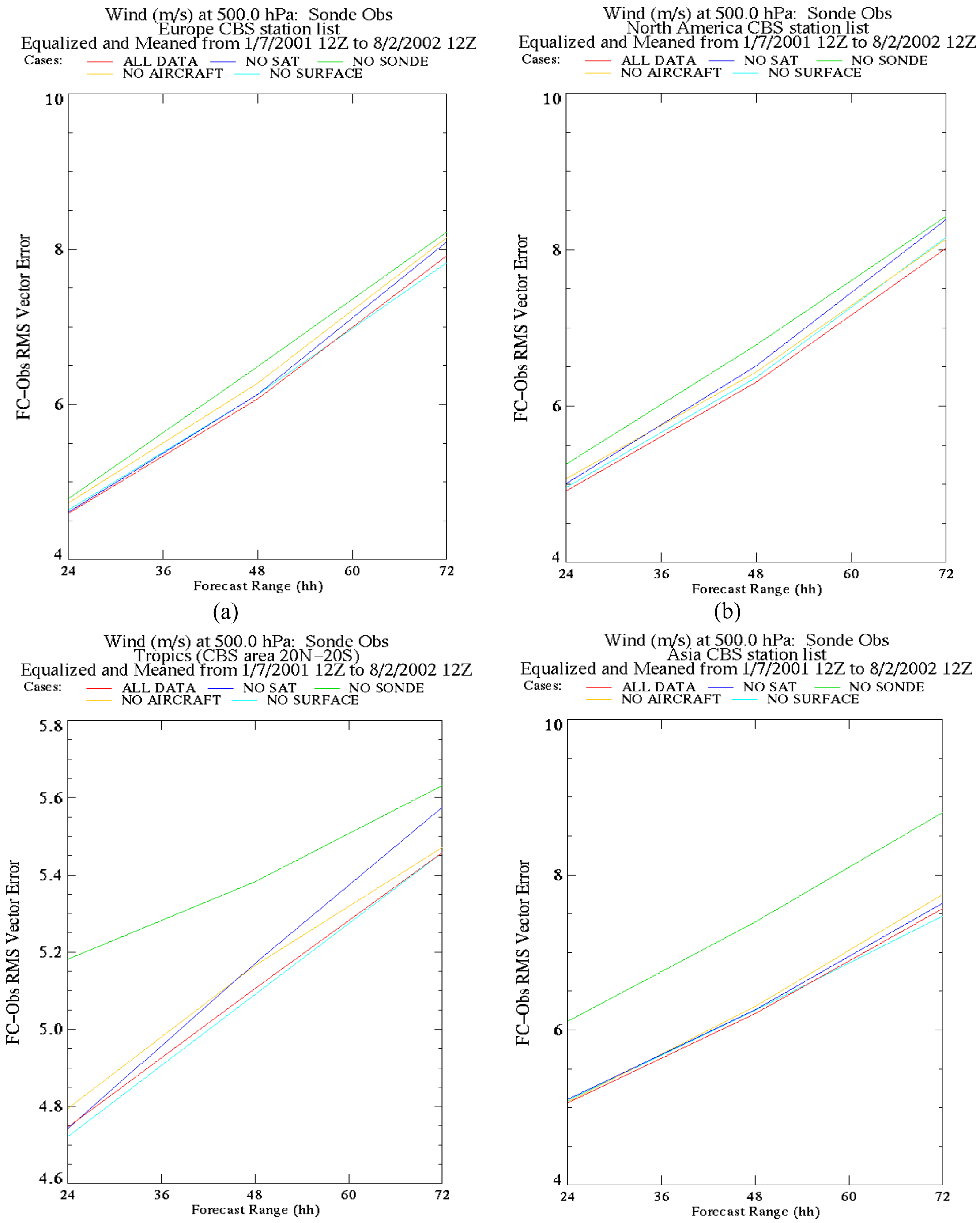
Radiosonde data have the largest impact on forecasts of wind and geopotential height in the northern hemisphere. For example, for 500 hPa geopotential height forecasts in the northern hemisphere, radiosonde data are the most important data source, followed by satellite data and aircraft data [Figure 4(a)]. For 500 hPa wind forecasts over Asia, radiosonde data give a benefit of about 24 hours whereas the other observation types have neutral impact [(Figure 5(d)]. For 500 hPa wind forecasts over North America, radiosonde data have the largest impact of up to 12 hours [Figure 5(b)].



(a) (b)

(c) (d)

Figure 4. Comparison of satellite and surface-based data. 500 hPa height anomaly correlation coefficient for (a) northern hemisphere (b) tropics (c) southern hemisphere (d) Europe.



(a) (b)

(c) (d)

Figure 5. Comparison of satellite and surface-based data. Root mean square vector wind error for 500 hPa wind versus radiosondes for (a) Europe (b) North America (c) tropics (d) Asia.

Satellite data have the largest impact on forecasts of wind and geopotential height in the southern hemisphere. For 500 hPa geopotential height forecasts in the southern hemisphere, satellite data have by far the largest impact of all observation types; the impact at all forecast ranges up to T+144 is about 40 hours [Figure 4(c)]. A similar large impact is seen on wind forecasts (not shown).

In the tropics satellite data have the largest impact on geopotential height forecasts, and both satellite and radiosonde data have a marked impact on wind forecasts. For 500 hPa geopotential height forecasts in the tropics, satellite data have the largest impact which varies between about 18-30 hours depending on forecast range. Radiosonde data are the next most important data source at almost all forecast ranges up to T+144. Up to T+72, aircraft data having a bigger overall impact than surface data whereas at longer forecast ranges surface data have a bigger impact [Figure 4(b)]. For wind forecasts at 500 hPa in the tropics, radiosonde data have the largest impact up to T+72 [Figure 5(c)] but at longer forecast ranges and other levels, satellite data have the largest impact [Figure 3(e)].

Aircraft data play a relatively important role, particularly for short-range wind forecasts in the northern hemisphere and tropics. For 500 hPa wind forecasts over Europe, aircraft data have an impact of up to 6 hours over the forecast ranges 24-72 hours compared with radiosonde data which have the largest impact of all observation types of up to 12 hours [Figure 5(a)]. For 500 hPa wind forecasts in the tropics, aircraft data have an impact of about 6 hours up to T+48 and a smaller impact at longer forecast ranges [Figure 5(c)].

Surface data have a positive impact on forecasts, particularly in the southern hemisphere. For the forecasts of 500 hPa height in the southern hemisphere, surface data are the second most important data source after satellite data with an impact of about 6 hours at all ranges up to T+144 compared with radiosonde data that have an impact of less than 3 hours [Figure 4(c)].

For forecasts over Europe, all observation types have a neutral impact on geopotential height forecasts, but radiosonde, aircraft and to a lesser extent satellite data, have a positive impact on wind forecasts. The mean impact of all observing systems on 500 hPa height over Europe is shown in Figure 4(d). It can be seen that the overall impact of any observing system is neutral. However, radiosonde data and to a lesser extent aircraft and satellite data, have a clear benefit on 500 hPa wind forecasts [Figure 5(a)].

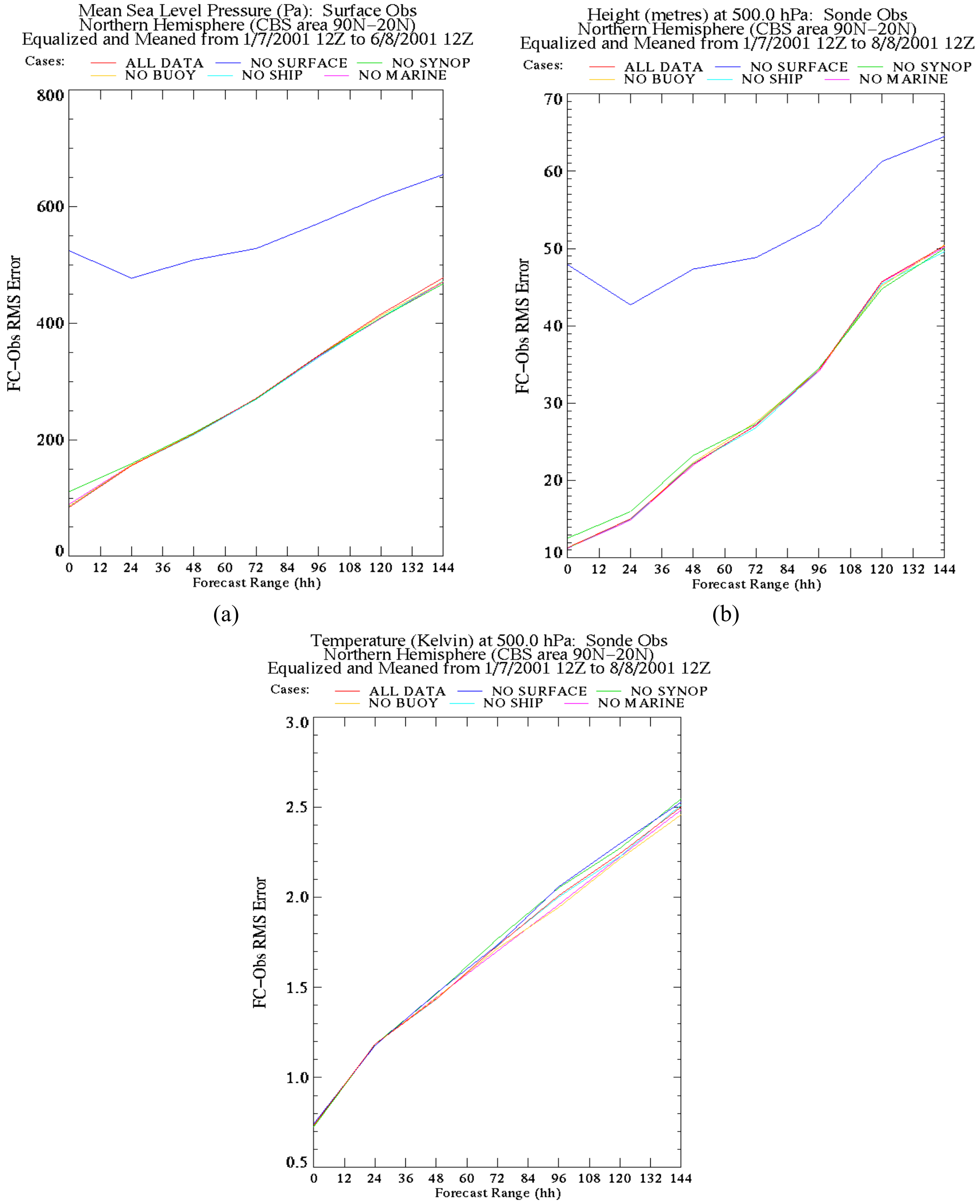
The impact of individual observing systems in the northern hemisphere is small. The impact of removing any one observing system is small. For example, for 500 hPa height forecasts, a maximum impact of less than 12 hours on 500 hPa height is seen when radiosonde data are removed [Figure 4(a)].

3.1.5. Impact of surface data

Some results showing in more detail the impact of surface data and components of the surface observing network are presented in this section. The mean scores plotted are averaged over 30 forecasts from July 2002 observations.

From Figure 6(a) it can be seen that surface data have a large impact on forecasts of mean sea level pressure. Similar results exist for the tropics and southern hemisphere, but are not shown here. It should be noted that the Met Office global data assimilation system uses surface pressure from all surface reports, whereas surface wind observations are used only from ships, moored buoys, platforms and rigs and no surface temperature observations are used. It is thus concluded that 'in-situ' surface pressure observations are essential for Met Office forecasts. It appears that the data assimilation scheme cannot produce a realistic surface pressure field using upper air observations only.

From Figure 6(a) it can also be seen that removing whole sub-components of the surface network makes very little difference to the mean scores, except for forecasts up to T+24 in the case when all SYNOP reports are removed. This result suggests that whilst some 'in-situ' surface pressure observations are essential, the Met Office global NWP system run at 90 km resolution cannot fully utilise all the information provided by a dense surface network [as shown in Figure 1(d)].



(a)

(b)

(c)

Figure 6. Impact of surface observations. Root mean square errors versus observations meaned over the northern hemisphere for (a) mean sea level pressure, (b) 500 hPa geopotential height, (c) 500 hPa temperature.

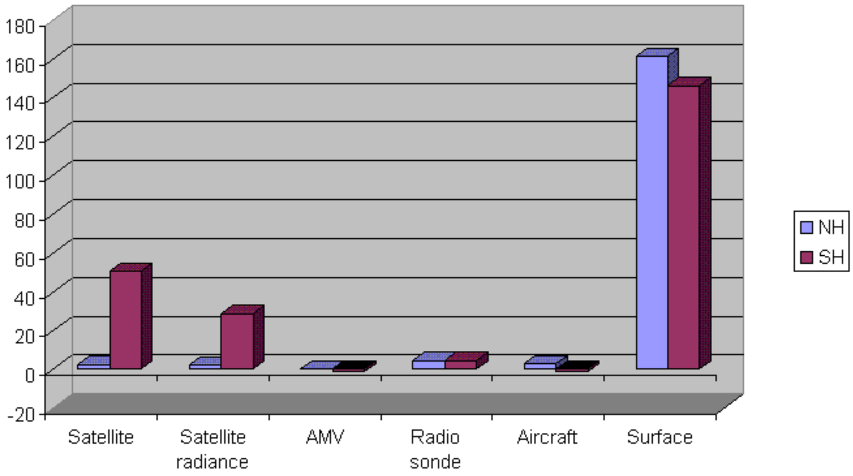


Figure 6 (d) Percentage change in RMS error of 24-hr forecast of mean sea level pressure meaned over the northern and southern hemisphere for all observation types.

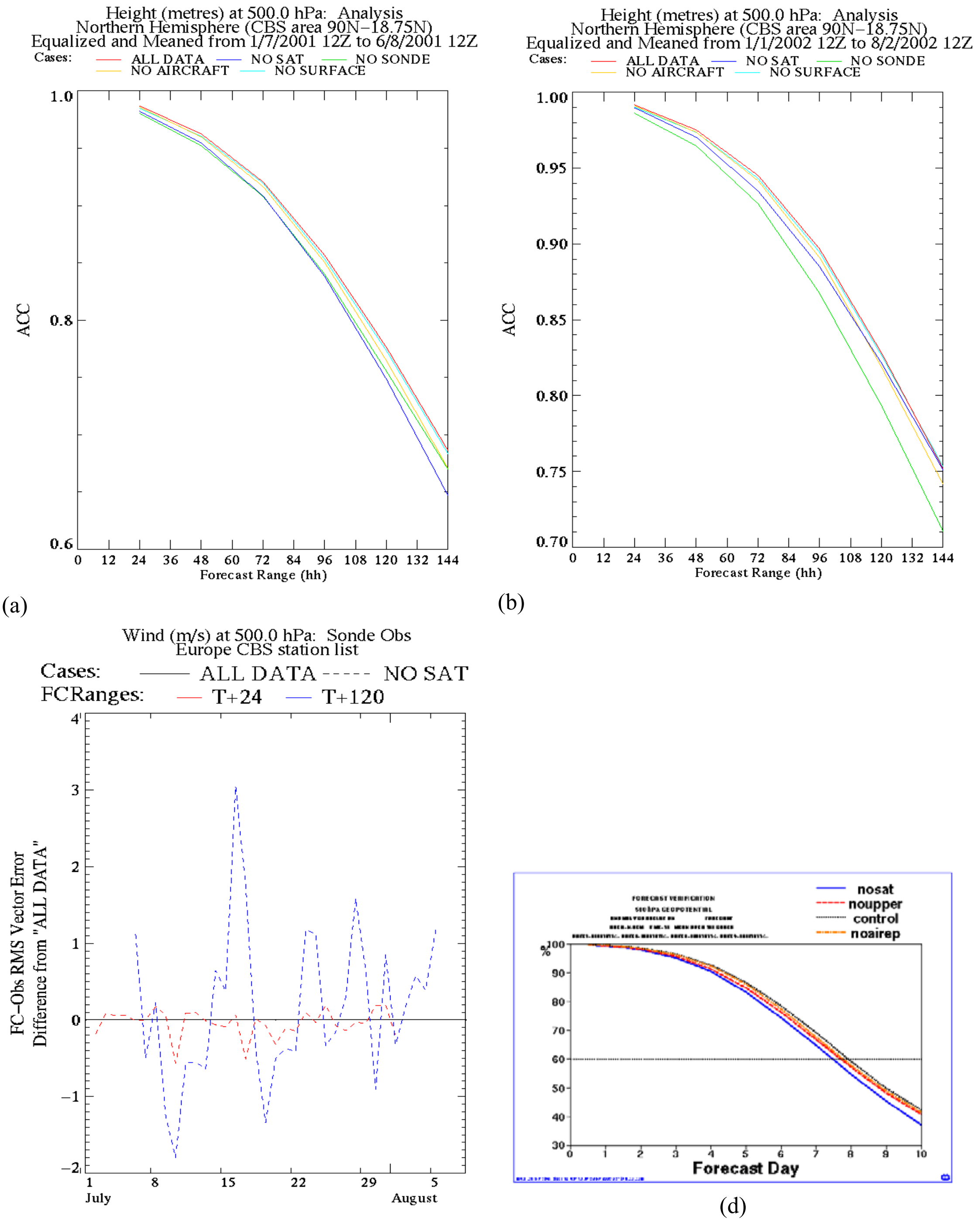
Surface observations have a large positive impact on 500 hPa geopotential height, when verifying against radiosonde observations [Figure 6(b)]. The impact on the geopotential height field is largely due to changes in the surface pressure field; the upper level temperature fields are not affected significantly by removing surface observations [Figure 6(c)].

The relative importance of the surface network on the 24-hour forecast of mean sea level pressure is indicated in Figure 6(d). Clearly surface observations have a much larger impact than any other observing system.

3.1.6. Influence of sample size, period and NWP system

Figures 7(a) & 7(b) show the differences in the impacts of satellite and surface-based observing systems on 500 hPa height. Both figures show the mean impact on the anomaly correlation coefficient over 30 forecasts. The values in Figure 7(a) were calculated from July 2001 forecasts and those in Figure 7(b) from January 2002 forecasts. It can be seen from the figures that satellite data have the largest impact on forecasts in July whereas radiosonde data have the biggest impact in January. Since the same NWP system was used to produce the forecasts from both periods, the differences in the results must either be due to differences in the observation coverage or flow regimes. It was shown in a previous section that most of the impact of satellite data in the northern hemisphere is due to the satellite radiance measurements from NOAA-15 and NOAA-16. Table 2 shows that the total number of satellite radiance profiles from these satellites and the number of radiosonde soundings. It can be seen that the number of satellite radiance and radiosonde soundings from each period was similar. Assuming that the distribution of the observations was also similar, then the differences in the results are likely to be due to the flow regime.

Figure 7(c) shows an example of the daily fluctuation in 500 hPa RMS vector wind errors over Europe during July 2001. This figure should be compared with Figure 5(a) which shows the mean impact. It can be seen from Figure 7(c) that the impact of satellite data is sometimes positive and sometimes negative, although Figure 5(a) shows that the mean impact is positive. Figure 7(c) shows how verifying small numbers of forecasts can give misleading results. The forecasts verifying between 7th and 14th July indicate a negative impact from satellite data, whereas those verifying between 23rd July and 6th August indicate a mainly positive impact. It should be noted that the mean RMS error for this period may be strongly influenced by the error of the forecast verifying on 17th July - the presence of satellite data for this forecast appears to have had a large positive impact.



(a)

(b)

(c)

(d)

Figure 7. Influence of sample size and NWP system on the results from OSEs. (a) Mean 500 hPa height anomaly correlation coefficient for northern hemisphere. Mean taken over 30 forecasts from July 2001. (b) As (a) but for January 2002. (c) Difference (NO SAT-ALL DATA) in 500 hPa RMS vector wind errors verified against European radiosondes for 24-hr and 120-hr forecasts in July 2001. (d) As (a), but from an OSE using the ECMWF NWP system (courtesy Graeme Kelly). Average taken over 120 forecasts from August/September 2002 and December 2002/January 2003.

	Radiosonde	Satellite (NOAA-15 & 16)
July 2001	63,965	38,138,928
January 2002	64,751	39,718,936

Table 2. Total (global) number of radiosonde and satellite radiance soundings available for assimilation.

By comparing Figure 7(d) and Figure 4(a), the potential impact of the NWP system can be assessed. The results from the ECMWF experiment show a bigger impact from satellite data in the northern hemisphere than in the Met Office experiment. This extra impact may indicate the benefit of ECMWF's 4D-Var data assimilation scheme since such schemes are better suited to assimilate data from asynoptic observations, such as those from satellites. Part of the difference may also be due to the fact that a later data cut off time, and hence more satellite data, was used in the ECMWF experiment. The difference in impact may also be explained by the difference in the period and length of each experiment.

3.1.7. Summary and conclusions

The results presented show that overall all observing systems are having a positive impact on Met Office forecasts apart from small negative impacts for some variables, levels and forecast ranges. Thus the Met Office 3D-VAR data assimilation system is working well. Satellite data have the largest impact of all observing systems in the southern hemisphere, with satellite radiance data having a larger impact than AMV data. Radiosonde data have an important impact, particularly in the northern hemisphere. These points are illustrated in Figure 8 which shows the impact of all observing systems on the root mean square vector wind error for 24-hr forecasts of 250 hPa wind.

Surface data are crucial for providing realistic forecasts of surface pressure; the Met Office data assimilation and forecast system cannot produce accurate surface pressure forecasts using upper air data alone.

Care is required when interpreting the results from these OSEs. It has been shown here that the results depend on the period chosen for study and its duration.

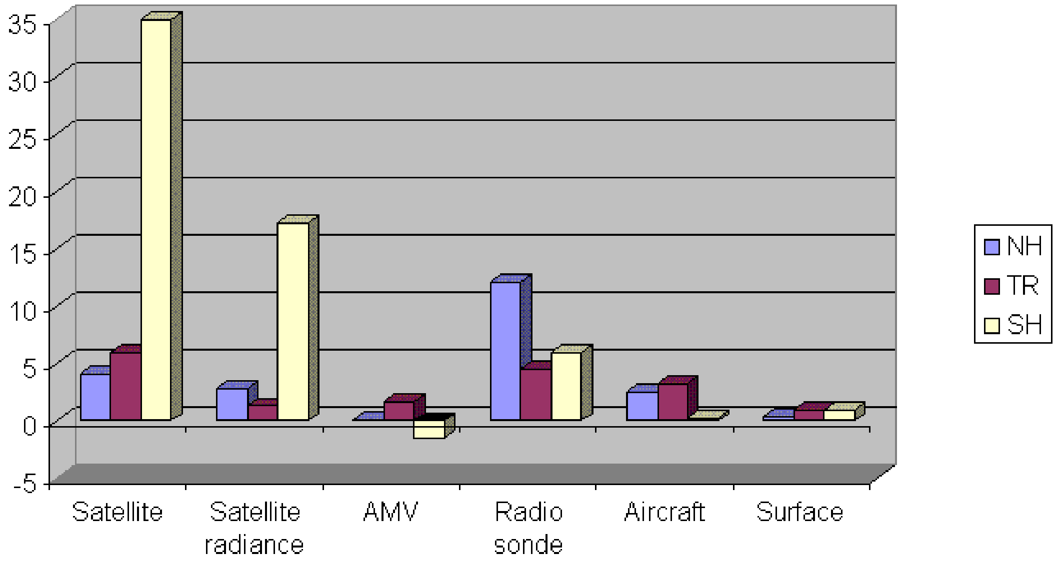


Figure 8. Percentage change in RMS errors (vs observations) for 24-hr forecasts of 250 hPa wind.

The results are in broad agreement to those of a similar study carried out at ECMWF. The reasons for differences cannot be easily identified using the results of this study but are likely to be due to differences in the periods chosen, in terms both of flow regime and observation availability, and the data assimilation methods employed.

3.2. Impact of 'in-situ' profile measurements in tropical regions

Surface-based profile observations have a significant impact on NWP forecasts, despite the use of increasing amounts of satellite data that can be effectively assimilated using variational techniques (Bouttier and Kelly, 2001). Currently, and for the next few years, improvements in the benefit of satellite data for NWP are likely to be limited due, for example, to the problems of obtaining useful information in cloudy areas which are highly correlated with initial condition sensitivity (McNally, 2000). A complimentary, global network of surface-based observations will be necessary to ensure continuing improvements to NWP (WMO, 2002).

It can be seen from recent plots of global radiosonde and aircraft distribution [Figures 9 & 1(e)] that there are large land areas of the tropics where the coverage of surface-based observations is sparse, particularly over tropical Africa. This led the WMO Expert Team on Observational Data Requirements and Re-design of the Global Observing System (ET-ODRRGOS) to request that leading NWP centres investigate the potential value of an enhanced surface-based profile network in the tropics. The results presented here attempt to answer the questions posed by the Expert Team.

Given the relatively dense coverage of radiosonde data in south-east Asia (see Figure 9), the Expert Team suggested that impact studies be carried out in which radiosonde data was denied from an area covering south-east Asia. Should it be found that the data have a positive impact on NWP forecasts, for the local region or outside, then it may be concluded that an improved surface-based profile network over say, tropical Africa, would have a similar benefit. The Expert Team suggested two experimental scenarios designed to assess the impact of adding either profile measurements from aircraft or radiosonde TEMP reports that include humidity data.

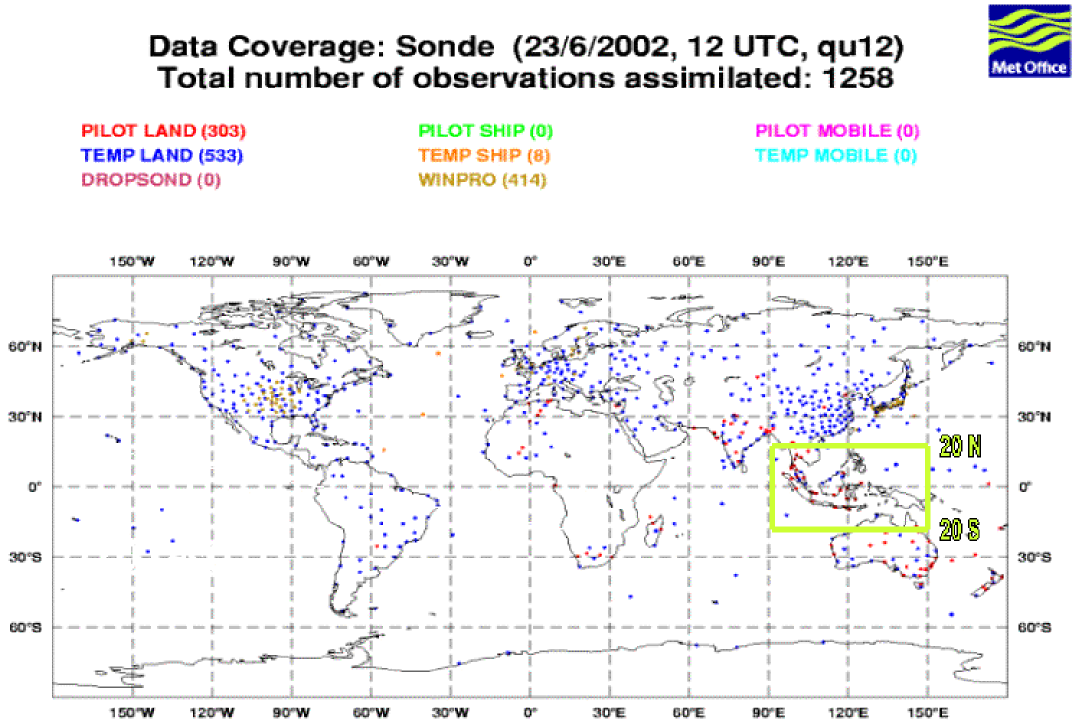


Figure 9. Global distribution of radiosondes. Box indicates the area of south-east Asia from which reports were denied.

3.2.1. Description of the Experiment

An Observing System Experiment was run using the Met. Office operational forecast model and 3-D variational data assimilation scheme. In order to reduce the computational expense, the forecast model was run used at reduced (90 km) horizontal resolution. A one month trial was performed using July 2001 observations and thirty 6-day forecasts were verified against both radiosondes and analyses. The verification areas used include those shown in the Appendix.

The area of south-east Asia over which data were denied is shown in Figure 9. Three runs were performed:

(i) using all available observations of all data types

(ii) as (i) but with no radiosonde or aircraft profile data from south-east Asia

(iii) as (i) but with no radiosonde humidity information or aircraft profile data from south-east Asia.

Scenario (ii) represents the current situation over some parts of the tropics, for example, Africa. Scenario (iii) represents the inclusion of AMDAR profile data, and scenario (i) the inclusion of sonde data. Note that the analysis fields from run (i) (the 'all data' run) were used in calculating the anomaly correlation coefficients.

3.2.2. Impact on wind forecasts

Verification scores against radiosondes within the south-east Asian region for two levels of wind forecasts are plotted in Figure 10(a) & 10(b). It can be seen that a both levels, but particularly 250 hPa, profile data have a positive impact on forecasts at most forecast ranges. Since the effect of removing humidity data is approximately neutral, it appears as though the benefit of the profile data comes largely from temperature and wind measurements. This point is further illustrated in Figure 10(c), which shows the vertical distribution of wind RMS errors for 24-hr and 120-hr forecasts. The impact of full profile information, including humidity, is positive or neutral at all levels and both forecast ranges. In contrast, the impact of the humidity profile information is either neutral or negative, being negative particularly at 120-hr range.

The impact of the tropical profile data on forecasts in a region (Asia) adjacent to where the observations were made is indicated in Figure 10(d). A small positive impact from the radiosonde data can be seen at some levels and forecast ranges. However, the impact of humidity data is neutral suggesting that the positive impact of the full profile is due to the temperature and wind components.

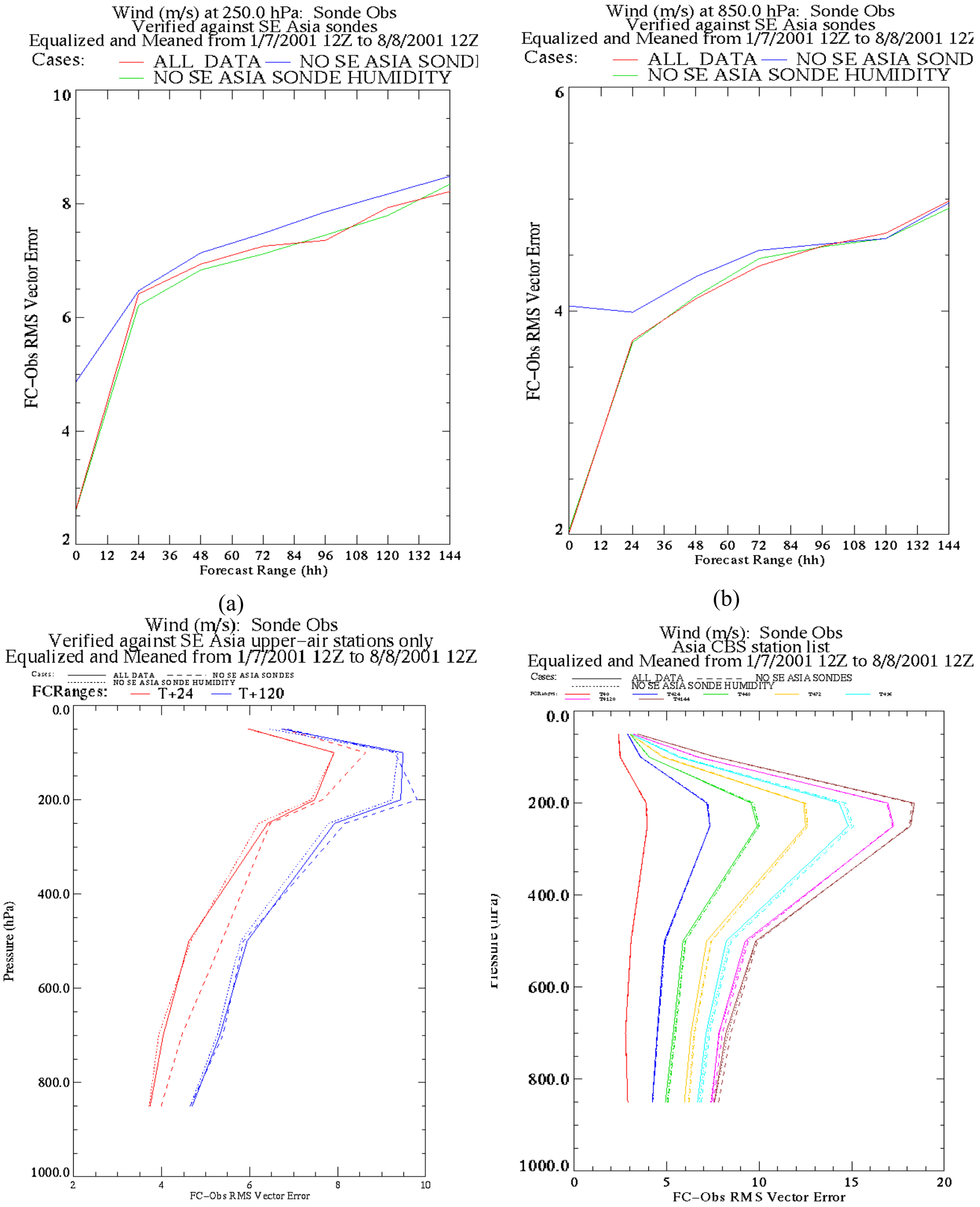
3.2.3. Impact on geopotential height forecasts

The results presented are the ACC scores using the analysis from the 'All data' run in the calculation. RMS errors versus radiosondes have also been examined and show qualitatively the same results.

Figure 11 show the forecast geopotential ACC scores versus pressure level and for different forecast ranges over the Asia. For a definition of this region see Table 1. The tropical profile data have a neutral or positive impact up to T+96, but a negative impact can be seen at T+120 and T+144.

3.2.4. Conclusions

It appears from these results that extra 'in-situ' temperature and wind profile measurements in the tropics would benefit wind and height forecasts in the regions where the observations are taken. However, the benefit of extra 'in-situ' humidity measurements is not clear. It is thus likely that extra AMDAR profile measurements over Africa would benefit forecasts for the region.



(a) (b) (c) (d)

Figure 10. Impact of 'in-situ' profile data from south-east Asia. Mean RMS vector wind errors (m/s). Mean values are calculated over 30 forecasts.(a) At 250 hPa verified against radiosondes in south-east Asia. (b) At 850 hPa verified against radiosondes in south-east Asia. (c) For 24-hr and 120-hr forecasts verified against radiosondes in south-east Asia plotted for selected pressure levels. (d) Against Asian radiosondes plotted for selected pressure levels.

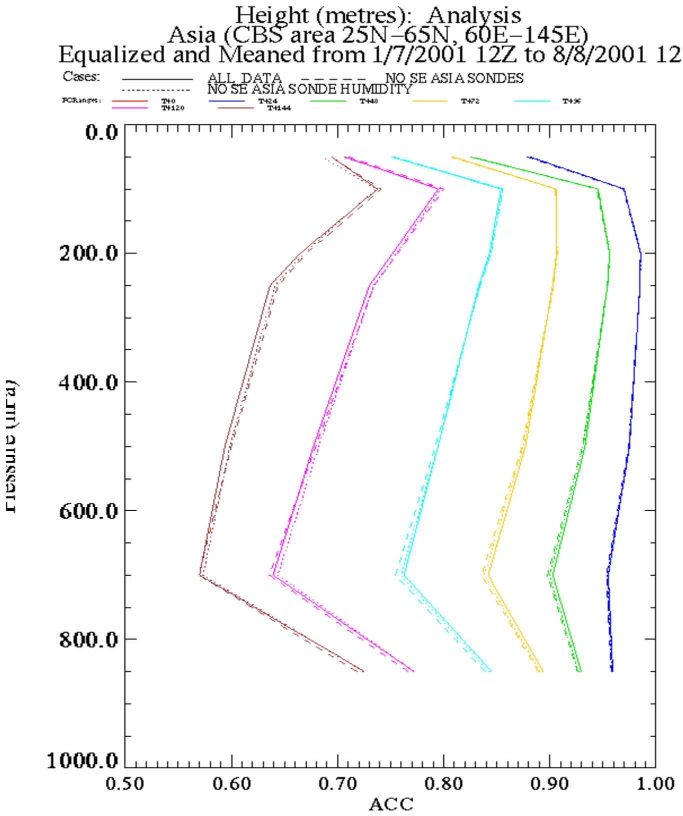


Figure 11. Mean anomaly correlation coefficient for geopotential height (using the 'All data' analysis) for the 'Asia' region. Mean calculated over 30 forecasts.

3.3. Impact of extra North Atlantic ASAP[2] reports

Various recent studies have suggested that improved observational coverage over the North Atlantic is needed for the production of better NWP forecasts over Europe. For example, Bader and Saunders (2001) found cases where a scarcity of observations to the west of France and Iberia may have resulted in poor short-range NWP forecasts over Europe. Such studies have encouraged EUCOS to consider the deployment of more 'in-situ' observations in the Atlantic.

As an initial step towards the enhancement of the North Atlantic observation network, an observational field campaign took place during September and October 2001. During this period, extra ascents from 12 Atlantic ASAP ships were made, and the Azores radiosonde (08508) reported four times each day. The aim of this pilot experiment was primarily to check the technical feasibility of producing extra ASAP reports at variable times, and secondly to evaluate the impact on NWP forecasts of the extra data (Gerard, 2000).

In order to assess the impact of the data, an Observing System Experiment (OSE) has been run using the Met Office's NWP system and the results assessed.

3.3.1. Availability of data during the trial period

During the field campaign, the number of temperature and wind reports that were assimilated by the Met Office model increased by 22% compared with the previous two month period. Figure 12 shows an example of the distribution of all 'in-situ' sounding data in the North Atlantic. At 12 UTC 21/9/01 six useful ASAP reports were available; five in the Atlantic and one in the North Sea.

---

[2] Automatic Ship Aerological Program

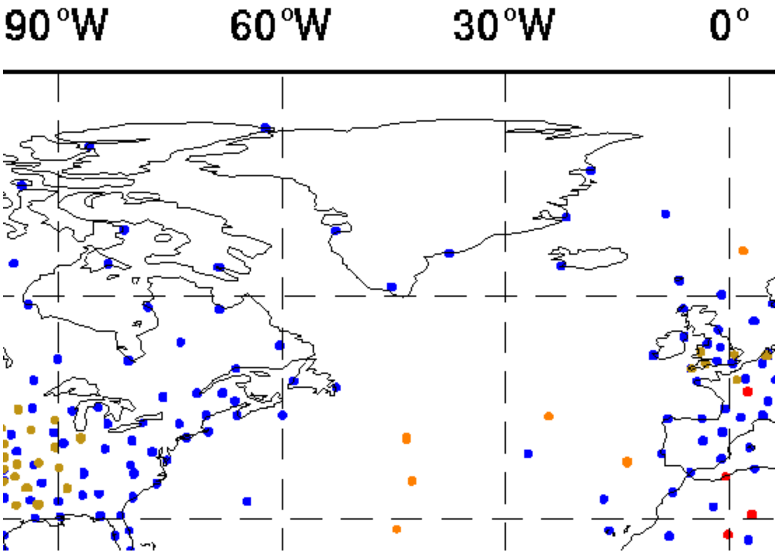


Figure 12: Distribution of useful reports for 12UTC 21/9/01. Orange dots denote the positions of ASAP ships.

Even with six ASAPs reporting, some regions of the flow containing notable features may not be observed by the ships. It should be noted that on many occasions during the trial less than seven E-ASAP reports were available; sometimes as few as three or four.

3.3.2. Description of the OSE

The Met Office operational global model (Cullen, 1993), which uses a three-dimensional variational assimilation scheme, was run for the two-month period of the E-ASAP field campaign. In order to reduce the resources required to complete the OSE, the forecast model was run at less than operational horizontal resolution (approximately 90km compared with 60km) but at full vertical resolution.

Two runs were performed:

(i) Using all available data, including all the extra data from the ASAPs and Azores radiosonde (the ‘All data’ run)

(ii) Using all data less all data from all ASAP ship reports, including those not part of the field campaign, and all Azores radiosonde reports (the ‘No ASAP’ run).

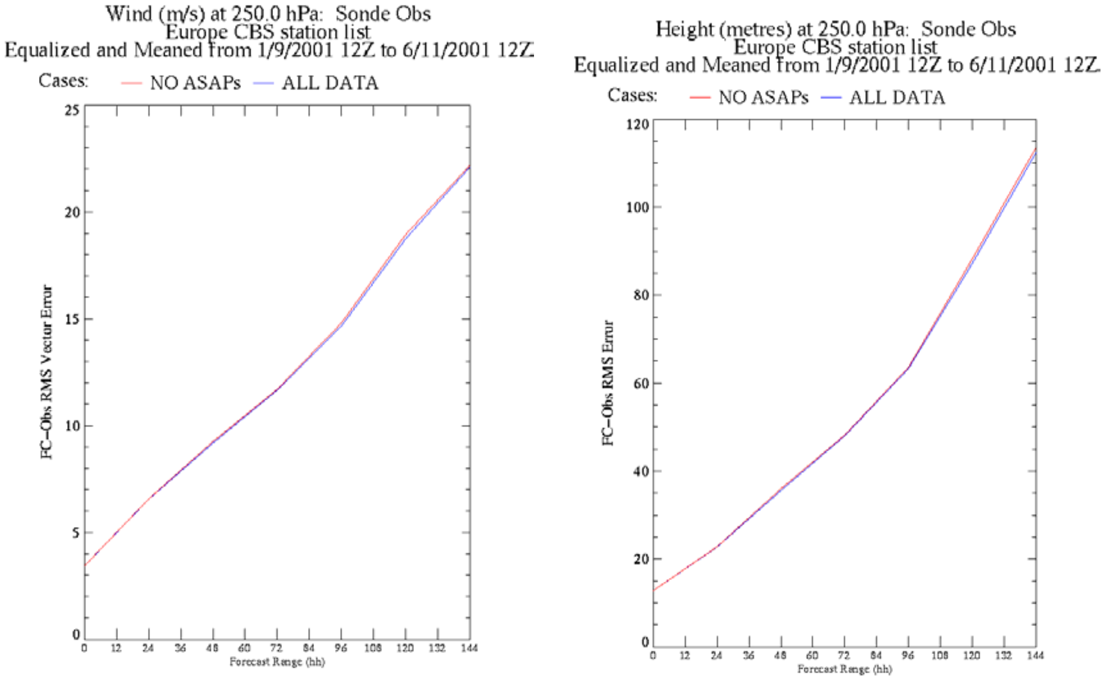
Scenario (ii) is designed to test the ***maximum*** impact that ASAPs and the Azores radiosonde could have.

For each run, 6-day forecasts were produced every day from 12UTC data. The forecasts and analyses were verified over a comprehensive set of regions covering the whole globe.

3.3.3. Results from the OSE

Only small differences in RMS errors, averaged over the trial period, have been found. Some results are presented here showing verification against European sondes.

Overall, the impact of the ASAP and Azores radiosonde data on RMS scores against European sondes, averaged over the two-month trial period, was neutral to slightly positive. Two examples of mean statistics indicating slight positive impact are shown in Figures 13 (a) & (b). Both the plots of RMS error of 250 hPa height and 250 hPa wind show that the ‘All data’ run is marginally better than ‘No ASAP’ run.



(a) (b)

Figure 13. Impact of ASAP data over Europe. Means calculated over 60 forecasts. (a) Mean RMS error (m/s) of the vector wind at 250 hPa verified against European radiosondes for different forecast ranges. (b) Mean RMS error (metres) of the height at 250 hPa verified against European radiosondes for different forecast ranges.

It should be noted that the RMS errors plotted have been averaged over 60 cases; hence any noticeable differences in RMS scores in individual cases will not be clearly seen in these averaged RMS errors.

3.3.4. *Case study: T+96 forecast from 12UTC 17/9/01*

For those variables, at levels and over regions where the mean statistics were observed to be slightly different, the time series of differences in RMS errors were examined to identify cases in which inclusion of ASAP data gave noticeable impact on the forecast. In such plots the RMS error in the ‘No ASAP’ run has been subtracted from the RMS error in the ‘All data’ run. Hence negative values indicate smaller errors in the ‘All data’ run, and positive values denote larger errors.

Such a time series, showing the difference in the RMS errors in the 4-day forecast of 250 hPa wind over Europe, is shown in Figure 14. It can be seen that two forecasts made in September from the ‘All data’ run have lower RMS errors than any others during the trial period.

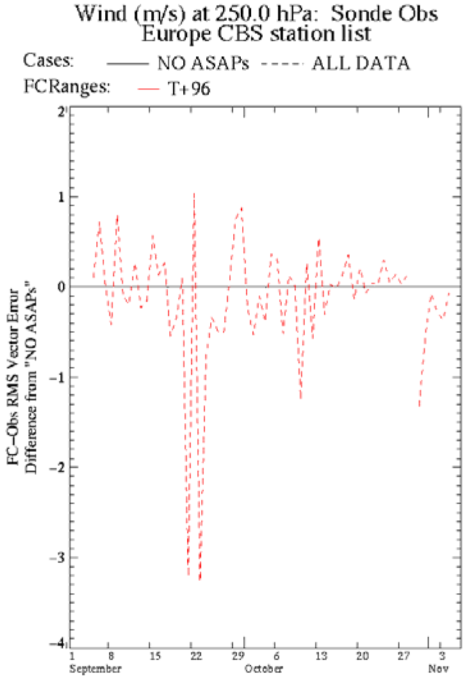
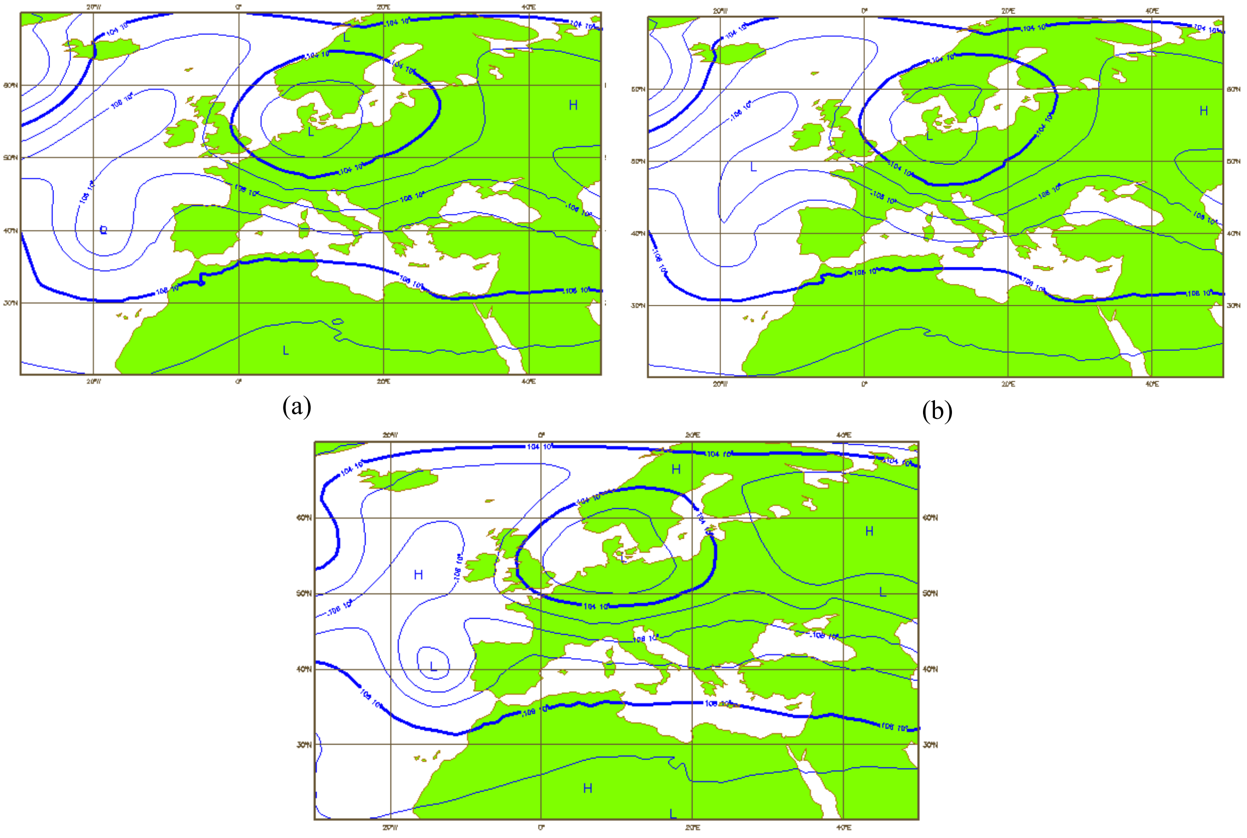


Figure 14. Time series of the difference ('No ASAP' – 'All data') in RMS errors (m/s) for 250 hPa vector wind verified against European radiosondes.

The difference in the wind forecasts is reflected by a difference in the 250 hPa height forecasts. Figures 15(a) and 15(b) show the T+96 forecast from the 'All data' and 'No ASAP' runs respectively. It can be seen that the trough to the west of Iberia is deeper in the 'All data' run than the 'No ASAP' run.



(a)

(b)

(c)

Figure 15. Impact of ASAP data on T+96 forecast of 250 hPa geopotential height valid at 12UTC 21/9/01. Contour interval is 100 metres. (a) T+96 forecast from the 'All data' run. (b) T+96 forecast from the 'No ASAP' run. (c) Analysis valid at 12UTC 21/9/01 from the 'All data' run.

The forecast from the 'All data' run is more similar to the 12UTC 21/9/01 analysis (Figure 15(c)) than is the forecast from the 'No ASAP' run.

3.3.5. Discussion and conclusions

The overall impact on forecasts up to 6 days ahead of the data from all ASAP ships plus data from the Azores radiosonde reporting four times a day was neutral to slightly positive. A few cases were found when the data had a significant impact at longer forecast ranges (T+96 - T+144) when verifying over the European region.

In this OSE, the data from up to 13 North Atlantic radiosondes were denied, although typically no more than about 6 were reporting simultaneously. However, Pailleux (1997) suggests that the data from 10 or more North Atlantic radiosondes need to be denied in order to get a measurable impact on forecasts over Europe. Moreover, Pailleux's conclusion was based on studies performed before the profile data from current satellites were available. The assimilation of these data using recently implemented variational techniques has led to satellite sounding data having an increased benefit on forecasts in the Northern Hemisphere (Bouttier & Kelly, 2001). Thus 'in-situ' sounding data over the North Atlantic are now unlikely to have as much impact as observed by Pailleux. Thus it would be expected that many more than 6 additional radiosonde reports over the North Atlantic would be required to obtain significant benefit on forecasts over Europe.

The relatively small impact may be due to an 'undersampling' of the synoptically sensitive parts of the flow over the Atlantic. The weather over the Atlantic during September and October 2001 was markedly anticyclonic with, in particular, anti-cyclones or weak flow persisting over the Azores for significant periods (Met Office Daily Weather Summary, September & October 2001). Sounding data taken in such conditions are unlikely to produce a large impact. Given that no more than 6 soundings were typically available at any one time, it is likely that the parts of the North Atlantic that were sensitive to synoptic development were not observed from the ASAP ships.

It is important that the results from this experiment are not interpreted to mean that in-situ profile data have little value in a GOS containing increasing amounts of profile information from satellites. The benefit of radiosonde data as a whole on global forecasts has been confirmed by recent studies carried out at the Met Office (see section 3.1) and ECMWF (Bouttier & Kelly, 2001). Despite the planned improvements in satellite sounding data, it is likely that in situ profile observations will still be necessary to provide observations especially where the satellite data are less accurate, such as at lower levels and in cloudy conditions.

In anticipation of the continuing limitations of satellite sounding data, it is planned that THORpex[3] (http://www.nrlmry.navy.mil/~langland/THORPEX_document/Thorpex_plan.htm) will provide improved coverage of in-situ profile data over the Northern Hemisphere oceans. In line with the THORpex strategy, it is recommended that more ASAP reports are required over the North Atlantic, preferably targeted at synoptically sensitive areas, in order to improve the benefit on forecasts for the European region.

4. Summary and conclusions

The results from Observing System Experiments can be used to assess data assimilation performance and for the designing of observation networks. Recent results suggest that variational data assimilation schemes are performing well overall.

---

[3] THe Observing system Research and Predictability Experiment

Future OSEs are planned to investigate the impact of the observations obtained during TOST (THORpex Observing System Test), a field campaign taking place in Autumn 2003. OSEs may also be used for investigating the development of the space and terrestrial components of the Global Observing System as part of the EUCOS studies program (see http://www.eucos.net/).

References

Bader, M.J. & Saunders, F.W. (2001). The Requirements for extra Observations to the West of France and Iberia for NWP over Europe. *The Met Office NWP Forecasting Research Technical Report* No. 345.

Bouttier, F. & Kelly, G. (2001). Observing System Experiments in the ECMWF 4D-Var data assimilation system. *Q.J.R. Meteorol. Soc.,* 127, pp 1469-1488.

Cardinali, C. (2000). EUCOS impact study. *ECMWF Technical Memorandum* No. 325.

Cress, A. & Wergen, W. (2001). Impact of profile observations on the German Weather Service's NWP system. *Met Zeitschrift,* Vol. 10, pp 91 - 101.

Cullen, M.J.P, (1993). The unified forecast/climate model. *Met. Mag.,* 122, 81-94.

Dalby, T. & Berney, A. (1999). Modifications to Aircraft Thinning and Observation Errors. *The Met Office. Forecasting Research Technical Report* No. 276.

Gerard, F. (2000). Targeted observation experiment from ASAP. Specifications of an EUCOS-COSNA feasibility study (Unpublished report).

Graham, R.J. and Bader, M.J. (1996). Impact of observations in NWP models: techniques and results of recent studies. *Forecasting Research Division Scientific Paper No. 42,* the Met Office.

Heming, J.T., Chan, J.C.L., & Radford, A.M. (June 1995). A new scheme for the initialisation of tropical cyclones in the UK Meteorological Office global model. *Meteorological Applications,* 2, 171-184.

Hirshberg, P. A. et al (2001). An observing system experiment with the west coast picket fence. *Mon Wea Rev,* Vol. 129 pp 2585 - 2599.

Ingleby, N.B. (1992). The new Meteorological Office quality control system. pp 219-224 in *Preprint Volume of the 12th AMS Conference on Probability and Statistics in the Atmospheric Sciences*. June 1992.

Lorenc, A.C., Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L.F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J., and Saunders, F.W. (2000). The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. *Q.J.R. Meteorol. Soc.,* 126, 2991-3012.

McNally, A.P. (2000). The occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *Technical proceedings of the 11th International TOVS Study Conference,* Budapest, Hungary, 20-26 Sep 2000.

Pailleux, J. (1997). Impact of observations in NWP models. Summary of recent results and perspectives. WMO, *World Weather Watch Report No. 18.* Impact of various observing systems on numerical weather prediction. Proceedings of CGC[4]/WMO Workshop, Geneva, 7-9 April 1997.

Rabier, F., Jarvinen, 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.

---

[4] Co-ordination Group for COSNA (Composite Observing System for the North Atlantic)

Simmons, A.J. (1995). The skill of 500 hPa height forecasts. *Proceedings of 1995 ECMWF Seminar on Predictability, Vol.,* 19-68.

Simmons, A.J. (1999). Objective Verification of deterministic forecasts pp 385 - 404. *ECMWF. Diagnosis of models and Data Assimilation Systems,* 6 - 10 September 1999

World Meteorological Organization (2001). Composite Observing System for the North Atlantic (COSNA). *Report of the meeting of the Scientific Evaluation Group (SEG),* Eleventh session, ECMWF, Reading, UK, 14-15 May 2001, p14.

WMO (2002). *Expert Team meeting on Observational Data Requirements and Redesign of the GOS,* Reduced Session (Oxford, United Kingdom), 1-5 July 2002, AnnexVI (see http://www.wmo.ch/web/www/reports.html#GOS)

Zapotocny, T.H. et al (2002). An impact study of five remotely sensed and five in-situ data types in the Eta data assimilation system. *Weather and Forecasting,* Vol. 17, pp 263 - 285.

Zapotocny, T. H. et al (2000). A case study of the sensitivity of the Eta data assimilation system. *Weather and Forecasting,* Vol. 15, pp 603 - 621.