Meteorology Research and Development

The impact of conventional observations on global and regional NWP forecasts

Technical Report No. 516

Richard Dumelow

email: nwp_publications@metoffice.gov.uk

Document History for NWP Technical Report No. 516

Date	Version	Action/comments	Approval
2/4/08	1.0	Minor modifications from initial draft.	Adrian Semple
3/4/08	1.0	Approved for release.	Andrew Lorenc

ABSTRACT

The impact of conventional observations is investigated through an Observing System Experiment (OSE) which uses the Met Office operational global and limited area (NAE) models. The experiment is run using observations from a two-month period during winter and another two-month period during summer. Global forecasts of up to 6-days are verified against the analyses from the run that uses all observations (COMB) and NAE forecasts up to 3-days are verified against radiosondes. The runs use several re-configurations of the Global Observing System (GOS) starting with a baseline network comprising all satellite data, observations from land stations that form the Global Climate Observing System (GCOS) plus observations from buoys (the BASE run). A second run uses the baseline system plus all remaining radiosonde temperature and wind observations (BPGTW); a third run the baseline system plus all remaining radiosonde wind observations only (BPNGW); a fourth run uses the baseline network plus observations from aircraft (BPAIR) and the final run uses all observations (COMB). Forecasts from the BASE run are compared with those from the other runs. The largest differences in forecast fields occur in the northern hemisphere where most conventional observations occur. Global model forecasts from the BASE run are degraded by 5-30% on average compared with forecasts from the COMB run and the probability of very poor forecasts increases. The differences in the objective verification scores between the COMB and BASE runs are statistically significant at the 90% level. On average, forecasts from the BPGTW, BPNGW and BPAIR runs are improvements over the BASE run forecasts in descending order. For NAE model forecasts verified against European radiosondes, the results are less consistent and vary between winter and summer although adding observations to the baseline network results in improved forecasts in all observation scenarios tested. The variability of the results over Europe is thought to be caused because the verification statistics are calculated over a relatively small geographical area. Global model forecasts give different verification statistics against European radiosondes compared with NAE model forecasts. These differences are thought to be partly caused by the differences in data assimilation schemes used: 3D-Var is used for NAE runs and 4D-Var for global runs. For both global and NAE model forecasts, none of runs tested (BASE, BPGTW, BPNGW, BPAIR) are able to completely recover the forecast skill obtained by using all observations in the GOS (COMB).

1. Introduction

The provision of observational data is an essential part of the NWP forecast system. These data are used to provide a four-dimensional picture of the atmosphere from which mathematical forecast models can estimate future atmospheric states. The world's observations, or Global Observation System (GOS), comprises a mixture of terrestrial-based and space-based observations and has evolved in response to national and international observation programs.

A group of twenty-two European countries have been collaborating on the re-design of the European component of the GOS for a number of years under the organisation of EUCOS[1]. These European countries have collectively maintained both a high-quality surface-based network throughout Europe and also funded a space-programme implemented by the European Space Agency (ESA[2]). Since funding for the space-based observing system

[1] See http://www.eumetnet.eu.org/
[2] See http://www.esa.int/esaCP/index.html

has already been allocated for the next few years, EUCOS asked the practical question “What additional terrestrial observations are required to compliment the space system?”

Previous studies have tended to examine the impact on forecast quality of denying whole observing systems from the full GOS either globally (e.g. Bouttier & Kelly 2001) or regionally (e.g. Zapotocny et al. 2005, Kelly et al. 2007). Such studies give valuable insight into the benefit of observations in the presence of all others, and also are a very good check on the overall performance of the data assimilation system used. Other studies have focussed on what happens if the GOS is enhanced by using existing observing systems at higher density (e.g. Cardinalli et al. 2003, Andersson et al. 2005) or the impact of extra observations taken during field campaigns (e.g. Peterson & Thorpe 2007). However, none of these types of study give a clear indication of how NWP performance might be affected if the GOS was entirely re-designed from a ‘baseline’ network comprising mainly of satellite data.

Bearing in mind these considerations, EUCOS proposed a systematic set of studies to investigate the incremental benefit of observations provided on top of an observing network consisting mainly of satellite data. It differs from other studies in that a new GOS is constructed from a baseline network of observations for which funding is already allocated so that the value of additional observations can be assessed by comparison with the impact of the baseline network (Andersson et al. 2004).

This paper reports on the study carried out by the Met Office in response to the EUCOS request for research into the correct mix of space and terrestrial observations. In section 2 the experimental set up and method of verification are described; in section 3 a description of the important results is given and the study is summarised and some conclusions drawn in section 4.

2. Description of the Experiment

An Observing System Experiment was performed in which observations were added to a ‘baseline’ network of observations comprising all available satellite observations plus a set of terrestrial observations. The conventional observations used in the baseline were those from stations in the Global Climate Observing System (GCOS) network. These stations have a climate monitoring role and are likely to remain part of a future GOS. Thus observations included in the baseline network were from the GCOS Upper Air Network (GUAN) radiosonde stations (Figure 2(a)) and from land surface stations that are part of the GCOS Surface Network (GSN, Figure 2(b)). Additionally, in order to ensure coverage over the oceans, observational data from drifting and moored buoys were used in the baseline.

The satellite data used in the baseline were those in operational use in the Met Office at the time the experiment was run, namely:

(i) radiance data from the HIRS/3, AMSU-A and AMSU-B radiometers on the NOAA-15 and NOAA-16 polar orbiting satellites
(ii) radiance data from the AIRS and AMSU-A radiometers on the AQUA polar orbiting satellite
(iii) atmospheric motion vectors (AMVs) from imagers on geostationary satellites Meteosat 5, 7 and GOES 9, 10, 12
(iv) surface wind speed over the sea from the SSM/I radiometer on the DMSP-F13 polar orbiting satellite
(v) surface vector wind over the sea from the scatterometer instruments on the QuikSCAT and ERS-2 polar orbiting satellites.

The exact use of the radiance data depends on the channel, height and geographical location of the sounding. Typically more channels are used over the sea where the modelling of surface emissions is more straightforward than over land.

Geostationary satellites derive AMV data from sequential images taken in different channels. The satellites are positioned over the equator and provide useful data in the range of about 60N to 60S outside which large zenith angles result in significant errors in the retrieved data. Within this latitude band, AMV data use is dependent on channel and level.

The scatterometer instruments measure microwave backscatter from small scale waves on the ocean surface from which surface vector wind can be derived.

Typical observation distribution maps are not presented here but can be seen on the Met Office web page (www.metoffice.co.uk under Research -> Weather Research -> NWP -> Observations)

Global and limited area runs were performed using the Met Office NWP system which was in operational use at the time the experiment was started. The global runs used an incremental 4D-Var data assimilation scheme with a single outer loop of a non-linear model providing linearization states and background fields for the assimilation of observations (Rawlins et al. 2007). At the time this experiment was run, 4D-Var had not been implemented in operations within the Met Office limited area model so these runs used the operational 3D-Var data assimilation scheme (Lorenc et al. 2000).

The Met Office runs a grid point forecast model which uses a non-hydrostatic form of the governing equations with extra terms to allow a complete representation of the Coriolis force. The semi-implicit time integration scheme has been designed with conservative properties that are necessary when undertaking long climate simulation integrations (Davies et al 2005). This ‘unified’ forecast model, with its current dynamics configuration, was introduced into operational use at the Met Office in August 2002.

The limited area used was the Met Office North Atlantic European (NAE) area shown in figure 1. The global experiment ran forecasts up to 6 days from 00 UTC and 12 UTC and the limited area experiment ran forecasts up to 3 days from the same times. The boundary conditions for the NAE run were taken from the corresponding global run.

In order to get a good sample of cases, the experiment used observations from a winter and summer period. The winter period was between 4th December 2004 and 27th January 2005 and the summer period between 5th July and 15th September 2005.

The runs performed used:

(i) all observations or the combined system (COMB)
(ii) observations in the baseline system (BASE)
(iii) observations in the baseline plus all aircraft data (BPAIR)
(iv) observations in the baseline plus all non-GUAN radiosonde temperature and wind observations (BPGTW)
(v) observations in the baseline plus all non-GUAN radiosonde wind observations (BPNGW).

Objective verification was carried out by comparing analyses and forecasts against surface and radiosonde observations, and the analyses from the COMB run which was assumed to produce the best analyses. To allow for spin-up, the first 10 days of each period were excluded when calculating the mean verification statistics.

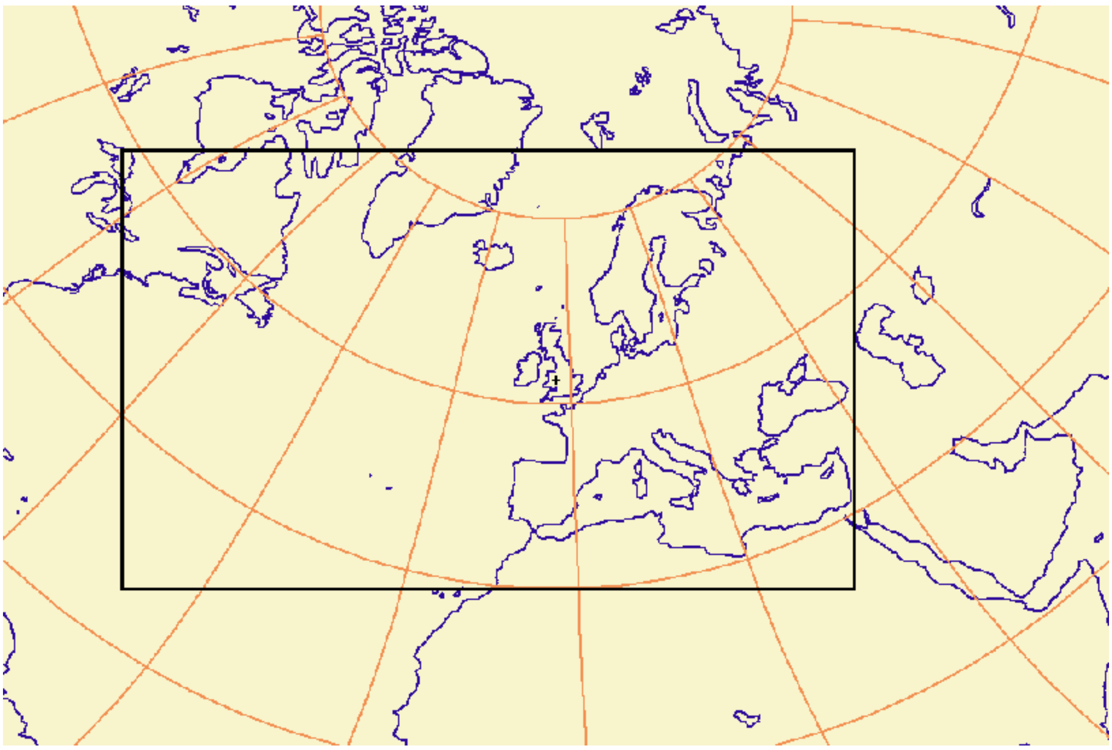
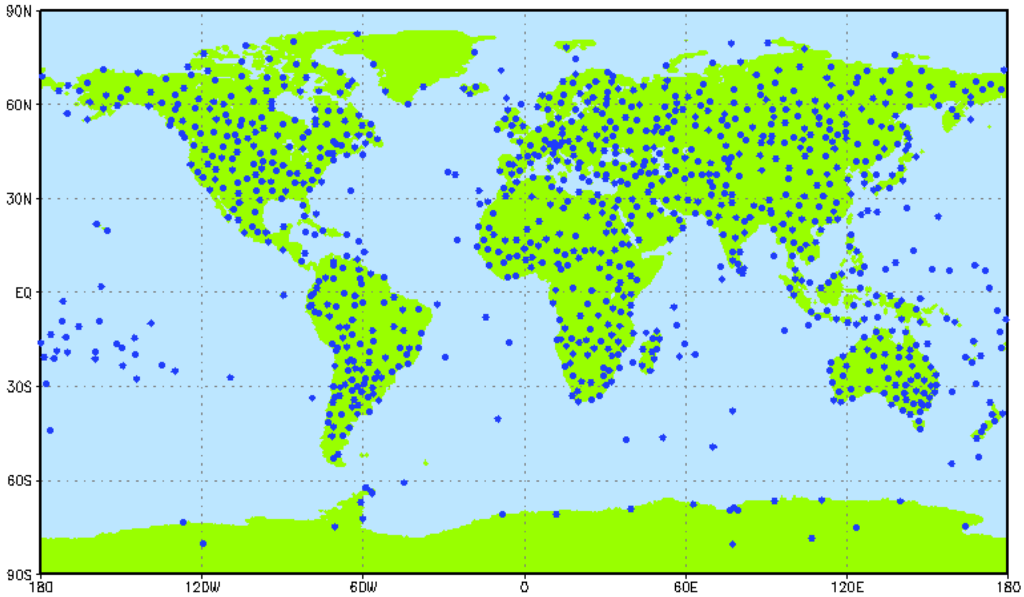
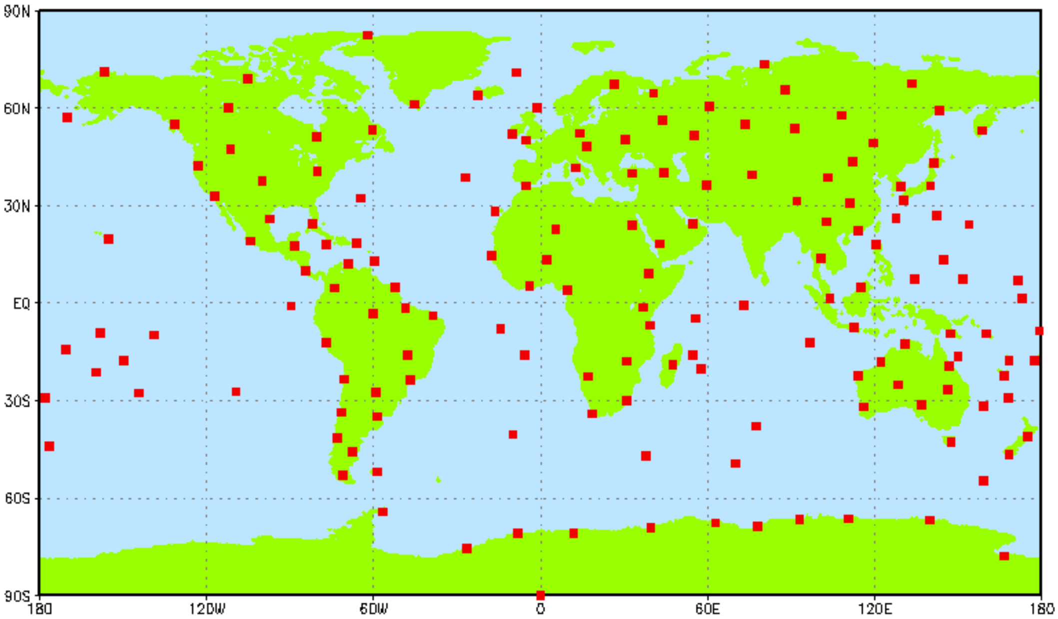


Figure 1. North Atlantic European (NAE) area.



(a)



(b)

Figure 2. Observations used in the baseline network: (a) upper air (b) surface. *(Courtesy Stefan Klink, EUCOS)*

3. Results

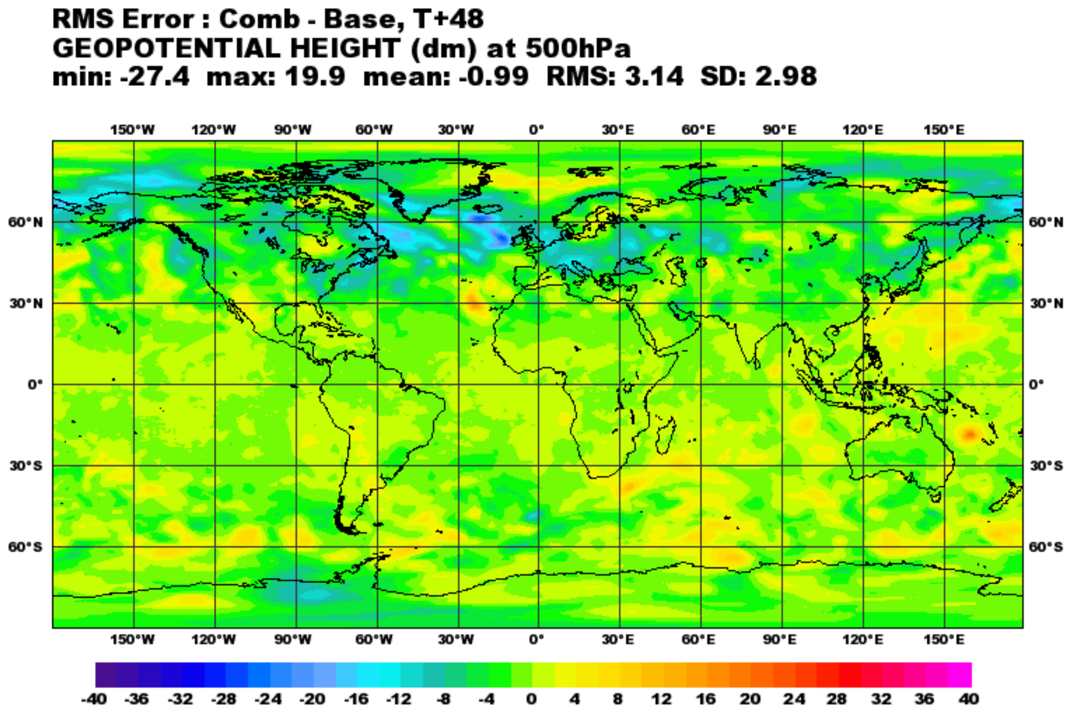
Many results are available from this experiment. A summary is presented in two sections that look at the effect of running with the baseline network only and the impact of adding in extra observations to the baseline.

The RMS error values are plotted in subsequent figures with error bars representing a 90% confidence interval. The errors bars were calculated using a Monte Carlo method (Robert and Casella 2004). The method used here is to generate a 100 element distribution of sample means by randomly selecting elements from each set of RMS forecast errors and calculating the mean of each sample. The sample size is equal to the number of forecasts available and so in general repeated elements will be used when calculating the sample means. Since 120 forecasts are available from the winter and summer periods, by the central limit theorem the distribution of sample means for the summer and winter period RMS errors is approximately Gaussian. Hence, if the sample means are sorted into ascending order, the 90% confidence interval is approximately the difference between the 5^{th} and 95^{th} elements.

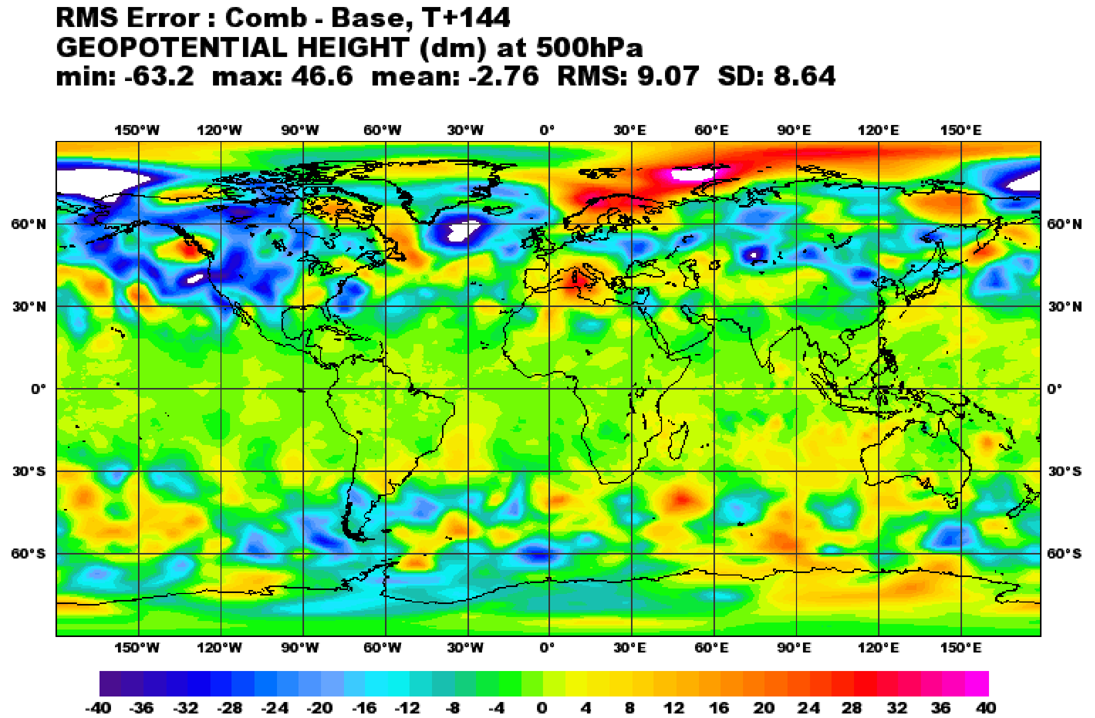
3.1 Baseline versus the full system

Figure 3 shows the distribution of the differences in the RMS errors of 500 hPa height forecasts between the COMB and BASE runs, averaged over the winter period. The largest differences in forecast quality due to observation use are likely to be in the short range forecasts. The differences in the T+48 forecasts are shown in figure 3(a). It can be seen that most of the positive impacts of the COMB run compared with the BASE run occur in the northern hemisphere where most of the conventional observations occur. The difference in conventional observation use between the COMB and BASE runs is less in the tropics and southern hemisphere than northern hemisphere so the difference in forecast quality is less marked. At the T+120 range (Figure 3(b)) the largest differences remain the northern hemisphere but can be either positive or negative as the influence of initial condition differences diminishes. Note that the differences in RMS errors grow with increasing forecast range in line with a growth in absolute RME errors. Because the differences are smaller outside the northern hemisphere, no further results from the tropics or southern hemisphere will be presented.

The difference in the forecasts in the northern hemisphere between the COMB and BASE runs is further examined in figure 4. Here the mean impact on the temperature and wind fields for different levels and forecast ranges are presented. To aid the comparison, the values on the plots are normalised with respect to the values from the BASE experiment and 90% error bars have been added. The differences are statistically significant at the 90% level if the error bars do not cross the zero line. The sign on the horizontal axes of the plots indicate whether or not the COMB run produces better forecasts than the BASE experiment. Negative values indicate improvements of the COMB run over the BASE.



(a)



(b)

Figure 3. Difference in the mean RMS errors versus analysis between the COMB and BASE runs for 500 hPa gepotential height for forecasts at (a) T+48 and (b) T+144. The mean is calculated over all winter period forecasts.

It can be seen from figure 4 that at almost all levels and forecast ranges the COMB run produces forecasts with better mean scores and that the improvements are statistically significant at the 90% level. Note that the influence of the difference in observation use declines with increasing forecast range.

Of particular interest to EUCOS is the impact of changing the observing network on short-range forecast quality over Europe for which the recommended verification method is against radiosonde observations. In figure 5 the mean impact on the short-range forecasts of temperature and wind is presented. Impacts on the higher resolution forecasts from the limited area NAE model are shown. Note that the results here are influenced by both the differences in observation use within the NAE area and differences in the boundary conditions used in the COMB and BASE run. The results are similar to those from the global model verified over the whole northern hemisphere in that all runs show a statistically significant improvement over the BASE at all forecast ranges except at high levels. However, the difference in the quality of forecasts at different forecast ranges is not always statistically significant possibly because the statistics are averaged over a small geographical area and against observations. When verifying against observations the sample size is smaller than when verifying against analyses (or all model grid points) for a given geographical area, particularly at high levels where there tend to be fewer observations. Also the sample size reduces as the size of the verification area reduces. Thus ideally a longer trial should be run in order to obtain statistically significant results when verifying over a limited area.

Results from OSEs depend not only on the observation usage but also on the data assimilation scheme. In this study the global runs used 4D-Var whereas the NAE runs used 3D-Var. To investigate what effect the use of different data assimilation schemes has on the results, the normalised RMS errors for global model forecasts, verified against the same set of European radiosondes that was used for verifying NAE model forecasts, have been calculated and are shown in figure 6. Comparing figures 5 and 6 it can be seen that the errors are not identical but show similar characteristics. The global forecasts from the COMB run show a statistically significant (at the 90% level) improvement over the BASE run forecasts at almost all forecast ranges and levels as do forecasts from the NAE model. Another similarity is that the difference in global model forecast quality between forecasts of different ranges is not statistically significant in general. It appears from these results that the use of 3D-Var and 4D-Var gives similar results when verifying forecasts against European radiosondes.

For both northern hemisphere and European area verification the deterioration in forecast quality in the BASE run compared with the COMB run is approximately in the range 5-30% depending on forecast period and level. For both regions, the impact tends to be greater at short forecast ranges than at longer ranges indicating that the effect on the initial conditions is less important as forecast period increases. The results from the winter and summer periods are similar.

The effect on individual forecasts of reducing the GOS to the baseline level can be seen by looking at time series of forecast scores. In figure 7 the anomaly correlation coefficient values of the 6-day forecast of 500 hPa height averaged over the northern hemisphere from the COMB and BASE runs is shown. It can be seen that there are several forecasts between the end of January and early February when the BASE run produced forecasts with anomaly correlation coefficient values below 0.6 whereas the COMB run forecasts scored much higher. Forecasts with anomaly coefficient scores below 0.6 are regarded as being of little value. Thus reducing the GOS to the baseline level is likely to increase the probability of very poor forecasts or 'busts'.

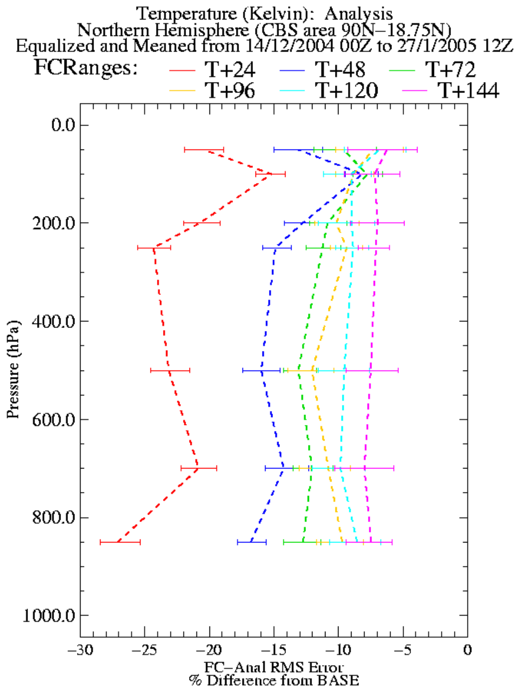
3.2 Impact of adding observations to the baseline

Given that using the baseline network of observations only would reduce overall forecast quality by a statistically significant amount, the next question to address concerns what is the minimum network of observations that needs to be added to the baseline in order to recover all or most of the skill of the full system. Results from the BPAIR, BPGTW and BPNGW runs help to answer this question. In the interests of brevity and to reduce clutter on the diagrams only the impact on the 500 hPa level is presented here – the results from other levels are similar.

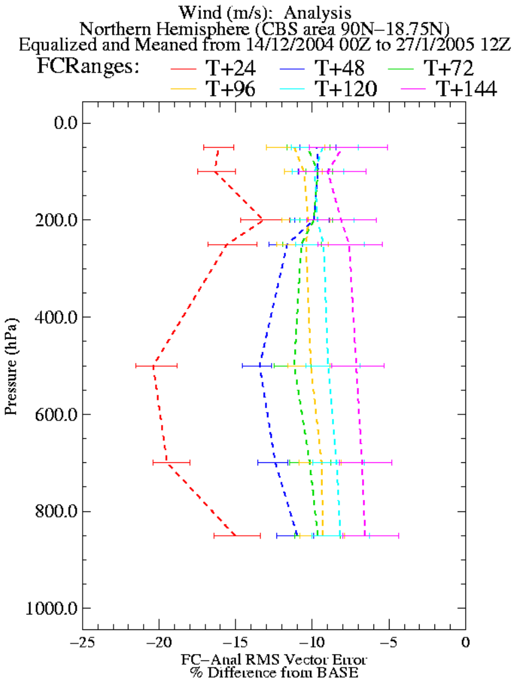
In figure 8 the impact of adding in extra observations to the baseline network can be seen in terms of impact on forecasts from the global model verified over the northern hemisphere. As in other figures, the normalised difference scores relative to the values from the BASE run are presented. The error bars show statistical significance at the 90% level.

For most forecast ranges adding in either radiosonde temperature and wind observations (BPGTW), radiosonde wind observations only (BPNGW) or aircraft temperature and wind observations (BPAIR) produces a statistically significant improvement in forecast quality compared with the BASE run. Generally speaking, adding in radiosonde temperature and wind observations produces the most benefit, followed by radiosonde winds only then aircraft temperature and wind observations. However, at forecast periods up to at least 72 hours there is a statistically significant difference between the scores for the COMB and BPGTW runs indicating that forecast quality cannot be recovered by adding in radiosonde temperature and wind observations only to the baseline network. It can be seen that the impact on winter and summer forecasts are similar.

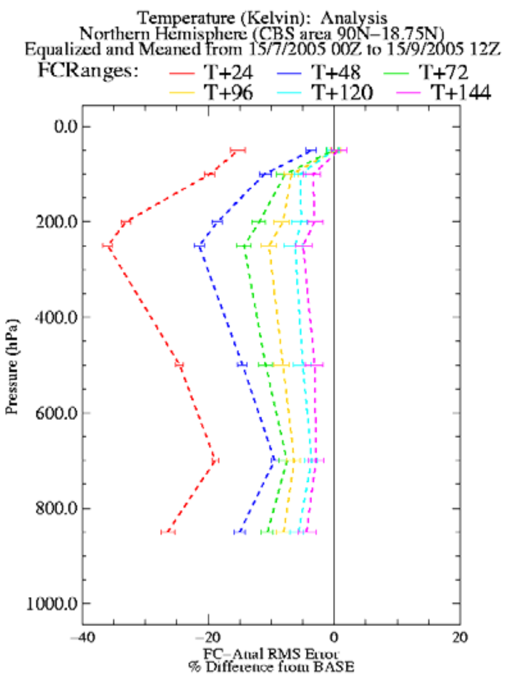
The relatively large positive impact from radiosonde data compared with aircraft data may be due to the fact that radiosonde observations provide profile information over a wide geographical area and at regular intervals in time, usually every twelve hours. In contrast, aircraft observations are not distributed evenly in either space or time. Most observations are concentrated over North America and Europe and taken at times determined by flying



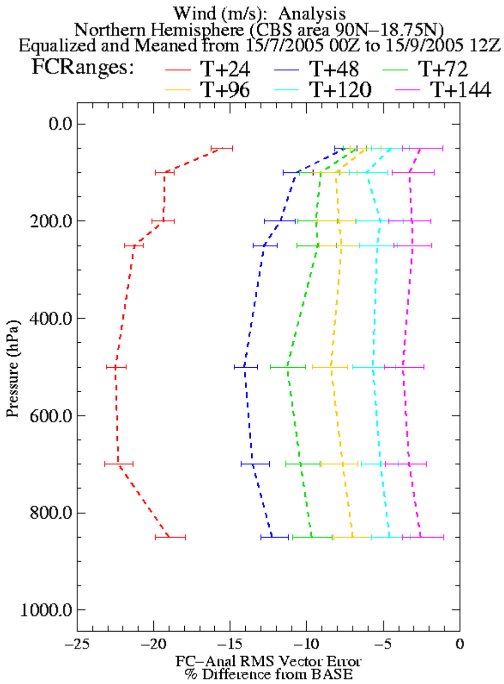
(a)



(b)



(c)



(d)

Figure 4. Differences in normalised RMS forecast error with pressure between the BASE and COMB runs for the global model. Errors calculated against the COMB analysis, normalised by BASE run values and averaged over the northern hemisphere. (a) temperature for winter; (b) vector wind for winter; (c) temperature for summer; (d) vector wind for summer. Error bars give statistical significance at the 90% level.

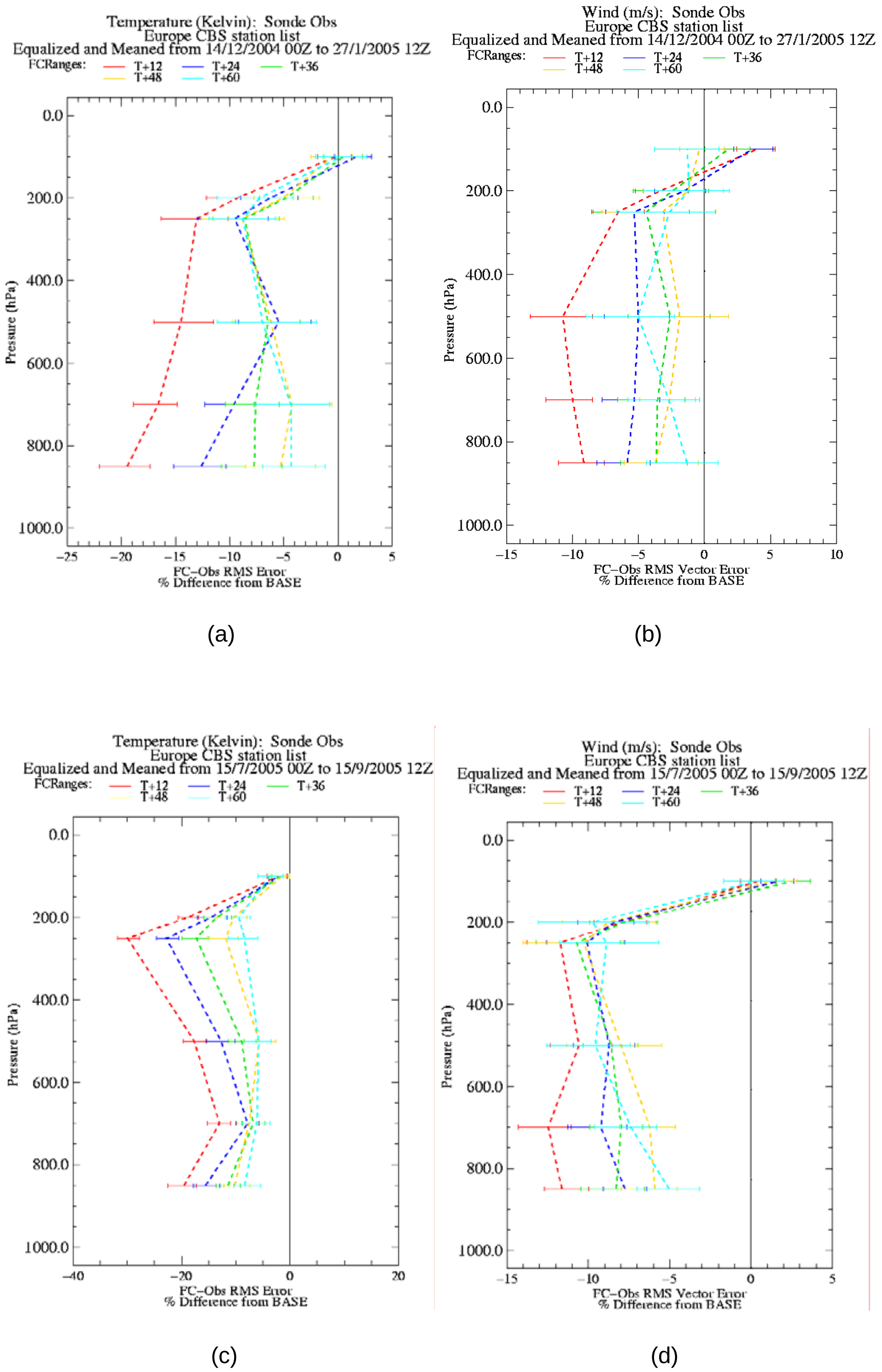


Figure 5. Differences in normalised RMS forecast error with pressure between the BASE and COMB runs for the NAE model. Errors calculated against European radiosonde observations and normalised by BASE run values. (a) temperature for winter; (b) vector wind for winter; (c) temperature for summer; (d) vector wind for summer. Error bars give statistical significance at the 90% level.

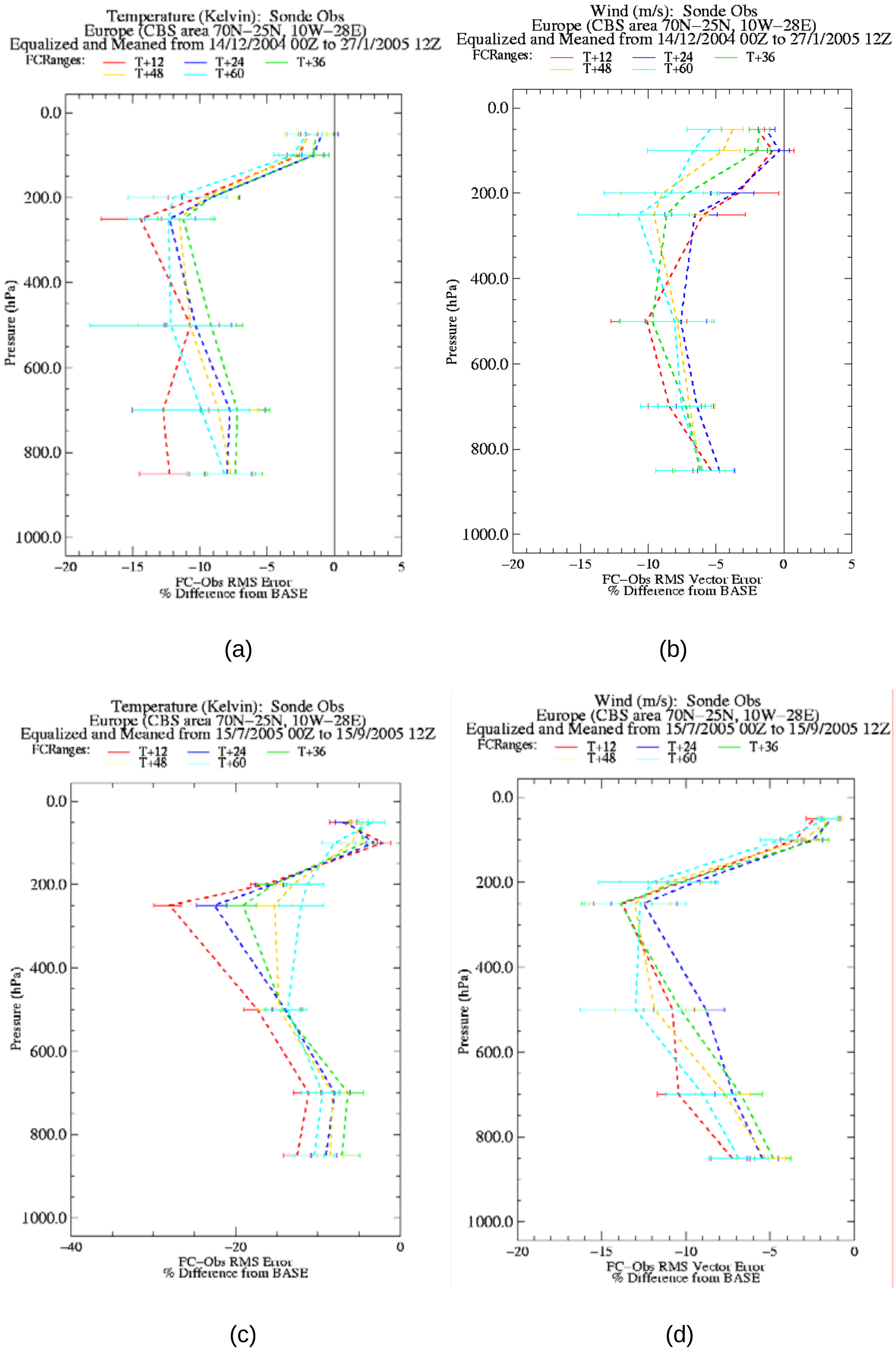


Figure 6. As figure 5, but showing errors for the global model forecasts.

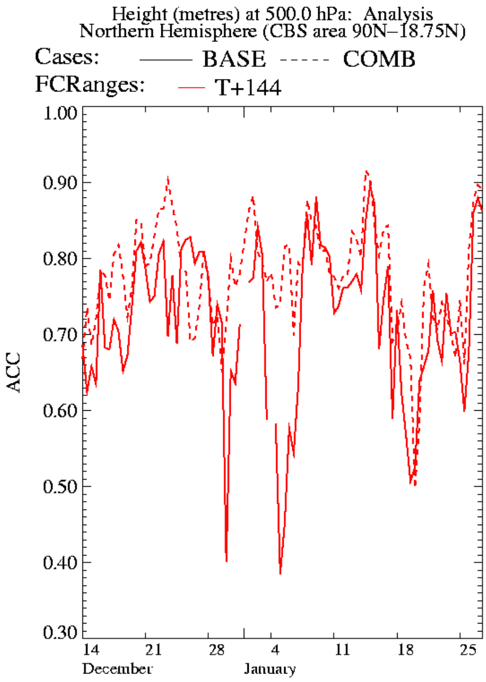


Figure 7. Time series of anomaly correlation coefficient of 6-day forecasts of 500hPa geopotential height averaged over the northern hemisphere for the winter period comparing the results from the COMB and BASE runs. Values calculated using the COMB run analysis.

schedules and tend not to be evenly distributed. Additionally, although ascent and descent profiles are available at airports, much aircraft data measures flight level weather parameters only. Thus a larger impact from radiosonde data than aircraft data is to be expected, and has been noted in many studies (e.g. Bouttier and Kelly 2001).

In figure 9 the impact on NAE model forecasts verified against European radiosondes is shown. Here winter and summer results are different. Over the winter period [figures 8(a) and 8(b)], all runs show a statistically significant improvement over the BASE run at most forecast ranges but, apart from at T+12, the difference in forecast quality between the COMB, BPAIR, BPGTW and BPNGW is not consistent or statistically significant. Over the summer period [figures 8(c) and 8(d)], all runs show a statistically significant improvement over the BASE run at all forecast ranges. The differences between the runs are more consistent and for some forecast ranges, statistically significant. On average, adding in aircraft temperature and wind observations (BPAIR) produces the most benefit followed by radiosonde temperature and wind observations (BPGTW) then radiosonde winds only (BPNGW).

In figure 10 verification of global model forecasts against European radiosondes is presented and so the figure can be compared directly with figure 9 which shows the verification of NAE model forecasts over the same area. From figure 10, it can be seen that all runs produce a statistically significant improvement over BASE for all winter period forecasts and most summer period forecasts. As for the NAE forecasts (figure 9), the difference

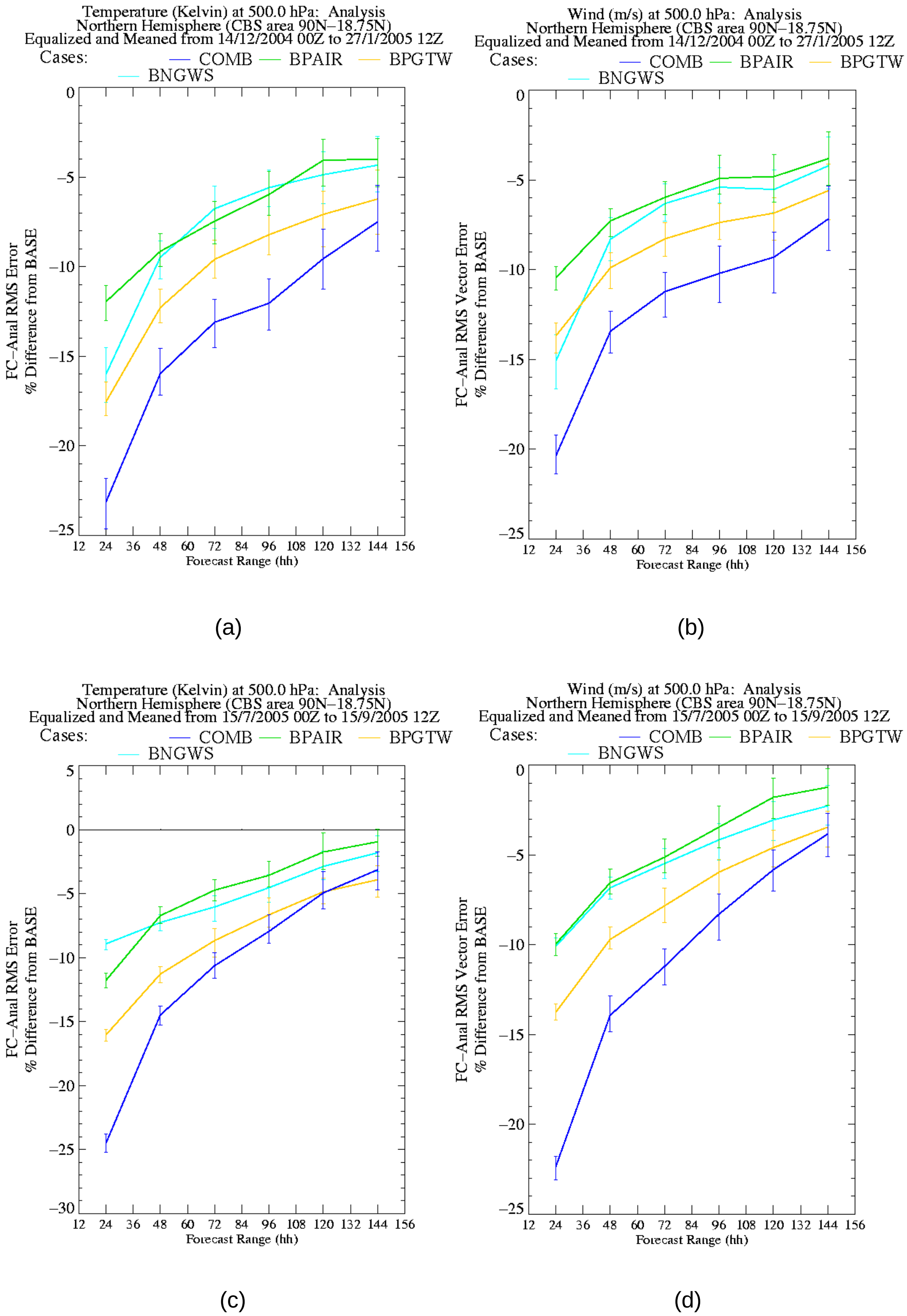


Figure 8. Differences in normalised RMS forecast error at 500hPa between the BASE and other runs for the global model. Errors calculated against the COMB analysis, normalised by BASE run values and averaged over the northern hemisphere. (a) temperature for winter; (b) vector wind for winter; (c) temperature for summer; (d) vector wind for summer. Error bars give statistical significance at the 90% level.

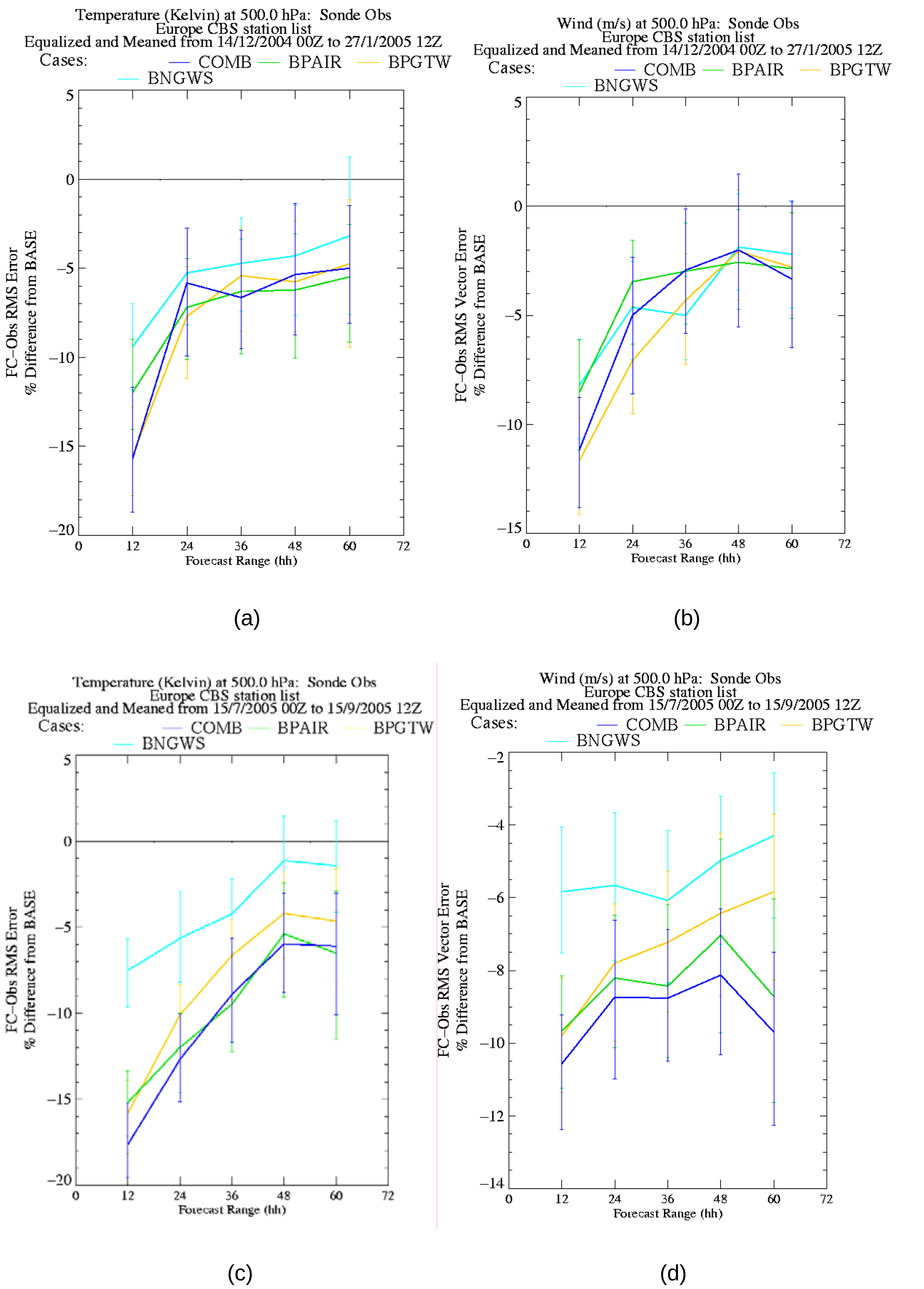
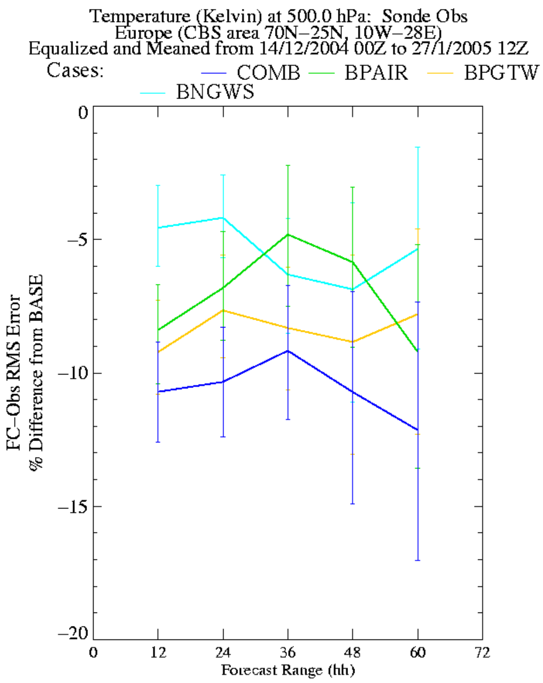
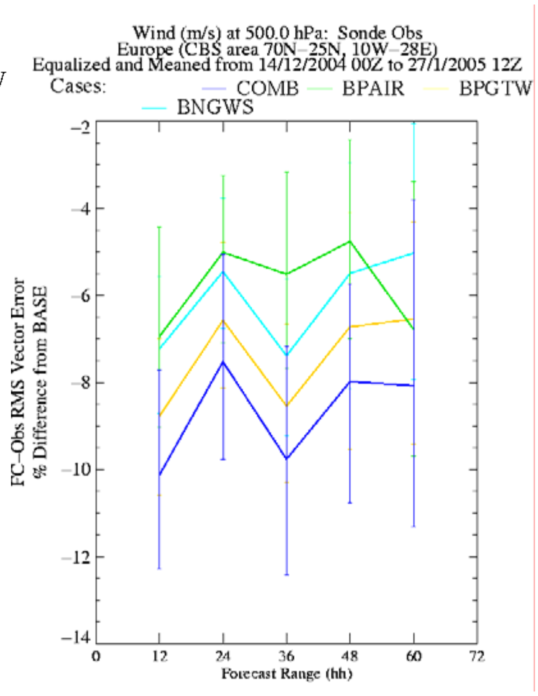


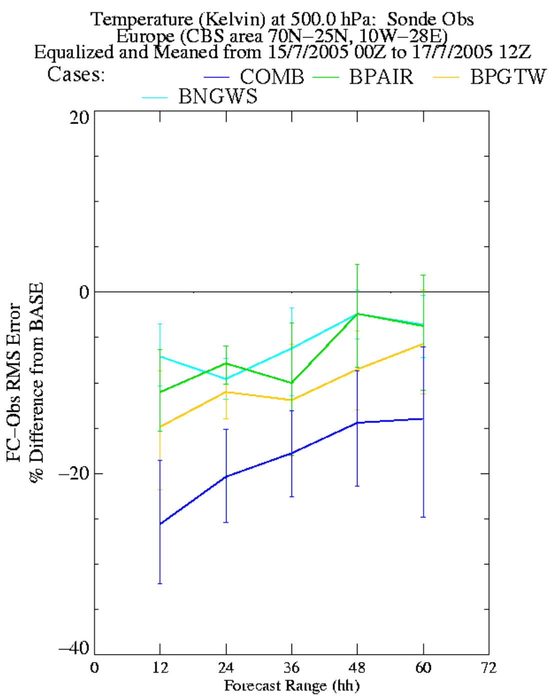
Figure 9. Differences in normalised RMS forecast error at 500hPa between the BASE and other runs for the NAE model. Errors calculated against European radiosonde observations, normalised by BASE run values. (a) temperature for winter; (b) vector wind for winter; (c) temperature for summer; (d) vector wind for summer. Error bars give statistical significance at the 90% level.



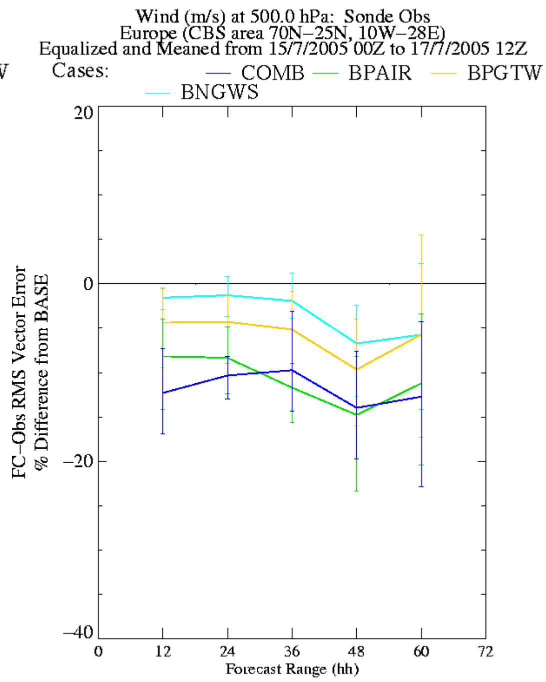
(a)



(b)



(c)



(d)

Figure 10. As figure 9, but showing errors for the global model forecasts.

between the runs is variable and not always statistically significant. Moreover, there is some inconsistency between the results for the two models. For example, the benefit over Europe of aircraft data is less significant for global model forecasts (Figure 10) than some NAE model forecasts (Figure 9(c), (d)).

As noted in 3.1, more variability in the results can be expected when verifying over a ‘small’ region such as Europe. The inconsistency of the results between figures 9 and 10 may be partly explained by the different data assimilation schemes used: 3D-Var in the NAE and 4D-Var in the global model. Others factors that may cause differences in the results include differences in model resolution, the effect of flow across the boundaries in the NAE model, and differences in observational data use. The NAE model uses the MOPS 3D-cloud analysis that is not used by the global model runs.

4. Summary and Conclusions

An Observing System Experiment is carried out to investigate the impact of conventional observations on NWP forecasts. The experiment is run over two-month periods in both winter and summer. The impact is measured by adding in conventional observations to a baseline network comprising all satellite data plus data from the land stations from the GCOS network and all buoy data. The observations added to baseline are radiosonde temperature and wind observations, radiosonde wind observations only, and aircraft temperature and wind observations. The experiment uses the Met Office global and limited area (NAE) models run in a configuration that was operational at the time of the experiment. For each experimental run the boundary conditions for the NAE model are taken from the corresponding global run. Global model forecasts are verified against the analyses from the ‘all data’ run and the NAE forecasts against European radiosondes.

Using the baseline network only results in degradation of forecast quality in the northern hemisphere and Europe of between 5 and 30% depending on level and time range compared to the run using all observations. Most of the differences in the objective verification scores are statistically significant at the 90% level. In addition to a significant reduction in the mean skill, running forecasts using the baseline network of observations only is likely to increase the probability of very poor forecasts. Much smaller differences in the quality of forecasts produced from the baseline and full system are seen in the tropics and southern hemisphere where there are less conventional observations.

For the global model forecasts, the most effective way of recovering forecast skill is to add radiosonde wind and temperature observations to the baseline. Next most effective method is to add in radiosonde wind observations only and the least effective method is to add in aircraft temperature and wind measurements. Adding in radiosonde data to the baseline is thought to be the most effective strategy since these observations add profile information over a wide geographical area and frequently in time.

For NAE model forecasts verified against European radiosondes the results are more variable most probably due to the smaller geographical area (sample size) used when calculating the statistics. For the winter period, there is no statistically significant difference in the objective verification scores for the observation use scenarios that add observations to the baseline. For the summer period, adding in aircraft observations appears to be the best strategy for improving on the skill of forecasts using the baseline network only and gives a similar result to using all observations.

For both global and NAE models and for winter and summer periods, none of the observation networks tested completely recover the forecast skill obtained when using all observations in the GOS.

The results obtained by an OSE such as this are influenced by the NWP system used as well as observation usage. In particular it should be noted that 3D-Var was used in the NAE model and thus the results for the NAE model might have been different if 4D-Var was used. For example, a notable advantage of 4D-Var over 3D-Var is its greater ability at using information from sparse observation networks (Kelly et al 2007). A comparison of verification results from the global and NAE model forecasts verified against European radiosondes indicates that the two sets of runs do not always give similar results over Europe.

The general conclusion from this study is that the conventional observation network still adds value to the quality of NWP forecasts and any re-design of it must include a comprehensive coverage of conventional profile observations.

Acknowledgements

This study was instigated and partly funded by EUCOS and the author is grateful for permission from EUCOS to publish the results.

References

Andersson E, Dumelow R, Huang H, Thepaut J-N, Simmons A 2004. Space-Terrestrial Link – Outline study proposal for consideration by EUCOS Management (available from EUCOS Secretariat).

Andersson E, Cardinali C, Truscott B, Hovberg T 2005. High-frequency AMDAR data - a European aircraft data collection trial and impact assessment. ECMWF Technical Memorandum No. 457.

Bouttier F, Kelly G 2001. Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Q. J. R. Meteorol. Soc.* **127**: 1469-1488.

Cardinalli C, Isaksen L, Andersson E 2003. Use and Impact of Automated Aircraft Data in a Global 4D VAR Data Assimilation System. *Mon. Weather Rev*. **131**: 1865-1877.

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA Wood N 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759-1782.

Kelly G, Thépaut J-N, Buizza R, Cardinali C 2007. The value of observations. I: Data denial experiments for the Atlantic and the Pacific. *Q. J. R. Meteorol. Soc.* **133**: 1803-1815.

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW 2000. The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. *Q. J. R. Meteorol. Soc.* **126**: 2991-3012.

Petersen GN, Thorpe AJ 2007. The impact on weather forecasts of targeted observations during A-TReC. *Q. J. R. Meteorol. Soc.* **133**: 417-431.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne, TJ 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347-362.

Robert CP, Casella G 2004. "Monte Carlo Statistical Methods" (second edition). *New York: Springer-Verlag*, ISBN 0-387-21239-6.

Zapotocny TH, Menzel WP, Jung JA, Nelson JP 2005. A four-season impact study of rawinsonde, GOES, and POES data in the Eta Data Assimilation System. Part I. The total contribution. Part II. Contribution of the components. *Weather and Forecasting*. **20**: 161-198.