Tellus (2004), 56A, 456–471
Printed in UK. All rights reserved

TELLUS

Sensitivity of the ERA40 reanalysis to the observing system: determination of the global atmospheric circulation from reduced observations

By L. BENGTSSON[1,2], K. I. HODGES[1,*] and S. HAGEMANN[2], [1]*Environmental Systems Science Centre, University of Reading, PO Box 243, Early Gate, Reading RG6 6AL, UK;* [2]*Max Planck Institute for Meteorology, Hamburg, Germany*

(Manuscript received 10 February 2004; in final form 17 May 2004)

ABSTRACT

The impact of selected observing systems on the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-yr reanalysis (ERA40) is explored by mimicking observational networks of the past. This is accomplished by systematically removing observations from the present observational data base used by ERA40. The observing systems considered are a surface-based system typical of the period prior to 1945/50, obtained by only retaining the surface observations, a terrestrial-based system typical of the period 1950–1979, obtained by removing all space-based observations, and finally a space-based system, obtained by removing all terrestrial observations except those for surface pressure. Experiments using these different observing systems have been limited to seasonal periods selected from the last 10 yr of ERA40. The results show that the surface-based system has severe limitations in reconstructing the atmospheric state of the upper troposphere and stratosphere. The terrestrial system has major limitations in generating the circulation of the Southern Hemisphere with considerable errors in the position and intensity of individual weather systems. The space-based system is able to analyse the larger-scale aspects of the global atmosphere almost as well as the present observing system but performs less well in analysing the smaller-scale aspects as represented by the vorticity field. Here, terrestrial data such as radiosondes and aircraft observations are of paramount importance. The terrestrial system in the form of a limited number of radiosondes in the tropics is also required to analyse the quasi-biennial oscillation phenomenon in a proper way. The results also show the dominance of the satellite observing system in the Southern Hemisphere. These results all indicate that care is required in using current reanalyses in climate studies due to the large inhomogeneity of the available observations, in particular in time.

1. Introduction

Understanding the atmospheric circulation and its role in determining the climate requires long-term consistent four-dimensional data sets preferably having a global coverage. Such data sets can be provided by reanalysis projects. Reanalyses of the Earth's atmosphere have contributed significantly towards an enhanced understanding of the general circulation of the atmosphere and provide an important contribution towards the long-term systematic study of the Earth's atmosphere (Bengtsson and Shukla, 1988). However, the large changes which have taken place in the global observing system during the last 50 yr need to be better quantified, because the amount and quality of available observations have a considerable effect on the resulting analyses.

From the middle of the nineteenth century, the global observing system has evolved from a surface-based system (prior to 1945/50), to a system essentially based on radiosondes and surface data (1950–1978) and finally to the present integrated system where space-based observing platforms play an increasingly dominant role. This is a very crude separation as the global observing system in reality changes from year to year. The surface-based system consisting of synop and ship data has evolved over some 200 yr from a small number of stations mainly in Europe to an essentially global network towards the first part of the twentieth century, albeit with significant data voids even now. The radiosonde network was rapidly built up after World War II, but has undergone numerous changes over the years – new stations have been set up and others have disappeared. At the same time, technology has evolved and present radiosondes are more accurate, reliable and more standardized than previously. A worrying trend is the reduced number of radiosonde observations from countries in the developing world and from

---

[*] Corresponding author.
e-mail: kih@mail.nerc-essc.ac.uk

the 1990s onward also from countries of the former USSR. Nevertheless, the radiosonde-based network has been the backbone of the atmospheric observational system for more than 50 yr and still is of key importance even though the use of satellite data has become ever more dominant.

The currently available reanalyses include the European Centre for Medium-Range Weather Forecasts (ECMWF) 15-yr reanalysis (ERA15; Gibson et al., 1997), the National Center for Environmental Prediction (NCEP) reanalyses – NCAR (Kalnay et al., 1996) and DOE (Kanamitsu et al., 1999) – and the NASA Goddard Earth Observing System 1 (GEOS1) reanalysis (Schubert et al., 1993). These have recently been joined by the ECMWF 40-yr reanalysis (ERA40; Simmons and Gibson, 2000). A reanalysis uses a fixed, modern numerical weather prediction (NWP) data assimilation system operating in successive mode to produce a physically consistent space–time interpolation of the historical observational data. Modern data assimilation systems are able to reduce the error by more than half (Bengtsson, 1999) in the initial state of the atmosphere. The fact that the analysis system is fixed excludes the type of secular biases often introduced with changes to the model or assimilation system in an operational system. However, because the observing system is not fixed, biases can be introduced into climatological quantities due to changes in the observing system. Although the reanalyses already mentioned use modern NWP analysis systems, this does not mean that the reanalysis products are identical; the use of different models and assimilation methods often leaves an imprint on the reanalysis products typical of the methodologies used (Hodges et al., 2003). This is particularly the case in data sparse regions where the observations provide less constraint on the GCM.

Reanalyses have now been provided for periods of some 50 yr (NCEP/NCAR; Kalnay et al., 1996) and plans have been put forward to extend reanalyses to the full twentieth century or longer, based on ensemble data assimilation methods for example (Whitaker et al., 2004). Whether such an undertaking is sensible or not is crucially dependent on the extent to which a limited set of atmospheric observations (before 1945 surface data only) is able to constrain the circulation of the atmosphere throughout the full atmospheric column. In this paper, reduced observing systems are used to compare and contrast the different forms of observing systems using the ERA40 system. This of course makes the results dependent on the ERA40 system, but it is believed the results presented here give a good indication of the likely value of reanalyses with a reduced observing system.

There are several indications that the global observing system has become more accurate, at least as indicated by NWP studies (Bengtsson, 1999; Simmons and Hollingsworth, 2002), particularly so for the Southern Hemisphere (SH). However, in assessing the accuracy and thus the information content in a given observing system the effects of models and data assimilation need to be better identified. For this reason, in this paper we report an integrative study by combining a state-of-the-art data-assimilation system with selected reduced observing systems, chosen to mimic observing systems of the past.

Three specific sets of observing systems have been explored by suitable changes to the basic ERA40 observational data base. For practical reasons, it was necessary to introduce a number of simplifications. The systems investigated include a surface-based system, which is selected to represent the typical observing system for the first half of the twentieth century. As a proxy for such a system, a typical surface synoptic network of the recent decade has been used, although this will necessarily be of a higher data density than a typical observing network from earlier in the twentieth century. A system, which consists of the surface and radiosonde network, is also explored and is assumed to broadly represent the observing system in the period 1950–1978. This will be called the terrestrial-based system. Finally, we also consider a system which is an assumed satellite-based system, consisting of the satellite data available in the recent decade together with surface pressure observations to provide a constraint at the surface. This is called the space-based system. Such a system could be seen as a fully global automated system as the surface observations could easily be provided automatically and the data interrogated via satellite systems. Whether such a system may exist in some form or another in the future is an interesting possibility but the purpose here is to specifically assess the information content in the satellite-based observations. It should be noted that the satellite and terrestrial systems are not completely independent of one another, not only via the use of the surface pressure in the satellite system to constrain the surface but also via the satellite radiance bias corrections. These are usually determined by examining the differences between the observed and computed brightness temperatures, of which the latter will depend on the assimilated terrestrial observations.

This paper addresses the following science issues related to the observing system and its impact on reanalyses. If the reference ERA40 system is considered as a substitute for the truth, how much variance reduction is obtained by the surface-based network, the terrestrial network and the space-based network, respectively? How large is the variance reduction for different variables, levels, geographical regions and for different time and space scales including selected phenomena? The key issue here is to determine to what degree a reduced network is able to constrain the time evolution of the atmospheric circulation. Is it at all feasible for a surface-based only system, together with the surface boundary conditions, to constrain the atmospheric circulation throughout the depth of the troposphere and lower stratosphere and at what scales? An application of particular interest is to use the reanalysis data in climate change studies. Is this at all feasible when the observing systems undergo successive changes? The present study contributes towards quantifying the effect of observing system changes over a wide range of atmospheric parameters. In another study (Bengtsson et al., 2004b), the results of this work are used to improve the estimates of long-term climate change by correcting some of the biases introduced by observing system changes.

The paper is organized in the following way. In Section 2 we explain the experimental design. In Section 3 we discuss the relative importance of the different observing systems and its information content. In Section 4 we explore the representation of extratropical cyclones, and in Section 5 the results of the 1990/91 period are compared with those from the later period in 2000/01. Finally, in Section 6 we summarize the results with further discussion.

2. Analysis system and experimental design

The experiments carried out for this study are based on the system used to produce the ERA40 reanalysis (Simmons and Gibson, 2000). This is a comprehensive data-assimilation system using a three-dimensional variational (3D Var) analysis and a semi-Lagrangian spectral-transform model. The spectral resolution has a triangular truncation of 159 (T159) and there are 60 levels in the vertical. A comprehensive system of physical parametrizations is also used, including the convective scheme of Tiedtke (1989), the rapid radiation transfer model (Mlawer et al., 1997) and a parametrization for subgrid-scale orographic processes (Lott and Miller, 1997). As well as the terrestrial observations of temperature, pressure, wind and humidity in the observational database, extensive use is made of satellite data either as derived products such as cloud motion winds or by the direct assimilation of the observed radiances (temperature and water vapour), for example from the TIROS Operational Vertical Sounder (TOVS) [High-Resolution Infrared Radiation Sounder (HIRS), Microwave Sounding Unit (MSU), Stratospheric Sounding Unit (SSU)].

In all the experiments that have been performed, the humidity observations have been excluded, as in the experiments conducted for the study described in Bengtsson et al. (2004a) (humidity observations have very little impact on the analyses). This entails excluding all surface and upper air observations of humidity as well as the HIRS channels 11 and 12 and the Special Sensor Microwave/Imager (SSM/I) precipitable water content. This means there are no direct humidity increments in the experiments; however, indirect increments are possible in particular via the temperature channels of HIRS and the super-saturation check. Also excluded in all the experiments are the PAOBS observations which are only available in the SH (these are given low weight in the ERA40 system in any case). The control experiment against which the other experiments are compared consists of the full observing system, excluding all the humidity observations as in Bengtsson et al. (2004a). Because of the computational and data handling volumes involved, the study is restricted to the seasons: December–February (DJF) 1990/1991 and DJF 2000/2001. Using two winter periods 10 yr apart allows the effects of recent changes in the observing systems to be explored.

In the surface-based system, only observations of pressure, temperature and wind velocity are used from land-based stations, ships and moored buoys. In the terrestrial system, the surface observations are used with the addition of upper air observations from radiosondes, aircraft and profilers. Finally, in the satellite-only system the surface is constrained by the surface pressure only and all other observations come from satellite systems alone, including HIRS, MSU, SSU and cloud motion winds derived from geostationary satellites. An experiment is also run as an extended forecast, where no observations are assimilated for the 1990/91 winter but using the same boundary forcing as in the original experiment, to provide a comparison with the assimilating model alone.

The choice of these observing systems for the experiments is driven by the historical context discussed in the introduction as well as studies based on the full ERA40 data set, which indicate significant jumps in climatological parameters with the introduction of the satellite observing systems (Bengtsson et al., 2004b). The contrast between the pre-satellite (1958–78) period and satellite (1979–2002) period is clearly seen in synoptic eddy statistics as determined by objective feature tracking using the full ERA40 data set. An example of this is shown in Fig. 1. This shows the mean intensity of mid-latitude cyclones for the two periods and for the DJF season based on the 500-hPa relative vorticity field using the same methodology as in Hoskins and Hodges (2002). In the NH (Figs. 1a and b) differences in the mean intensity of cyclones are seen to be relatively small. The differences in the NH Atlantic are negligible, larger differences are seen in the NH Pacific, indicating a possible sensitivity to the introduction of the satellite data. In the SH (Figs. 1c and d) the mean intensity of cyclones also shows larger differences highlighting the significant influence of the satellite observations, introduced after 1979. These results are also seen for other fields, for example the mean sea level pressure (MSLP). This disparity between the pre-satellite and satellite periods is also seen in global climate measures (Bengtsson et al., 2004b) and provides the motivation in part for this study.

3. Information content of the experiments

An important question is whether the differences between the experiments are randomly distributed or whether they also include a systematic component. In this section, the general performance of the different experiments when compared against the control experiment for the 1990/91 DJF period will be presented. This will highlight the relative importance of the different observing systems in different regions and vertically. This will take the form of some selected maps including the difference in the time-averaged geopotential ($\overline{Z}_{\mathrm{dif}}$) at 500- and 50-hPa, i.e. tropospheric and stratospheric levels; the root-mean-square error (RMSE) of the geopotential (Z) and relative vorticity (ξ), standardized by the standard deviation (StD) of the control experiment, at the same levels. These will also be summarized as line plots of $\overline{Z}_{\mathrm{dif}}$ and the RMSE against height, averaged over the NH, SH and the tropics (TR). Other quantities related to the general circulation have also been explored in relation to the information content of

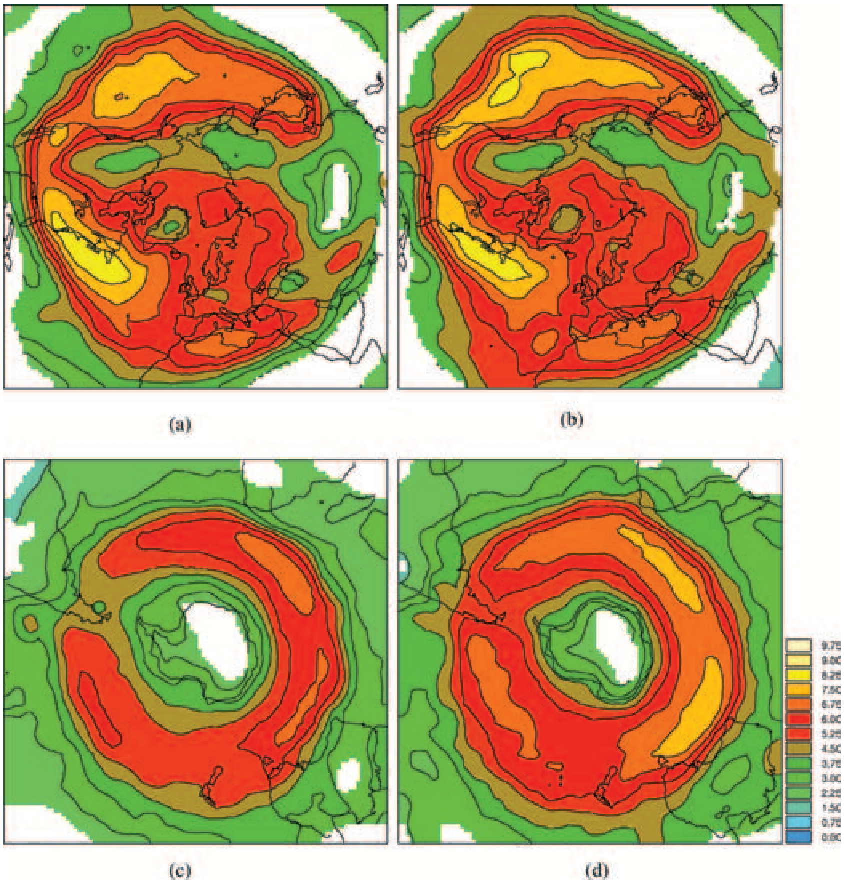


Fig 1. NH and SH cyclone mean intensity for the DJF 1958–1978 and 1979–2002 periods as determined from the 500-hPa relative vorticity field from ERA40: (a) NH, DJF 1958–1978; (b) NH, DJF 1979–2002; (c) SH, DJF 1958–1978; (d) SH, DJF 1979–2002. Units are × 10^{-5} s^{-1}.

the different experiments, such as eddy kinetic energy, horizontal eddy momentum flux and horizontal eddy heat and moisture transport. These are not presented here as they only confirm the results from the core diagnostics in this section and the following section.

3.1. *Mean geopotential*

In Fig. 2, $\overline{Z}_{\mathrm{dif}}$ at the 500-hPa level for the different experiments is compared with the control. Clearly the model-only experiment (Fig. 2d) shows the greatest errors with large differences in the NH and SH with some absolute values greater than 100 m in the NH. In particular, a region of the northern Pacific shows a large systematic bias compared with the control. Differences are smaller but still large in the SH, probably associated with the period being a SH summer. This situation is reversed when the surface-only system is considered (Fig. 2c) with the largest differences now occurring in the SH (∼100 m) with differences in the NH now typically 50 m, reflecting the influence of the denser surface observations in the NH providing some constraint even at this mid-tropospheric level. For the two other experiments, the terrestrial (Fig. 2a) and satellite (Fig. 2b) observing systems, in the NH the differences are comparable with absolute values typically less than 10 m although the satellite system does show larger differences over the largest land masses in the NH of North America and Siberia, whilst the terrestrial system shows larger differences over the oceans. These differences emphasize the fact that the radiosondes are more dominant over the land regions whilst the satellite data are more dominant over the oceans (the only satellite data available over land being the cloud motion winds). In the SH, larger differences are seen for the terrestrial system (>25 m) with the satellite system showing differences comparable with the NH. This reflects the importance of the satellite observing system in the SH where the surface and radiosonde observations are relatively sparse compared with the NH. It also highlights the equal importance of both the satellite and radiosonde observing systems in the NH. For the tropics, differences are smaller than in the extratropics for all experiments, but this is probably due to the smaller scale nature of the dynamics in the tropics, which are not well represented in the geopotential field.

Figure 3 shows results at the stratospheric level of 50 hPa. Larger differences are more generally seen with a change in some of the patterns compared with the mid-tropospheric results. The model-only experiment (Fig. 3d) shows very large differences, in particular in the NH with absolute values as high as 350 m, although differences are once again smaller in the SH. The surface system (Fig. 3c) now shows equally large differences in both hemispheres and larger than for the 500-hPa level,

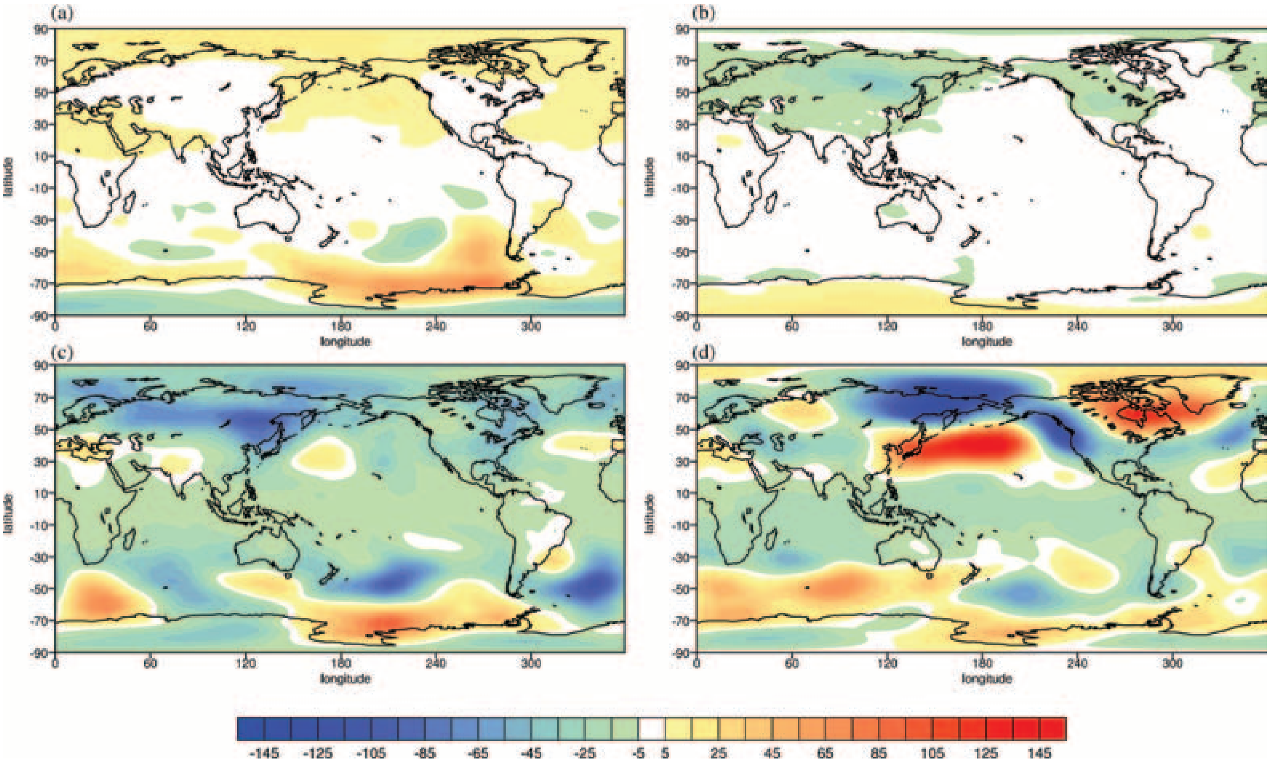


Fig 2. Difference plots of the time mean (DJF 1990/91) geopotential ($\overline{Z}_{\mathrm{dif}}$) at 500 hPa: (a) terrestrial–control, (b) satellite–control, (c) surface–control, (d) model–control. Units are m.

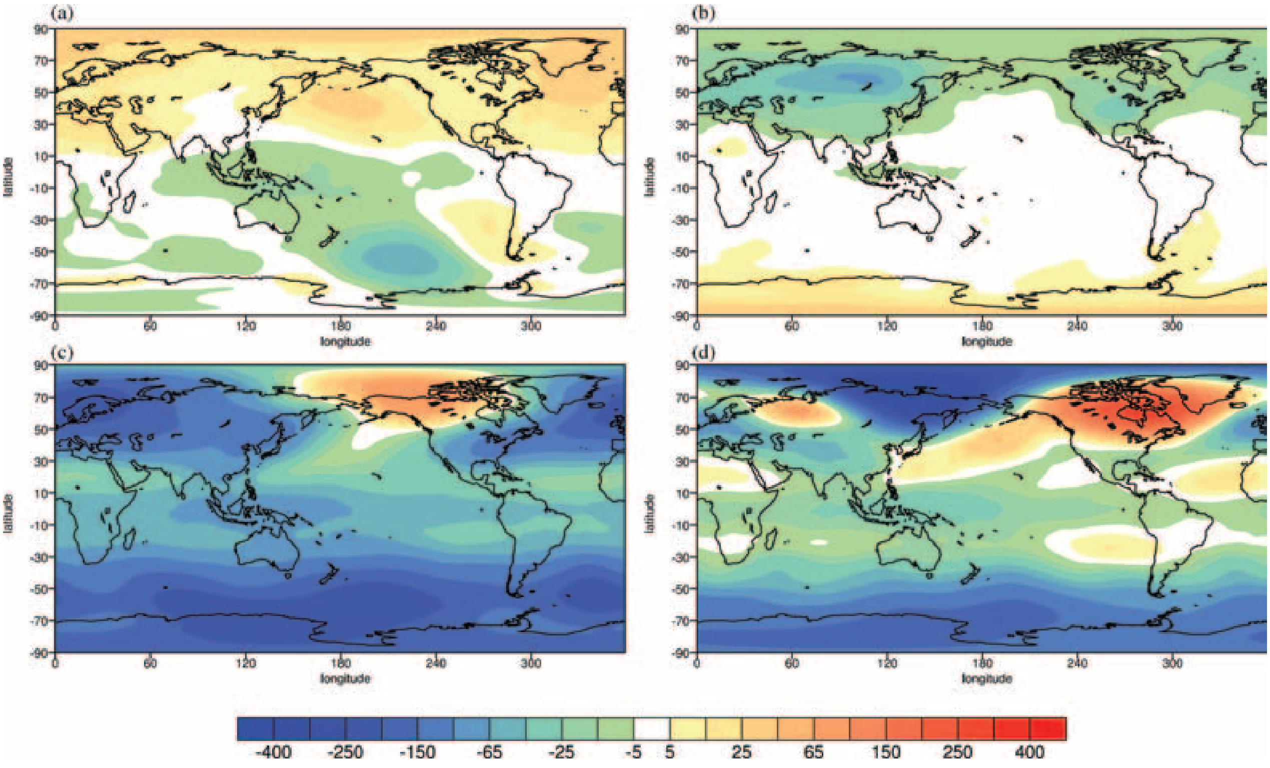


Fig 3. As for Fig. 2 but for 50 hPa. Note that the colour scale is non-linear.

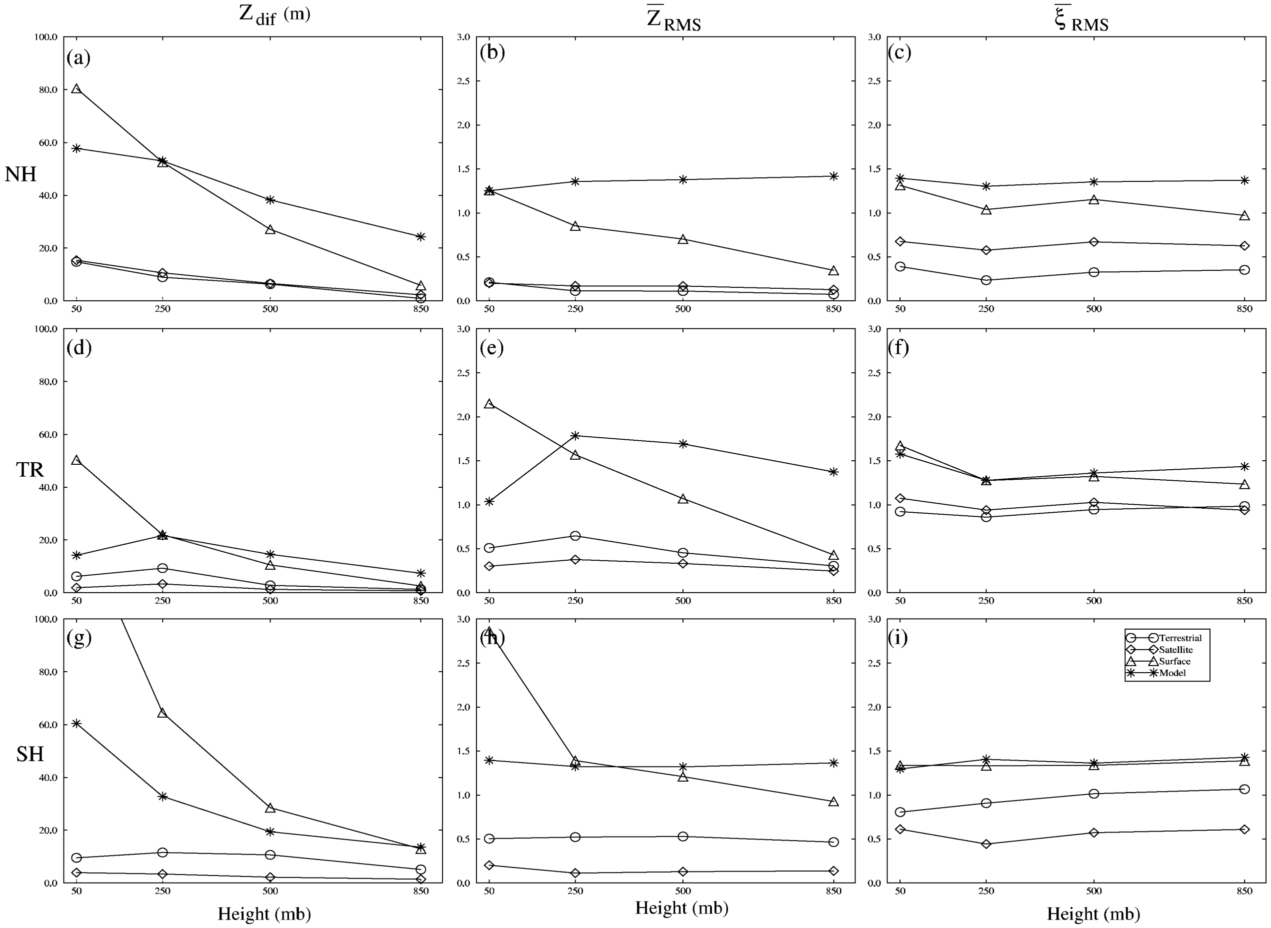


Fig 4. Time–area-averaged errors relative to the control experiment with respect to height (hPa) for the NH, tropics and SH for DJF 1990/91: (a) NH, absolute $\overline{Z}_{\rm dif}$; (b) NH, Z, RMSE; (c) NH, relative vorticity (ξ), RMSE; (d) tropics, absolute $\overline{Z}_{\rm dif}$; (e) tropics, Z, RMSE; (f) tropics, ξ, RMSE; (g) SH, absolute $\overline{Z}_{\rm dif}$; (h) SH, Z, RMSE; (i) SH, ξ, RMSE. The RMSE is normalized by the control experiment and so is dimensionless. Units are m for absolute differences in Z.

reflecting the situation that the surface observations have very little impact in the stratosphere. The terrestrial and satellite (Figs. 3a and b) systems also show larger differences than at the 500-hPa level but with similar patterns of differences. This confirms the importance of the satellite system in the SH and the equal importance of the satellite and radiosonde systems in the NH. In the NH, where the differences are large for the satellite system, i.e. over the large land masses, this highlights the greater importance of the radiosondes to the complete system. Note also that the terrestrial system has a negative bias in the NH extratropics, while the satellite system has a positive bias. The reason for this is not clear and could be due to a systematic bias in any of the data sets or both. However, since the comparison is with the full system, which contains both the terrestrial and satellite observations, this may just reflect the relative balance of the two sets of observations in the full system. There is also a marked positive bias in the surface system and in the model-only run covering most of the globe.

A different view of the relative importance of the different observing systems with respect to height can be seen in Figs. 4a, d and g, which show $\overline{Z}_{\rm dif}$ averaged over the NH (north of 20° latitude), the SH (south of -20° latitude) and the tropics (between -20° and $+20^\circ$ latitude) for the levels 850, 500, 250 and 50 hPa. This highlights the previously mentioned equal importance of the radiosonde and satellite observing systems in the NH, as apparent from the very similar plots in Fig. 4a and the importance of the satellite system in the SH (Fig. 4g). In fact, in the SH the satellite system dominates throughout the depth of the atmosphere. For the surface-only experiment, it is obvious that these observations provide only limited constraint away from the surface in both the NH and SH and, in fact, in the SH deviate from the control more than the model run. It is not clear why this is the case but it could be that the ERA40 system does not handle very sparsely distributed data in an optimal way, possibly associated with the background error statistics. Alternatively, it could be a sampling problem. In the tropics (Fig. 4d) the geopotential is not a good indicator of the smaller-scale dynamics in this region, as shown by the similarity between the terrestrial, satellite and surface differences.

A more detailed examination of the different observing systems at smaller temporal and spatial scales is provided by the RMSE of both Z and ξ, as presented in the next section. An even more detailed view is provided by examination of the individual weather events as discussed in Section 4.

3.2. RMSE of Z and ξ

To explore the instantaneous performance of the different observing systems, the RMSE is used to compare the experiments with the control experiment, with the RMSE scaled by the StD of the control experiment. As well as considering maps of the RMSE, the area-averaged values are computed as previously discussed for the time-averaged Z.

Figure 5 shows the RMSE of the 500-hPa geopotential (Z_{500}) for the different experiments. For the terrestrial system (Fig. 5a) the NH once again shows small differences relative to the StD of the control, reflecting the influence of the radiosonde network, whilst in the SH there are now larger differences, close to unity in some regions. Because the StD is large in the SH storm track region, these differences are also large; in fact, in Section 4 it will be shown that the weather systems being generated by the terrestrial system in the SH are often not directly comparable in terms of position and intensity with those in the control, highlighting the greater dependence on the underlying model in the SH for the terrestrial system due to the lack of observations. The tropics also show a larger degree of difference. These results show strong similarities with those of $\overline{Z}_{\mathrm{dif}}$ (see Fig. 3).

The satellite system (Fig. 5b) shows results comparable with those of the terrestrial system in the NH but with much smaller differences in the SH, once again highlighting the dominance of the satellite observing system in the SH. The surface and model experiments (Figs. 5c and d) show much larger differences, with some values more than twice the StD of the control. These represent quite large differences in the storm track regions where StD values are also large.

The surface-only observations provide little constraint on the mid-troposphere and upper troposphere, as previously noted, and as this is an instantaneous comparison it should not be expected that the model experiment would compare very well. These results more or less carry through to the 50-hPa level (not shown)

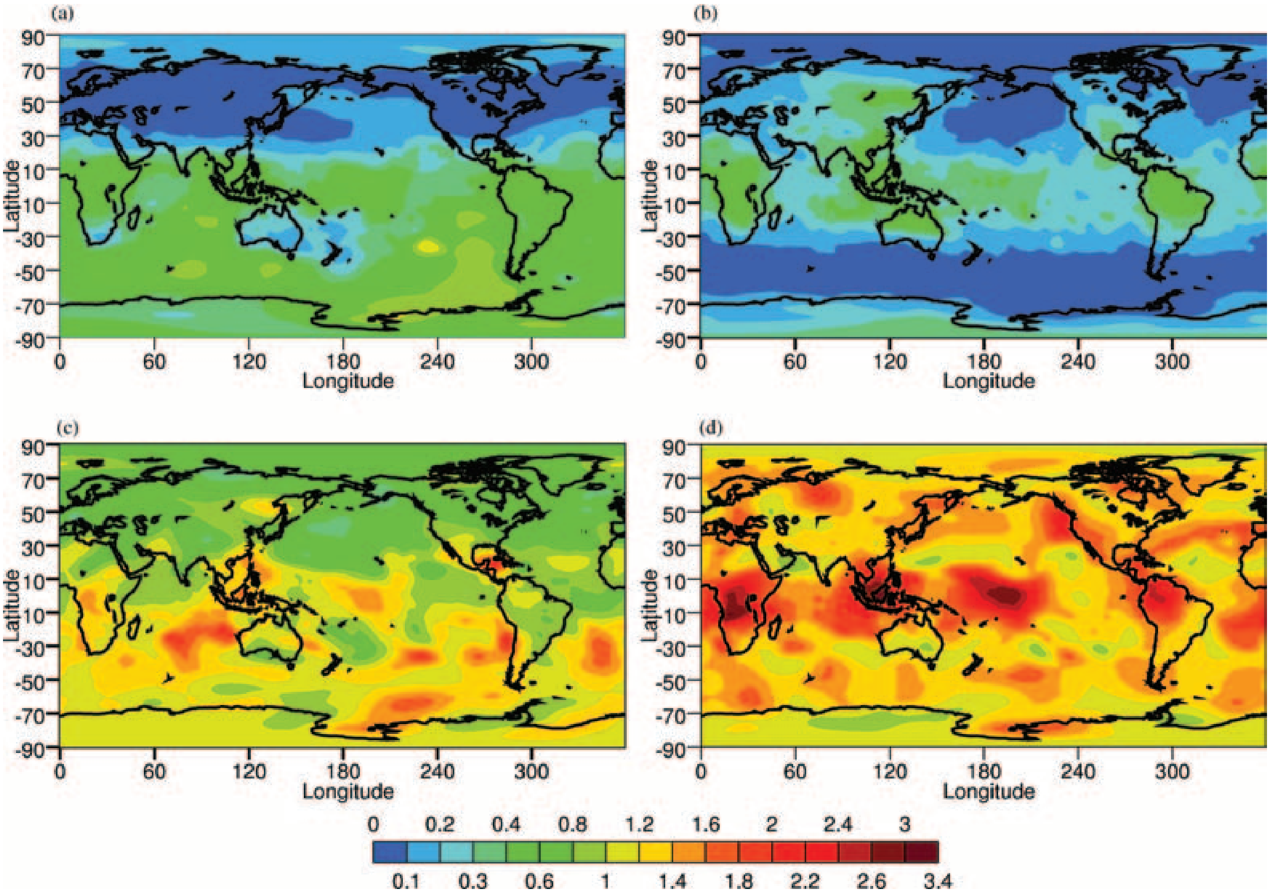


Fig 5. RMSE of Z_{500} for DJF 1990/91, normalized by the StD of the control experiment for (a) terrestrial, (b) satellite, (c) surface and (d) model.

except for the surface system. This has even greater differences, in particular in the SH, which suggests that even the sparse surface observations that are assimilated in the SH are having an adverse effect on the stratospheric circulation.

A more concise summary of the height dependence on the RMSE can be seen in Figs. 4b, e and h for the NH, tropics and SH, respectively. This shows that, for the terrestrial and satellite systems, the differences in the NH are small throughout the depth of the atmosphere whilst for the surface system the differences increase with height and the model shows uniformally large differences with height. The tropics show a similar behaviour to the NH, although in the stratosphere the surface system performs worse than the model. In the SH once again, the dominance of the satellite system is obvious with the terrestrial system showing larger differences. The surface system actually produces significantly larger differences than the model, indicating possible problems with the use of the surface-only data.

The relative vorticity field (ξ) focuses on smaller spatial scales (in a geostrophic sense). Figure 6 shows the RMSE for vorticity at the 500-hPa level (ξ_{500}). This shows much large differences than for Z_{500} consistent with the smaller-scale structures that are typical of this field. However, patterns of difference similar to those for Z_{500} are apparent for the terrestrial and satellite systems, albeit with much larger values. However, the terrestrial system now has much smaller differences over the NH compared with the satellite system, whilst in the SH the satellite system shows much smaller differences in the main storm track region. These results are summarized further in Figs. 4c, f and i as a function of height and show that the differences are more uniform with height for this field, again reflecting the smaller spatial scale and time-scale of structures in this field.

3.3. Quasi-biennial oscillation

In addition to the standard diagnostics discussed above, the quasi-biennial oscillation (QBO) has been explored as an example of the general circulation in the tropical stratosphere, which is very sensitive to the use of the available observations. The results from the different experiments compared with observations (zonal winds at Singapore kindly provided by Ms Naujokat, Free University of Berlin) are shown in Fig. 7. The model used in ERA40 does not succeed in simulating a regular QBO in free running mode, but the bias (mainly easterly) is small enough so that the assimilation of a few measurements of the equatorial wind can force a realistic analysis of the QBO. This is seen in

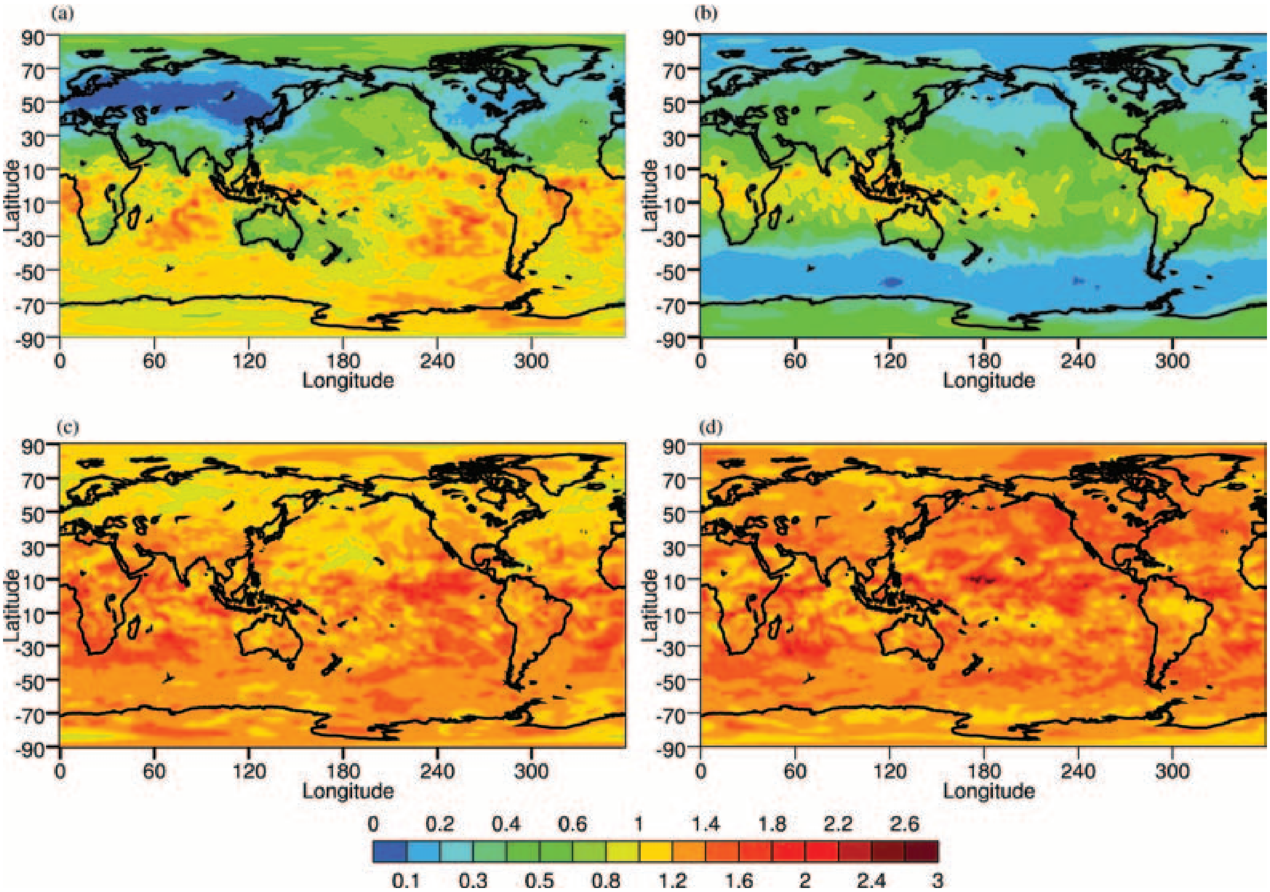


Fig 6. As in Fig. 5 but for ξ_{500}.

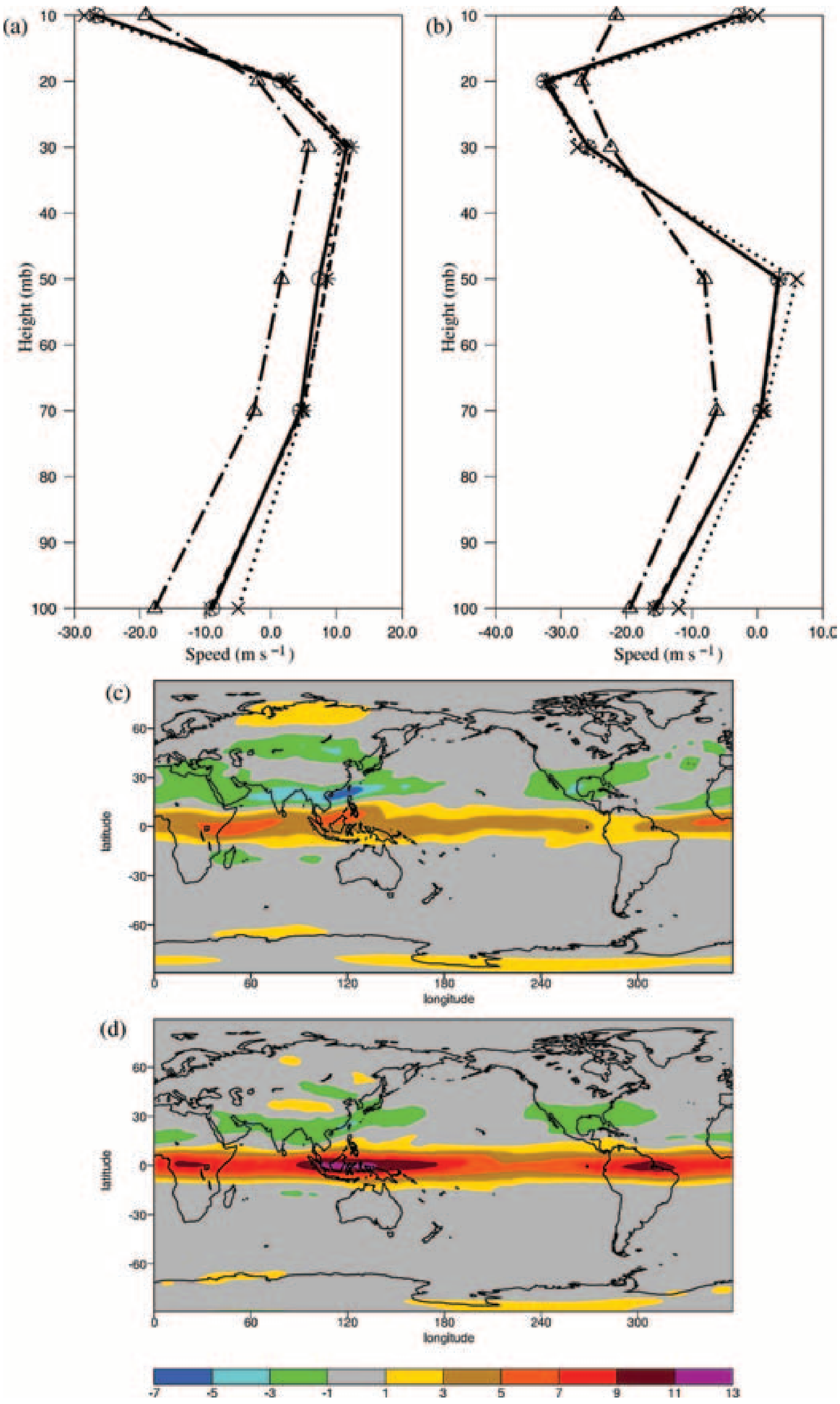


Fig 7. (a) Time-averaged zonal wind for DJF 1990/91 at Singapore as determined from the control (solid line, circle), the terrestrial-based system (dashed line, star) and the spaced-based system (dot-dashed line, triangle), respectively compared with the observed values (dotted line, cross) (b) the same for DJF 2000/01: (c) control-spaced-based system at 50 hPa for DJF 1990/91; (d) the same for DJF 2000/01.

Figs. 7a and b. During the winter of 1990/91 there were westerlies between the levels 20 and 90 hPa, while during DJF 2000/01 there were weak westerlies around 50 hPa and a sharp transition at about 40 hPa to strong easterlies above. This is well handled by the control experiment. The terrestrial system also handles this well, while the satellite system has large errors. This is further highlighted in Figs. 7c and d for the two periods respectively, which show the spatial variation of the differences in the zonal wind at 50 hPa between the control and satellite observing system experiment. This shows that the bias exists throughout the tropics in both periods. These results indicate that there is an interesting and important contribution by a small number of *in situ* wind measurements in the tropics. It certainly highlights the need for maintaining selected radiosonde observations in the equatorial belt or alternatively arranging for satellite wind measurements in the equatorial stratosphere.

4. Extratropical cyclones

A detailed study of the representation of extratropical cyclones has also been undertaken. This entails identifying and tracking the cyclones for each experiment using both the MSLP and 850-, 500- and 250-hPa vorticity (ξ_{850}, ξ_{500}, ξ_{250}). The methodology used for the tracking is the same as that used by Hoskins and Hodges (2002). Using both the MSLP and vorticity allows different scales of weather system to be considered; the MSLP tends to focus on the larger-scale cyclones whilst the vorticity tends to focus on smaller spatial scales. A comparison has been performed between the experiments and the control by directly comparing the track ensembles as in Hodges et al (2003, 2004). This entails identifying the same weather systems for each pair of experiment–control comparisons based on a set of matching criteria for the separation distance and temporal overlap between pairs of tracks (typically tracks have lifetimes of 4–5 d in the main ocean basins). However, instead of using a single set of matching parameters, a hierarchy of matching parameters is used. The first set consists of fairly strict values of 2.0° mean separation distance (pairs of tracks must have a mean separation less than or equal to this to be considered a good match) and a temporal overlap of 60% of the points of a pair of tracks. A second set consists of a minimum separation distance of 2.0° (i.e. for a single pair of points) and 60% temporal overlap. The final set consists of a minimum separation distance of 2.0° and a temporal overlap of 30%. If a large proportion of the tracks match for the strict constraints, then it would be expected that this would not change significantly for the other constraints. However, if the proportion of tracks that match is small for the strict parameters, this does not mean that the same weather systems are not present in both ensembles but that there maybe a greater degree of error in location and/or temporal overlap. This is apparent if the number of cyclones that match increases as a proportion of the total as the matching parameters are relaxed (cf. second and third sets of parameters). Cyclones often do not match, either because of larger differences in location or temporal overlap than the chosen threshold or a different degree of temporal coherence that may result in a track appearing in one ensemble and not in the other. The fact that it is the weaker cyclones that often do not match well indicates that these are more sensitive to the assimilation of the available observations. The results of this analysis are summarized in Table 1 for both the NH and SH, and are discussed below.

As well as the summary statistics, several distributions can be determined which also help in the comparison. These are determined using the middle set of matching parameters, i.e. neither too strict nor too lax. The first form of distribution is computed from the point-by-point intensity differences for those tracks that match. Note that intensities are relative to the removed background (see Hoskins and Hodges, 2002). The second form of distribution is for the mean separation distance between matching tracks (note that the matching is performed with the minimum separation criteria). These two distributions are computed as probability density functions (pdfs). The distributions are presented only for the lower tropospheric fields for conciseness (MSLP and ξ_{850}). The mean intensity number distribution for those cyclones that match and those cyclones that do not match, as used by Hodges et al. (2003, 2004), is not used here as Table 1 provides a suitable summary of these statistics.

Table 1. Summary of track ensemble matching statistics for all fields, and for the two hemispheres separately. # is the total number of systems in the three-month DJF period, %(1) is the percentage of systems from the second ensemble (experiment) that match with the first ensemble (control) for the strict set of matching parameters, %(2) is for the second set and %(3) for the third most relaxed set

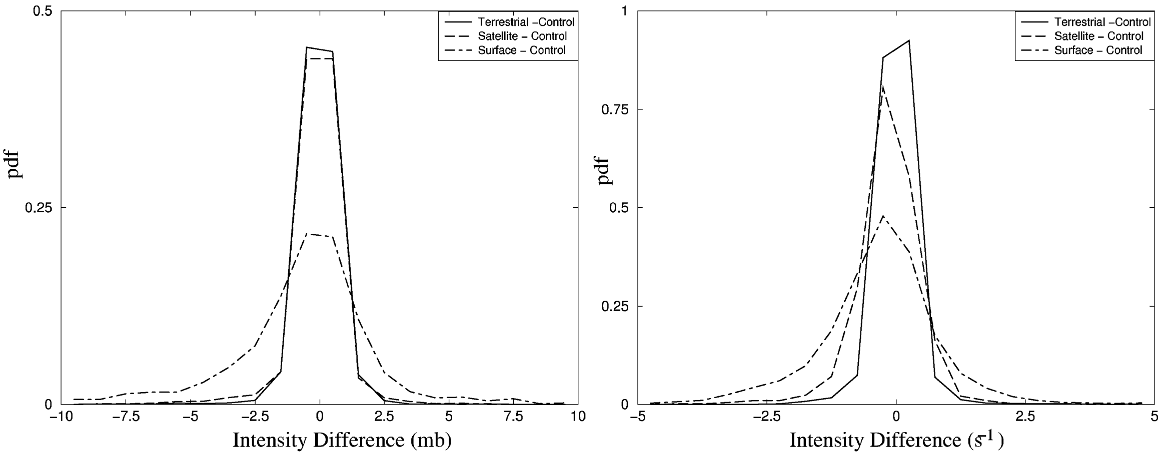
NH													
	CNTL	Terrestrial				Satellite				Surface			
	#	#	%(1)	%(2)	%(3)	#	%(1)	%(2)	%(3)	#	%(1)	%(2)	%(3)
MSLP	163	175	84.5	86.3	86.3	165	86.0	87.9	89.1	159	48.4	70.4	76.1
ξ_{850}	426	463	71.7	74.3	76.5	458	58.1	67.0	71.2	438	25.6	52.9	60.3
ξ_{500}	546	563	64.8	69.3	71.7	568	40.1	58.9	67.6	551	3.6	27.9	39.7
ξ_{250}	428	439	55.1	57.8	61.5	455	25.3	38.4	47.7	450	2.4	16.0	23.1
SH													
	CNTL	Terrestrial				Satellite				Surface			
MSLP	154	148	15.5	47.9	55.4	146	86.3	88.4	89.0	145	4.8	23.4	31.7
ξ_{850}	364	391	9.7	35.5	45.8	352	66.5	77.3	80.1	412	2.2	17.5	26.2
ξ_{500}	457	456	7.6	33.8	44.1	457	57.9	66.1	71.3	454	0.9	13.6	23.8
ξ_{250}	446	461	12.5	35.1	45.5	456	57.9	67.9	72.1	469	1.1	11.5	17.7

Because for the model–control comparison common cyclones are unlikely to occur, the model experiment is excluded from this analysis. Ideally, the spatial statistics discussed in Hoskins and Hodges (2002) should be computed; however, one winter provides insufficient cyclone tracks to provide reliable statistics for this.

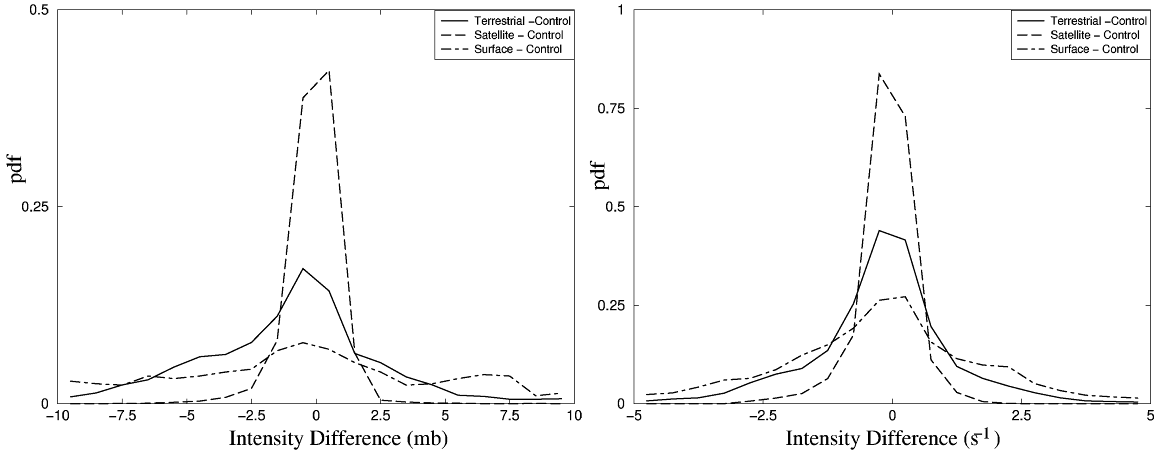
4.1. Northern Hemisphere

Table 1 shows that for the MSLP there is a large proportion of the tracks (>80%) for both the terrestrial and satellite systems that match with the control and that this is consistent for all three of the matching regimes with only a change of 2–3% in going from the strict matching to the most relaxed. The cyclones that do not match tend to be weak in nature. If the distributions for the instantaneous intensities and mean separation distances are considered, Figs. 8a and 9a, respectively, it is apparent that the intensity differences are strongly peaked at zero between ±2.5 hPa and that the mean separation of cyclones is predominately less than 1.0° for both terrestrial and satellite comparisons with the control. This implies a very good correspondence between the MSLP cyclones in these two experiments with the control both in terms of intensity, location and temporal coherence. These results are consistent with the previous mean Z and RMSE results that the radiosonde network and satellite observing systems are of equal importance in the NH at this scale.

For the surface system, Table 1 indicates that the number of MSLP cyclones that match with those in the control for the strict matching parameters is much reduced (<50%). However, using the more relaxed matching parameters, the number of cyclones that match increases to >70%. Hence, the surface system does appear to capture many of the same cyclones as in the control but there is a greater difference in location and/or temporal coherence than was the case with the terrestrial and satellite systems. As

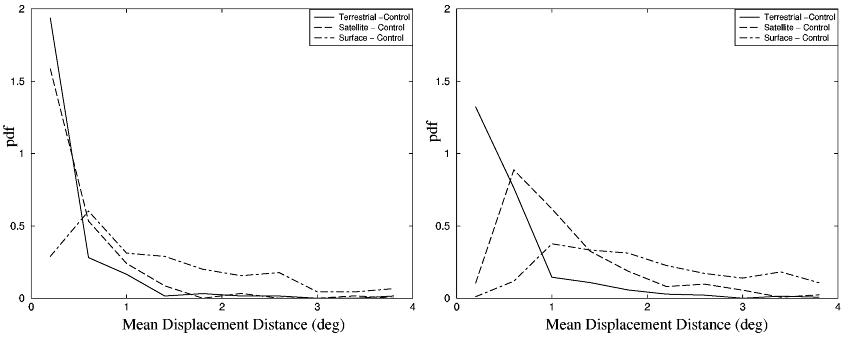


(a) NH, MSLP (b) NH, ξ_{850}

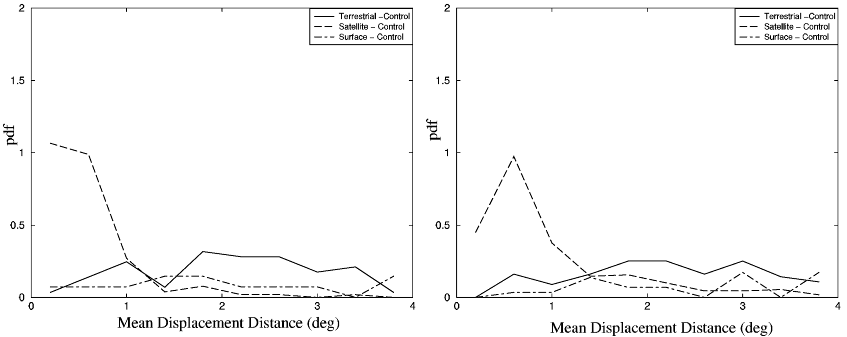


(c) SH, MSLP (d) SH, ξ_{850}

Fig 8. Cyclone instantaneous intensity difference pdfs for the NH and SH for the three experiments compared with the control (experiment–control) for (a) NH, MSLP, (b) NH, ξ_{850}, (c) SH, MSLP and (d) SH, ξ_{850}. Intensities are relative to the removed background. Units are hPa for the MSLP and $\times$ 10^{-5} s^{-1} for ξ. Note that some probability may occur outside the range plotted.



(a) NH, MSLP (b) NH, ξ 850



(c) SH, MSLP (d) SH, ξ 850

Fig 9. Cyclone mean track separation pdfs for the NH and SH for the three experiments compared with the control for (a) NH, MSLP, (b) NH, ξ_{850}, (c) SH MSLP and (d) SH, ξ_{850}. Distance measured in degrees (geodesic). Note that some probability may occur outside the range plotted.

with the terrestrial and satellite experiments, the cyclones that do not match tend to be weak, although there are a larger number of stronger cyclones that do not match depending on the matching regime used. Figures 8a and 9a for the intensity differences and mean separation distance distributions highlight the greater error in the representation of the cyclones in the surface system. The intensity differences show a broader distribution than for the terrestrial or satellite systems and there is some indication of a negative bias. The separation distances also show a broader distribution with some mean separation distances as large as 4.0°.

For ξ_{850}, the terrestrial system shows similar results to the MSLP (Table 1), although there are now many more cyclones, which is typical of the smaller-scale nature of the weather systems identified in this field. The majority of cyclones identified in this field at this level match well ($>70\%$) for all three matching regimes, with a smaller number of cyclones that do not match, mostly at the weaker end of the intensity scale. The satellite system shows much less agreement with the cyclones in the control for the strict matching parameters ($<60\%$). However, the proportion of cyclones that match increases rapidly as the matching parameters are relaxed, to $>70\%$ for the most relaxed matching regime which is comparable with the terrestrial system. This indicates that, for the satellite system, the location and temporal coherence of the cyclones is not as good as for the terrestrial system. This is apparent in Fig. 9b, where for the terrestrial system the cyclones predominately have mean separations less than 1.0° whilst for the satellite system a broader distribution is shown. Figure 8b shows a similar type of behaviour for the intensity differences with the terrestrial system having a strongly peaked distribution centred on zero between $\pm$ 1.0 $\times$ 10^{-5} s^{-1}, whilst the satellite system has again a broader distribution. This sensitivity at the smaller spatial scales is typical of vorticity and was also seen in the previously discussed results for the RMSE. This highlights the importance of the terrestrial observing system as smaller-scale structures are considered. For the surface system, the contrast between the strict and relaxed matching regimes is even more striking indicating the larger degree of error in location and temporal coherence than for the other systems. This is most apparent in Fig. 9b where the mean separation statistic is now very flat and broad and with some probability occurring outside the plotted range. The intensity differences also show a broader distribution than for the other experiments and, as with the MSLP, there is the indication of a negative bias. The greater errors in intensities, location and temporal coherence indicate that surface-only observations supply insufficient information content to constrain the upper troposphere at these scales anywhere away from the surface.

If higher levels in the troposphere are considered, it is apparent that at 500 hPa the terrestrial and satellite systems are still providing results similar in behaviour with those at 850 hPa for the vorticity (Table 1); this is also true for the intensities and mean separation statistics (not shown). For the surface system however, it is apparent that the matching results are poorer than at 850 hPa with less than 50% cyclones matching even for the most relaxed matching regime. This further indicates the lack of constraint provided by the surface observations in the upper troposphere further away from the surface. The results at the 250-hPa level confirm the results at the lower levels, although the degree to which cyclones match between the experiments and the control is reduced compared with the lower levels, even for the terrestrial system. This deterioration in the comparison at the tropopause is probably due to dynamic instability at this level with large wind speeds and shear resulting in a greater sensitivity to the available observations and how they are assimilated. This can also lead to greater difficulty in performing the tracking (Hodges et al., 2003). For the surface system, the degree to which cyclones match is now very poor. Examining the intensity difference distributions at these higher levels (not shown) indicates that the bias to negative difference values becomes more prominent with height for the satellite and surface observing systems, whereas the terrestrial system distributions are still centred on zero at the 500- and 250-hPa levels. This is probably related to the impact of the radiosondes, which provide information on winds and temperature throughout the depth of the atmosphere at much higher resolution than the satellites for example, and effectively adjust the model which is biased to weaker cyclones. This argument is supported by the fact that in the SH, where there are fewer radiosondes, the bias appears weaker with height for the satellite system.

4.2. Southern Hemisphere

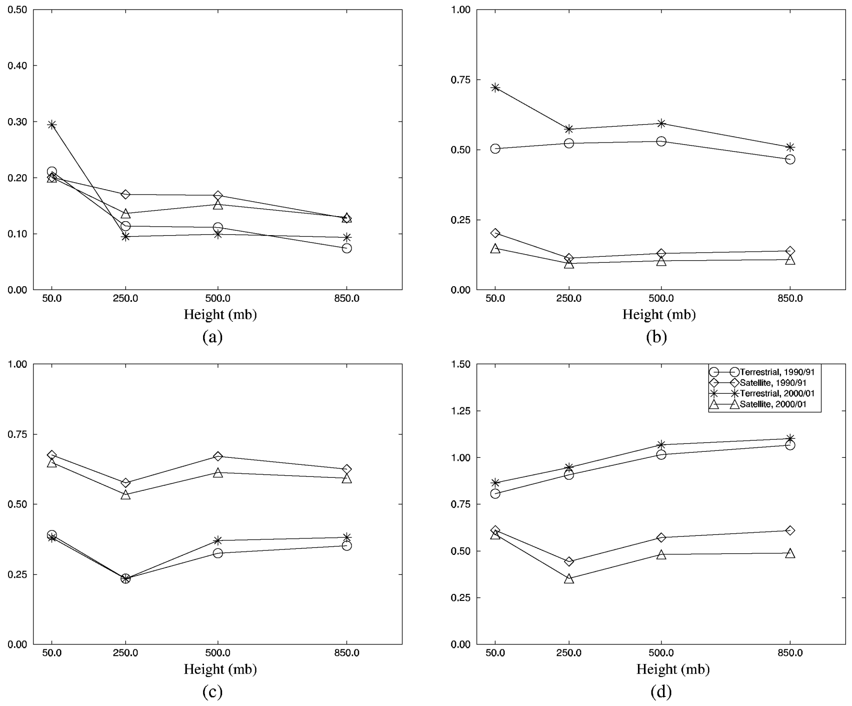
The main conclusion that can be derived from the matching results in the SH is the superiority of the satellite observing system compared with the other observing systems. Table 1 shows this for both the MSLP and vorticity with the vorticity appearing to provide a better comparison than in the NH. However, this is deceptive due to the dominance of the satellite observing system in the SH, which may result in a too incestuous comparison. The MSLP comparison for the satellite system is comparable with that in the NH. For the intensity differences shown in Figs. 8c and d for the MSLP and ξ_{850}, respectively, the performance of the satellite system is comparable with the NH with a strongly peaked distribution with a similar spread. The mean separation distance statistics shown in Figs. 9c and d for the MSLP and ξ_{850}, respectively, show results similar to those in the NH. Thus, the satellite system is seen to perform with similar levels of correspondence between cyclones identified in the experiment and the control in both hemispheres. However, for the terrestrial and surface observing systems the lack of sufficient observations results in very poor levels of correspondence between cyclones in the experiments and the control (Table 1) in terms of intensity (Figs. 8c and d), location (Figs. 9c and d) and temporal coherence throughout the depth of the troposphere. It should be remembered, however, that the satellite system used here is constrained at the surface by the pressure data. The lack of these data in the SH probably makes the analysis of the satellite data more dependent on the model than is the case in the NH. This is seen to some extent in the results of Hodges et al. (2003) where the comparison of different reanalyses showed greater uncertainty in the representation of cyclones in the SH between the different reanalyses when contrasted with the NH.

5. Effect of observational changes in the 1990s

During the 1990s, the global observing systems underwent significant changes including increasing data from satellite observations. This included the use of scatterometer winds and Advanced Microwave Sounding Unit (AMSU) soundings. There has also been an increase in the number of satellite temperature and moisture profiles. At the same time, the radiosonde network has diminished, particularly so in the countries of the former Soviet Union but also in Africa and South America. The number of radiosondes was reduced from 641 in December 1990 to 548 in December 2000 (compared at 12 UTC; S. Uppala, ECMWF, private communication). As a result, it is to be expected that the reanalyses have become gradually more dominated by space-based observations through the 1990s. Figure 10 shows the relative information from the spaced-based and the terrestrial-based systems for the two periods. As in Fig. 4, it is apparent that the terrestrial observations dominate in the NH and satellite data in the SH. Generally, the weight of information from the space-based system has increased and the terrestrial system correspondingly has decreased in both hemispheres between 1990 and 2000; this is apparent both for the geopotential and vorticity results. The NH vorticity field behaves slightly differently with reduced weight for the terrestrial system in the stratosphere. This suggests that this may be due to a smaller number of soundings reaching stratospheric levels. The slight improvement of the terrestrial system in the upper and middle troposphere may be due to the large increase in wind and temperature measurements from aircraft data, which are included in the terrestrial system, particularly over the major air routes of the Atlantic and Western Europe.

6. Discussion and summary

The main objective of a reanalysis is to create a homogeneous four-dimensional data set to be used for the study of climate and climate variability. As demonstrated in the original paper by Bengtsson and Shukla (1988), data sets from operational numerical weather prediction are inadequate because of the many changes in the model and the data assimilation over time. Repeating the exercise with a frozen system, as has been done in ERA40 and other similar studies, has made it possible to eliminate such problems. However, when a reanalysis is extended over long periods of time which include major changes in the observing systems, great caution must be exercised when using the analysed data sets. When comparing analysed fields prior to 1979 with later data, it must be kept in mind that the analyses of the tropics and the SH for the early period are dominated by



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

Fig 10. Area-averaged RMSE of the terrestrial and satellite systems relative to the control for geopotential and vorticity for the two winter periods (DJF) 1990/91 and 2000/01 and for each hemisphere with respect to height (hPa). The errors are normalized by the standard deviation of the relevant control experiment. (a) RMSE geopotential in the NH extratropics (20°N–90°N), (b) RMSE geopotential in the SH extratropics (90°S–20°S), (c) RMSE vorticity in the NH extratropics and (d) RMSE vorticity in the SH extratropics.

model information. Any model deficiency will therefore have a major impact on the analysed fields in these areas. This is seen clearly in the results from ERA40 shown in Fig. 1 as well as the results presented by Bengtsson et al. (2004b) and Hodges et al. (2003, 2004).

In this paper, an attempt to quantify these deficiencies is made. This is done in an integrative way by rerunning the ERA40 analysis for two periods of 1990/91 and 2000/01 with reduced observing systems, mimicking the observing systems typical of the past. Such an integrative approach is required as it is not possible to quantitatively assess the effect of the complex interaction of model and data information during the assimilation in any other way. Because of plans to extend reanalyses to the period before 1945/50, a study has also been carried out with an experiment where only the surface data are used. Finally, because of the potential evolution towards a fully satellite-based system in the future, this has been investigated by excluding all the terrestrial observations except surface pressure. This experiment provides an opportunity to assess the information content of the space-based observing systems.

For practical reasons, it was found necessary to limit the evaluation of the experiments. Here the study concentrates on quantifying the information content of the different observing systems and to calculate selected aspects of the general circulation of the atmosphere. This includes aspects of the reduced observing systems to depict transient dynamical processes.

The deficiencies of the reduced systems are of two kinds. First, there are clear systematic deficiencies. These include a colder and dryer atmosphere as well as slightly weaker extratropical weather systems, with the bias becoming stronger with height. Such systematic errors are particularly worrisome when trying to assess possible long-term climate trends. In an accompanying paper (Bengtsson et al., 2004b), the results from this study have been used to correct the systematic errors in global climate trends.

The main results of the study are summarized as follows.

6.1. *Surface-based system*

(i) The surface-only network is insufficient to produce reliable fields aloft. For the vorticity field, the RMS error is equal to or larger than the standard deviation of the control field everywhere. The normalized error of the geopotential field is generally below 1 for the NH troposphere. For the SH and the tropics, the normalized error is larger than 1 for all variables.

(ii) The surface-based system produces extratropical vortices which, in general, are in poor correspondence with those in the control, particularly in the upper troposphere and the SH.

(iii) The mean errors include a too cold upper troposphere and lower stratosphere, mainly at higher latitudes. This also implies too strong mid-latitude westerlies.

(iv) Based on these findings, it is clear that much care is required when undertaking reanalysis with a system which only consists of surface-based observations. With the data assimilation used in ERA40, surface observations cannot effectively constrain the large-scale tropospheric circulation and provide virtually no constraint on the stratospheric circulation. Some of these problems may be related to the fact that the ERA40 system is probably not in an optimal configuration when using surface observations alone, in particular in the data-sparse SH. This is likely to occur due to the use of stationary background error statistics derived from the complete system using all the available observations. This could be circumvented to some extent if a more optimal system was used where the background error statistics are derived based solely on the surface observing system. The use of more sophisticated assimilation systems such as 4D Var or the Kalman filter, where the background errors are treated in a more dynamic way, may also help. Experimental integrations with a 4D Var system have been performed but with no noticeable improvement. The introduction of better error statistics and new techniques is only likely to have an impact in the NH where a suitably dense surface observing network exits. Even here, it should not be expected that a surface-only system could provide sufficient constraint on the whole troposphere, and it is unlikely to provide any in the stratosphere.

To a certain degree, the results presented here for the surface system can be seen as an upper bound on what could be expected from reanalysing the atmosphere prior to 1945, as the surface network used here is more comprehensive than the observing system during the first half of the twentieth century. That other assimilation techniques may perform better is not excluded, as suggested in recent results reported by Whitaker et al. (2004). However, the definition of better should always be qualified by the scale at which the quality of a reanalysis is determined, as shown in this study.

6.2. Terrestrial system

(i) This is the most accurate system for the NH extratropics in particular with respect to the vorticity field. In the tropics, the terrestrial and the space system have the same accuracy; the space-based system is slightly better for geopotential but the terrestrial system is better for wind and vorticity. For the SH extratropics, the normalized RMS is between 0.5 and 1 and is largest for vorticity and smallest for geopotential. Weather systems compare very well in the NH for both the MSLP and vorticity with the control.

(ii) For the SH extratropics, the vorticity fields are almost as poor as the surface-based system and the comparison of weather systems is poor.

(iii) The terrestrial system is the only system which is able to analyse the QBO. This is because the model cannot reproduce the QBO per se and there is no information in the satellite observations of the existence of the QBO, as it cannot be inferred from available satellite information. It seems that the very few equatorial radiosondes, which extend into the lower stratosphere, are crucial. When the QBO signal has entered the system, it remains for long enough times to be subsequently activated by the few equatorial radiosondes reaching the stratosphere.

6.3. Spaced-based system

(i) There are minor differences in the mean field over the NH with slightly lower geopotential values than in the control.

(ii) Variance reduction for geopotential and temperature is rather similar to the terrestrial system for the NH extratropics but better in the tropics and significantly better in the SH extratropics. The error reduction for the vorticity fields on the other hand is smaller and hardly gets below 0.5.

(iii) The spaced-based system underestimates the vorticity pattern aloft for the NH but the MSLP is practically identical to the terrestrial system. For the SH extratropics, there is hardly any difference from the control.

(iv) Extratropical cyclones in the NH are less well analysed in the space-based system than in the terrestrial system, especially in the upper troposphere. However, in the SH the extratropical cyclones compare better with those in the control than for the terrestrial system, although this is probably due to the dominance contribution of the satellite data to the full system in the SH.

6.4. Concluding remarks

As a general comment we believe that the integrated methodology used here is a rational approach to assess the value of different observing systems. Due to computational and other technical restrictions, it was necessary to restrict the assimilation periods to a few seasons. Furthermore, only what is believed to be the main characteristic global observing systems in the twentieth century have been explored. It has not been possible, for example, to investigate the specific influence of aircraft reports which are likely to have added value to the terrestrial system.

The results reported here depend on the use of the ERA40 system and might be different when using another model with different biases and a more advanced assimilation system. However, it is not expected that the more general results achieved here would be very different.

We believe this study will be of interest for the planning of future global observing systems. Such a system will, in all likelihood, be more spaced-based and automated. It goes without saying that such observational changes should be explored beforehand using a methodology similar to that applied in this study to determine their utility.

7. Acknowledgment

The authors would like to thank the ECMWF for making the ERA40 system available and for providing support in running the experiments.

References

Bengtsson, L. 1999. From short range barotropic modelling to extended-range global weather prediction. *Tellus* **51A**, 13–32.

Bengtsson, L. and Shukla, J. 1988. Integration of space and *in situ* observations to study climate change. *Bull. Am. Meteorol. Soc.* **69**, 1130–1143.

Bengtsson, L., Hodges, K. I. and Hagemann, S. 2004a. Sensitivity of large scale atmospheric analyses to humidity observations and its impact on the global water cycle and tropical and extratropical weather systems. *Tellus* **56A**, 202–217.

Bengtsson, L., Hagemann, S. and Hodges, K. I. 2004b. Can climate trends be computed from reanalysis data? *JGR*. vol. 109, D11111, doi:10.1029/2004JD004536, 2004.

Gibson, J., Kållberg, P., Uppala, S., Nomura, A., Hernandez, A. and Serrano, E. 1997. ERA description. ECMWF Reanalysis Final Report Series, 1, ECMWF, Shinfield Park, Reading, UK, 71 pp.

Hodges, K. I., Hoskins, B. J., Boyle, J. and Thorncroft, C. 2003. A comparison of recent re-analysis data sets using objective feature tracking: storm tracks and tropical easterly waves. *Mon. Wea. Rev.* **131**, 2012–2037.

Hodges, K. I., Hoskins, B. J., Boyle, J. and Thorncroft, C. 2004. Corrigendum to “A comparison of recent re-analysis data sets using objective feature tracking: storm tracks and tropical easterly waves”. *Mon. Wea. Rev.* **132**, 1325–1327.

Hoskins, B. J. and Hodges, K. I. 2002. New perspectives on the northern hemisphere winter storm tracks. *J. Atmos. Sci.* **59**, 1041–1061.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. et al. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471.

Kanamitsu, M., Ebisuzaki, W., Woolen, J., Potter, J. and Fiorino, M. 1999. An Overview of Re-Analysis 2. In: *Proc. 2nd Int. Conf. on Reanalysis*, Reading, UK, WCRP.

Lott, F. and Miller, M. J. 1997. A new subgrid-scale ororgraphic drag parameterization: its formulation and testing. *J. Atmos. Sci.* **31**, 118–133.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J. and Clough, S. A. 1997. Radiative transfer for inhomogeneous atmospheres: RRTM a validated correlated-k model for the longwave. *J. Geophys. Res.* **102D**, 16 663–16 682.

Schubert, S. D., Rood, R. B. and Pfaendtner, J. 1993. An assimilated dataset for earth science applications. *Bull. Am. Meteorol. Soc.* **74**, 2331–2342.

Simmons, A. J. and Gibson, J. K. 2000. The ERA-40 Project Plan, ERA-40 Project Report Series No. 1 ECMWF. Shinfield Park, Reading, UK, 63 pp.

Simmons, A. J. and Hollingsworth, A. 2002. Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.* **128**, 647–677.

Tiedtke, M. 1989. A comprehensive mass flux scheme for cumulas parametrization in large-scale models. *Mon. Wea. Rev.* **117**, 1779–1800.

Whitaker, J. S., Compo, G. P., Wei, X. and Hamill, T. M. 2004. Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.* **132**, 1190–1200.