

Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature

A. J. Simmons

European Centre for Medium-Range Weather Forecasts, Reading, UK

P. D. Jones

Climatic Research Unit, University of East Anglia, Norwich, UK

V. da Costa Bechtold, A. C. M. Beljaars, P. W. Kållberg, S. Saarinen,

S. M. Uppala, P. Viterbo, and N. Wedi

European Centre for Medium-Range Weather Forecasts, Reading, UK

Received 2 August 2004; revised 1 October 2004; accepted 21 October 2004; published 31 December 2004.

[1] Anomalies in monthly mean surface air temperature from the 45-Year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) and the first National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis are compared with corresponding values from the Climatic Research Unit (CRU) CRUTEM2v data set derived directly from monthly station data. There is mostly very similar short-term variability, especially between ERA-40 and CRUTEM2v. Linear trends are significantly lower for the two reanalyses when computed over the full period studied, 1958–2001, but ERA-40 trends are within 10% of CRUTEM2v values for the Northern Hemisphere when computed from 1979 onward. Gaps in the availability of synoptic surface data contribute to relatively poor performance of ERA-40 prior to 1967. A few highly suspect values in each of the data sets have also been identified. ERA-40's use of screen-level observations contributes to the agreement between the ERA-40 and CRUTEM2v analyses, but the quality of the overall observing system and general character of the ERA-40 data assimilation system are also contributing factors. Temperatures from ERA-40 vary coherently throughout the boundary layer from the late 1970s onward, in general, and earlier for some regions. There is a cold bias in early years at 500 hPa over the data-sparse southern extratropics and at the surface over Antarctica. One indicator of this comes from comparing the ERA-40 analyses with results from a simulation of the atmosphere for the ERA-40 period produced using the same model and same distributions of sea surface temperature and sea ice as used in the ERA-40 data assimilation. The simulation itself reproduces quite well the warming trend over land seen in CRUTEM2v and captures some of the low-frequency variability. **INDEX TERMS:** 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; 3309 Meteorology and Atmospheric Dynamics: Climatology (1620); 1610 Global Change: Atmosphere (0315, 0325); **KEYWORDS:** reanalysis, temperature trends

Citation: Simmons, A. J., P. D. Jones, V. da Costa Bechtold, A. C. M. Beljaars, P. W. Kållberg, S. Saarinen, S. M. Uppala, P. Viterbo, and N. Wedi (2004), Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature, *J. Geophys. Res.*, 109, D24115, doi:10.1029/2004JD005306.

1. Introduction

[2] Comprehensive reanalyses derived by processing multidecadal sequences of past meteorological observations using modern data assimilation techniques have found widespread application in many branches of meteorological and climatological research. Their utility for helping to document and understand climatic trends

and low-frequency variations is nevertheless a matter of some debate. Atmospheric data assimilation comprises a sequence of analysis steps in which background information for a short period, typically of 6-hour duration, is combined with observations for the period to produce an estimate of the state of the atmosphere (the “analysis”) at a particular time. The background information comes from a short-range forecast initiated from the most-recent preceding analysis in the sequence. Problems for climate studies arise partly because the atmospheric models used to produce these “background forecasts” are prone to

biases. If observations are abundant and unbiased, they can correct the biases in background forecasts when assimilated. In reality, however, observational coverage varies over time, observations are themselves prone to bias, either instrumental or through not being representative of their wider surroundings, and these observational biases can change over time. This introduces trends and low-frequency variations in analyses that are mixed with the true climatic signals. Progress in the longer term depends on identifying and correcting model biases, accumulating as complete a set of historic observations as possible, and developing improved methods of detection and correction of observational biases. In the shorter term, awareness of how these factors influence a particular reanalysis can aid the interpretation and application of its results.

[3] In this paper, processed station values of monthly mean surface air temperature are compared with corresponding values derived from the products of two reanalyses, the 45-Year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) and the first National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis. A global view is taken, and upper air data from ERA-40 and data from a simulation using the ERA-40 model are used as part of the evaluation. ERA-40 is the most recent comprehensive reanalysis to be completed and the first to provide an alternative to the earlier NCEP/NCAR reanalysis [Kalnay *et al.*, 1996; Kistler *et al.*, 2001] for the years before 1979. Observations from September 1957 to August 2002 were analyzed. General information on ERA-40 can be found in a series of project reports (available online at <http://www.ecmwf.int/publications>) and from the project's own Web pages (<http://www.ecmwf.int/research/era>).

[4] The ERA-40 analyses of temperature at a height of 2 m were produced every 6 hours as part of the data assimilation but not directly by its primary three-dimensional variational analysis of atmospheric fields. Instead, a separate analysis of screen-level measurements of dry bulb temperature was made using optimal interpolation (OI). The background 2-m temperature for this analysis was derived from 6-h background forecasts of skin temperature and temperature at the lowest model level (located at a height of ~ 10 m), using Monin-Obukhov similarity profiles consistent with the model's parametrization of the surface layer [Beljaars and Viterbo, 1999]. Details of the optimal interpolation analysis are given in Appendix A.

[5] The 2-m temperature analysis was not used to modify the model level atmospheric fields from which the background forecast for the next analysis in the data assimilation sequence was initiated. It was, however, used together with a similar analysis of 2-m humidity as input to an analysis of soil moisture and temperature [Douville *et al.*, 2000]. It thus influenced the background forecast through the resulting adjustments to the model's soil moisture and soil temperature fields, although the impact of this on the results presented here is assessed to be small.

[6] Monthly mean 2-m temperature anomalies from ERA-40 and NCEP/NCAR have been compared with corresponding gridded values from the CRUTEM2v data set produced directly from monthly station data [Jones and Moberg, 2003], referred to in subsequent sections simply as

Climatic Research Unit (CRU) data. CRUTEM2v is based on anomalies computed for all stations that provide sufficient data to derive monthly climatic normals for the period 1961–1990. Station values are aggregated over $5^\circ \times 5^\circ$ grid boxes and adjustments made for changes over time in station numbers within each box [Jones *et al.*, 1997, 2001].

[7] The ERA-40 analyses are not fully independent of CRUTEM2v. CRUTEM2v uses monthly average (World Meteorological Organization (WMO) CLIMAT message) data provided by 1500–2000 stations. The averaging was carried out by the original data providers using individual temperature measurements, many of which would also have been assimilated in ERA-40. ERA-40 used data from the much larger number of stations that made synoptic observations (most of which were transmitted originally as WMO SYNOP messages) but suffers in early years from gaps in data coverage not present in CRUTEM2v, as discussed in section 3.1. The number of surface observations used per day in ERA-40 varied from $\sim 10,000$ to $\sim 40,000$ during the period of the reanalysis. The basic temperature measurements are processed very differently in CRUTEM2v and ERA-40.

[8] Another surface air temperature data set (HadCRUT2v) is available from the Climatic Research Unit Web site (<http://www.cru.uea.ac.uk>). It includes sea surface temperature (SST) anomalies derived from the analyses of Rayner *et al.* [2003]. As the same SST analyses were used in ERA-40 until late 1981, and similar though not identical analyses from NCEP [Reynolds *et al.*, 2002] were used thereafter [Fiorino, 2004], emphasis in this paper is placed on the predominantly terrestrial regions covered by the CRUTEM2v data set.

[9] Screen-level temperature measurements were not used in the NCEP/NCAR reanalysis; the surface air temperature product was derived instead from analyzed atmospheric values that were constrained primarily by observations of upper air variables and surface pressure. This fact was used by Kalnay and Cai [2003] to interpret the results of comparing surface air temperature over the United States from the reanalysis with corresponding observations from stations below 500-m elevation. They reported quite good agreement as regards interannual variability [see also Kistler *et al.*, 2001] but found significantly less net warming over time in the reanalysis data. They argued that the warming in the surface station data caused by urbanization and land-use change could be a significant factor in explaining the difference between the trends in reanalysis and station values. Their study attracted criticism [Trenberth, 2004; Vose *et al.*, 2004] to which Cai and Kalnay [2004] responded. CRUTEM2v provides a fully independent validation data set in the case of the NCEP/NCAR surface air temperature product.

[10] Some computational details of the present study are given in section 2. Section 3 presents the main comparisons of the temperature analyses. Time series are presented for different geographical regions, maps of the linear trends are discussed, and suspect data values are identified. The consistency between the 2-m temperature analyses, corresponding background values, and analyzed boundary layer temperatures from ERA-40 is discussed in section 4. Section 5 provides further insight through comparison with results from a single atmospheric simulation for the ERA-40

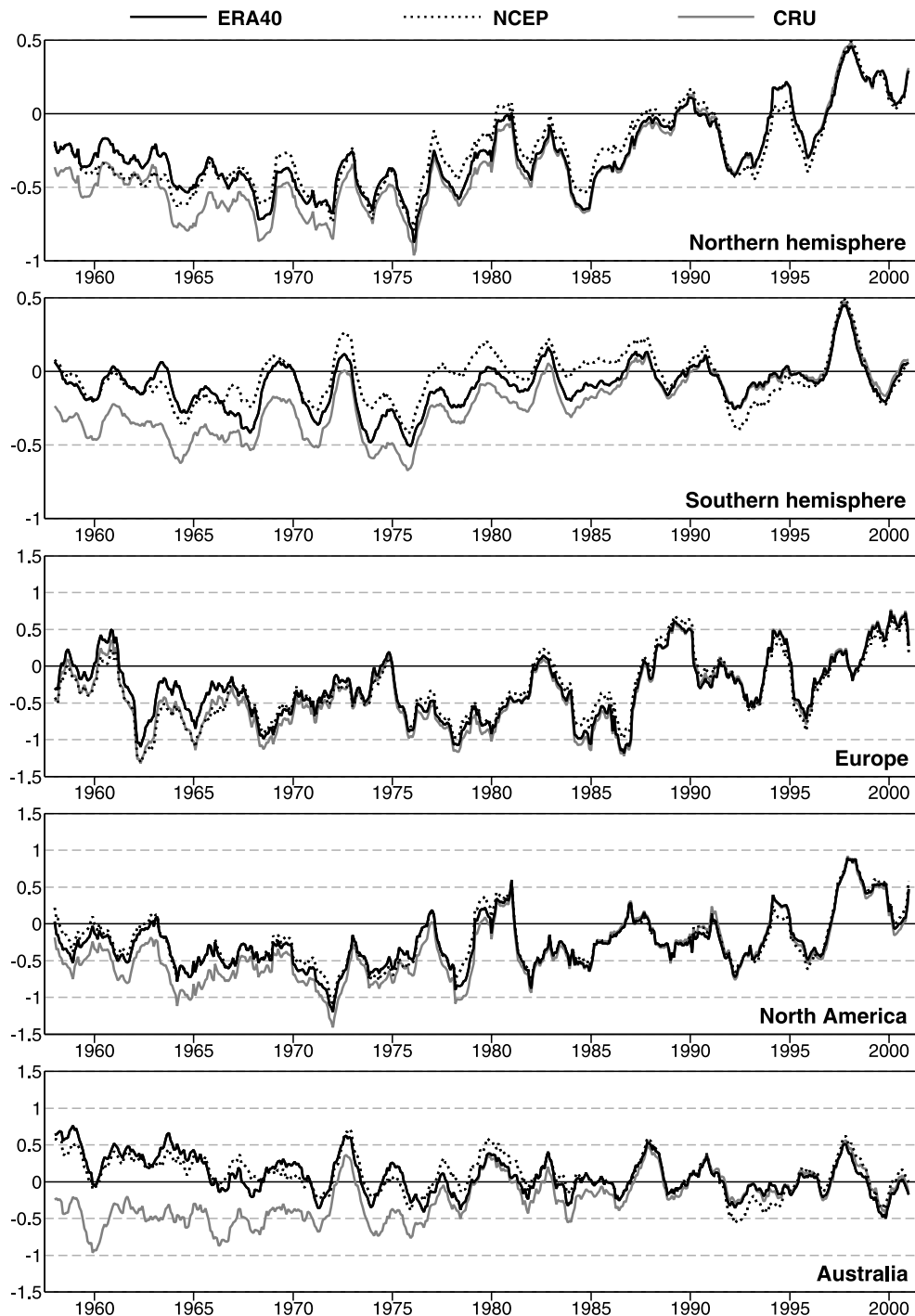


Figure 1. Twelve-month running means of 2-m temperature anomalies from the ERA-40 (black solid), CRU (grey), and NCEP/NCAR (black dotted) analyses, averaged over all CRU grid boxes in the Northern and Southern Hemispheres (first and second panels) and for European, North American, and Australian grid boxes (third, fourth, and fifth panels). Anomalies are defined with respect to monthly climate normals for the reference period 1961–1990. Values are adjusted further to have zero mean over the period 1987–2001. Units are $^{\circ}\text{C}$.

period produced by the model used for the ERA-40 data assimilation. Conclusions are presented in section 6.

2. Some Computational Details

[11] Values of the reanalyses for $5^\circ \times 5^\circ$ grid boxes were needed for comparison with the CRU data. For ERA-40, linear interpolation was used to transform 2-m temperatures from the irregular computational grid of the assimilating model (which has ~ 125 -km resolution) to a finer 0.5° regular latitude/longitude grid. Model level temperatures were evaluated on the 0.5° grid directly from their native T159 spherical harmonic representation. Results presented here used $5^\circ \times 5^\circ$ ERA-40 values formed by averaging over the 0.5° grid. Several calculations were repeated using linear interpolation of 2-m temperatures directly from the ERA-40 to the CRU grid. Little difference was seen. Monthly means of the NCEP/NCAR surface air temperature product defined on a 2.5° grid were downloaded from the National Oceanic and Atmospheric Administration–Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES) Climate Diagnostics Center (<http://www.cdc.noaa.gov>) and averaged onto the $5^\circ \times 5^\circ$ CRU grid.

[12] For any particular month the CRU data set contains values only for grid boxes for which there was at least one station reporting in the box. Except where stated otherwise, comparisons are made using only those reanalysis values for which there is a corresponding CRU value for the month and box in question. The reanalysis values are averages over the analysis times of 0000, 0600, 1200, and 1800 UTC. No account has been taken of model land-sea distributions in producing averages for the CRU grid boxes. Thus for some grid boxes the CRU values are derived from island stations, whereas the reanalysis values are derived from model sea points. For coastal grid boxes the CRU data are based only on observations from land or offshore island stations, whereas the reanalysis values are derived from a mixture of model land and sea points. Air temperatures measured aboard the fixed-position ocean weather ships operated from the 1950s to the late 1990s are included in the CRU analysis; values for the grid boxes in which these ships are located are included in the calculation of domain averages.

[13] The CRU data are anomalies computed with respect to station normals for 1961–1990, a period chosen because station coverage declined during the 1990s. The reanalyses have accordingly been expressed as anomalies with respect to their own monthly climatic means for 1961–1990. Anomalies for the ERA-40 background forecasts and the simulation have been computed with respect to the climate of the ERA-40 analyses. Working with anomalies rather than absolute values avoids the need to adjust for differences between station heights and the terrain heights of the assimilating reanalysis models.

[14] In the time series displayed here, each set of monthly anomalies based on analyzed fields has been adjusted by subtracting the mean value (averaging over all months of the year) for the period 1987–2001. The mean value of the ERA-40 analyses was subtracted in the case of fields from the ERA-40 background forecasts and simulation. The reference period was chosen to be 1987–

2001 to enable time series to be compared in what is arguably the fairest way. Observational counts (except for radiosondes) and quality are generally highest for the most recent years, data assimilation and forecast statistics for ERA-40 indicate best performance then, and temporal variations in the time series are most coherent then. Evidence will be presented of biases in ERA-40 that are relatively large before 1967 and still quite significant over the Southern Hemisphere until the 1980s. Without adjustment, time series of monthly anomalies with respect to 1961–1990 would show a misleading mean discrepancy between CRU and ERA-40 for recent years. Linear trend calculations are hardly affected at all by the adjustment; with or without it they differ only because of missing data for some CRU grid boxes.

3. Comparison of Surface Air Temperature Analyses

3.1. Time Series of Area Averages

[15] Figure 1 shows time series from 1958 to 2001 of 12-month running means of the monthly CRU, ERA-40, and NCEP/NCAR anomalies averaged over all CRU grid boxes in the Northern and Southern Hemispheres and averaged over European, North American, and Australian domains defined by selecting all CRU grid boxes within the regions (35° – 80° N, 10° W– 40° E), (20° – 80° N, 170° – 50° W), and (50° – 10° S, 110° – 160° E). Averages were made with area weighting by the cosine of the central latitude of each grid box. By construction the mean of each time series is zero over the period 1987–2001.

[16] Although all three data sets display similar interannual variability, ERA-40 is the closer of the two reanalyses to the CRU data from 1967 onward. The ERA-40 and CRU curves are especially close toward the end of the period for all regions and are close from the late 1960s onward for Europe and to a lesser degree for North America. The NCEP/NCAR reanalysis is much closer than ERA-40 to the CRU analysis over Europe prior to 1967, whereas for North America and Australia in the early years the two reanalyses are much closer to each other than either is to the CRU analysis. Overall warming trends are smaller for the reanalyses than for CRU.

[17] Table 1 shows least squares linear trends derived from the monthly mean data. They are shown for the full period (1958–2001) and for 1979–2001, which, as discussed in section 4, is when the upper air data used in the reanalyses are best and which is often chosen for study of trends in upper air temperature. It can be seen from Table 1 that the ERA-40 trend in 2-m temperature for the Northern Hemisphere (and for Europe and North America separately) is $\sim 30\%$ smaller than the CRU trend for the full period but within 10% of the CRU trend for 1979–2001. Agreement is less good for the Southern Hemisphere. For Australia, ERA-40 cools over the period as a whole, whereas CRU warms, albeit at a lower rate than for other regions [see also *Jones and Moberg, 2003*]. Whilst there is little to choose between ERA-40 and NCEP/NCAR as regards trends over the whole period, ERA-40 is the closer of the two to CRU trends for 1979–2001. *Santer et al. [2004]* show correspondingly that ERA-40 is the closer of the two reanalyses to upper air (deep tropospheric and lower stratospheric)

Table 1. Linear Trends for the CRU, ERA-40, and NCEP/NCAR Analyses and for the ERA-40 Background Forecasts^a

	Northern Hemisphere	Southern Hemisphere	Europe	North America	Australia
1958–2001					
CRU	0.19	0.13	0.17	0.21	0.14
ERA-40	0.13	0.04	0.11	0.14	−0.10
NCEP/NCAR	0.14	0.03	0.16	0.11	−0.08
Background	0.09	0.03	0.06	0.14	−0.09
1979–2001					
CRU	0.30	0.11	0.46	0.30	0.01
ERA-40	0.27	0.04	0.42	0.28	−0.10
NCEP/NCAR	0.19	−0.06	0.31	0.25	−0.12
Background	0.23	0.04	0.37	0.30	−0.07

^aLinear trends are in °C/decade.

temperature trends derived from the microwave sounding data available from 1979 onward.

[18] The discrepancy between the ERA-40 and CRU curves in Figure 1 is most marked before 1967. This is related to limited availability of surface observations for ERA-40 combined with a near-surface warm bias in the background forecasts. Most of the observations from before 1979 were supplied for assimilation in ERA-40 by NCAR, whose holdings for the early years of ERA-40 had some serious deficiencies at the time of data supply, with very few synoptic reports from Australia and several European countries, for example. Coverage in fact declines from 1958 to 1966. Data from many countries can be seen to be missing in the data coverage for the 1200 UTC analysis for 1 July 1966 shown in Figure 2 (left). This example is typical for the years 1965 and 1966. Many more observations were supplied from 1967 onward, initially from NCAR's copy of a U.S. Air Force archive. They not only filled national gaps but also increased the density of coverage generally. The number of observations jumps on 1 January 1967 and increases during subsequent months; the coverage for 1 July 1967 is shown in Figure 2 (right).

[19] Antarctic observations for the early years were provided not only by NCAR but also by the Australian Bureau of Meteorology and the British Antarctic Survey. Data from the latter source were used by *Jones and Moberg* [2003] in producing the CRUTEM2v data set; see also *Turner et al.* [2004]. Because of a technical problem, not all were assimilated in ERA-40.

[20] A complete set of data coverage maps showing observation frequencies month by month for each observation type can be viewed on the project Web site (<http://www.ecmwf.int/research/era/>) in the section on monitoring. The Web site plots of surface synoptic data coverage, unlike those in Figure 2, include stations reporting only snow depth. Snow depth observations were analyzed in ERA-40 to provide initial snow conditions for the background forecasts, and thereby they locally influence background 2-m temperatures and hence the 2-m temperature analyses. Snow data were limited to Canada for the early years of ERA-40. Data for the former Soviet Union were available from 1966 onward, but data for other countries could be used only from 1976 onward.

[21] Month-to-month variability in the ERA-40 and CRU analyses is very similar throughout the period. Table 2 shows correlations and standard deviations between the time series of monthly anomalies, after removing linear trends. Results are presented for the periods 1958–2001 and 1979–2001. Over the full period there is much better agreement between ERA-40 and CRU for the Northern than for the Southern Hemisphere, with correlations of 99.6% for Europe and 92.5% for Australia. Agreement is distinctly better for the Southern Hemisphere when the comparison is restricted to 1979–2001; the correlation for Australia increases to 97.3%. The standard deviation for Australia reduces from 0.22°C for the full period to 0.13°C for 1979–2001. Correlations between NCEP/NCAR and CRU are substantially lower than between ERA-40 and CRU, and standard deviations are substantially higher, except for Australia over the full period.

[22] Figure 3 presents time series of monthly CRU anomalies and of differences between ERA-40 and CRU and between NCEP/NCAR and CRU, averaged over Europe, North America, and Australia. The differences for Europe and North America are small compared with month-to-month variations in the CRU values, indicating good general performance of both reanalyses, although ERA-40 tends to be the closer of the two to CRU from 1967 onward. In contrast, the two sets of differences are similar in overall magnitude for Australia until the late 1970s, after which ERA-40 again tends to be the closer to CRU data. There are also larger intraannual variations in the NCEP/NCAR differences. As the mean annual cycle for 1961–1990 is subtracted from each data set, this indicates greater vari-

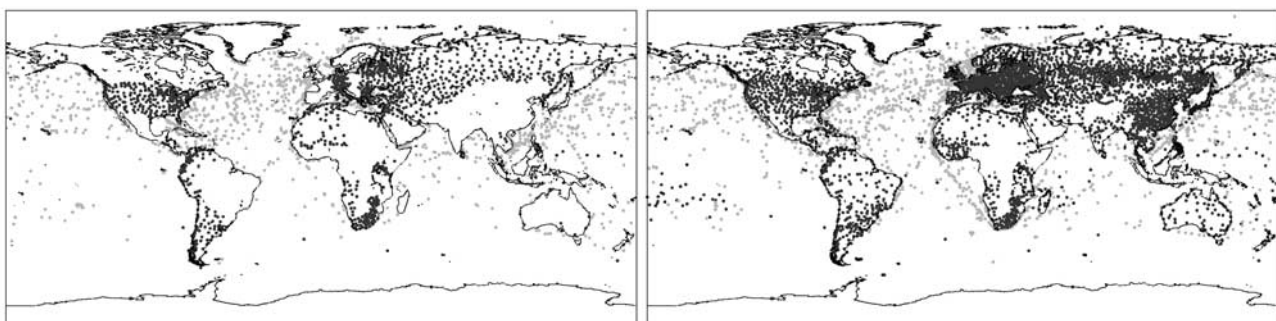


Figure 2. Coverage of surface synoptic observations from land stations (black dots) and ships (grey dots) supplied to the ERA-40 data assimilation for (left) 1200 UTC 1 July 1966 and (right) 1200 UTC 1 July 1967.

Table 2. Correlation and Standard Deviation Between the Monthly CRU Analyses and the ERA-40 Analyses, the NCEP/NCAR Analyses, and the ERA-40 Background Forecasts, With Linear Trends Removed^a

	Northern Hemisphere	Southern Hemisphere	Europe	North America	Australia
Correlation, 1958–2001					
ERA-40	98.3	92.6	99.6	98.7	92.5
NCEP/NCAR	93.4	87.6	98.1	96.7	94.1
Background	96.0	88.9	99.1	97.2	90.4
Standard deviation, 1958–2001					
ERA-40	0.06	0.09	0.09	0.12	0.22
NCEP/NCAR	0.12	0.11	0.18	0.19	0.19
Background	0.10	0.11	0.14	0.18	0.24
Correlation, 1979–2001					
ERA-40	98.8	96.3	99.6	99.1	97.3
NCEP/NCAR	95.2	88.0	98.8	97.1	94.0
Background	97.7	91.9	99.3	98.0	95.4
Standard deviation, 1979–2001					
ERA-40	0.05	0.06	0.08	0.11	0.13
NCEP/NCAR	0.10	0.11	0.15	0.20	0.20
Background	0.07	0.09	0.13	0.17	0.17

^aCorrelation is in %, and standard deviation is in °C.

ability over time in the annual cycle in the NCEP/NCAR analyses.

[23] *Betts and Beljaars* [2003] document a subset of near-surface ERA-40 data for 1986–1995 produced in support of

the Second International Land Surface Climatology Project (ISLSCP-II). They briefly discuss the agreement between the 2-m temperature analyses included in the data set and a gridded ISLSCP-II data set of monthly mean surface tem-

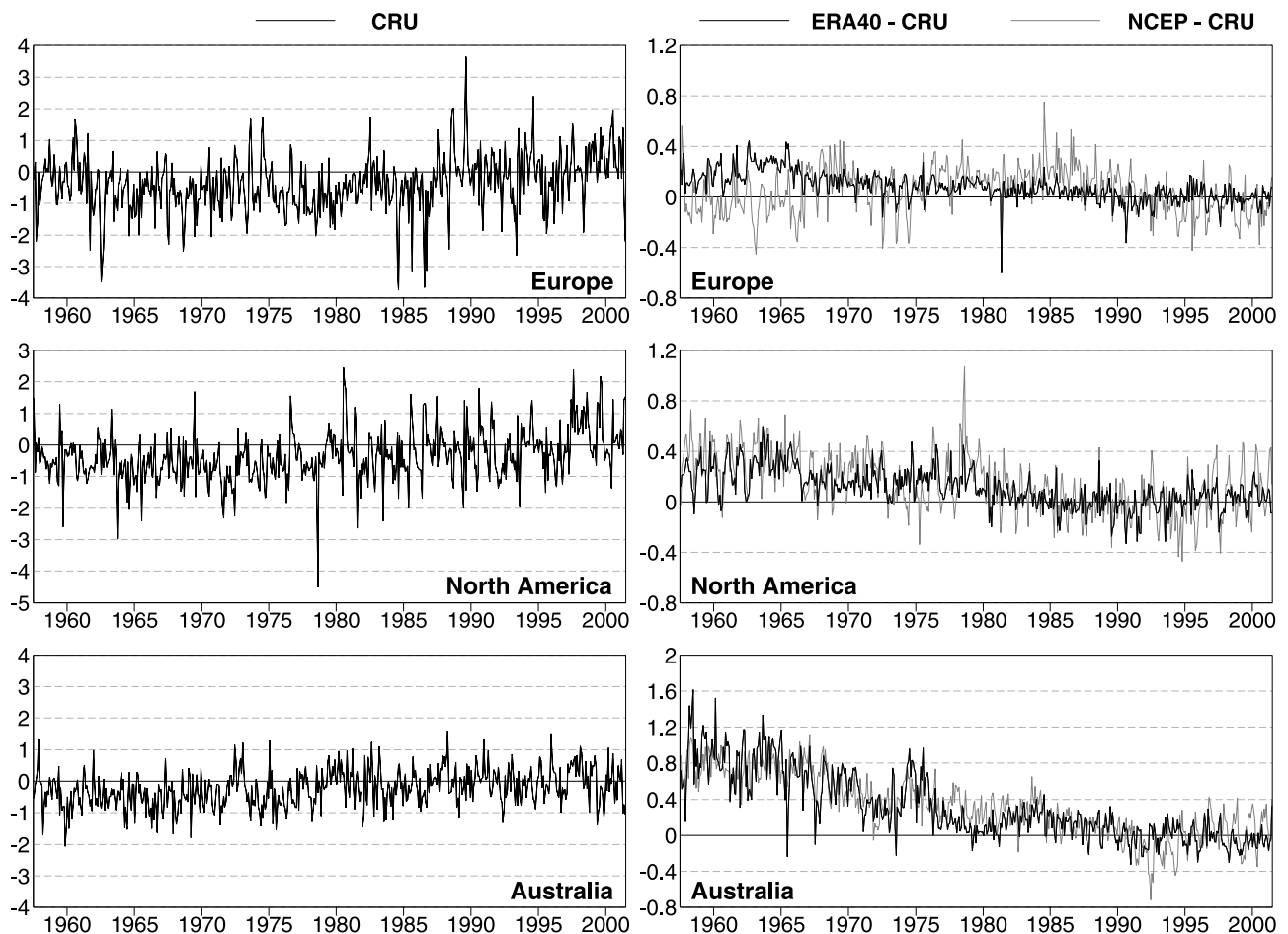


Figure 3. (left) Two-meter temperature anomalies from the monthly CRU analyses, averaged over all (top) European, (middle) North American, and (bottom) Australian grid boxes. (right) Corresponding differences between the ERA-40 and CRU analyses (black) and between the NCEP/NCAR and CRU analyses (grey). Values are adjusted to have zero mean over the period 1987–2001. Units are °C.

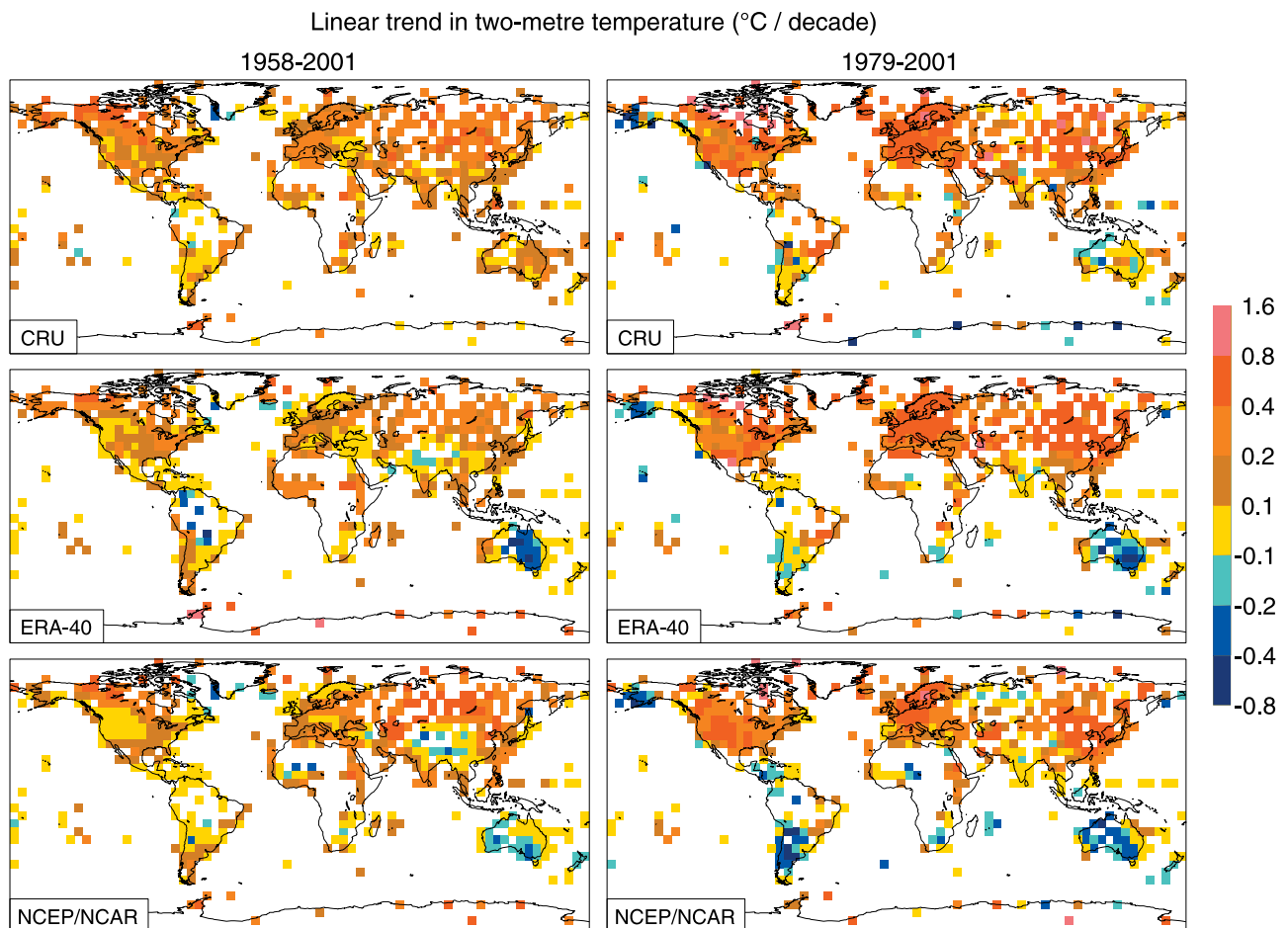


Figure 4. Linear trend over the periods (left) 1958–2001 and (right) 1979–2001 in 2-m temperature computed using (top) CRU data and equivalents from the (middle) ERA-40 and (bottom) NCEP/NCAR reanalyses. Units are $^{\circ}\text{C}/\text{decade}$.

peratures derived from an earlier Climatic Research Unit data set [New *et al.*, 1999, 2000]. Although absolute seasonal mean values show some differences, particularly in regions of high terrain, close agreement is seen in the sample maps of seasonal anomalies presented by Betts and Beljaars.

3.2. Geographical Distribution of Trends

[24] Maps of least squares linear trends of the CRU, ERA-40, and NCEP/NCAR anomalies are shown in Figure 4 for the periods 1958–2001 and 1979–2001. Values are plotted only for grid boxes for which there is a quite complete temporal record in the CRU data set, excluding grid boxes if data from more than 48 months were missing for the 1958–2001 trend or if more than 24 months of data were missing for 1979–2001. As linear trend calculations can be highly sensitive to data values close to the end points of time series, differences between 11-year means for 1958–1968, 1969–1979, 1980–1990, and 1991–2001 have also been examined. This confirmed the findings for linear trends reported in this section.

[25] There is reasonable agreement between the three data sets for many features of the linear trends for 1958–2001. All exhibit predominant warming over Eurasia and North America. ERA-40 shows a pronounced (and almost certainly

erroneous) cooling over much of Australia. There is also strong cooling in ERA-40 for tropical South American grid boxes east of the Andes where, as for Australia, there were few surface observations available for assimilation prior to 1967. NCEP/NCAR is closer than ERA-40 to CRU full period trends for Australia and tropical South America. Both ERA-40 and NCEP/NCAR show more warming than CRU at several of the small number of Antarctic grid boxes.

[26] ERA-40 does not show the warming over the United Kingdom, Norway, and Sweden that is seen in the CRU data for 1958–2001. Here too there was relatively poor or nonexistent coverage of synoptic observations prior to 1967 in the data sets supplied to ERA-40. Similar behavior is seen, however, for the NCEP/NCAR reanalysis, which did not use screen-level temperature measurements at any time.

[27] Figure 4 shows much closer agreement between ERA-40 and CRU trends for 1979–2001, as has already been discussed for the regional means. This is particularly the case for northwestern Europe, tropical South America, and Antarctica. The disagreement for Australia is less marked, but ERA-40 still has many more Australian grid boxes with cooling, as does NCEP/NCAR. The CRU trends for 1979–2001 are matched more closely by ERA-40 than by NCEP/NCAR in several regions: around the Mediterra-

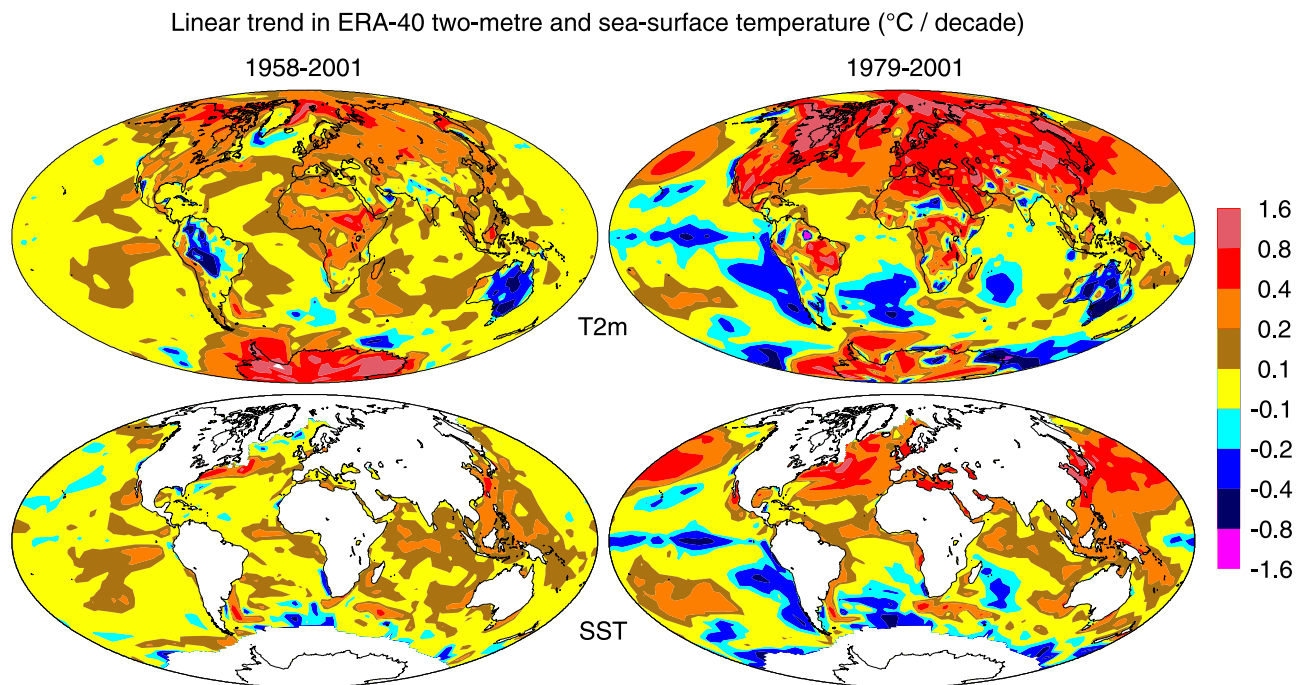


Figure 5. Linear trends over the periods (left) 1958–2001 and (right) 1979–2001 in the (top) ERA-40 2-m temperature analyses and in the (bottom) sea surface temperature analyses used by ERA-40. Trends in sea surface temperature are shown only for points that are free of ice for every month in each period. Units are $^{\circ}\text{C}/\text{decade}$.

nean and over Chile and Argentina, for example. Overall, ERA-40 is the closer of the two reanalyses to CRU at 64% of the CRU grid boxes for the 1979–2001 trend. The corresponding value is 55% for the 1958–2001 trend.

[28] The reanalysis data assimilation systems provide a complete spatial and temporal record of 2-m temperature. Figure 5 (top) shows maps of the complete global trends from ERA-40 for 1958–2001 and 1979–2001. There is spatially coherent warming at northern latitudes, encompassing both land and sea ice areas where there are missing values in the CRUTEM2v data set. There is also a pronounced warming over Antarctica, particularly for 1958–2001. The warming over high-latitude northern regions, over southern Africa, and over Antarctica causes the global mean warming to be larger when averages are taken over all land areas than when taken only over the grid boxes where there are CRU data. It has, however, already been noted that ERA-40 warms more than CRU at several of the few Antarctic grid boxes for which comparison can be made. Other results that cast doubt on the ERA-40 trend over Antarctica are presented later.

[29] The trends in ERA-40's 2-m temperature analyses over the oceans are not surprisingly very similar to the trends in the externally produced sea surface temperature analyses used by the ERA-40 data assimilation. Maps of the trends in SST are presented in Figure 5 (bottom). The regions of warming and cooling over the oceans for 1979–2001 match quite well regions of warming and cooling seen both in measured layer average tropospheric temperatures from channel 2 of the satellite-borne Microwave Sounding Unit and in equivalents derived from the ERA-40 analyses [Santer *et al.*, 2004, Figure 11]. Cooling

earlier in the period south of Greenland is consistent with temporal shifts in the analyzed flow patterns over the northwestern Atlantic (not illustrated).

[30] The 2-m temperature analysis shows a weaker warming trend than the SST analysis over the Indian and tropical western Pacific Oceans. An intermediate trend is found for the background field, indicating that the screen-level temperature observations assimilated in ERA-40 counteract some of the trend imposed by the SST analysis in this region. Analyzed daytime temperatures are higher than background values where air temperature observations from ships were used, which is probably because no correction was applied for the unrepresentative nature of the daytime measurements due to solar heating (S. Tett, personal communication, 2004). As there are changes over time both in ship size and in the amount of ship and buoy data, differences in trend between ERA-40's 2-m marine temperature analyses and the SST analyses cannot be regarded as reliable.

3.3. Identification of Suspect Values

[31] Comparison of the CRU, ERA-40, and NCEP/NCAR analyses has identified a small number of highly suspect values in these data sets. For example, the difference between the ERA-40 and CRU anomalies for Europe plotted in Figure 3 is atypically large for November 1981. The similar difference between NCEP/NCAR and CRU is masked by the dip in the ERA-40 plot. The cause of the difference has been identified as erroneous values from several Turkish stations entering the CRU analysis for this month. The same wrong data are also found in the standard WMO publications *Monthly Climatic Data for the World* and *World Weather Records*.

Table 3. North American Grid Boxes for Which the Difference Between NCEP/NCAR and CRU Anomalies Exceeds 10°C in Magnitude^a

Month	Latitude, °N	Longitude, °W	CRU	ERA-40 – CRU	NCEP – CRU
Jan. 1963	75–80	70–65	24.9	–19.6	–20.2
Feb. 1964	75–80	70–65	–13.8	17.8	18.5
May 1961	75–80	70–65	14.3	–16.1	–16.7
May 1973	55–60	55–50	16.5	–16.1	–16.2
Jan. 1995	55–60	160–155	12.5	–11.3	–12.2
Jan. 2001	70–75	80–75	–3.0	1.0	12.2
Feb. 2001	70–75	80–75	–1.7	0.8	12.0
Dec. 1980	70–75	125–120	–2.8	1.3	10.7
Dec. 1980	70–75	130–125	–2.1	0.1	10.2
Dec. 2000	70–75	80–75	0.0	2.4	10.2

^aEntries are ordered by the magnitude of these differences. Values of the CRU anomaly and its differences from ERA-40 and NCEP/NCAR values are given in °C for each grid box.

[32] Table 3 shows the ten North American points and data values for which the NCEP/NCAR anomaly differs from the CRU anomaly by more than 10°C. All are at high latitudes in winter or spring, suggesting problems with snow or sea ice fields. The points clearly divide into two groups of five. The first is characterized by highly anomalous CRU values and ERA-40 values that are similar to NCEP/NCAR values, suggestive of erroneous (or highly unrepresentative) CRU values. The second is characterized by highly anomalous NCEP/NCAR values and ERA-40 values that are similar to CRU values. This suggests a problem in the NCEP/NCAR reanalysis for these five values. Three refer to a single grid box for the winter months of 2000/2001. Two refer to neighboring grid boxes for December 1980. There is no North American data point other than the first five shown in Table 3 at which the ERA-40 anomaly differs from the CRU anomaly by more than 10°C.

[33] Overall, there are 14 points in the whole data set where the ERA-40 anomaly differs from the NCEP/NCAR anomaly by more than 10°C. Ten are over North America, and it is the ERA-40 anomaly that is closer to the CRU anomaly for each of them. Four are over Antarctica, and for three of them it is the NCEP/NCAR anomaly that is closer to the CRU anomaly. There are 35 points in the whole data set where the ERA-40 anomaly differs from the CRU anomaly by more than 10°C. For all but four points over Antarctica it is the CRU anomaly that is the larger of the two. The total number of data points is 388,315. Several errors in station values entering the CRU analysis other than those from Turkey in 1981 have been identified by this three-way comparison of CRU and reanalysis values. The Climatic Research Unit is currently working to correct values that are clearly in error and revise station normals and is collaborating with the Hadley Centre of the Met Office to produce a new HADCRUT3 data set by the end of 2004.

4. Comparisons of Background and Analyzed ERA-40 Temperatures

[34] The significance of the agreement between ERA-40 and CRU would be limited if it resulted overwhelmingly from ERA-40's explicit analysis of screen-level temperature measurements. If it did, there would be little significance to the differences between the ERA-40 and NCEP/NCAR

reanalyses, as the latter did not utilize such observations. It is thus desirable to establish the extent to which the agreement between ERA-40 and CRU comes from information brought forward in the background forecasts of the data assimilation and the extent to which variations in the ERA-40 temperature analyses are coherent through the boundary layer. Evidence that trends and low-frequency variability in these quantities are insensitive to the analysis of screen-level temperatures is presented at the end of this section.

[35] Twelve-month running means for the background forecasts are plotted in Figure 6 together with corresponding values for the ERA-40 and CRU analyses. The background values are plotted as anomalies with respect to mean analyzed values to show the sign of the “analysis increments,” the differences between analysis and background forecast, or, equivalently, the direct impact of the analyzed observations. The results shown are averages over all Northern and Southern Hemispheric grid boxes at which there are CRU data, over European, North American, and Australian subsets as previously, and also over the small number of Antarctic grid boxes (boxes south of 60°S) in the CRU data set.

[36] Background values are generally warmer than or similar to analyzed ERA-40 values early in the period. Quite pronounced cooling analysis increments persist throughout in the average for North America, and smaller cooling increments persist for Australia. Conversely, the increment shifts from cooling to warming for the Northern Hemisphere as a whole and for Europe in particular. Antarctica differs from other regions in that the ERA-40 analysis is relatively cold early on, warming by around 1.5°C in the late 1970s. Here analysis increments warm the ERA-40 background systematically from 1973 to 1980.

[37] Figure 6 shows much better agreement between ERA-40 and CRU for Australia as well as Antarctica from the late 1970s onward. The overall observing system for the Southern Hemisphere was dramatically improved around the end of 1978, with better satellite temperature and humidity sounding, new wind estimates from geostationary satellite imagery, new surface observations from drifting buoys, and increased data from commercial aircraft. Characteristics of the ERA-40 data assimilation and the accuracy of medium-range forecasts initiated from the ERA-40 analyses are much improved after 1978 [e.g., Simmons, 2003]. Others have noted improved agreement between

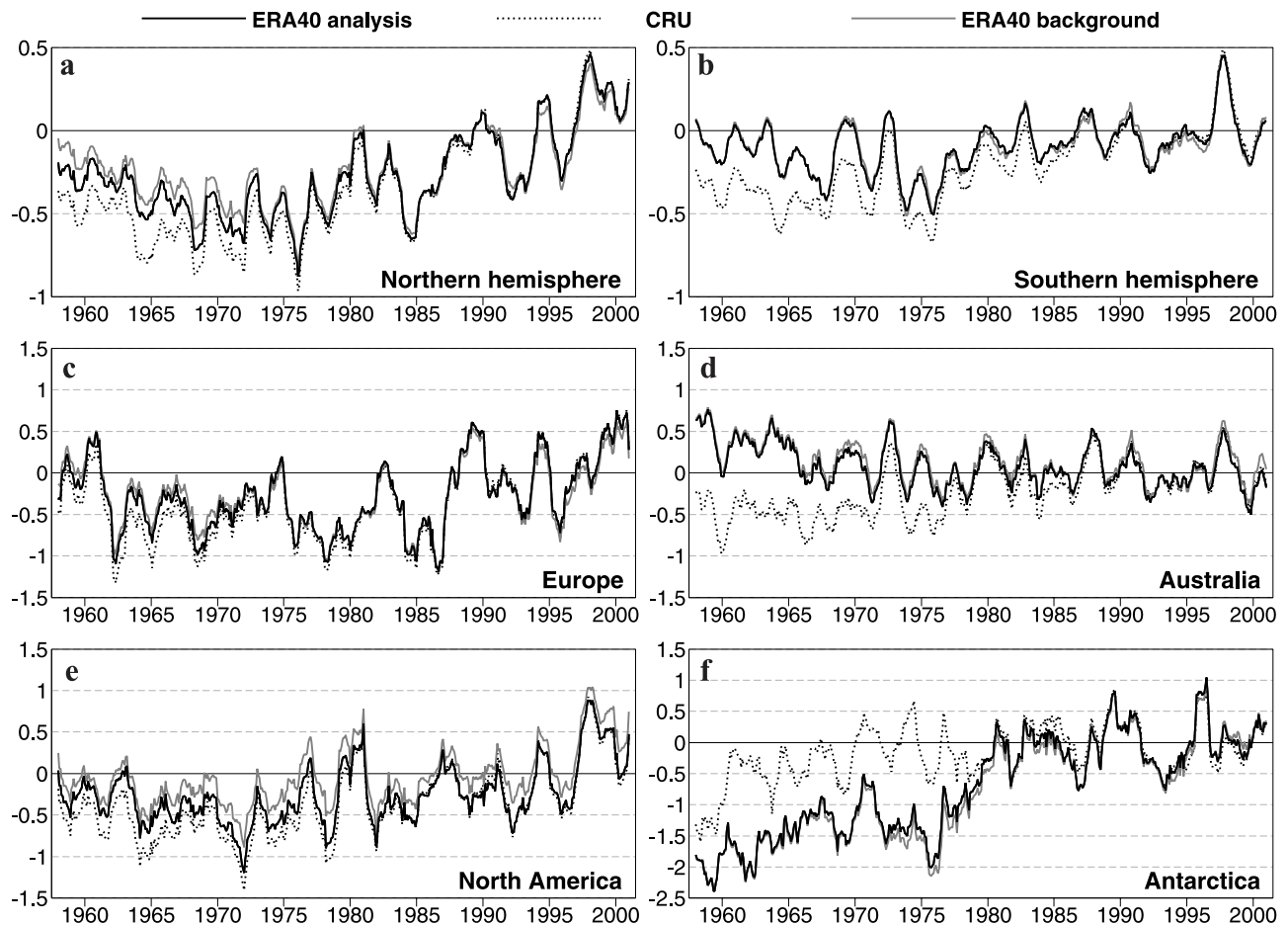


Figure 6. Twelve-month running means of 2-m temperature anomalies from the ERA-40 (black solid) and CRU (black dotted) analyses and from the ERA-40 background forecasts (grey), averaged over all CRU grid boxes in the (a) Northern and (b) Southern Hemispheres and averaged over (c) European, (d) Australian, (e) North American, and (f) Antarctic grid boxes. Time series based on analyses are adjusted as before to have zero mean for the period 1987–2001. ERA-40 background anomalies are computed with respect to the monthly climate derived from the ERA-40 analyses for the reference period 1961–1990 and then adjusted by the same single amount applied to adjust the ERA-40 analysis time series to zero mean value for 1987–2001. Units are $^{\circ}\text{C}$.

ERA-40 and extratropical Southern Hemispheric station data [Bromwich and Fogt, 2004] and improved agreement between ERA-40 and NCEP/NCAR [Sterl, 2004] from 1979 onward.

[38] Statistics for the ERA-40 background temperatures are included in Tables 1 and 2. The background trends are generally weaker than the trends in the ERA-40 analyses, although North America is an exception. The CRU trends for 1979–2001 are nevertheless matched more closely by the ERA-40 background trends than by the trends in the NCEP/NCAR analyses. Month-to-month variations are quite well captured in the ERA-40 background. The correlations with CRU values are mostly higher for the ERA-40 background than for the NCEP/NCAR analysis, and standard deviations are correspondingly mostly smaller. The better fit that ERA-40 provides to the CRU data is not entirely a direct consequence of ERA-40's analysis of surface observations; information carried in the background forecasts is a factor also.

[39] Figure 7 presents maps of annual mean analysis increments in temperature at 2 m, at the lowest model level (level 60, at $\sim 10\text{-m}$ height), and at level 49, which is the model level closest to 850 hPa for a surface pressure close to 1000 hPa. Results for 1958 and 2001 are shown. The separate OI analysis of surface observations generally produces much larger local mean increments at 2 m than those that are produced by the main variational analysis either at the lowest model level or close to the top of the boundary layer. The increased availability of surface observations results in more widespread mean 2-m temperature increments over Australia, Antarctica, and Brazil in 2001 than in 1958. The warming over the oceans from the analysis of shipboard measurements can also be seen.

[40] Greater availability of observations over the Northern Hemisphere results in widespread mean increments at 2 m in 1958 as well as 2001 for this hemisphere. Increments are more widespread away from the surface also. The increments from the variational analysis at level 60 are

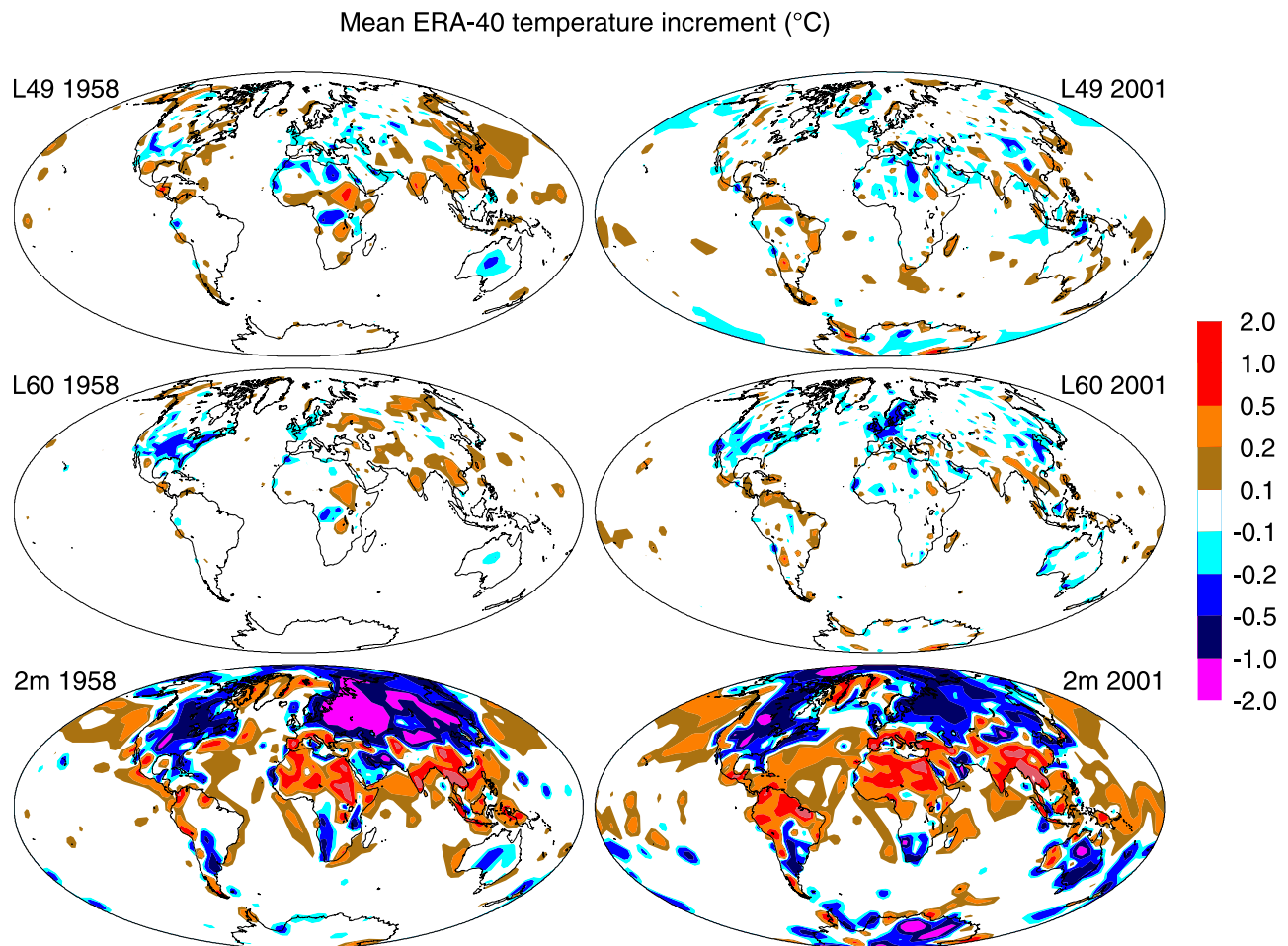


Figure 7. Annual mean ERA-40 temperature analysis increments (analysis minus background values) for (left) 1958 and (right) 2001 at (top) model level 49 (level located close to 850 hPa for a surface pressure of 1000 hPa), at (middle) model level 60 (level located at a height close to 10 m), and at (bottom) a height of 2 m. Units are $^{\circ}\text{C}$.

largely consistent in sign with those from the OI 2-m analysis for 2001. This is seen also over North America in 1958, but over Russia in 1958, there is a large mean cooling increment at 2 m but mostly a (weaker) warming at level 60. The general pattern of cooling increments at 2 m over the United States, Canada, and northern Eurasia and warming further south is characteristic of other years examined. Over Eurasia, there is a decrease over time in the extent and intensity of the cooling increment and an increase in the warming increment. This results in the shift over time of the net increment from cooling to warming in the averages for Europe and the Northern Hemisphere shown in Figure 6.

[41] The increase in the warming increment over southern Europe occurs primarily in summer daytime analyses, whereas that over southern Asia occurs primarily in winter nighttime analyses. A shift in bias of the background forecasts could, following *Kalnay and Cai* [2003], be associated with unmodeled changes in land use and urbanization, but as *Trenberth* [2004] has pointed out, there are other reasons why such a shift might occur. In particular, ERA-40 exhibits a marked upward trend in water vapor at low latitudes and an increasingly excessive tropical rainfall,

associated with assimilation of increasing volumes of satellite data [*Andersson et al.*, 2004]. Changes in water vapor, cloud, and circulation are all candidates for changing biases in background temperature.

[42] The low-latitude warming increments extend throughout the tropics. As analysis increments do not fully compensate for bias in background temperatures, the pattern of increments implies an overall warm bias in the ERA-40 analyses at middle and high latitudes and an overall cold bias at low latitudes. Just such a pattern is seen in differences in mean temperatures for 1986–1995 between the ISLSCP-II data sets from ERA-40 and *New et al.* [1999, 2000], as illustrated by *Betts and Beljaars* [2003]. An exception to the picture of high-latitude warm bias occurs around Greenland, where the annual mean increment maps show warming and the ERA-40 analyses are colder than those of *New et al.* in wintertime.

[43] Figure 8 shows times series of ERA-40 temperature anomalies at 2 m, level 60, and level 49, averaged over CRU grid boxes for comparison with plots shown in Figures 1, 3, and 6. Variations at 2 m are matched quite closely by variations throughout the planetary boundary layer, apart from an overall shift in values during the 1970s

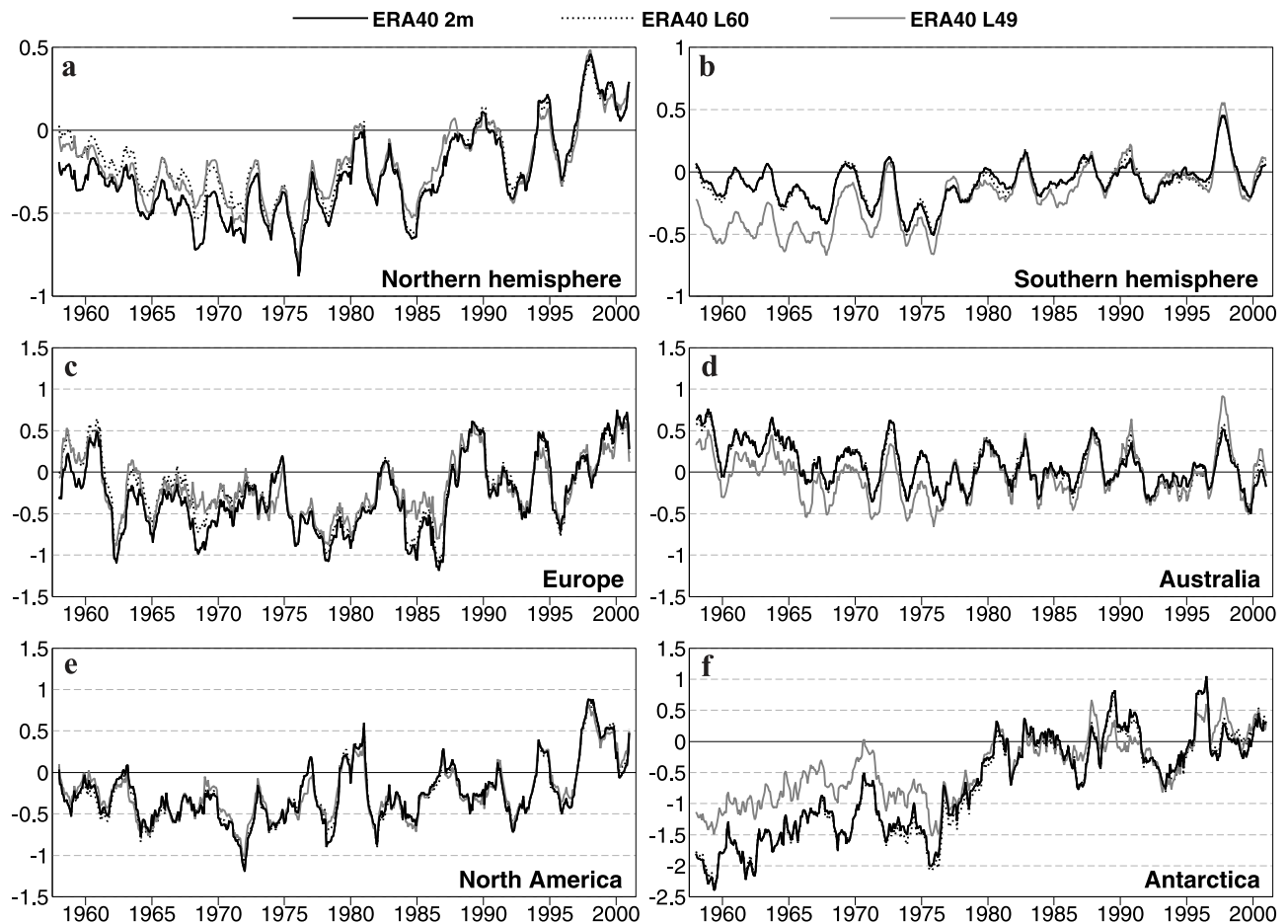


Figure 8. Twelve-month running means of temperature anomalies from the ERA-40 analyses at 2-m height (black solid) and at model levels 60 (black dotted) and 49 (grey). Values are averaged over all CRU grid boxes in the (a) Northern and (b) Southern Hemispheres and averaged over (c) European, (d) Australian, (e) North American, and (f) Antarctic grid boxes. Values for each level are adjusted to give zero mean value over the period 1987–2001. Units are $^{\circ}\text{C}$.

for some regions. Agreement is close throughout the period for North America. Both level 60 and level 49 temperatures are relatively warm early in the period for the Northern Hemisphere as a whole, as seen also for background temperatures. Two-meter and level 60 temperatures vary similarly throughout for the Southern Hemisphere, but temperature differences between level 49 and the surface are larger earlier in the period than later for this hemisphere, associated with a bias which, in the early years at least, switches from warm near the surface to cold (as shown in section 5) at higher levels. Antarctica is an exception.

[44] *Kalnay and Cai* [2003] attributed the underestimation of surface warming over the continental United States in the NCEP/NCAR reanalysis to an unmodeled and unanalyzed effect of urbanization and land use change. Such changes are also not modeled in ERA-40, but their effect may be felt through the analysis of screen-level temperature measurements. If significant net warming were to have been caused by changes in surface character, it would be expected that there would be a warming trend in the analysis increment over the course of ERA-40, as the observations force a warming that would otherwise have been underestimated. This does not happen over North America, where

the trend in the background forecast is very similar to that in the analysis. Moreover, if there were to be significant surface warming due to urbanization and land use change, warming would not be expected to be as strong throughout the planetary boundary layer. *Kalnay and Cai* [2003] themselves noted that weaker warming measured at upper levels could be partially explained by a predominance of land use effects over greenhouse warming near the surface. For ERA-40 the temperature changes at a model level close to the top of the boundary layer are very similar to those at the surface over North America.

[45] Averages have also been calculated for CRU grid boxes covering only the eastern United States from 100° to 70°W and 25° to 45°N , a region that contains most of the U.S. stations below 500 m examined by *Kalnay and Cai* [2003]. Following these authors, temperature differences between means for 1980–1999 and 1960–1979 have been computed. Values are 0.44° , 0.34° , and 0.20°C for the CRU, ERA-40, and NCEP/NCAR analyses, respectively. Other differences for ERA-40 are 0.38°C for the background forecasts, 0.40°C for the level 49 analysis, and 0.37°C for the 500-hPa analysis. The larger warming aloft and in background forecasts for ERA-40 makes it difficult to

ascribe much of the discrepancy between the CRU and ERA-40 surface warmings to unmodeled urbanization and land surface change. Some such effect cannot be ruled out, but the results presented here do not provide confidence in the estimate made by Kalnay and Cai. The increase in 500-hPa temperature in the NCEP/NCAR reanalysis is in fact higher still, 0.54°C . This is larger than for ERA-40 because of a shift of $\sim 0.2^{\circ}\text{C}$ in the difference between the two 500-hPa analyses around 1979, when there was a substantial change to the observing system.

[46] The ERA-40 background forecasts of 2-m temperature and analyses of temperature higher in the boundary layer could only have been affected by the separately analyzed screen-level observations through the influence of the screen-level analysis on the soil moisture and soil temperature analyses. Evidence indicates that any such effect is unimportant in the context of this paper. A trial assimilation was carried out for a 34-day period during the boreal summer of 1999 immediately prior to implementation of the screen-level temperature analysis in the operational ECMWF forecasting system. The new analysis caused a mean shift in analyzed temperature near the top of the boundary layer of only 0.01°C over Europe and 0.02°C over North America, and corresponding changes in background 2-m temperatures were 0.06° and 0.08°C , respectively. By design the soil moisture analysis used for ERA-40 was insensitive to the screen-level temperature analysis in winter at middle and high latitudes [Douville *et al.*, 2000], and the temperature of the soil did not influence atmospheric fields where there was snow cover, yet the results presented for 12-month running means in Figures 6 and 8 are found to hold separately for winter as well as summer when more detailed plots without the 12-month averaging are examined. Twelve-month mean analysis increments in 2-m temperature for North America are seen to be relatively large in Figure 6, but the increments occur mostly in winter. The soil analysis for Antarctica could have had no effect on atmospheric fields, as the assimilating model had permanent snow cover there.

5. Comparisons of Analyzed and Simulated ERA-40 Temperatures

[47] It is instructive to compare the ERA-40 analyses with a simulation of the atmosphere for the ERA-40 period that has been carried out using the same model and the same analyses of SST and sea ice cover as employed for the ERA-40 data assimilation. This provides evidence of shifts in the analyses that can be related to changes in the observing system or in the treatment of observational biases and evidence of the extent to which variability and trends in the analyses can be regarded as forced either by variability and trends in the SST/sea ice analyses or by the trends in specified, radiatively active gases that were included in the ERA-40 model based on the 1995 Intergovernmental Panel on Climate Change assessment [Houghton *et al.*, 1996]. The model did not include any aerosol trend or variability due to volcanic eruptions, and there was no interaction between the model's radiation scheme and ozone fields. Instead, a fixed geographical distribution of aerosol and a climatological ozone distribution were used in the radiation calculation.

There was also no variation over time in the model's vegetative characteristics.

[48] Figure 9 compares time series of 2-m temperature anomalies from the simulation and from the ERA-40 and CRU analyses. The simulation is presented as an anomaly relative to mean analyzed ERA-40 values. In the hemispheric means the simulation is generally quite close to the analyzed values, rarely deviating by more than 0.5°C in the 12-month running means. The simulation is at most times warmer than the ERA-40 analysis for the Northern Hemisphere, as are background and analyzed level 60 values early in the period. A warm near-surface bias of the assimilating model, incompletely corrected by the analyzed observations, thus appears to be at least partly responsible for near-surface temperatures being relatively warm in the ERA-40 analyses early in the period.

[49] Some larger differences between analysis and simulation are seen regionally, reaching up to 1.5°C in the 12-month means for Europe and Antarctica. The differences over Europe are quite similar throughout the period. In contrast, over Antarctica the ERA-40 analysis differs little from the simulation prior to the late 1970s and is substantially warmer thereafter. By construction the time series for the CRU and the ERA-40 analyses both have zero mean for the period 1987–2001, but they in fact agree quite closely from 1979 onward for Antarctica to within 0.5°C for each 12-month running mean.

[50] A different behavior occurs for Australia. Here, the simulation is warmer than the ERA-40 analysis later in the period, but it is colder for the early years and closer then to the CRU analysis. The warm bias of the early ERA-40 analyses over Australia cannot be ascribed simply to an inherent bias in the climate of the assimilating model and must be related to a characteristic either of the data assimilation or of the observations. Further work is needed to understand this feature of ERA-40 and the similar behavior of the NCEP/NCAR reanalysis for Australia.

[51] There is, not surprisingly, poor agreement between the simulation and CRU as regards short-term variability of monthly means. There are nevertheless quite substantial correlations between the time series of 12-month running means from the simulation and the CRU data, 76 and 82% for the Northern and Southern Hemispheric averages, respectively, and 51% for the European average, with the linear trend removed. This almost certainly reflects a significant net influence of SST anomalies on anomalies of surface air temperature over land for sufficiently large space and time averages. The agreement between the simulation and the CRU data thus provides partial validation of the variability of the SST analyses used in ERA-40 [see also Folland *et al.*, 2001].

[52] Maps of the linear trend from the simulation for 1958–2001 and 1979–2001 are presented in Figure 10. The simulation reproduces the predominant signal of strong warming over Northern Hemisphere landmasses seen in the CRU and ERA-40 analyses, including the larger values seen at higher latitudes and for 1979–2001. Moreover, consistent with Figure 9, the maps from the simulation show neither the strong cooling over Australia (and tropical South America, for 1958–2001) nor the very strong warming over Antarctica seen in the ERA-40 analyses.

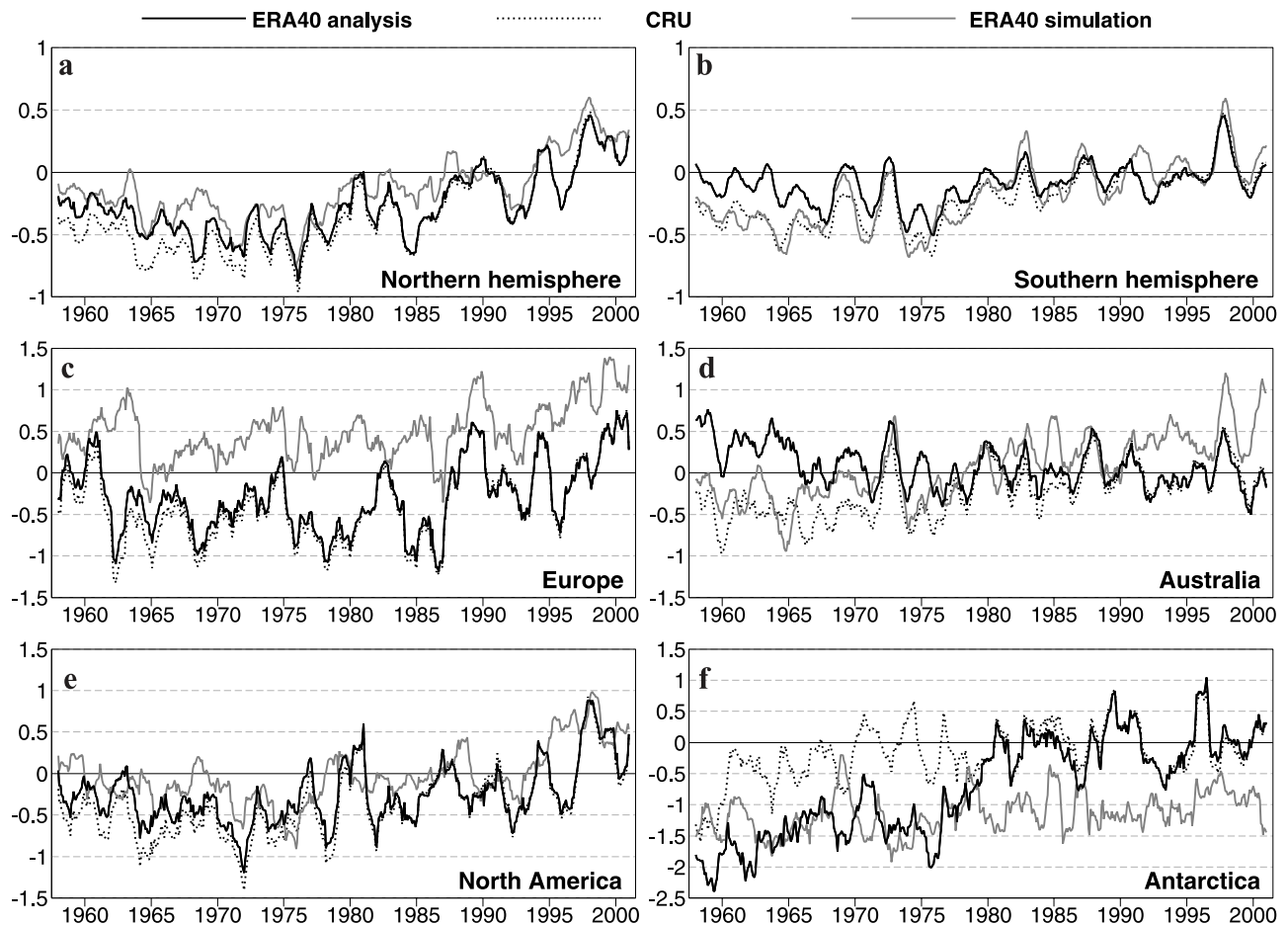


Figure 9. Twelve-month running means of 2-m temperature anomalies from the ERA-40 (black solid) and CRU (black dotted) analyses and from the simulation using the ERA-40 model (grey), averaged over all CRU grid boxes in the (a) Northern and (b) Southern Hemispheres and averaged over (c) European, (d) Australian, (e) North American, and (f) Antarctic grid boxes. Time series based on analyses are adjusted as before to have zero mean for the period 1987–2001. Simulated anomalies are computed with respect to the monthly climate derived from the ERA-40 analyses for the reference period 1961–1990 and then adjusted by the same single amount applied to adjust the ERA-40 analysis time series to zero mean value for 1987–2001. Units are $^{\circ}\text{C}$.

[53] Midtropospheric model bias is illustrated in Figure 11 (left), which shows time series of 12-month running means of 500-hPa temperature anomalies from the simulation and the ERA-40 analyses. Anomalies for the simulation are again defined with respect to mean analyzed values. Averages over the entire extratropical Northern and Southern Hemispheres and the tropics are shown. The simulation is colder than the analysis at 500 hPa for almost all months and both hemispheres.

[54] The ERA-40 analysis shows almost no overall temperature trend at 500 hPa for the Northern Hemisphere, whereas the simulation warms at a rate of a little over $0.1^{\circ}\text{C}/\text{decade}$, the two curves becoming close toward the end of the period. As is the case for terrestrial 2-m temperatures, the simulation captures much of the variability in the 12-month running mean temperature analysis at 500 hPa. Results for the Southern Hemisphere differ in that the analysis warms much more than the simulation. This analyzed warming is almost certainly exaggerated by im-

provement of the observing system for the Southern Hemisphere. Radiosonde coverage in the early years appears insufficient to counter the cold bias of the assimilating model over the hemisphere as a whole, resulting in analyses that are biased cold; the advent of satellite data and other enhancements to the observing system then result in a warming of analyses during the 1970s. A similar conclusion has been drawn recently by *Bengtsson et al.* [2004].

[55] The warming due to assimilation of the early satellite data appears in fact to have been too strong in ERA-40 for several years from 1975 onward. Error in the bias correction of vertical temperature profile radiometer (VTPR) sounding data from the NOAA 4 satellite during 1975 and the first half of 1976 accounts for the sharp divergence between simulation and analysis that is seen around 1975 in Figure 11. Bias correction coefficients derived for NOAA 3 were inadvertently used in the adjustment of NOAA 4 data. The problem period is marked not only by particularly large monthly mean differences between analysis and sim-

Linear trend in simulated two-metre temperature ($^{\circ}\text{C} / \text{decade}$)

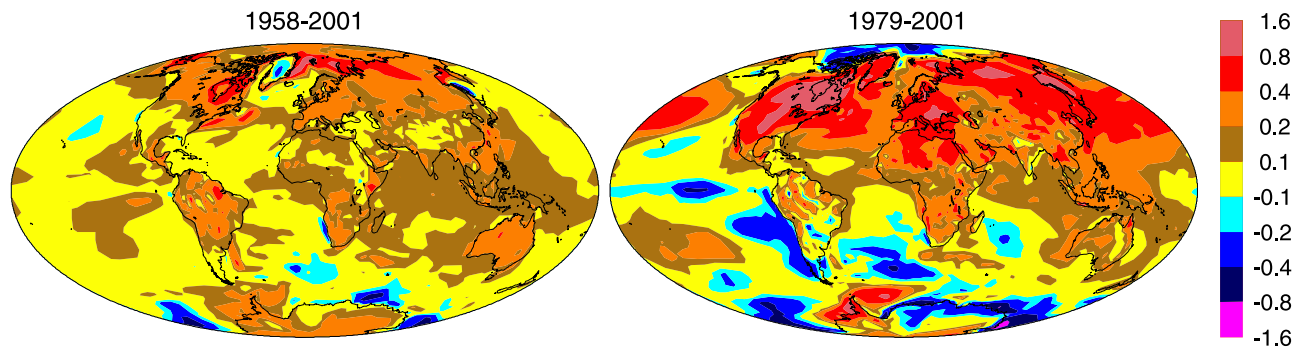


Figure 10. Linear trends over the periods (left) 1958–2001 and (right) 1979–2001 in 2-m temperature from the simulation using the ERA-40 model. Units are $^{\circ}\text{C}/\text{decade}$.

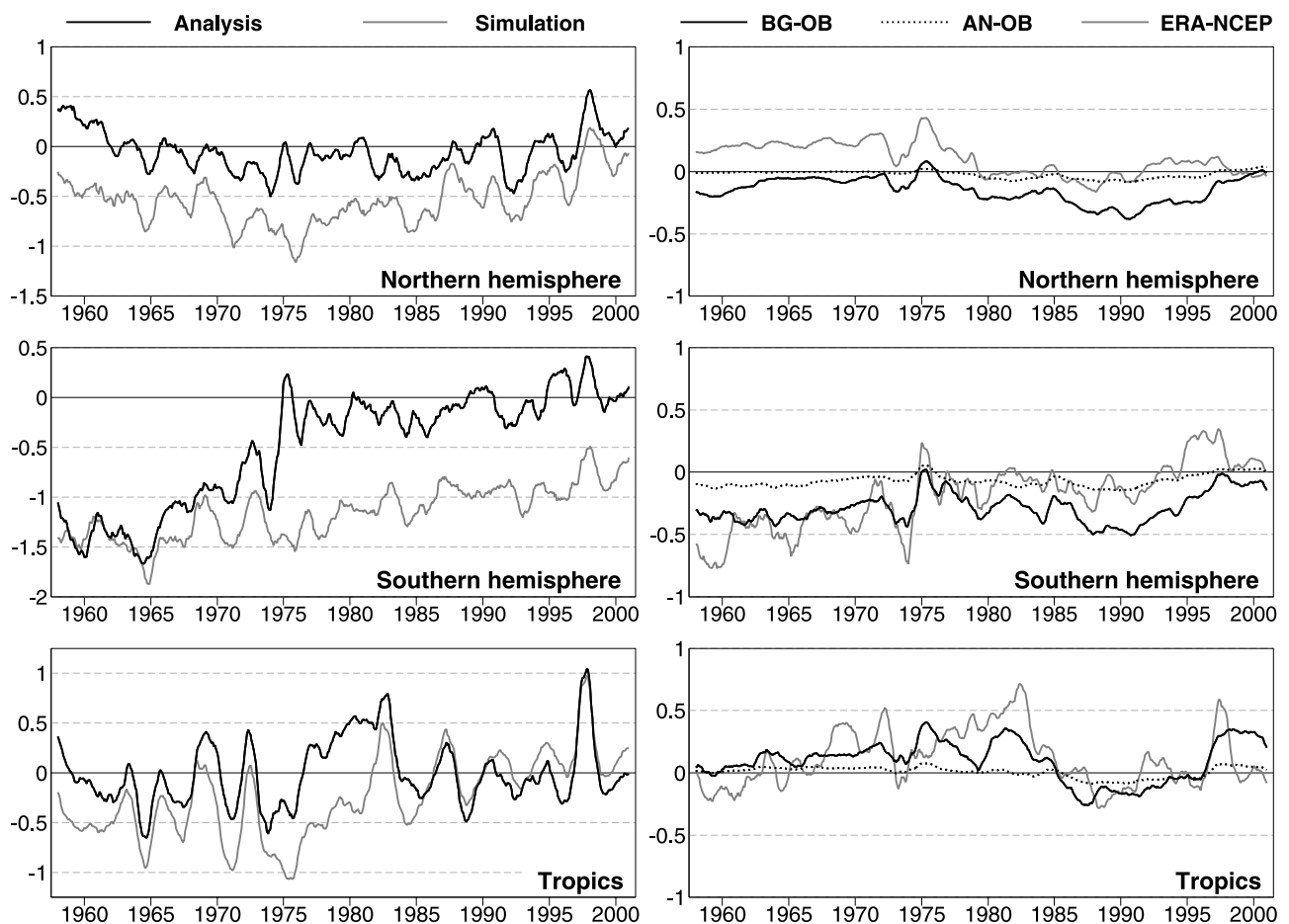


Figure 11. (left) Twelve-month running means of 500-hPa temperature anomalies from the ERA-40 analyses (black) and from the simulation using the ERA-40 model (grey). Values are averaged over the extratropical (top) Northern and (middle) Southern Hemispheres and over the (bottom) tropics, here defined as the region from 20°N to 20°S . Values for the analyses are adjusted as before to give zero mean value over the period 1987–2001. Simulated anomalies are computed with respect to the monthly climate derived from the 500-hPa ERA-40 analyses for the reference period 1961–1990 and then adjusted by the same single amount applied to adjust the analyses to zero mean value for 1987–2001. (right) Corresponding running means of the differences between the ERA-40 background forecasts and radiosonde observations (BG-OB (black solid, averaged over each domain without area weighting)), between the ERA-40 analyses and radiosonde observations (AN-OB (black dotted, likewise without area weighting)), and between the ERA-40 and NCEP/NCAR analyses (grey). Units are $^{\circ}\text{C}$.

ulation but also by a jump in the difference between the ERA-40 and NCEP/NCAR analyses and a maximum in the time series of differences between ERA-40 background temperatures and radiosonde measurements, as shown in Figure 11 (right). Differences between the ERA-40 analyses and the radiosonde measurements (shown also in Figure 11) are much smaller, as the analyses benefit from assimilation of the measurements. It is, however, to be expected that away from radiosonde locations the ERA-40 analyses assume more of the bias characteristics of the background forecasts.

[56] The difference plots in Figure 11 show a general extratropical cold bias in the ERA-40 background forecasts as measured against radiosonde data, consistent with the cold bias of the assimilating model indicated by the simulation. The NCEP/NCAR analyses show less bias in the extratropical Southern Hemisphere early in the period, when they are some 0.5°C warmer than the ERA-40 analyses. Conversely, they are colder than the ERA-40 analyses in early years over the Northern Hemisphere. The difference plots and simulation results also indicate that the ERA-40 analyses are biased warm at 500 hPa in the tropics in the late 1970s and early 1980s. The most likely explanation for this is that the bias correction of the early TIROS Operational Vertical Sounder satellite data was poorer for these years than later in the reanalysis period.

[57] It is beyond the scope of this paper to delve further into the characteristics of the upper level ERA-40 analyses. Figure 1 does, however, show a relatively large discrepancy between the ERA-40 and CRU anomalies around 1975 in the time series for Australia and the Southern Hemisphere. A corresponding comparison for tropical CRU grid boxes (not shown) exhibits poorer agreement around 1975 and in the late 1970s and early 1980s. This suggests that the highly erroneous bias adjustment of NOAA 4 VTPR data had some detrimental impact right down to the surface, as did the source of the warm 500-hPa tropical bias a few years later.

6. Conclusions

[58] There is a good measure of agreement between the CRUTEM2v data set of surface air temperature anomalies derived from monthly mean station data and corresponding results from the comprehensive ERA-40 reanalysis. Variability on short timescales is similar in the two data sets throughout the reanalysis period, although agreement is better in the second half of the period than the first, especially for the Southern Hemisphere. Linear trends computed over the full period of the comparison, 1958–2001, are generally lower in ERA-40, but there is agreement to within $\sim 10\%$ in the rate of warming of the terrestrial Northern Hemisphere since the late 1970s.

[59] There are insufficient upper air data early in the ERA-40 period to prevent contamination of the midtropospheric Southern Hemisphere analysis by a cold model bias, and there are problems also with the bias correction of early satellite data. The cold bias early in the period extends to near-surface temperatures over Antarctica, where the ERA-40 analysis warms by some 1.5°C in the late 1970s, behavior not seen in the CRU data. Temperatures appear also to be biased cold around Greenland.

Elsewhere, evidence points to a warm low-level bias in ERA-40 temperatures over land at middle and high latitudes and a cold bias at low latitudes. The warm bias is larger early in the period, consistent with underestimation of the warming trend.

[60] ERA-40 suffers from significant gaps in the coverage of synoptic screen-level data available for assimilation prior to 1967. Improved retrieval of pre-1967 data from national or other collections would clearly benefit future reanalyses. Analysis of the Southern Hemisphere is likely, however, to remain a challenge for the data-sparse years before the introduction of comprehensive satellite, buoy, and aircraft observations, as would also be analysis of the Northern Hemisphere for the first half of the twentieth century, for example. Progress in this may require specific developments in data assimilation, either alternative approaches [Whitaker *et al.*, 2004] or at least retuning of error statistics and quality control, as direct application of systems developed to work effectively in the comparatively data-rich present may well not be the best approach when data coverage is poor. Nevertheless, more than 25 years have now passed since the global observing system was very significantly upgraded by the additional types of observation that are an enduring legacy of the work of the Global Atmospheric Research Programme in the 1970s. For these years and into the future, there is already a clear role for comprehensive reanalysis to play alongside specific analyses of station data and other individual data sets in monitoring variations in climate.

[61] Capability to produce better analyses of the global atmosphere stems from improvements in the observing system, improvements in the technique of data assimilation, and improvements in the realism of the assimilating model. These have been substantial over recent years [Simmons and Hollingsworth, 2002], and ERA-40 has benefited from many of them. Particularly important in the present context has been work by Viterbo *et al.* [1999] to address the substantial cold bias in winter temperatures that was evident in ECMWF's earlier ERA-15 analysis [Källberg, 1997]. Comparing ERA-40, which may be regarded as the first of a second generation of reanalyses, with the first-generation NCEP/NCAR reanalysis, it is reassuring that ERA-40 is the closer of the two to the CRU analysis for all but the earliest years. Nevertheless, the value of having three diverse analyses available, albeit not of equal overall quality, has been demonstrated by the way they have been used here to identify both a relatively small number of erroneous station values that entered the CRUTEM2v analysis and an even smaller number of highly suspect values in the NCEP/NCAR and (to a lesser extent) the ERA-40 reanalyses.

[62] ERA-40's 2-m temperature analysis was derived by analyzing surface synoptic observations; NCEP/NCAR's was not. This contributed toward the better agreement between ERA-40 and the CRU analysis, although it also exposed ERA-40 to spurious trends associated with changes in synoptic data coverage. The match with the CRU analysis benefits also from the general quality of the observing and data assimilation systems used by ERA-40. Continental-scale trends and variability of background and analyzed temperatures are quite similar, although trends are lower in the background for many regions. After the 1970s upgrade

of the observing system, there is overall consistency in trends and variability between the analyses of temperature at 2-m height and the analyses of temperature at model levels throughout the planetary boundary layer. Results for North America cast doubt on *Kalnay and Cai's* [2003] estimate of the effect of urbanization and land use change on surface warming.

[63] The simulation of the atmosphere for the ERA-40 period produced using the same model, sea surface temperatures, and sea ice cover as used in the ERA-40 data assimilation has provided extra insight into the differences between the ERA-40 and CRU analyses. The simulation is of intrinsic interest in that it captures many features of the variations and trends in the CRU data, and this could be explored further by additional model runs to isolate the extent to which SST variations and the model's greenhouse gas trends are responsible for the agreement with the CRU data. The simulation has also helped to identify problems in the ERA-40 assimilation. It was carried out for this purpose after production of the ERA-40 analyses had been completed. In future it could be advantageous to carry out such a simulation prior to a new reanalysis, as a preliminary check of both the model and the ocean boundary conditions, and to provide a baseline for use in monitoring subsequent production.

Appendix A: ERA-40 Analysis of Screen-Level Temperature

[64] Two-dimensional univariate statistical interpolation was used for the analysis of screen-level temperature in ERA-40. The background temperature at 2-m height derived from the 6-hour forecast of the data assimilation cycle was interpolated horizontally to the observation locations using bilinear interpolation, and "background increments" (the deviation between the observation and the background value) ΔX_i^b were calculated at each observation location i .

[65] The analysis increments ΔX_k^a that were added to the background field at each model grid point k to form the analysis were then derived by linearly combining the background increments at N observation points:

$$\Delta X_k^a = \sum_{i=1}^N W_{ki} \Delta X_i^b.$$

The weights W_{ki} were computed by solving a matrix equation for each grid point k :

$$(\mathbf{B} + \mathbf{O})\mathbf{W}_k = \mathbf{b}_k, \quad (\text{A1})$$

where the column vector \mathbf{b}_k (of dimension N) represents the error covariances between background values at the observation points i and the model grid point k , and the $N \times N$ matrix \mathbf{B} describes the error covariances of background values at pairs of observation points i and j . The horizontal correlation coefficients (structure functions) of background error were specified to have the form

$$\mu_{il} = \exp\left[-\frac{1}{2}\left(\frac{r_{il}}{d}\right)^2\right],$$

where r_{il} is the horizontal separation between observation point i and point l , which is either an observation point j in the case of \mathbf{B} or the model grid point k in the case of \mathbf{b}_k . The horizontal scale d was set to the value 300 km. \mathbf{B} and \mathbf{b}_k were thus given by

$$B_{ij} = \sigma_b^2 \mu_{ij} \quad \text{and} \quad b_{ki} = \sigma_b^2 \mu_{ki},$$

where σ_b is the assumed fixed standard deviation of background errors.

[66] The covariance matrix of observation errors \mathbf{O} was set to $\sigma_o^2 \times \mathbf{I}$, where σ_o is the assumed fixed standard deviation of observation errors, and \mathbf{I} is the identity matrix. The standard deviation σ_o has to take account of both measurement error and how representative a point measurement is of the grid square mean estimate provided by the background model.

[67] The variables σ_b and σ_o were set to 1.5° and 2°C, respectively. The maximum number of observations N used to solve (A1) was 50. The observations located nearest to the model grid point in question were chosen, provided they lay within a radius of 1000 km. The analysis was performed over land and ocean, but only land (ocean) observations were used for model land (ocean) grid points.

[68] Gross quality checks were applied to the observations. Observations taken within a 6-hour period centered on the analysis time were considered, and in the case of multiple reports from a station, only the report closest to the analysis time was used. Observations from stations whose heights differed by more than 300 m from the background model orography were rejected. Any observation that differed sufficiently from the background value was also rejected, using the criterion

$$|\Delta X_i^b| > 3\sqrt{\sigma_o^2 + \sigma_b^2}.$$

It was assumed that all screen-level measurements were made at a height of 2 m; in practice, instruments may be sited at a height ranging from 1.25 to 2 m. No bias correction was applied to the observations.

[69] **Acknowledgments.** The ERA-40 project was partially funded by the European Union under contract EVK2-CT-1999-00027 and was supported by Fujitsu Ltd. through provision of additional computing capacity to ECMWF. P.D.J. acknowledges support of the Office of Science (BER), U.S. Department of Energy, grant DE-FG02-98ER62601. The description of the 2-m temperature analysis in Appendix A is adapted from forecasting system documentation written by Jean-François Mahfouf.

References

- Andersson, E., et al. (2004), Assimilation and modeling of the hydrological cycle in the ECMWF forecasting system, *Bull. Am. Meteorol. Soc.*, in press.
- Beljaars, A. C. M., and P. Viterbo (1999), The role of the boundary layer in a numerical weather prediction model, in *Clear and Cloudy Boundary Layers*, edited by A. A. M. Holtslag and P. G. Duynkerke, pp. 287–403, North-Holland, New York.
- Bengtsson, L., S. Hagemann, and K. I. Hodges (2004), Can climate trends be calculated from reanalysis data?, *J. Geophys. Res.*, *109*, D11111, doi:10.1029/2004JD004536.
- Betts, A. K., and A. C. M. Beljaars (2003), *ECMWF ISLSCP-II Near-Surface Dataset From ERA-40, ERA-40 Proj. Rep. Ser. 8*, 33 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- Bromwich, D. H., and R. L. Fogt (2004), Strong trends in skill of the ERA-40 and NCEP/NCAR reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958–2001, *J. Clim.*, *17*, 4603–4619.

- Cai, M., and E. Kalnay (2004), Climate (communication arising): Impact of land-use change on climate, *Nature*, *427*, 214.
- Douville, H., P. Viterbo, J.-F. Mahfouf, and A. C. M. Beljaars (2000), Evaluation of the optimum interpolation and nudging techniques for soil moisture analysis using FIFE data, *Mon. Weather Rev.*, *128*, 1733–1756.
- Fiorino, M. (2004), *A Multi-decadal Daily Sea-Surface Temperature and Sea-Ice Concentration Data Set for the ERA-40 Reanalysis, ERA-40 Proj. Rep. Ser. 12*, 16 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- Folland, C. K., et al. (2001), Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.*, *28*, 2621–2624.
- Houghton, J. T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell (Eds.) (1996), *Climate Change 1995, The Science of Climate Change*, 572 pp., Cambridge Univ. Press, New York.
- Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and update to 2001, *J. Clim.*, *16*, 206–223.
- Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *J. Clim.*, *10*, 2548–2568.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner (2001), Adjusting for sampling density in grid-box land and ocean surface temperature time series, *J. Geophys. Res.*, *106*, 3371–3380.
- Källberg, P. (1997), *Aspects of the Re-analysed Climate, ECMWF Re-Analysis Final Rep. Ser. 2*, 89 pp., Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- Kalnay, E., and M. Cai (2003), Impact of urbanisation and land-use change on climate, *Nature*, *423*, 528–531.
- Kalnay, E., et al. (1996), The NCEP/NCAR reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471.
- Kistler, R., et al. (2001), The NCEP-NCAR 50-year reanalysis: Monthly means CD-Rom and documentation, *Bull. Am. Meteorol. Soc.*, *82*, 247–267.
- New, M., M. Hulme, and P. Jones (1999), Representing twentieth-century space-time climate variability. part I: Development of a 1961–90 mean monthly terrestrial climatology, *J. Clim.*, *12*, 829–856.
- New, M., M. Hulme, and P. Jones (2000), Representing twentieth-century space-time climate variability. part II: Development of 1901–96 monthly grids of terrestrial surface climate, *J. Clim.*, *13*, 2217–2238.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *J. Clim.*, *15*, 1609–1625.
- Santer, B. D., et al. (2004), Identification of anthropogenic climate change using a second-generation reanalysis, *J. Geophys. Res.*, *109*, D21104, doi:10.1029/2004JD005075.
- Simmons, A. J. (2003), Observations, assimilation and the improvement of global weather prediction—Some results from operational forecasting and ERA-40, in *Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8–12 September 2003*, pp. 1–28, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
- Simmons, A. J., and A. Hollingsworth (2002), Some aspects of the improvement in skill of numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *128*, 647–677.
- Sterl, A. (2004), On the (in-)homogeneity of reanalysis projects, *J. Clim.*, *17*, 3866–3874.
- Trenberth, K. E. (2004), Rural land-use change and climate, *Nature*, *427*, 213.
- Turner, J., S. R. Colwell, G. J. Marshall, T. A. Lachlan-Cope, A. M. Carleton, P. D. Jones, V. Lagun, P. A. Reid, and S. Iagovkina (2004), The SCAR READER project: Towards a high-quality data base of mean Antarctic meteorological observations, *J. Clim.*, *17*, 2890–2898.
- Viterbo, P., A. C. M. Beljaars, J.-F. Mahfouf, and J. Teixeira (1999), The representation of soil moisture freezing and its impact on the stable boundary layer, *Q. J. R. Meteorol. Soc.*, *125*, 2401–2426.
- Vose, R. S., T. R. Karl, D. R. Easterling, C. N. Williams, and M. J. Menne (2004), Impact of land-use change on climate, *Nature*, *427*, 213–214.
- Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill (2004), Reanalysis without radiosondes using ensemble data assimilation, *Mon. Weather Rev.*, *132*, 1190–1200.

A. C. M. Beljaars, V. da Costa Bechtold, P. W. Källberg, S. Saarinen, A. J. Simmons, S. M. Uppala, P. Viterbo, and N. Wedi, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berks RG2 9AX, UK. (adrian.simmons@ecmwf.int)

P. D. Jones, Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.