ERA report series

15 Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim (Revised 2014)

A.J. Simmons, P. Poli, D.P. Dee, P. Berrisford, H. Hersbach,
S. Kobayashi and C. Peubey

Series: ERA Report Series

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Abstract

Low-frequency variability and trends in temperature from 1979 to 2012 are examined. Observational improvements are noted and near-surface behaviour of the ERA-Interim reanalysis is reviewed. Attention is then focussed on how closely ERA-Interim fits the upper-air data it assimilates, the bias adjustments it infers for satellite data, and its agreement with the ERA-40, MERRA and JRA-55 reanalyses and with model simulations.

Global-mean fits to independently homogenised radiosonde temperatures and variationally adjusted satellite brightness temperatures are mainly within 0.1K in the troposphere, with some degradation over time from assimilating varying amounts of aircraft and rain-affected microwave-radiance data, and from a change in source of sea-surface-temperature analysis. Lower-tropospheric warming appears to be somewhat underestimated. Temperature variations in the tropical upper troposphere correlate well with those at the surface, but amplitude is more than doubled, in agreement with modelling. Specific humidity varies in concert; relative humidity is largely uniform, but dips during El Niño events.

Agreement with the other reanalyses is particularly close in the lower stratosphere, where radiance data and the background model constrain cooling to be slightly slower than in the homogenised radiosonde data. Perturbations to global-mean temperatures from underestimating warming following the El Chichón and Pinatubo volcanic eruptions and from assimilating recent GPSRO data are at most 0.2K, less than 20% of the net change since 1979 at 50hPa. Middle-stratospheric variations are more uncertain. Recent cooling appears to be underestimated by assimilating increasing amounts of unadjusted radiosonde data, but results do not support a recent reprocessing of earlier sounding data that suggests stronger middle-stratospheric cooling than previously indicated. Strong analysed upper-stratospheric cooling agrees quite well with model simulations if occasional jumps due to adjusted bias changes in high-sounding satellite data are discounted.

Producing ERA-Interim in two separate streams caused only minor discontinuities where streams join at the start of 1989.

To appear in Q. J. R. Meteorol. Soc., DOI:10.1002/qj.2317

1 Introduction

Obtaining reliable estimates of recent multi-annual variability and trends in atmospheric temperature has proved to be challenging. Measurement by radiosonde ascents was enhanced by operational vertical profiling from space in late 1972, and more comprehensive sounding from several types of satellite-borne instrument has provided increasing coverage and quality of data since 1979. The challenge for estimating multi-annual temperature changes over this period stems from changes in the geographical coverage and biases of the radiosonde data, in the biases of the data from individual satellite instruments, in the relative amounts of data from radiosondes and satellites, and in the information provided by other observing systems. The latter include upper-air data from commercial aircraft, near-surface data over land from the synoptic meteorological network and the information on near-surface marine air temperature implicit in observations of sea-surface temperature (SST).

The reliability of estimates of temperature change based on the records of single types of observation varies with height. Comparisons such as presented in annual climatic assessments (e.g. Blunden and Arndt, 2013) or in specific studies reporting new versions of datasets (e.g. Jones et al., 2012; Haimberger et al., 2012) give most confidence in the estimates for near-surface air temperature. There

is increasing agreement between various datasets over temperature change for the bulk of the troposphere and lower stratosphere, but appreciable uncertainty remains as to rates of change and variations with height. The upper stratosphere is more problematic still, as reliance has to be placed on satellite data alone for global monitoring, and instrumental changes make it difficult to construct a coherent record for the whole of the satellite era. Moreover, substantial differences exist between two data records for the Stratospheric Sounding Unit (SSU) operated on a series of seven satellites between October 1978 and May 2006 (Thompson et al., 2012). Such processed data records differ in the adjustments made not only for changes in bias of instrumental origin but also for changes in how the atmosphere is sampled as satellites drift away from their nominal sun-synchronous orbits, as discussed by Christy et al. (2003) and Mears and Wentz (2009) for microwave sounding and by Wang et al. (2012) for the SSU. Sampling in the vertical changes as carbon dioxide levels increase in the case of the SSU and other types of infrared sounding instrument (Shine et al., 2008; Chung and Soden, 2010).

Atmospheric reanalysis provides, in two stages, a route to reconcile data quality issues and establish more reliable estimates of trends. In the first stage, the reanalysis process confronts diverse types of observation with a single atmospheric model, constructing from multiple observational datasets a more comprehensive picture than can be obtained from any single type of data. It employs data assimilation, a sequence of analysis steps in which background information for a short period, typically 6 to 24 hours long, is combined with observations 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. In the second stage, the topic of the present report, the outputs of reanalysis are diagnosed to expose the inconsistencies among components of the process (observations, model, forcings) and identify the most likely sources of systematic error. For example, background information in the reanalysis chain may be biased due to earlier assimilation of biased observations or due to bias in the atmospheric model used to produce the background forecasts. If observations are abundant and have been adjusted to remove any biases, they can correct bias in the background forecasts when assimilated. In practice, changes over time in observational coverage and residual biases introduce low-frequency variations and trends in analyses that are mixed with the true climatic signals. Diagnosis of one or more reanalyses can help identify artefacts and establish the signals that are reliable.

Reanalysis has already contributed to the estimation of climatic variability and multi-decadal trends. For near-surface temperature there is increasingly good agreement between direct analyses of monthly climatic data and reanalyses that assimilate synoptic data (Simmons et al., 2004, 2010; Onogi et al., 2007; Jones et al., 2012). Reasonable agreement is also found with the near-surface products from reanalyses that assimilate no near-surface temperature observations (Vose et al., 2012; Compo et al., 2013) and from runs of reanalysis models that are constrained by SST analyses but not by assimilation of atmospheric observations (Simmons et al., 2004; Hersbach et al., 2013). Santer et al. (2004) showed that multi-decadal lower-stratospheric temperature change and shorter-term variability from ERA-40 (Uppala et al., 2005) were in reasonable agreement with estimates based on microwave sounding data. Corresponding temperature changes from ERA-Interim (Dee et al., 2011a) have been shown to be broadly comparable with estimates from both microwave sounding and radiosondes, for the

troposphere as well as the lower stratosphere (Dee et al., 2011b; Haimberger et al., 2012). Comparisons with other reanalyses can be seen in Blunden and Arndt (2013).

The combined use by reanalysis of multiple types of observation facilitates detection and correction of individual observational biases. Increasing use has been made of a variational approach to achieve this (Derber and Wu, 1998; Dee and Uppala, 2009), especially for radiance measurements from satellites. In implementing the approach, some observations may be subject to independent bias-correction or none at all, to limit any tendency for observations to be adjusted to have the same biases as the assimilating model. Radiosonde data in particular are generally subject to separate adjustments. Indeed, three of the most recent reanalyses, ERA-Interim, JRA-55 (Ebita et al., 2011) and MERRA (Rienecker et al., 2011), each use variants of the RAOBCORE adjustments (Haimberger, 2007; Haimberger et al., 2008), which in turn use information from the earlier ERA-40. Reanalysis also provides supplementary data that may be used in processing satellite data records (Mears and Wentz, 2009; Wang et al., 2012).

It thus becomes vital to diagnose how data are fitted and biases inferred by the assimilation systems used in reanalysis. This is needed to understand the differences and interdependencies between reanalyses, to indicate the reliability of particular products and to identify features of reanalysis systems that are open to improvement prior to new production runs. It provides feedback for further improvement of observational data records, for use both in future reanalyses and in quasi-independent estimates of climate variability and trends. It also identifies improvements over time in the underlying observing systems and the impacts of different types of observation, information needed to improve the design of the composite global observing system for climate (Houghton et al., 2012; Trenberth et al., 2013).

These considerations motivate the present study. Its focus is on the performance of the ERA-Interim reanalysis as reported for the period from 1989 by Dee et al. (2011a) and since extended to include 1979 to 1988. The domain studied stretches from the surface to the stratopause. Emphasis is placed on the temperature record, but some related aspects of the humidity analysis and precipitation are discussed. The report presents fits to the assimilated observations, inferred or specified bias adjustments, consistency of products over time and space, and comparisons with ERA-40, MERRA and JRA-55. Comparison is also made with atmospheric model runs that used analysed SST and sea-ice distributions, and varying trace gases and aerosol, but no assimilation of atmospheric observations. Conclusions are drawn on the evolution of the observing system and diagnosis and improvement of reanalysis, and on the nature and reliability of the estimated temperature changes.

2 The ERA-Interim reanalysis and other reanalysis and model data used in this study

ERA-Interim (Dee et al., 2011a) is ECMWF's current comprehensive atmospheric reanalysis. It is based on a version of the ECMWF forecasting system operational in 2006, with horizontal grid resolution of around 80km, and a 60-level vertical resolution extending to 0.1hPa. The earlier ERA-40 (Uppala et al., 2005) was based on a 2001 version of the 60-level ECMWF system, using six-hourly

three-dimensional variational (3D-Var) analysis with ~125km grid resolution. The new reanalysis was termed "interim" as it was started quite soon after ERA-40 was completed and run from 1989 onwards to provide improved products for the 1990s, which had proved problematic for ERA-40. In the event, the improvement was quite substantial, and with widespread use being made of ERA-Interim and time needed to establish a fully-fledged ERA-40 replacement, it was decided to extend ERA-Interim back to 1979, and to continue running in close to real time. The 34-year period from 1979 to 2012 is studied here.

ERA-Interim inherited its SST analyses and many of its input observational datasets from ERA-40, supplemented by ECMWF operations after 2001. A consequence of particular relevance for this report is a shift to cooler SSTs in mid-2001 arising from a change in source of analysis discussed by Simmons et al. (2010). Other reflections of the interim nature of the reanalysis include its use of fixed carbon dioxide both in the background model forecasts and in calculating background equivalents of measured radiances, and its use of an aerosol climatology that has seasonal variations but no longer-term variations of volcanic or other origin. Implications are discussed subsequently at various places in this report; here it is noted in summary that information on low-frequency variability and trends enters ERA-Interim through variations in SST and sea-ice concentration and through the assimilated observations, and that this information appears to be damped to only a rather small extent by the absence of inter-annual and longer-term changes in radiatively active constituents in the background model.

ERA-Interim nevertheless improves significantly on ERA-40, through its use of twelve-hourly 4D-Var analysis, variational bias adjustment of satellite radiances (Dee and Uppala, 2009), higher horizontal resolution and other improvements to the assimilating model (Dee et al., 2011a). It also uses homogenised radiosonde data (version 1.3 of the RAOBCORE adjustments), supplemented by an adjustment for the annual cycle of diurnally-varying biases (Haimberger and Andrae, 2011) that built on the scheme developed for ERA-40. As such, ERA-Interim can be viewed as the first of the next generation of reanalysis beyond ERA-40, notwithstanding its interim features.

Comprehensive inter-comparison of all available reanalyses was not an objective of this study, but it was considered important to make detailed comparison of ERA-Interim with a contemporary alternative product. MERRA (Rienecker et al., 2011), a recent reanalysis produced by NASA's Global Modelling and Assimilation Office (GMAO), was chosen for this. The alternative CFSR (Saha et al., 2010) was ruled out because it had been produced in six parallel streams spanning the period from 1979 and substantial drift in stratospheric temperature over the course of each stream resulted in large changes in temperature at stream boundaries. Moreover, CFSR used bogus observations derived from ERA-40 winds above 20hPa, between 30S and 30N from 1981 to 1998, as its assimilation system was incapable of determining quasi-biennial oscillations from the available radiosonde and satellite data. Output from the recently completed JRA-55 reanalysis (Ebita et al., 2011) had not been released at the time most comparisons were carried out. Monthly-mean data became available, however, shortly before revisions were made to this report to account for reviewers' comments, and this opportunity was taken to include JRA-55 results in section 5.1 and the concluding discussion.

Production of MERRA started in 2008, almost two years after ERA-Interim. It uses GMAO's GEOS-5 assimilation system, run from 1979 with ½^0 x ⅔^0 horizontal resolution and 72-level vertical resolution extending to 0.01hPa. Its use of variational radiance bias adjustment and homogenised radiosonde data is largely as in ERA-Interim, but its later start enabled use of the newer version 1.4 of the RAOBCORE adjustments and some pre-processed inter-calibrated satellite data. Major differences from ERA-Interim are in the assimilating model and six-hourly 3D-Var analysis method.

ERA-Interim and MERRA are both compared with ERA-20CM (Hersbach et al., 2013), an ensemble of ten atmospheric-model runs with prescribed SSTs. ERA-20CM is a component of the ERA-CLIM project (Dee et al., 2014), and in particular part of the preparation for ERA-20C, an ensemble reanalysis for the period that assimilates surface pressure and marine wind data (Poli et al., 2013). Key details are provided later.

This study uses monthly-mean area averages and fields produced from archived ERA data as described by Simmons et al. (2010). Monthly means were first evaluated on a regular 0.25^0 x 0.25^0 latitude/longitude grid using data archived either on the model's coarser irregular grid or with the model's native spectral representation (T159 for ERA-40 and ERA-20CM; T255 for ERA-Interim). Values on the 1^0, 1.25^0 and 5^0 grids of datasets with which comparison is made were derived from the 0.25^0 sub-grid values using area-weighted averaging. Except where stated otherwise, maps were plotted for ERA from 1^0 gridded values, and global and other large-area averages were evaluated directly using the 0.25^0 sub-grid values. Monthly-mean pressure-level data from MERRA were used on the 1.25^0 grid on which they were downloaded from http://disc.sci.gsfc.nasa.gov/daac-bin/DataHoldings.pl; ERA-Interim values were produced on this grid for plotting maps of the differences between the two reanalyses. Corresponding data from JRA-55, also supplied on a 1.25^0 grid, were downloaded from http://jra.kishou.go.jp/JRA-55/index_en.html.

3 Changes in observational coverage

The year 1979 was a landmark in the evolution of the atmospheric observing system. It saw (1) routine temperature sounding extended throughout the troposphere and stratosphere using new satellite-borne sensors of infrared and microwave radiance, (2) winds from tracking features observed by geostationary satellites first becoming available in significant numbers, (3) drifting ocean buoys deployed more widely and in much larger numbers, and (4) a substantial increase in the number of measurements from commercial aircraft. Considerable further improvements have been made since then, in particular from the late 1990s onwards. Germane to this study are increases and other improvements in sounding data from satellites and direct measurements by radiosondes. Increasing amounts and quality of other observations, either directly of temperature or of related variables, also contribute to the changing quality of the temperature fields provided by reanalysis.

Figure 1 shows the distribution from 1979 to 2012 of the sounding instruments on polar-orbiting satellites from which temperature-sensitive radiance data were assimilated in ERA-Interim. Data originate from five types of instrument. They are the SSU, the High-resolution Infrared Radiation Sounder (HIRS), the Microwave Sounding Unit (MSU), the Advanced Microwave Sounding Unit A

(AMSU-A) and the Atmospheric Infrared Sounder (AIRS). These instruments were or are still being flown on a total of sixteen satellites, as labelled in Figure 1. HIRS data span the whole period. Data from an increasing number of satellites were assimilated later in the period, and more still would have been used had ERA-Interim not been restricted by employing the fixed 2006 version of the ECMWF forecasting system. This prevented use of data from the Infrared Atmospheric Sounding Interferometer (IASI) on the Metop-A satellite launched in October 2007 and the sounders on the US Suomi National Polar-orbiting Partnership satellite launched in October 2011. The launch of Metop-B in September 2012 was too late for any of its data to be used in 2012.

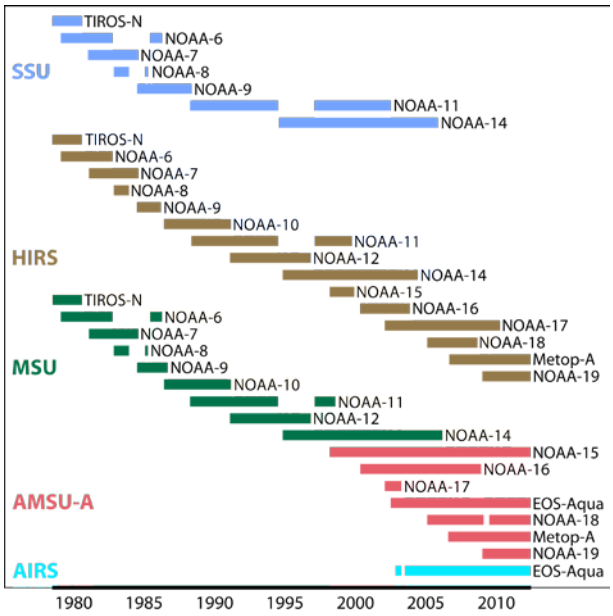

Figure 1: Temperature-sounding instruments (SSU, HIRS, MSU, AMSU-A and AIRS) from which data were used in ERA-Interim from 1979 to 2012. The satellites were TIROS-N and NOAA-6 to NOAA-19, operated by the US National Oceanic and Atmospheric Administration (NOAA), EOS-Aqua, operated by the US National Aeronautics and Space Administration (NASA) and Metop-A operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). Coverage over time from each instrument varies somewhat from channel to channel. The horizontal bars denote time periods over which data were assimilated from channels sounding primarily the lower to middle stratosphere: SSU-1, HIRS-2, MSU-4, AMSU-A-10 and AIRS-40.

These instruments make measurements for sets of channels associated with various frequencies of radiation, and provide radiance data that are sensitive to the temperature distribution over layers upwards of several km thick. The height of maximum sensitivity and layer thickness vary from channel to channel and from one type of instrument to another, and depend to a lesser extent on the temperature and composition of the atmosphere itself. Channel numbers vary from three for SSU to 2378 in the thermal infrared range for AIRS, although many of the latter channels are sensitive to variables other than temperature. The horizontal footprint or pixel size of the measurement also varies with the type of instrument, but is smaller than 100km or so at nadir for all instruments except SSU. Specific detail will be given later where relevant; more detail in general may be found in original publications (e.g. Miller et al. (1980) for SSU), in papers that discuss the formation of long-term data records (e.g. Christy et al. (2003) and Mears and Wentz (2009) for the merging of information from MSU and AMSU-A), and on space-agency websites for instruments currently being flown (e.g.

aqua.nasa.gov/about/instruments.php, www.eumetsat.int/Home/Main/Satellites/Metop/Instruments and www.nasa.gov/mission_pages/noaa-n/spacecraft).

Some of the temperature-sensitive data from satellites that could in principle have been used were not assimilated in ERA-Interim, because the reanalysis was constrained in practice to use only data for which there was some experience of radiance assimilation, either in ECMWF operations or in ERA-40. This includes data from some channels of the instruments already being used, notably the deep stratospheric sounding channel HIRS-1 and the upper-stratospheric sounding channels of AIRS. Data from the series of SSM/T sounders flown by the US Defense Meteorology Satellite Program (DMSP) also have potential for long-term assimilation. Other types of satellite data have potential for diagnostic purposes, either through short periods of parallel assimilation such as developed and tested for MIPAS limb-sounded radiances by Bormann and Thépaut (2007) or through comparison of independent retrievals with reanalysis outputs.

The left-hand panels of Figure 2 show the amounts of radiosonde temperature data assimilated each month from the standard (or mandatory) pressure levels of 20, 50, 100, 300 and 500hPa and from the

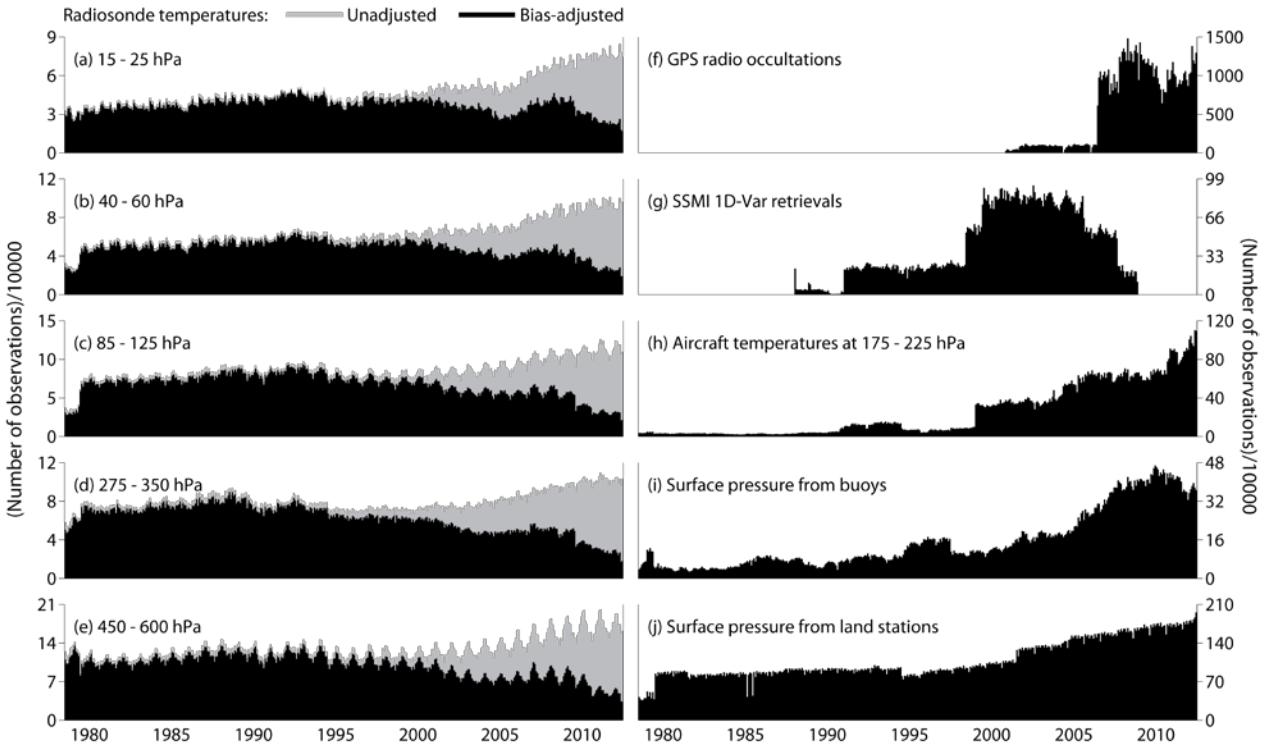

Figure 2: The number of data assimilated each month (divided by 104) in ERA-Interim for radiosonde temperatures at or near (a) 20hPa, (b) 50hPa, (c) 100hPa, (d) 300hPa and (e) 500hPa, and for (f) GPSRO, (g) retrievals from rain-affected SSMI radiances, (h) temperature measurements from aircraft between 225 and 175hPa, and surface-pressure measurements from (i) buoys and (j) land stations reporting in SYNOP code. The counts for radiosondes include data from significant levels in layers surrounding the standard levels, as labelled in the figure. Grey shading indicates the number of radiosonde observations that were not subject to bias adjustment.

nearby levels (the significant levels) at which additional data are provided by the radiosonde operator to characterize the vertical structure of the ascent more fully. Grey shading denotes the number of observations not adjusted for bias, which increases later in the period as the corrections based on data usage in ERA-40 become increasingly inapplicable. Amounts fall in the 1990s following dissolution of the Union of Soviet Socialist Republics (USSR), but increase later to reach their largest values near the end of the period. This is especially so at the uppermost levels, where a sharp increase can be seen since 2005 at 20hPa. The pronounced annual cycle for 500hPa is mainly due to variable amounts of data from nearby significant levels.

The radiosonde data used from 1995 onwards are those received operationally by ECMWF, apart from a small amount rejected by quality control. Some of the increases shown in Figure 2 are likely due to better identification and transmission of significant-level data, but data availability from standard levels has also improved. Figure 3 shows the spatial coverage of all 500hPa and 20hPa radiosonde temperature data received for 2000 and 2012. Improvements over time in both frequency of reporting and number of active stations are evident over the former USSR, and there are also substantial improvements over Central and South America and West Africa. Improvements at 20hPa are more widespread than at 500hPa, with increased frequency of data at this level over the large land masses of China and North America as well as the former USSR, and for a number of island stations. Europe remains largely data-rich despite some decline. Frequency of ascents has declined over Australia; the decline over India is of little consequence as the temperature and humidity data from these ascents are not used due to persistent poor quality.

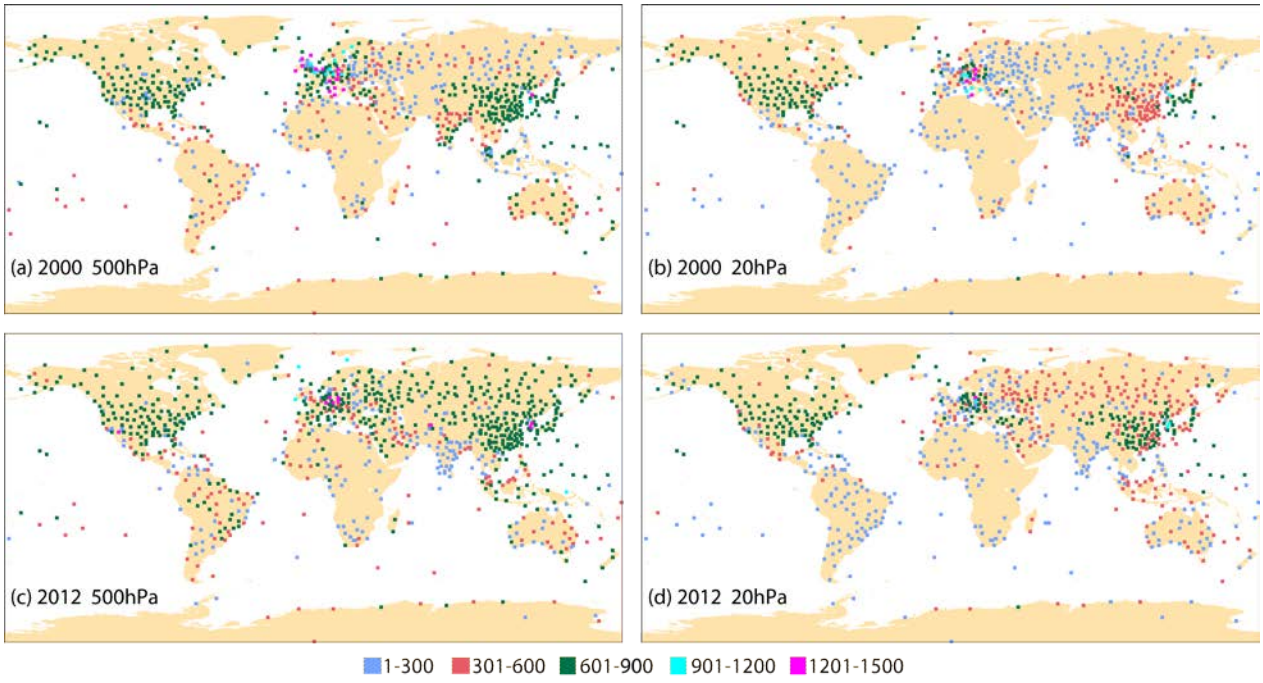

Figure 3: Annual counts of radiosonde temperature data received operationally by ECMWF per 1O grid box for 2000 (top) and 2012 (bottom), at the 500hPa (left) and 20hPa (right) levels. Coloured markers are used to indicate ranges of numbers. A grid box containing a single station reporting regularly twice a day would be denoted by a green marker. The few instances of cyan and magenta markers, mostly over Europe, indicate either more frequent reporting or greater station density.

Counts of assimilated observations from several other sources are presented in the right-hand panels of Figure 2. Panel (f) is for Global Positioning System radio occultation (GPSRO) data, which are available in significant numbers from late 2006, from the joint Taiwanese-US COSMIC constellation of six receivers and later the receiver on Metop-A. GPSRO provides substantial information on temperature in the upper troposphere and lower and middle stratosphere (Poli et al., 2010). Panel (g) shows the number of humidity adjustments over sea derived by retrieval from rain-affected radiances (Bauer et al., 2006) measured by the Special Sensor Microwave Imager (SSMI) instruments flown on up to three DMSP satellites at any one time. The temperature analysis is influenced by assimilating these data through resulting changes in diabatic heating in background forecasts. Also shown are counts of (h) the near-tropopause temperature measurements from aircraft, which increase substantially in 1999 and subsequently, (i) the surface-pressure measurements from buoys, which also increase significantly before declining a little since 2010, and (j) the surface pressure from land stations providing SYNOP messages, for which the rise over time is mainly due to more-frequent observations. The changes in the counts for the latter two types of observation have implications for temperature analyses beyond any benefits that may accrue synoptically from assimilating more surface-pressure data, as the buoys also provide data for the SST analyses used by ERA-Interim and the air temperature measurements included in SYNOP messages are used in ERA-Interim's analysis of 2m temperature over land.

4 Near-surface temperature

Some aspects of the temperature analyses for 2m height are considered first. The representation of near-surface air temperature over land has been examined in some detail for ERA-40 by Simmons et al. (2004) and updated for ERA-Interim from 1989 to 2008 by Simmons et al. (2010), making comparison respectively with the CRUTEM2V (Jones and Moberg, 2003) and CRUTEM3 (Brohan et al., 2006) analyses of monthly station climatological temperatures for 5O grid boxes. The two ERA reanalyses use the same method for analysing synoptic surface observations separately from upper-air observations (Simmons et al., 2004), but differ in resolution and background forecast.

Jones et al. (2012) compared the extended ERA-Interim for 1979-2010 with their new CRUTEM4 dataset. Agreement was excellent for averages over the land areas of the northern hemisphere. It was good for the land areas of the southern hemisphere north of 60OS, but much poorer for the data-sparse Antarctic. ERA-Interim warms slightly more than CRUTEM4 over the northern hemisphere, most likely because of larger than average warming in Arctic grid boxes that contribute to the ERA-Interim all-land average but not to the CRUTEM4 average. The discrepancy is nevertheless smaller than found for comparisons with CRUTEM3 (Simmons et al., 2010), which has poorer data coverage than CRUTEM4. CRUTEM4 also agrees more closely with ERA-Interim for the grid boxes where CRUTEM3 also provides values: standard deviations of the differences are 0.55K for CRUTEM3 and 0.50K for CRUTEM4 over the period 1979-2012.

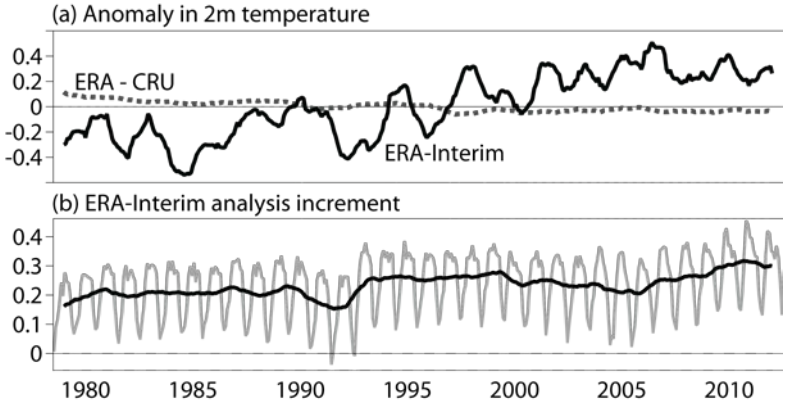

Figure 4(a) Twelve-month running-mean anomalies relative to 1981-2010 in monthly-mean 2m temperature (K) from ERA-Interim, averaged with area-weighting over all 5O grid boxes where CRUTEM4 (version 2.0.0 from www.metoffice.gov.uk/hadobs) provides values (solid), and corresponding differences between ERA-Interim and CRUTEM4 (dotted). (b) Corresponding analysis increment for ERA-Interim shown as individual monthly values (grey) and twelve-month running means (black).

ERA-Interim warms slightly less than CRUTEM4 early in the period if averages are taken only over the grid boxes for which CRUTEM4 provides values. This is shown in the upper panel of Figure 4 for twelve-month running means of the differences between the two datasets. These differences are, however, small compared with the analysed temperature changes, which are shown for ERA-Interim. The corresponding analysis increments, the differences between the analysis and the background forecast, are shown in the lower panel. The assimilating ECMWF model has a seasonally varying cold bias in near-surface air temperature over land, and the background forecast is generally warmed by the analysis process. The increment generally increases in magnitude over the period. Underestimation of warming in the background model, expected due to its use of fixed greenhouse-gas concentrations and inconsistent SST analyses, is thus seen to be largely corrected by the 2m temperature analysis. The long-term variation over time of the increment is nevertheless quite small compared with the variation in temperature itself, showing that much of the latter is captured in the background forecast. This forecast depends only weakly on observations of 2m temperature (Simmons et al., 2004) but benefits from the assimilation of much other data and from the forcing provided by prescribing SST, for which the inconsistency in the analyses used is small compared with the net change over the period.

Two fluctuations in analysis increment also merit comment. Firstly, the dip in the early 1990s may be partly due to colder actual temperatures following the volcanic eruption of Mt Pinatubo in June 1991, as the background model included no cooling effect of volcanic aerosols. This cannot be the whole explanation, however, as the dip begins prior to the eruption. Secondly, there is a slight fall in the increment from 2000 to 2005 and a larger rise thereafter. Similar behaviour will be shown in the following section for the fit of background tropospheric forecasts to radiosonde data, and will be discussed subsequently in the light of contrasting fits to satellite radiance data.

Figure 5 maps the change in decadal-mean temperature from 1981-1990 to 2001-2010 for the median of the HadCRUT4 ensemble (Morice et al., 2012) and ERA-Interim. Over land, HadCRUT4 is very close to CRUTEM4 apart from a few coastal grid boxes. ERA-Interim and HadCRUT4 show similar variations in warming from one decade to the other across the land masses of the extratropical northern hemisphere. Both show the much-discussed strong warming at high latitudes and only a few

regions with little warming or slight cooling. ERA-Interim underestimates warming over Australia compared with CRUTEM4; here the behaviour of the analysis scheme requires further investigation as it is anomalous in cooling the background and thereby degrading the fit to HadCRUT4. The values ERA-Interim provides where HadCRUT4 has data voids are spatially coherent except over Antarctica, where unrepresentative local values can result from changes over time in sparsely distributed data. Evidence for the large warming over central West Antarctica is discussed by Bromwich et al. (2013), who show good agreement between near-surface temperature observations from Byrd Station and essentially independent short-range ERA-Interim forecasts, and superior performance of ERA-Interim compared with several other reanalyses.

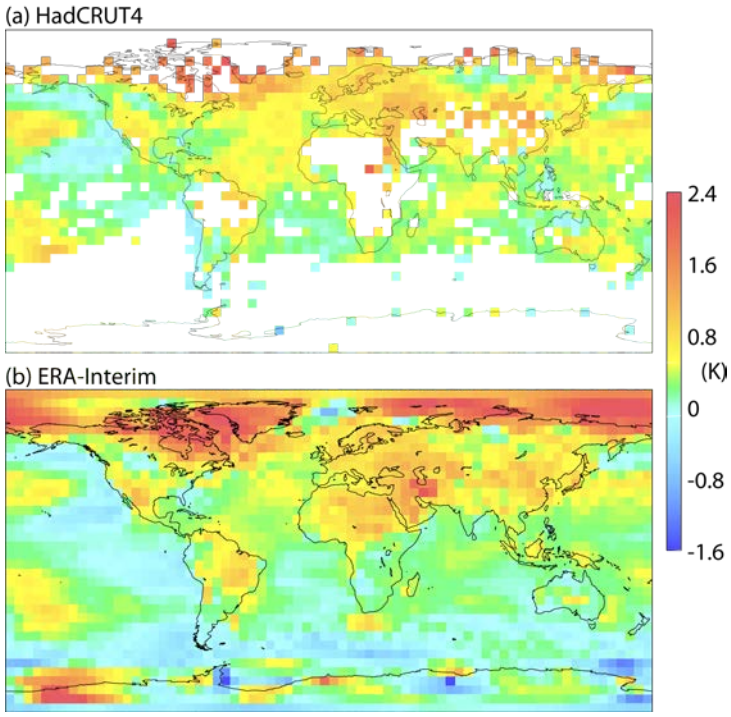

Figure 5: Changes in decadal-mean temperature (K) from 1981-1990 to 2001-2010 for (a) the median of the HadCRUT4 (version 2.0.0 from www.metoffice.gov.uk/hadobs) ensemble and (b) ERA-Interim. Temperatures are blended values for air at 2m height over land and sea-ice and for the sea surface otherwise. HadCRUT4 values are plotted only for 5O grid boxes that had no more than five months of missing values for each decade. ERA-Interim values are averages over the HadCRUT4 grid boxes.

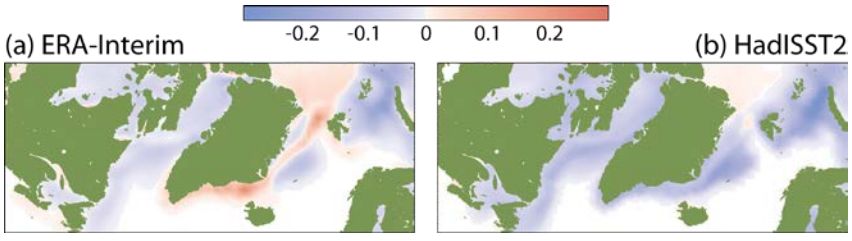

Figure 6: Difference in fractional sea-ice cover between the average for 2001-2010 and that for 1981-1990, for (a) ERA-Interim and (b) HadISST2 (Version 1.0.0). Operational ECMWF data were used instead of HadISST2 data for November and December 2010 as HadISST2 data were not available for these months.

Figure 5 shows a widespread small difference between ERA-Interim and HadCRUT4 over the oceans, where ERA-Interim cools by an average 0.11K relative to HadCRUT4 due to its use of cooler SST analyses from mid-2001. Consequences can also be seen of inconsistency over time in the sea-ice distributions, which for ERA-Interim are specified from a mix of sources that also includes a change in mid-2001 (Uppala et al., 2005; Dee et al., 2011). ERA-Interim shows strong warming over most of the high Arctic that is coherent with the warming over neighbouring land areas, but cooling occurs west of Svalbard and between Iceland and Greenland. Here differences in decadal means of the sea-ice concentrations used by ERA-Interim show increases that are barely seen in the HadISST2 dataset (Titchner and Rayner, 2014) used in ERA-20C (Figure 6). The local cooling seen north of Canada in Figure 5 similarly appears to be associated with a questionable increase in sea-ice concentration in ERA-Interim. No such cooling regions are seen in the 850hPa ERA-Interim analyses for the Arctic presented in the following section.

5 Upper-air temperature from ERA-Interim and ERA-40, and fits to assimilated observations

5.1 Anomalies in analysed temperatures

Figure 7 shows time series from 1979 to 2012 of monthly anomalies in global-mean temperature from ERA-Interim (solid lines) and ERA-40 (dotted lines) at a set of tropospheric and stratospheric levels. Shading denotes corresponding values from JRA-55. Anomalies are relative to 1979-2001, when ERA-Interim, ERA-40 and JRA-55 all overlap.

ERA-Interim shows net warming over the period at tropospheric levels up to 300hPa. Superimposed on this is substantial shorter term variability, of which the peak associated with the 1997/98 El Niño is the most prominent feature. The trend and variability increase coherently with height from the surface to 300hPa, as discussed further for the tropics in section 9. The 200hPa and 100hPa near-tropopause levels are characterised by little trend but considerable variability. Short-term warming peaks associated with the volcanic eruptions of El Chichón in 1982 and Pinatubo in 1991 are evident in the lower to middle stratosphere, with cooling in steps around these peaks, and some subsequent warming. Cooling increases strongly with height in the upper stratosphere, where there are discontinuities in ERA-Interim associated with the initial introduction of SSU-3 data in mid-1979, the change from NOAA-7 to NOAA-9 SSU-3 data in early 1985, and the introduction of AMSU-A data in mid-1998. Several of the features shown in Figure 7 are familiar from other plots of time series of global temperature from reanalyses or direct analyses of data from specific types of instrument, as shown for example in the annual Bulletin of the American Meteorological Society articles on the state of the climate (e.g. Blunden and Arndt, 2013). Much of what follows in this report is devoted to establishing the reliance that can be placed on the details of what is provided by ERA-Interim and other datasets, and to identifying where future improvement is expected.

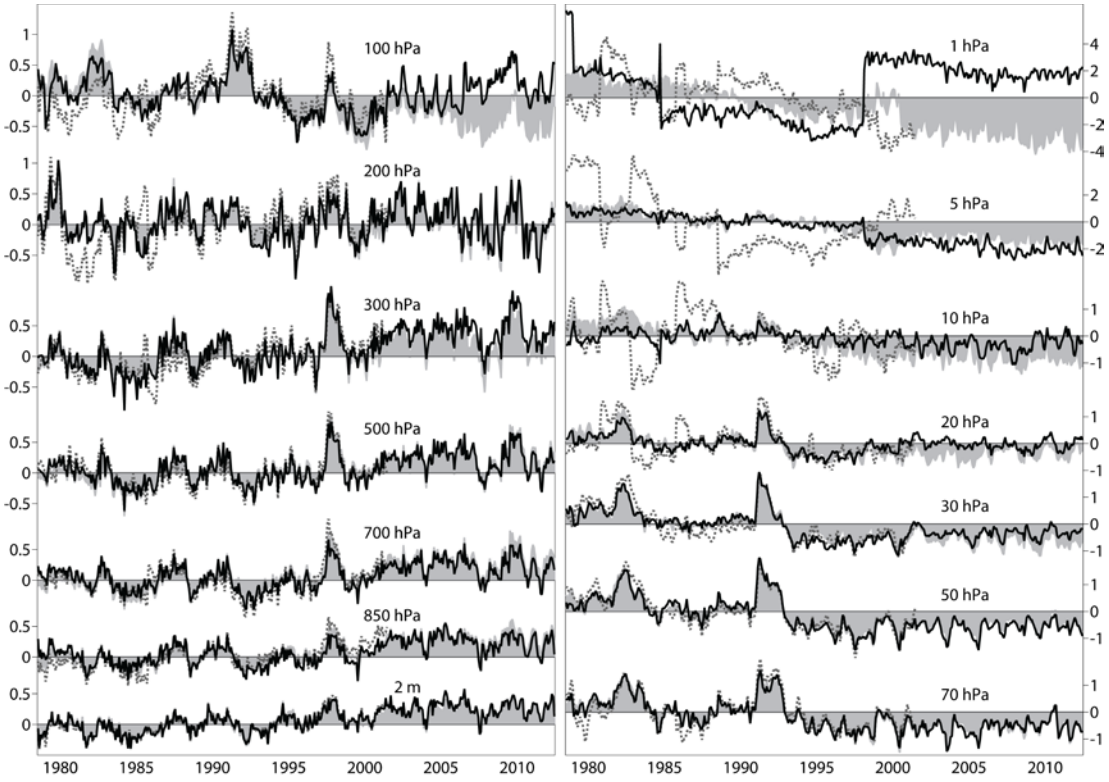

Figure 7: Anomalies in monthly and globally averaged temperatures (K) relative to 1979-2001 from ERA-Interim (black lines), ERA-40 (dotted) and JRA-55 (shading), at the indicated tropospheric and stratospheric levels. Common vertical scales are used for levels from 2m to 100hpa, from 70 to 10hPa, and for 5 and 1hPa.

Consistency between ERA-Interim and its predecessor, ERA-40, varies quite considerably with height. Agreement between the two reanalyses is generally good at 2m, in the middle to upper troposphere and in the lower to middle stratosphere. It is poorer in the lower troposphere, where ERA-40 suffered from problematic assimilation of data from several HIRS channels from 1989 to 1997 (Uppala et al., 2005). It is poorer also near the tropopause, particularly early in the period, and deteriorates above 30hPa, where ERA-40 exhibits many more discontinuities associated with changes in satellite data.

JRA-55 is also very close to ERA-Interim at many levels, more so than ERA-40 in general. In the lower troposphere it exhibits larger short-term warming associated with the 1997/98 El Niño than ERA-Interim, being closer to ERA-40 for this particular feature, and its lower-tropospheric warming trend over the period as a whole is somewhat larger than that of ERA-Interim. At 100hPa, low-frequency variability is similar for JRA-55 and ERA-Interim, but there is poorer agreement for annual variability. Moreover, ERA-Interim shifts quite substantially to warmer values in late 2006. This is linked to the assimilation of GPSRO data that correct a cold bias at the tropical tropopause in ERA-Interim, as discussed below for the fit of ERA-Interim to radiosonde data and in section 6 when comparing ERA-Interim with MERRA. Conversely, JRA-55 is cooled a little in the tropical upper troposphere by assimilating GPSRO data, as this provides additional correction of a warm model bias that is insufficiently constrained by the observations assimilated in earlier years.

JRA-55 agrees especially well with ERA-Interim at 70 and 50hPa, and also at 30 and 20hPa until around the late 1990s, but does not warm to the extent that ERA-Interim does beyond then at the latter two levels. It generally cools more over time than ERA-Interim at 10hPa. The use of full variational bias correction to the uppermost satellite sounding channels in JRA-55, something not done in ERA-Interim as discussed in section 5.4, avoids the discontinuities seen in ERA-Interim at 5 and 1hPa. However, the relatively large cooling rates at these levels in the periods either side of discontinuities in ERA-Interim are of similar magnitude to those in JRA-55.

5.2 Fits to radiosonde temperature data

Figure 8 shows times series of the differences between assimilated (bias-adjusted) radiosonde observations and corresponding values from the ERA-Interim analyses and background forecasts. Differences are evaluated at observation points during the assimilation process, and the plotted values are from a database of averages taken over all values for layers centred on standard pressure levels, without area weighting. Values are also shown (using shading) for the ERA-40 background forecasts. The vertical axes are chosen to display detail for ERA-Interim; the larger ERA-40 values often exceed the axis limits.

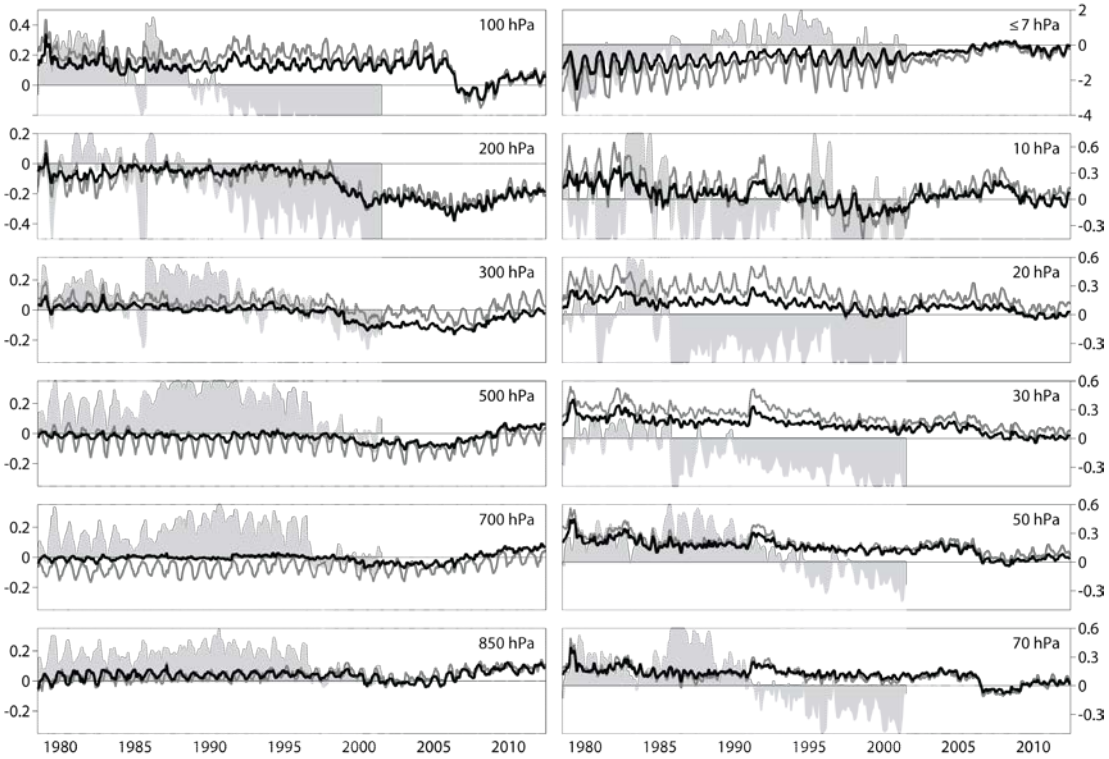

Figure 8: Monthly mean ERA-Interim observation-minus-analysis (black lines) and observation-minus-background (grey lines) differences for radiosonde temperatures (K) at the indicated tropospheric and stratospheric standard and nearby significant levels. Shading denotes the corresponding ERA-40 background differences. Averages are taken globally over all assimilated data. No area weighting is applied.

The ERA-Interim analyses fit the radiosonde data particularly closely in the lower and middle troposphere, especially for the first twenty years. The fit is to within about 0.1K or better at 500hPa and below for the period as a whole. The background forecasts at 500 and 700hPa show almost no drift in boreal winter, but warm by 0.1-0.2K in boreal summer. The distribution of ascents (Figure 3) is such that these averages are dominated by the contributions from the land masses of the northern hemisphere. Analyses and background forecasts are biased slightly cold relative to the radiosondes at 850hPa. Later in the period, analysis and to a lesser extent background values warm slightly relative to the radiosondes up to around 2005, and then cool. The corresponding lower- and middle-tropospheric background fits for ERA-40 are much poorer prior to the 1997 change in HIRS usage, and generally vary more over time than the ERA-Interim fits.

A much more marked warming of analysis and background occurs in the upper troposphere, especially at 200hPa, beginning in the late 1990s. This is associated with assimilating increasing amounts of warm-biased temperature data from commercial aircraft (Dee and Uppala, 2009). Implementation of a variational bias correction scheme for aircraft data in the operational ECMWF assimilation system in November 2011 (Isaksen et al., 2012) significantly improved the mean fits to both radiosonde and GPSRO data, and paves the way for better future reanalyses.

The impact of assimilating substantial amounts of GPSRO data from late 2006 onwards (Poli et al., 2010) is seen most clearly at 100hPa. Here the fits to radiosonde data are quite stable in earlier years, with analyses biased cold relative to the assimilated radiosonde data and a drift to colder values of around 0.2K in the course of the background forecasts. GPSRO provides a weight of data in later years that brings the analysis into much closer agreement with the radiosonde data.

Several features of the fits to radiosonde data in the lower to middle stratosphere merit comment. The first concerns the trend over the period. This is generally downwards, from values for analyses and background that are positive. This corresponds to analysis and background values that are colder than the radiosonde values, more so early than later in the period. The downward trend is much larger for the ERA-40 background than for ERA-Interim, although the two reanalyses themselves vary similarly in the lower stratosphere. This is consistent with ERA-Interim's use of radiosonde data that have been subject to much stronger homogenisation than those used in ERA-40. It is also consistent with Haimberger et al.'s (2012) finding that ERA-Interim cools less in the lower stratosphere than their RAOBCORE and RICH sets of homogenised radiosonde data, a result they ascribe in part to remaining uncorrected breaks in the radiosonde time series and in part to the warming late in the period in ERA-Interim from assimilating GPSRO data. Before then, the implication is that the background forecast and assimilated satellite data draw ERA-Interim away from too tightly fitting the homogenised radiosonde temperatures in the lower stratosphere.

Figure 8 also shows a spike in the fit to radiosonde temperatures from 20 to 200hPa in 1979, and poorer lower stratospheric fits for periods immediately following the El Chichón and Pinatubo eruptions. The early spike occurs at a time of sharp shifts (illustrated later) in the bias adjustments made to radiances from the instruments flown on the TIROS-N satellite, which occur when data from a second satellite, NOAA-6, begin to be assimilated. The poorer fits following the volcanic eruptions

are due to absence of the associated warming by stratospheric aerosols in the background model. As discussed by Dee and Uppala (2009) and illustrated later in this report, a fraction of the warming signal in the radiances is taken up by the variational bias adjustments of the data from channels sensitive to the lower to middle stratosphere. Much of the signal survives however: at 50 and 70hPa the radiosonde-analysis difference increases by less than 10% of the analysed warming following the Pinatubo eruption. At 30hPa, where there are fewer radiosonde data, the underestimation of warming is somewhat larger. ERA-40, which also lacked aerosol heating but used a more static radiance bias adjustment scheme, exhibits slightly stronger warmings than ERA-Interim.

Included in Figure 8 are the mean fits to all radiosonde temperature data received from 7hPa and above. Data counts vary both within each year and over the longer term, ranging from below 2000 for several months between 1979 and 1985 to above 10000 for several months from 2004 to 2007. The numbers are generally small enough for the observations to have little direct effect on global-mean analyses, even though the analyses may fit the data locally. Variations in spatial coverage and data quality also do not make for straightforward interpretation. It is striking nevertheless that the mean fits of background forecasts to these high-level radiosonde data improve substantially over time, becoming generally within 1K or less over the past ten years or so. Root-mean-square fits also improve significantly. This most likely indicates beneficial assimilation of upper-level satellite data whose quality and quantity increase over time.

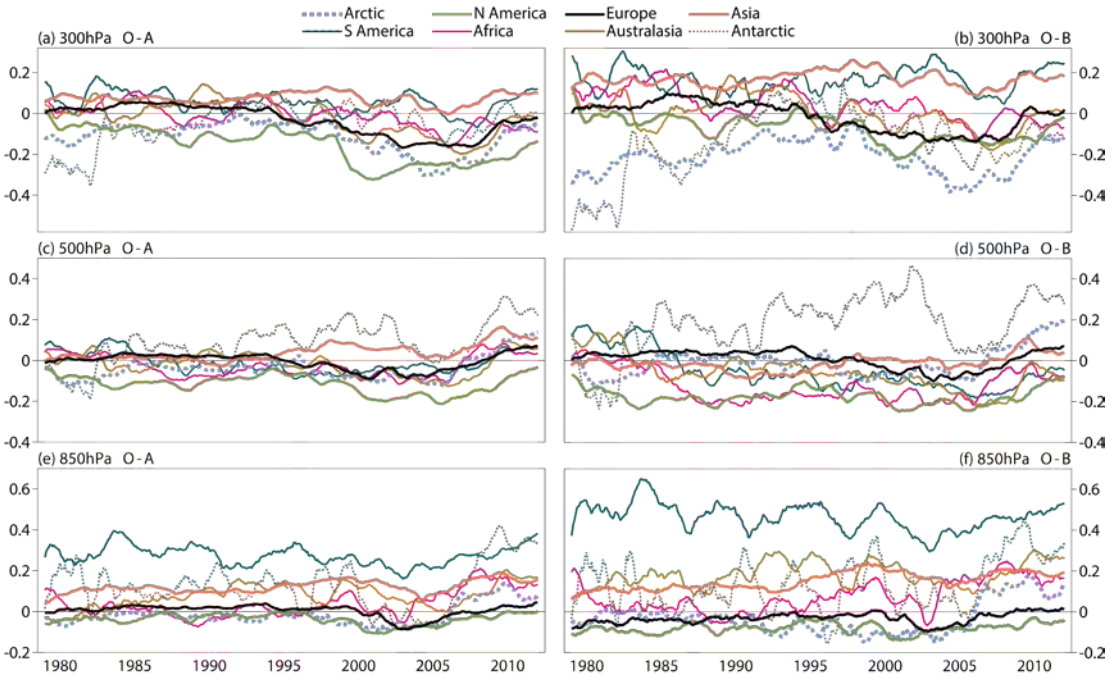

Figure 9 Twelve-month running means of observation-minus-analysis ((a),(c),(e)) and observation-minus-background ((b),(d),(f)) differences for radiosonde temperatures (K) at ((a),(b)) the 300hPa standard level and significant levels between 275 and 350hPa, ((c),(d)) the 500hPa standard level and significant levels between 450 and 600hPa, and ((e),(f)) the 850hPa standard level and significant levels between 775 and 887hPa. Averages are taken without area weighting for all assimilated measurements from the Arctic [70N-90N], North America [10N-70N; 170W-20W], Europe [30N-70N; 20W-60E], Asia [Eq-70N; 60E-170W], South America [50S-10N; 90W-20W], Africa [40S-30N; 20W-60E], Australasia [50S-Eq; 110E-170W] and the Antarctic [90S-70S].

5.3 Regional variations

Figure 9 shows regional-mean analysis and background fits to radiosonde data for 300, 500 and 850hPa. Seasonal variations are more marked for such averages, and are suppressed here by applying twelve-month running means so that longer term variations can be seen more easily. Observation numbers for the Arctic and Antarctic are low, so variations over time for these regions must be regarded with caution. Nevertheless, the fits for the Arctic, where the largest lower-tropospheric warming occurs, are within 0.2K throughout at 850hPa and 500hPa. There is generally little long-term trend in the fits for the other regions shown, especially for North America, Europe and Asia, although at 300hPa the effect of assimilating increasing amounts of warm-biased aircraft data from 1999 onwards is evident for North America and to a lesser extent Europe. Many regions show a slight dip in the time series around 2005, as seen in Figure 8 for the global average. Overall, these results give confidence in the extent to which ERA-Interim captures the main temperature changes over the period, especially in the lower to middle troposphere over the land masses of the northern hemisphere that are relatively well covered by the radiosonde network.

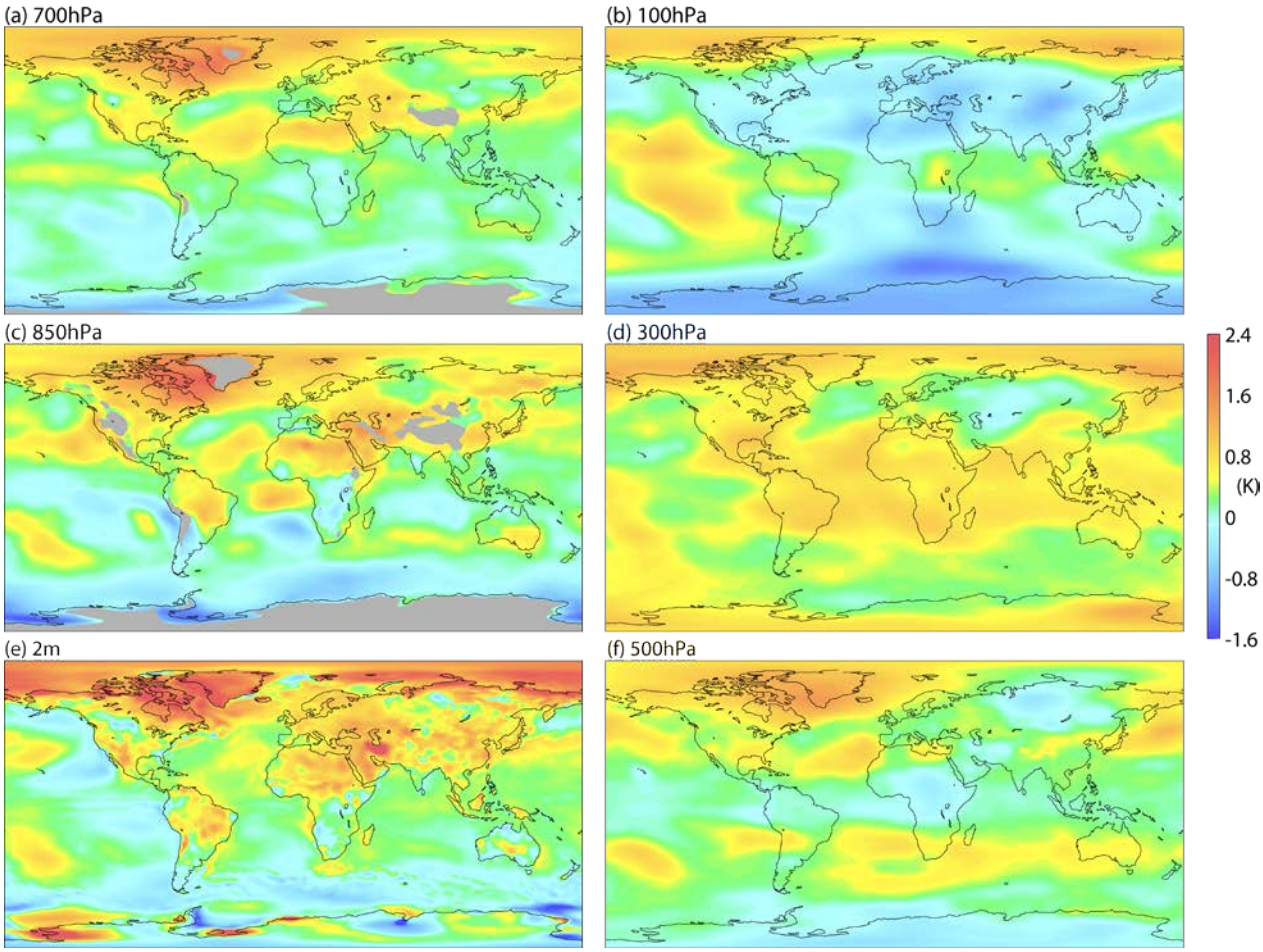

Figure 10: Changes in decadal-mean temperature (K) from (1981-1990) to (2001-2010) at (a) 700hPa, (c) 850hPa and (e) 2m, and at (b) 100hPa, (d) 300hPa and (f) 500hPa. Grey shading denotes where the 850 and 700hPa surfaces lie beneath the orography of the assimilating model.

The analysed changes in decadal-mean temperature from 1981-1990 to 2001-2010 are shown in map form in Figure 10 for levels from 2m to 100hPa. There is general consistency in the patterns of change over land from the surface to the middle troposphere, with a decrease in the strength and extent of warming with increasing height, most evident in the fall off in warming over northeast Canada and Greenland, and the expansion of the region of weak warming or cooling over Russia. In contrast, the band of weak near-surface warming or cooling stretching from western Canada to the south-eastern USA is overlain by stronger warming. More extensive warming over Australia at 850hPa than at 2m may be further evidence of the analysis problem for surface data surmised from the comparison with HadCRUT4 in section 4. The 850hPa analysis over Antarctica does not show the marked localised warming and cooling seen at 2m. It should be recalled that the latter is based on a separate analysis of synoptic surface observations not used in the upper-air analysis (Simmons et al., 2004).

Warming increases with height in the lower troposphere over sea, on average, in contrast to what happens over land. Changes over sea are nevertheless not as coherent in the vertical as over land between the surface, 850hPa and 700hPa. This may be due in part to assimilating satellite data that compensate in the lower troposphere for the cooling shift in SST discussed earlier, but there is warming southwest of the USA, west of central Africa and between Australia and southern Africa that is quite pronounced at 850hPa but not at either 2m or 700hPa. Warming over the high Arctic is much weaker at 850hPa than 2m, but a little stronger at 700hPa than 850hPa. Here it must be noted that the observational control brought about by assimilation of satellite radiance data is limited, and that MERRA (compared further in section 6) shows stronger warming at 850hPa than ERA-Interim.

Warming is quite marked in the tropics at 300hPa, as discussed further in section 9. For this level the overestimation over North America by several tenths of a Kelvin from assimilating warm-biased aircraft data must be recalled. The extratropics at 100hPa mostly show the cooling characteristic of the lower stratosphere, but warming still occurs over the north polar cap. Values in the tropics must be regarded with caution, as will be discussed further in the context of the comparisons with MERRA.

5.4 Fits to radiance data from satellites

The assimilation of satellite radiance data exercises large-scale control on the temperature analyses. As discussed by Dee and Uppala (2009) and illustrated further below, ERA-Interim draws very closely to radiances that are fully adjusted for perceived biases, and growth of error in the background forecasts is generally slow. Key to the reliance that can be placed on long-term temperature variations from reanalysis is the reliance that can be placed on the bias adjustments made to radiance data. These adjustments are dependent on such factors as the coverage and homogenisation of anchoring radiosonde data, biases in the assimilating model or the fast radiative transfer calculations used to derive background equivalents of the measured radiances, major changes in observational coverage that cause background errors to shift, such as the introduction of GPSRO data, and the extent to which data have been subject to separate inter-satellite cross-calibration prior to assimilation.

The fits to bias-adjusted satellite data and the bias estimates are presented below for averages over all used data, or later for all used data from the tropical belt from 20N to 20S, without taking account of spatial variations in observation density. At its simplest, the density of soundings from polar orbit is highest near the poles and lowest in the tropics, with some 30% of the SSU soundings coming from

poleward of 60O latitude. Use of data from lower-sounding channels depends on surface type and height, and on the presence of cloud in the case of infrared radiances and of precipitation or high liquid-water content in the case of microwave data. Thus about three times more data are used from the stratospheric HIRS-2 channel than from the tropospheric HIRS-5 channel, whilst about 50% more data are used for the stratospheric AMSU-A-10 channel than for the tropospheric AMSU-A-6 channel. With AMSU-A data available from as many as five satellites late in the period, these data are thinned prior to assimilation to avoid giving them excessive weight. This happens predominantly near the poles.

Figure 11 shows time series of the global analysis fits for selected groups of channels from the various sounding instruments. Corresponding background fits are shown in Figure 12 and the bias estimates (the values that are subtracted from the unadjusted radiances) are shown in Figure 13. Archived monthly-mean values are plotted, and for each of these figures the value for the first month in which data from a particular satellite were used has been discarded, as it can be unrepresentative if based on only a few assimilation cycles while the variational bias scheme was determining the general level of adjustment. Also not shown are SSU-1 and SSU-2 values for NOAA-8 for a short period in 1985 when data reappear in the record, as data numbers were small for much of the time. Figure 1 provides the key that links the segments of plotted data with the satellites from which the data originated.

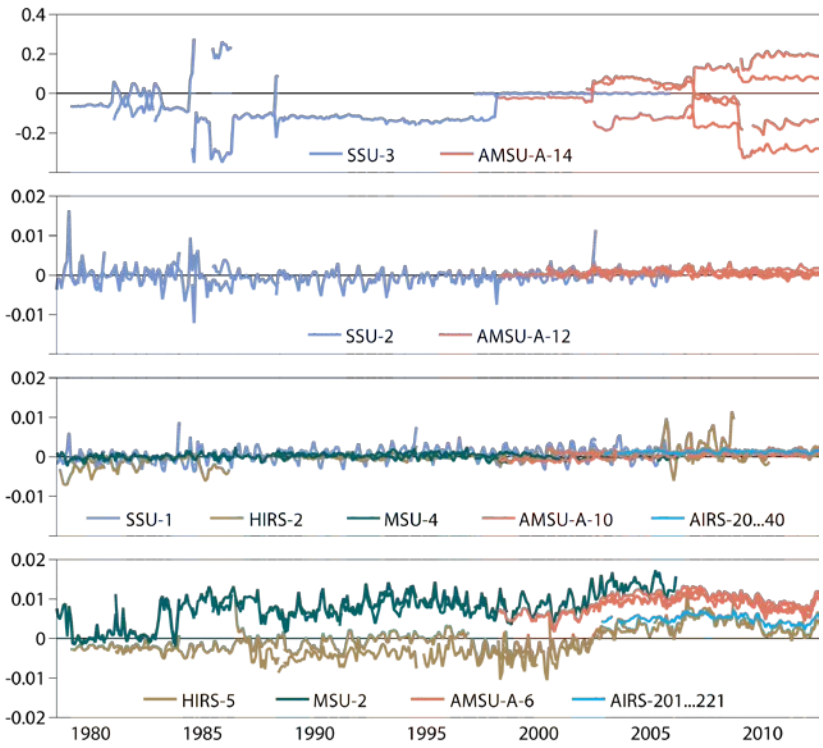

Figure 11: Monthly-mean observation-minus-analysis differences in brightness temperature (K) for selected channels from the SSU, HIRS, MSU, AMSU-A and AIRS sounding instruments. Observations have been adjusted for the estimated biases shown later in Figure 13. For the high-resolution AIRS data, results from channels in bands (20-40 and 201-221) from which data were assimilated have been amalgamated. The channels selected span from the upper stratosphere (top panel) to the middle troposphere (bottom panel), as discussed further in the text.

Data from only the two highest-sounding channels shown in Figure 11, SSU-3 and AMSU-A-14 (top panel), were not completely adjusted for bias. All radiances from these two channels received an adjustment for scan angle, but only the SSU-3 data from August 1998 (when AMSU-A-14 data started to be assimilated) received full adjustment. Before then, analysis equivalents of the SSU-3 brightness temperatures are mostly warmer than the observations, though by less than 0.2K, and quite stable over time apart from instances prior to 1989 when data from new instruments begin to be assimilated. The analysis draws quite closely to the AMSU-A-14 radiances from NOAA-15 when it is the only satellite flying the instrument, and closer still to the fully adjusted SSU-3 radiances. Later, the analysis compromises between AMSU-A-14 brightness temperatures that differ by several tenths of a K from one instrument to another.

The middle two panels of Figure 11 relate to channels that primarily sound the stratosphere, but for which full bias adjustment is applied. Here the fits in brightness temperature are mostly to within less than 0.01K, and are particularly close for the microwave and AIRS instruments. Values are amalgamated from several neighbouring channels for AIRS, but fits for individual channels are similarly good. SSU-1 and SSU-2 fits are generally a little noisy, while those for HIRS-2 vary from instrument to instrument, being similar to the microwave instruments in some cases but much worse in others, especially for the NOAA-18 data assimilated from mid-2005 to early 2010. The standard deviation of the analysis fit for the HIRS-2 on NOAA-18 is larger than for any other satellite1; values for HIRS-2 vary from satellite to satellite from about 0.2 to 0.8K, whereas corresponding values for SSU-1 lie mostly in the range from 0.4 to 0.6K. Aside from varying levels of instrumental noise, it is known that the HIRS spectral response functions differ significantly from satellite to satellite and some of the functions specified from pre-launch measurement are subject to significant error (Cao et al., 2009; Shi and Bates, 2011). Revised spectral response functions are available in a new release (version 11; Saunders et al., 2013) of the RTTOV fast radiative transfer model. This provides a ready route to improvement of future reanalyses, as previous versions of RTTOV are currently used in both the ERA and the JRA assimilation systems.

The bottom panel of Figure 11 shows analysis fits for a set of tropospheric sounding channels. Whilst still close, with model-equivalent brightness temperatures within 0.02K of the bias-adjusted observations, these fits are nevertheless poorer than for the bias-adjusted stratospheric channels. Two features are of note. The first is the clear distinction between values for the microwave (MSU and AMSU-A) and infrared (HIRS and AIRS) instruments, with better fits for the infrared. This most likely reflects where data are assimilated for the two types of instrument, as the fits for the microwave instruments include cloudy areas where background temperature errors are likely to be larger than in the clear-sky regions where the data from both microwave and infrared instruments are assimilated. The second is the cooling and subsequent warming of the analysis relative to the observations that occurs for both infrared and microwave data starting around the year 2000. This upward bowing of the observation-analysis curves is opposite in sign to the bowing in the lower-tropospheric radiosonde fits seen in Figure 8. It comes primarily from the tropics: the 20N-20S mean fit for brightness temperature

[1] An internet search readily reveals information from NOAA's post-launch investigations into the source of the noise.

for the tropospheric sounding channels reaches a maximum around the beginning of 2007 that is about 0.03K higher than typical pre-2000 values. It is also more pronounced in the corresponding fits of the background forecast presented in Figure 12. Further discussion is given in section 5.7.

Figure 12 shows that the background forecast cools relative to the tropospheric sounding radiances, although differences in brightness temperature are mainly below 0.05K, as they are for the lower-sounding stratospheric channels. Differences increase with height in the stratosphere, with warm background biases predominant at higher levels. They are relatively large for SSU data between 1985 and 1998. The unadjusted data for this period from SSU-3 on the NOAA-9, -11 and -14 platforms are more inconsistent with the background model than either the SSU-3 data from earlier satellites or the AMSU-A-14 data from later ones. The background forecasts accordingly warm more between 1985 and 1998 in the upper stratosphere due to assimilation of incompatible SSU-3 data in a region where no direct control is provided by radiosonde data. The background fits to SSU-2, SSU-1 and HIRS-2 are affected by this to varying degrees, as the weighting functions for these channels extend into the upper stratosphere. Figure 12 shows that the background fit for HIRS-2 varies from satellite to satellite during this period. The satellites for which the fit is poorer are NOAA-9, -11 and -14. The HIRS-2 fit is thus poorer for the satellites that carry an SSU instrument than for those (NOAA-10 and -12) that do not. As NOAA-9, -11 and -14 flew in “afternoon” orbits separated geographically from the “morning” orbits of NOAA-10 and -12, and as the background forecast evolves over the twelve-hour assimilation window and has a significant large-scale tidal component in the upper stratosphere, the forecast is sampled differently by the HIRS instruments on one set of satellites than by those on the other set.

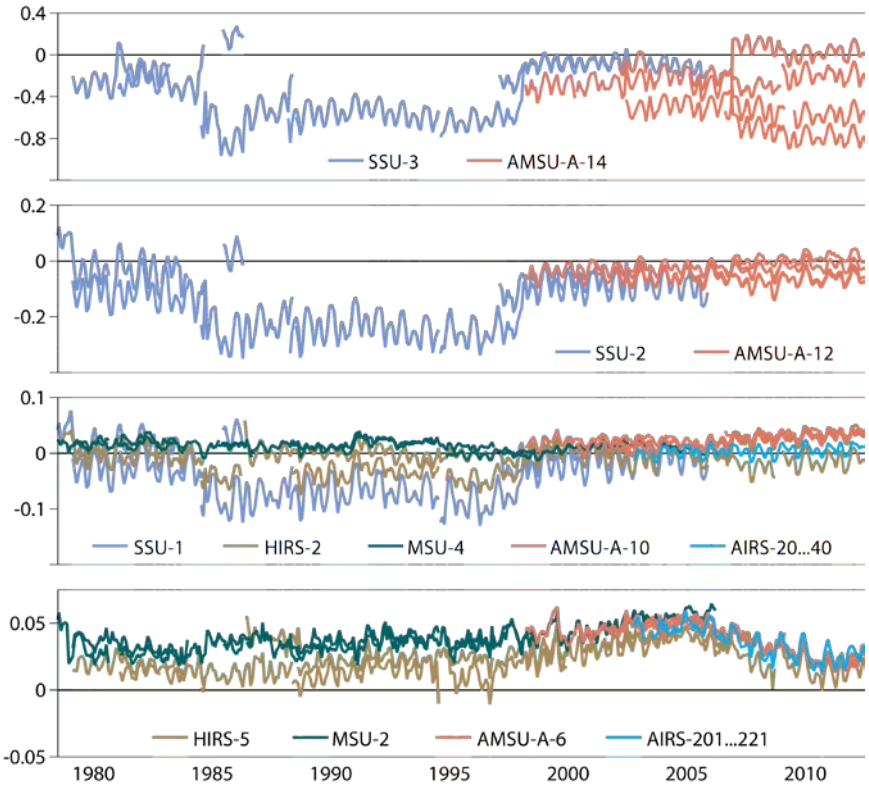

Figure 12: As Figure 11, but for corresponding observation-minus-background differences.

The change in character of the SSU-3 data near the start of 1985 can be linked with the higher mean pressures of the cells of carbon dioxide in the later instruments carried on NOAA-9, -11 and -14. This was not accounted for in the version of RTTOV used to compute background equivalents of the measured radiances in ERA-Interim. Kobayashi et al. (2009) identified the problem and showed that improvement can indeed be brought about by taking cell-pressure differences into account in the radiative transfer modelling.

Figure 12 also shows a slight drift in the observation-background differences for AMSU-A-10 and AMSU-A-12 late in the period. This indicates some contention between the background forecasts and drifts in the bias adjustments made over this period. These drifts are discussed later.

5.5 Radiance bias adjustments

The estimates of bias used to make the adjustments to brightness temperatures are shown in Figure 13. They are generally much larger than the analysis and background differences discussed above, and the range of the adjustments applied is similar to or larger than the range of analysed climatic variations. One thus looks in the first instance for estimates that vary little over time and represent primarily the adjustments (or calibrations) needed to produce continuity of values from one satellite to another for a particular instrument and channel.

Oscillations and drifts of values are nevertheless to be expected. One source of longer term variation is the changing solar heating of a satellite whose orbit drifts, as illustrated by Dee and Uppala (2009) in the case of the adjustment for MSU-2 on NOAA-14, the last of the MSU segments plotted (from 1995 to 2006) in the bottom panel of Figure 13. Another is the use of fixed CO2 in modelling the measured infrared radiances. Insofar as other observations and the prescribed SSTs control the tropospheric warming trend, a downward drift in the bias estimates for tropospheric-sounding CO2-absorption channels such as HIRS-6 or the selected AIRS channels should be inferred by the variational bias adjustment scheme, because the decrease in observed brightness temperatures due to increasing CO2 (Chung and Soden, 2010) is not matched by the background equivalents derived using fixed CO2 in RTTOV. In practise, however, these infrared instruments provide data that are part of the mix that determines the trend, so use of fixed CO2 in this component of the data assimilation system is a factor that likely contributes to some underestimation of tropospheric warming.

Dee and Uppala (2009) identified drift of possibly instrumental origin in bias estimates for the tropospheric sounding channels of the early AMSU-A instruments, which can be seen in the bottom panel of Figure 13. Conversely, there is no significant drift in the corresponding estimates for the later instruments on NOAA-18, Metop-A and NOAA-19. Lu and Bell (2013) argue that significant drift in measurements from channels 6-8 of the earlier AMSU-A instruments, but not from higher-sounding channels, arises from in-orbit changes in the frequencies that are sensed by the instruments, which are actively stabilised for channels 9-14 but not 6-8. Drift in bias estimates may also occur due to changes in the sampling of diurnally varying model bias as orbits drift, and cyclical shorter term variations can arise from annual or quasi-biennial variation in systematic model bias.

Notwithstanding the difficulties in interpretation caused by such effects, a number of conclusions can be drawn from Figure 13. The scan-angle corrections applied to SSU-3 prior to August 1998 and to the

AMSU-A-14 data thereafter are small, especially for AMSU. The ~2K bias estimate for SSU-3 brightness temperatures in the period of overlap with AMSU is very similar for the two satellites supplying SSU data at the time, NOAA-11 and NOAA-14. The bias adjustments applied to SSU-2 and AMSU-A-12 are mostly quite stable over time and less than 1K in magnitude. One exception is the sharply increasing bias estimated for NOAA-7 from mid-1981 onwards, which is not shown fully due to the choice of axis limit but grows to almost 8K by the time data cease to be assimilated early in 1985. This is consistent with the known high leakage of CO2 from the pressurized cell for this particular instrument and channel; much smaller shifts for other instruments may be due to leakage of water vapour (Nash and Saunders, 2013). The small downward trend from around 2003 onwards in the AMSU-A-12 bias estimates is discussed later in the context of differences between ERA-Interim and MERRA. The jump in the bias estimates early in 2007 for one AMSU instrument is due to recalibration by EUMETSAT of Metop-A radiances, which were assimilated from the beginning of 2007, prior to the data being formally declared operational by EUMETSAT.

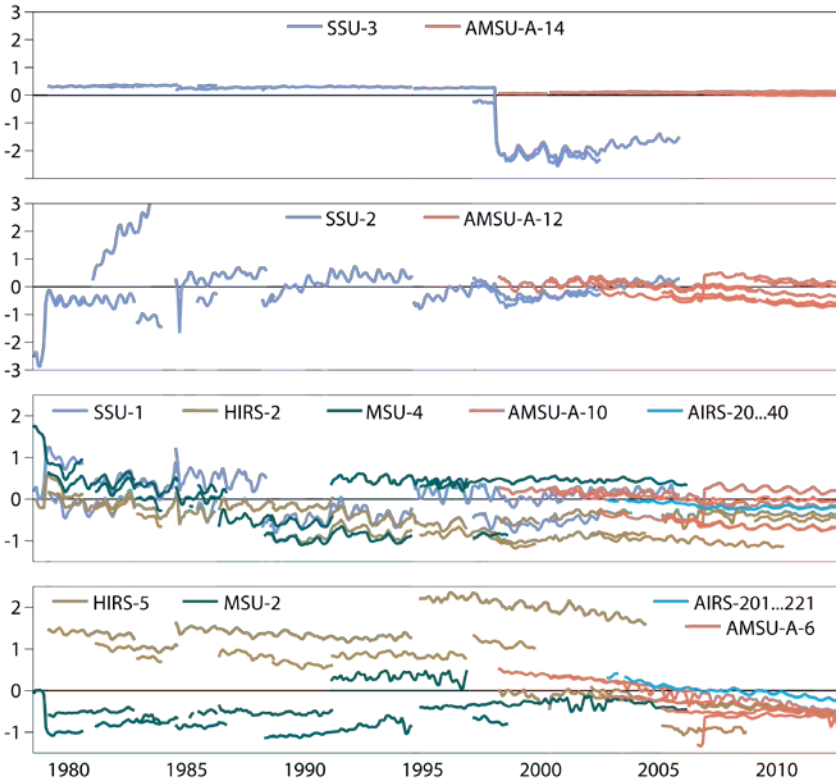

Figure 13: As Figure 11, but for corresponding estimates of biases in the data from each channel or group of channels.

The bias estimates for the channels that sound the lower to middle stratosphere are stable over time for much of the period apart from inter-satellite differences, which in the case of SSU-1 can again be linked to the cell-pressure differences discussed by Kobayashi et al. (2009). Variations do occur from time to time that are common to all types of instrument and thus indicative of issues in the data assimilation. Bias estimates for SSU-1, HIRS-2 and MSU-4 all rise in the second half of 1991 and then fall back by mid-1993 due to the absence of background-model warming from the Pinatubo

eruption. The estimates for SSU-1 and HIRS-2 (and SSU-2) drop in 1998 when full bias correction is applied to SSU-3, but eventually recover. Those for HIRS-2, AMSUA-10 and AIRS-20…40 shift to accommodate changed bias in the background forecast due to the assimilation of GPSRO data, mostly over about two years from late 2006 when the amount of occultation data increases substantially.

Particular discussion is needed for the initial period up to mid-1979 when TIROS-N was the only source of sounding data used in ERA-Interim. The bias adjustment to the SSU-2 data from this satellite shifts substantially when data subsequently become available from NOAA-6, making the adjustments for the two instruments almost identical. At the same time, shifts occur in the bias estimates shown for SSU-1, MSU-2 and MSU-4, and radiosonde data fits for the lower stratosphere deteriorate as shown earlier in Figure 8. These changes may be related to limitations in the data provided by TIROS-N: no data was assimilated from its SSU-3, HIRS-4 and HIRS-5 channels due to instrument problems. Absence of SSU-3 data from TIROS-N means that one of the key anchors of the variational bias-adjustment scheme was not operating for the first six or so months of 1979. This likely explains the shift in bias adjustment of the stratospheric-sounding channels and poorer fits to radiosonde data. For the troposphere, the number of assimilated soundings increased substantially when NOAA-6 started providing HIRS-4 and -5 data as well as additional data for the channels for which TIROS-N provided data. This change appears to have been sufficient to shift the bias adjustment of the TIROS-N MSU-2 data. Shifts also occurred at the same time in the adjustments applied to the lower-sounding TIROS-N channels not included in Figure 13. Although 1979 is often taken as the starting point for calculating trends over the modern satellite era, a start at least one year later would be appropriate given the deficiencies in TIROS-N data.

The plots discussed above are from 1989 onwards for the main ERA-Interim production stream and from 1979-1988 for the second production stream. It is nevertheless hard to discern in them any mismatches from 1988 to 1989 in the plotted data fits and bias estimates. The second production stream was in fact continued until the end of 1989, and Figure 14 presents examples of overlaps, showing bias estimates for HIRS-2 on NOAA-10, and SSU-1 and MSU-4 on NOAA-11. Differences are very small indeed by the end of 1989, but more evident in the first few months of the year. The main production stream started at the beginning of December 1988; Figure 14 and other evidence suggest that a warm-up period of three or more months would have been preferable. It is clear nevertheless that the variational bias adjustment scheme is robust to a change in starting point, which is important in practice as reanalyses are time consuming to produce and typically executed in a number of parallel streams in order to speed up production.

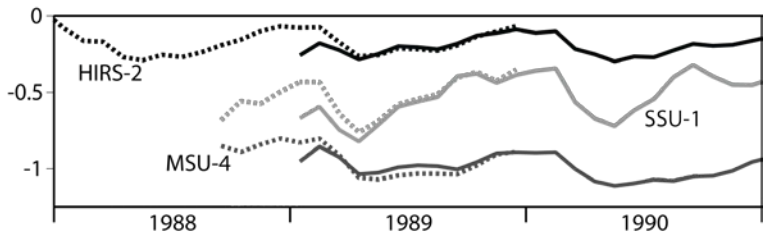

Figure 14: Examples of bias estimates from the original production stream (solid lines) from 1989 onwards and from the later production stream (dotted lines) for 1979-1988, which was extended through 1989 to study overlap.

5.6 Bias-adjusted stratospheric radiance data

It is instructive also to examine the time series of the bias-adjusted brightness temperatures themselves. Figure 15 displays HIRS-2, SSU-1 and AMSU-A-11 values, using a twelve-month running mean to remove pronounced seasonal cycles. It must be recalled that these plots are averages over all soundings, not estimates of true global-average temperatures. Also, the three channels have different weighting functions, though the values for HIRS-2 and AMSU-A-11 differ by only around 1K. To ease comparison 5.5K is subtracted from the values for SSU-1 as this channel is more sensitive to the warmer temperatures found higher in the stratosphere. Values also may absorb some bias from the background forecasts or the radiative transfer used to compute background equivalents of the measured radiances.

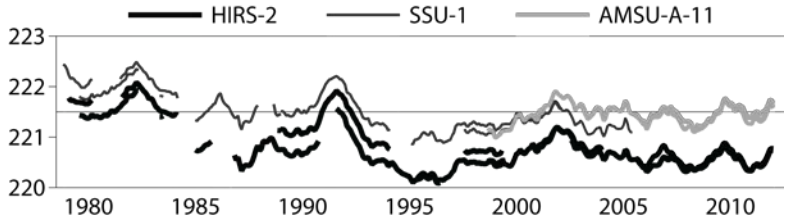

Figure 15: Twelve-month running means of bias-adjusted brightness temperatures (K) from HIRS-2 (black), SSU-1 (dark grey) and AMSU-A-11 (light grey). Averages are over all soundings, and 5.5K is subtracted from SSU-1 values.

Nevertheless, the variations over time seen in Figure 15 can be clearly identified in the height-resolved reanalysis values shown as monthly anomalies in Figure 7 (and later as twelve-month running means in Figure 17). They are also similar for the three types of instrument. There is, however, a quite distinct mismatch of the absolute values for HIRS-2 for some instruments, with values from NOAA-11 (which come in two segments, one in the early 1990s and one late in the decade) lying above those for NOAA-10, NOAA-12 and later satellites. The mismatch can occur because channel-2 spectral response functions differ from one HIRS instrument to another, but may include an effect of bias adjustments that compensate for poorly specifying the functions in the radiative transfer modelling.

5.7 Interpretation of the variation in lower-tropospheric data fits

The upper panel of Figure 16 shows time series of twelve-month running means of analysis increments, the analysis-background differences, for temperature at 850, 700 and 500hPa. The increments are calculated from complete model fields and are true global averages. The background is the average of forecasts at six- and twelve-hour range made twice-daily from 00 and 12UTC. The data assimilation generally increases background model temperatures that are perceived to be biased cold. It has already been noted that this is generally the case for the near-surface over land, and that the background is biased warm in the mid-troposphere relative to overwhelmingly land-based radiosondes, but biased cold relative to the globally-sampled tropospheric soundings from satellite. The warming increment reaches a maximum around 2005 and shows a bowed form similar to that seen in the analysis and background fits to satellite data, and thus opposite in sign to the otherwise similar feature seen in the radiosonde fits.

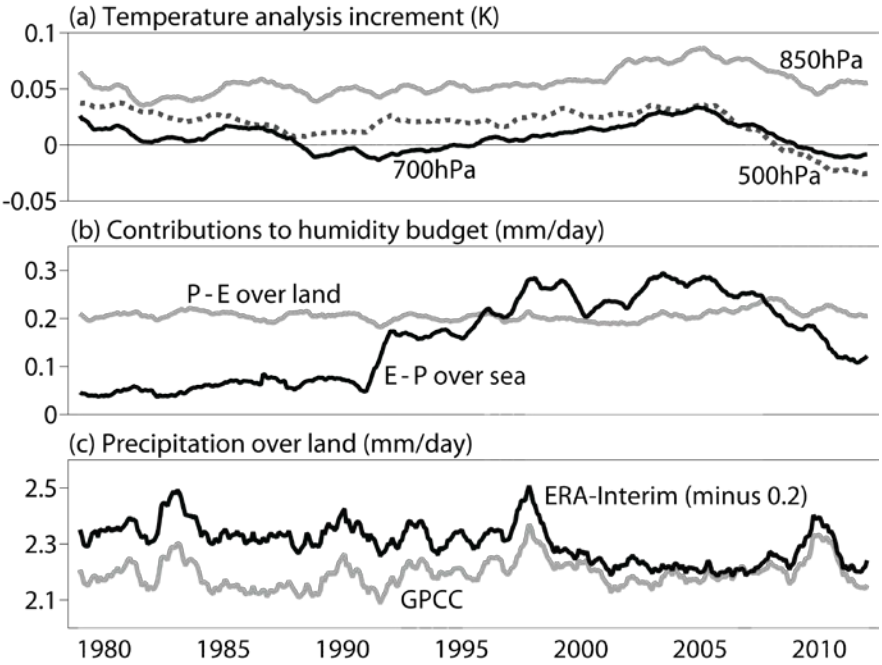

Figure 16: Twelve-month running means of (a) the global-mean ERA-Interim temperature (K) analysis increment at 850hPa, 700hPa and 500hPa, (b) the contributions to the global-mean atmospheric humidity budget (mm/day) of ERA-Interim from the difference between precipitation and evaporation averaged over land and from the difference between evaporation and precipitation averaged over sea (based on values from twice-daily 12h forecasts), and (c) precipitation (mm/day) over land from ERA-Interim and from the GPCC Full Data Reanalysis Version 6.0 up to the end of 2010 and the GPCC Monitoring Product for 2011 and 2012; ERA-Interim values are reduced by 0.2mm/day for clearer comparison with GPCC.

Evidence related to the increasing and then decreasing cold bias in background forecasts over sea is presented in the middle panel of Figure 16. It shows twelve-month running means of the contributions to the global-mean atmospheric humidity budget from the differences between precipitation and evaporation averaged over land and between evaporation and precipitation averaged over sea, from the twelve-hour forecasts (which are not generally in hydrological balance). The values for land are quite stable over time, but the excess of evaporation over precipitation increases over sea from 1991, reaching a broad maximum between 1999 and 2006, followed by a quite rapid subsequent decrease. This variation over time is comparable with that of the counts of 1D-Var retrievals from rain-affected SSMI radiances shown in Figure 2. The early version of the scheme for assimilating rain-affected radiances used in ERA-Interim had a pronounced and erroneous drying effect (Geer et al., 2008; Dee et al., 2011a), with a consequent reduction in latent heating in background forecasts and likely additional radiative effects, particularly in the tropics. Background temperatures also shift over sea because of the 2001 shift to cooler SST analyses.

From the above it can be inferred that although the background forecast model is held fixed in ERA-Interim, the biases in background temperature change due to changes in the humidity analysis and the externally provided SST analyses. Background cooling over sea is counteracted by warming analysis increments, primarily from satellite radiances. A data assimilation system in which there are systematic warming increments can be viewed as a free-running model in which an additional heating is added to represent the systematic effects of the increments, with implications for the circulation that

maintains balance (Uppala et al., 2005; Fueglistaler et al., 2009). In the case of warming tropospheric increments over sea a balancing thermally-driven circulation would be set up, with descent and thus warming of the background forecasts over land. This would in turn be partly compensated by cooling increments from assimilating radiosonde data and a reduction in the warming increment from analysing synoptic screen-level temperatures. The converse happens if the background forecasts warm over sea. This is a plausible explanation for the difference in sign of the bowing seen in the data-fit time series for radiosondes and satellite radiances, and the dip centred around 2005 in the increments shown in Figure 4 for the near-surface analysis.

Precipitation would be expected to decrease over land in ERA-Interim when moisture supply from the ocean is reduced by the drying of marine air that results from assimilating rain-affected radiances and from shifting to cooler SSTs that give less evaporation. Additional suppression of rainfall would be expected due to the descent over land forced by systematic warming increments over sea. Precipitation over land would increase when drying of the marine atmosphere is replaced by moistening.

This is indeed what is found. The third panel of Figure 16 compares time series of twelve-month running means of precipitation over land from ERA-Interim and from the Global Precipitation Climatology Centre (GPCC; Becker et al., 2013). To counter inhomogeneity over time in the distribution of gauge data available for analysis by GPCC, the land-average for both ERA-Interim and GPCC has been made by averaging only over the 2938 1O grid squares for which GPCC had access to data from at least one station for every month of the period. The resulting time series for GPCC shows a slight trend for precipitation to increase over time and maxima that are presumably related to increased moisture supply from the ocean associated with the El Niño events of 1982-83, 1997-98 and 2009-10. The successive twelve-hour ERA-Interim forecasts produce more precipitation than analysed by GPCC, but the discrepancy varies over time. It is about 0.3mm/day on average for the first decade or more before declining slowly at first and then more quickly to reach about 0.2mm/day from around 2004 to 2007. It then increases slightly. Notwithstanding this, ERA-Interim reproduces the shorter-term variability in the GPCC mean very closely, in line with the agreement reported for regional averages by Simmons et al. (2010). Implications for soil-moisture trends are discussed by Albergel et al. (2013).

6 Comparison with MERRA

6.1 Area means

Figure 17 compares time series of global-mean temperature from ERA-Interim, MERRA and ERA-40. To emphasise differences between the three reanalyses, twelve-month running means are shown, with all values presented relative to the ERA-Interim average for 1981-2010. This, plus the availability now of ERA-Interim data from 1979 onwards, makes the distinction between ERA-Interim and MERRA clearer than shown for 1989-2010 by Rienecker et al. (2011) for monthly anomalies expressed relative to the individual climate means of each reanalysis. It similarly provides a clearer distinction between ERA-Interim and ERA-40 than presented in Figure 7. This is achieved at the expense of blurring sharp

changes, which are spread over twelve months. Figure 18 shows corresponding differences between ERA-Interim and MERRA, separated into averages for the extratropical hemispheres and the tropics.

The reanalyses agree quite closely at 300hPa in the global mean. Here ERA-Interim and MERRA vary similarly over time, with MERRA quite consistently warmer than ERA-Interim, though by only 0.1-0.3K for most of the time. Agreement is better still at 70hPa and especially 50hPa. At this and higher levels MERRA shows slightly more warming due to the El Chichón and Pinatubo eruptions. Like ERA-Interim, it did not include anomalous radiative forcing due to volcanic stratospheric aerosols, but it did assimilate MSU radiances that had been cross-calibrated (Zou et al., 2006) to remove the gross inter-satellite radiance differences that had to be accounted for by variational bias-correction in ERA-Interim. The variational bias-correction scheme for MSU radiances in MERRA may thus have been set more conservatively than in ERA-Interim, and consequently may have absorbed less of the volcanic signal in the radiances.

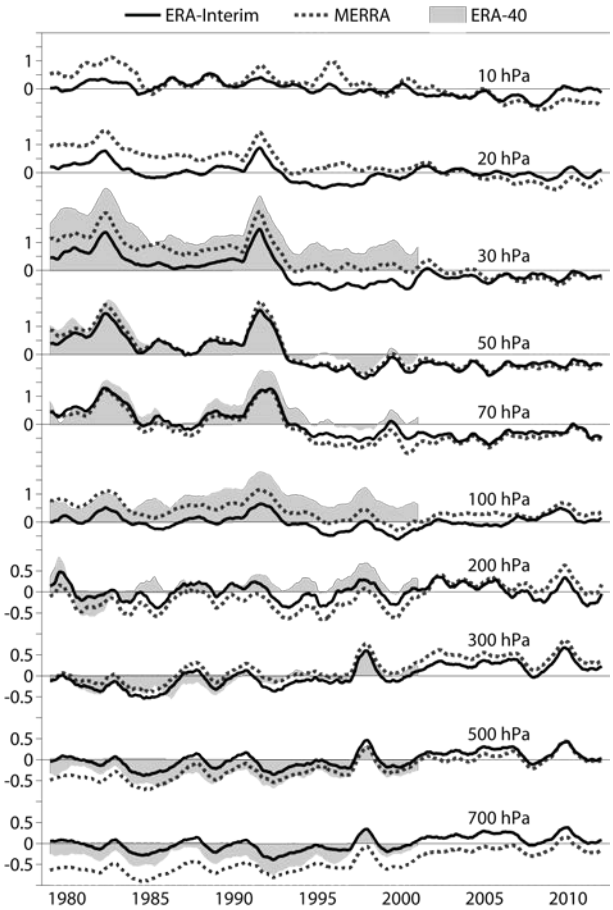

Figure 17: Twelve-month running mean anomalies in global-mean temperature (K) from ERA-Interim (solid), MERRA (dotted) and ERA-40 (shaded) at selected levels. Anomalies are calculated in each case relative to the ERA-Interim monthly means for 1981-2010. The global averages for 700hPa exclude points for which MERRA does not provide values due to the presence of orography. ERA-40 results are not shown at 10 and 20hPa as differences from ERA-Interim and MERRA are large at these levels, and the change in axis needed to display them would detract from the differences between ERA-Interim and MERRA.

MERRA generally cools more than ERA-Interim at 30, 20 and 10hPa, somewhat more so in the tropics than the extratropics. There is a short period in the mid-1990s where MERRA is more than 1K warmer than ERA-Interim in the tropics at 20hPa, and a corresponding though smaller difference is seen at 30hPa. Otherwise, the two reanalyses are within 1K of each other for the three levels, apart from the first few years at 20 and 10hPa. Comparisons above 10hPa are made in section 7.

ERA-Interim is generally closer to MERRA than it is to ERA-40. This is the case for all levels other than 700 and 300hPa for the root-mean-square differences of the running-mean global averages for 1979-2001 plotted in Figure 17. For root-mean-square differences computed over all points of MERRA's 1.25O product grid and all individual monthly means, it is only at 700hPa that MERRA is not the closer to ERA-Interim. This is despite the degree of commonality between the assimilating models used for the two ERA reanalyses. It indicates that ERA-Interim and MERRA are both subject to substantial control by the assimilated radiosonde and radiance data.

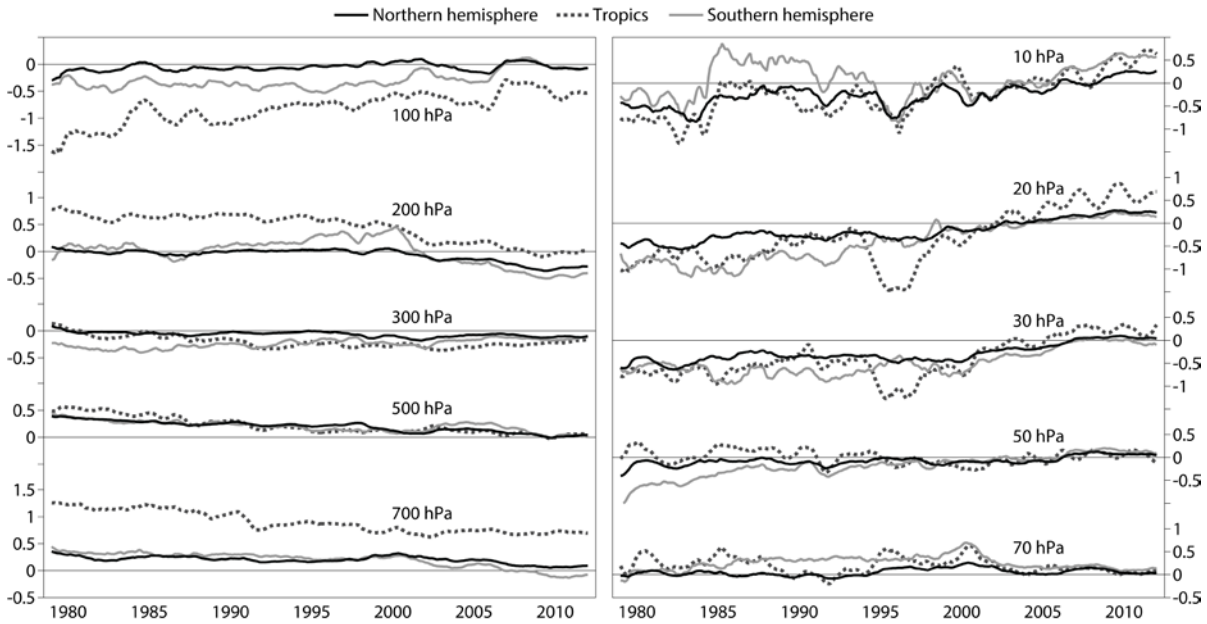

Figure 18: As Figure 17, but showing the differences between ERA-Interim and MERRA temperatures separately for the extratropical northern hemisphere (20N-90N; solid black), the tropics (20S-20N; dotted) and extratropical southern hemisphere (90S-20S; solid grey).

The differences at 700hPa between ERA-Interim and MERRA are particularly large in the tropics, especially early in the period, and differences in extratropical averages come mostly from the subtropics at this level. MERRA's tropical warming at 700hPa is much larger than ERA-Interim's, and is larger relative to near-surface warming than expected from considerations to be presented in section 9. At 850hPa MERRA exhibits substantial warming over central Africa and cooling over central South America that have no counterparts in ERA-Interim. Corresponding trends in surface soil moisture from associated land reanalyses are for pronounced drying and moistening respectively over these regions in MERRA, behaviour that is seen neither in ERA-Interim nor in analyses of microwave data (Albergel et al., 2013).

6.2 Regional differences

Maps of the change in decadal-mean temperature from 1981-1990 to 2001-2010 at 500, 300 and 100hPa are shown for MERRA in Figure 19. Differences between ERA-Interim values presented earlier in Figure 10 and these MERRA values are also shown. At 500hPa ERA-Interim changes less over time than MERRA over much of the tropics and subtropics; Figure 18 shows a decline over time in the differences between ERA-Interim and MERRA that is similar for tropical and extratropical averages. Differences are particularly small at 500hPa over the land masses of the northern hemisphere where radiosonde coverage is good, and the main patterns of change are largely similar for the two reanalyses at this level and at 300hPa, as they are lower down in middle to high latitudes (not shown).

The largest difference at 300hPa in the northern hemisphere lies over North America, despite the abundance of radiosonde data there. ERA-Interim thus appears to be more susceptible than MERRA to warming from assimilating increasing amounts of biased aircraft data that are especially dense over North America. At 200hPa (not shown), ERA-Interim warms a little more than MERRA over the main North Atlantic flight routes as well as North America. The largest difference at this level occurs in the tropics (Figure 18), where MERRA warms quite substantially and ERA-Interim changes little; this shows in the global averages presented in Figure 17. Much of the 200hPa warming in MERRA occurs from the late 1990s onwards. Further discussion is given in the following section.

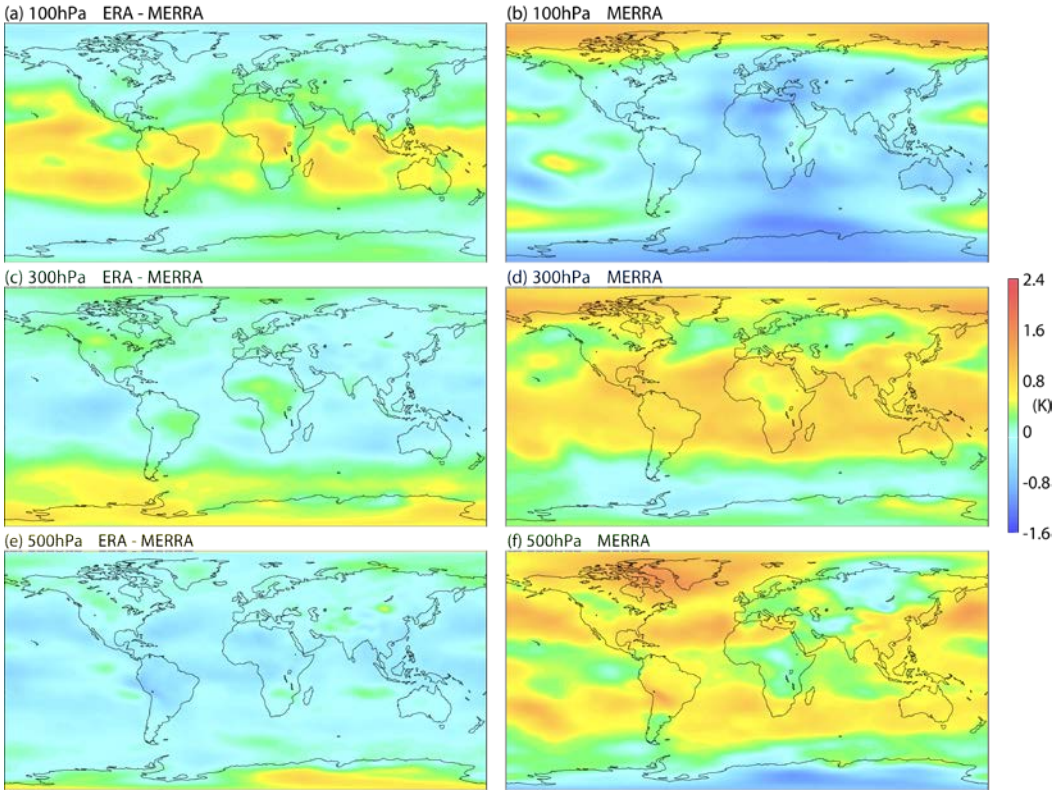

Figure 19: Changes in decadal-mean temperature (K) from (1981-1990) to (2001-2010) from MERRA at (b) 100hPa, (d) 300hPa and (f) 500hPa, for comparison with ERA-Interim results shown in the corresponding panels of Figure 10. The differences between ERA-Interim and MERRA are also shown, for (a) 100hPa, (c) 300hPa and (c) 500hPa.

Differences between ERA-Interim and MERRA change sign in the tropics between 200 and 100hPa, and are slightly larger at 100hPa (Figure 18). Tropical differences in fact oscillate in the vertical between 500hPa and 70hPa over most of the period. This is to be expected as an analysis that is subject to error in a shallow, near-tropopause layer must have error of opposite sign in neighbouring layers if it is to fit the broader-scale data from nadir-sounding satellite-borne instruments. As at 200hPa, the 100hPa differences decrease over time. Figure 19 maps the change, showing large differences over much of the tropics and southern subtropics. Here ERA-Interim warms and MERRA cools up to around 2007, when both warm, ERA-Interim more so because of its assimilation of GPSRO data. Differences are quite small elsewhere in the extratropics, where both reanalyses show large areas of the cooling that characterises the lower-stratospheric change from the one decade to the other, but warming of the northern polar cap.

6.3 Tropical data fits and bias adjustments

Further information on the tropical differences discussed above may be gleaned from data fits and bias adjustments. Figure 20 shows twelve-month running means for the tropics of ERA-Interim fits to assimilated, bias-adjusted radiosonde temperatures, and the corresponding bias estimates, for 700, 200, 100 and 20hPa.

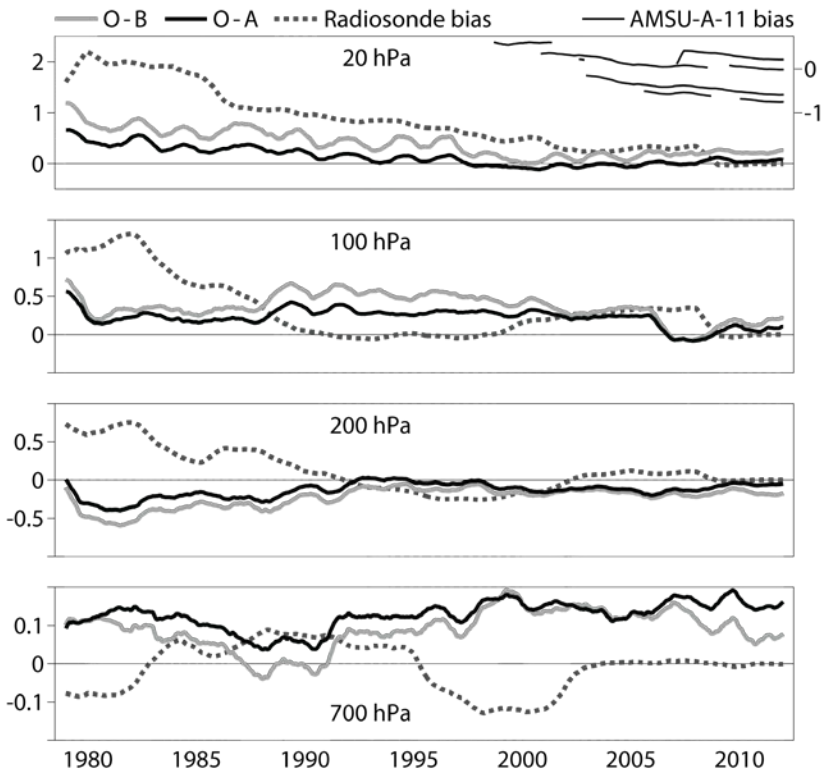

Figure 20: Twelve-month running means from ERA-Interim of observation-analysis (solid black) and observation-background (solid grey) differences for radiosonde temperatures (K; left-hand scale), and of the radiosonde bias estimates (dotted) applied to adjust a subset of the measurements (dotted lines), for the indicated standard and nearby significant levels. Also shown (top panel) are the bias estimates used to adjust AMSU-A-11 brightness temperatures (K; right-hand scale, offset for clarity). Averages are taken over all data from the tropics, with no area weighting.

At 700hPa the analysis is biased cold relative to the radiosonde observations, though by no more than 0.2K. The assimilation system does not move the background closer to the observations, as the bias in the background is generally smaller than that of the analysis. From the late 1990s onwards the tropical background cools and then warms relative to the analysis as discussed in section 5.7. The net increase over time of the differences from radiosondes suggests a slight underestimate of the overall warming of the tropical lower troposphere in ERA-Interim. However, what is seen in the fit of ERA-Interim to radiosondes is far from enough to explain the differences between ERA-Interim and MERRA, given that the 700hPa differences of around 1K or more shown for the tropical average in Figure 18 originate from most of the tropical belt, including the regions with denser radiosonde coverage.

Variations over time in the homogenising radiosonde bias adjustment are sufficiently large to complicate the picture at 700hPa, as they are at 200 and 100hPa also. Bias estimates at these levels tend to be anti-correlated with observation-analysis and observation-background differences: the data assimilation system tends to counteract fluctuations in the adjustments made to the radiosonde observations prior to assimilation. The variations in bias estimates may well reflect variations over time in the ERA-40 and operational ECMWF background forecasts used to estimate the RAOBCORE adjustments, but changes in radiosonde type not accounted for in the adjustments may also be a factor (L. Haimberger, personal communication). It should also be recalled that an increasing fraction of radiosonde data was not subject to adjustment from around the year 2000 onwards.

Aside from this, analysis and background fits at 200hPa are similar, with little net variation over time from the early 1990s. There is no evidence from the ERA-Interim radiosonde fits to support the tropical warming of about 0.5K that occurs at 200hPa in MERRA but not ERA-Interim. This warming happens when an increasing amount of AMSU data is assimilated, and may be linked with a consequential large increase in precipitation that is known to occur in MERRA but not ERA-Interim (Rienecker et al., 2011).

Conversely, the differences in tropical temperature at 100hPa may be ascribed, in significant part at least, to bias in ERA-Interim. Figure 20 shows that the ERA-Interim analysis is systematically colder than indicated by radiosondes, and the background more so, up until the time GPSRO data begin to be assimilated. The biases against radiosondes are smaller than the tropical mean differences between ERA-Interim and MERRA, but these differences are largest in regions of sparse radiosonde coverage where bias may build up over successive cycles of background forecast. The cold-biased tropical tropopause in ERA-Interim leads to a dry-biased stratosphere, and was a problem known from the outset of ERA-Interim. Test assimilation of MIPAS limb-sounded radiances using a version of the ECMWF forecasting system similar to that used in ERA-Interim was characterised by considerable moistening of the stratosphere (Bormann and Thépaut, 2007). Schoeberl et al. (2012) and Fueglistaler et al. (2013) provide recent discussion of this topic.

The ERA-Interim analysis fits the tropical radiosonde data quite closely at 20hPa from the early 1990s, and the background forecast has a cold bias that is smaller from the late 1990s. The latter may be associated with the better background fit to AMSU-A than to SSU data shown earlier in Figure 12, but the cold bias of the background compared to radiosondes is in contrast to the warm bias compared to radiances. This may be indicative both of residual warm bias in the adjusted radiosonde data

assimilated earlier in the period and of unadjusted warm bias in much of the radiosonde data assimilated for recent years.

Aside from this, the fits to tropical radiosonde data at 20hPa, or indeed to corresponding radiance data (not shown), give no sign of anomalous behaviour in the mid-1990s, suggesting that the relatively large 20hPa differences between ERA-Interim and MERRA at this time (Figure 18) stem from a spell when MERRA did not fit the observations as well as usual.

6.4 Two other differences

ERA-Interim is systematically colder than MERRA at both 20 and 30hPa for more than half the period, but warms over a few years from the late 1990s to become similar to MERRA. This is clearly seen in the global averages presented in Figure 17, and is found individually for the tropics and extratropics. The change occurs around the time AMSU-A data start to be assimilated from first one then two satellites. ERA-Interim's behaviour is consistent with warming over this period in the bias-adjusted radiance data (Figure 15) and in the radiosonde data, for which good analysis and background fits have been demonstrated (Figure 8). Some warming in both ERA-Interim and MERRA is seen at 50hPa at this time. This behaviour is not, however, supported by the model behaviour shown in the following section.

MERRA cools more than ERA-Interim at 10, 20 and 30hPa over the final decade. Again, this can be seen in Figure 17 and occurs separately for tropics and extratropics. In this case the evidence is quite clearly in favour of MERRA. ERA-Interim fits the radiosonde data closely over the final decade, but there is an accompanying drift in the bias adjustment of AMSU-A radiances that is questionably common to all satellites. This can be seen in the bias estimates inferred for channel 11 (included in the 20hPa panel of Figure 20) and channel 12 (Figure 13). The 20hPa level is characterised by a substantial increase over recent years both in the amount of radiosonde data assimilated and in the fraction of the data not subject to bias adjustment. ERA-Interim draws to these data rather than unadjusted AMSU-A data. MERRA draws instead to AMSU-A data that appear to receive little bias adjustment: radiosonde fits viewable at http://gmao.gsfc.nasa.gov/research/merra/catalog/ show a shift in the MERRA analysis and background from warm in 2001 to cold in 2011 relative to 20hPa radiosonde data. The model data presented next support MERRA in this regard.

7 Comparison with model integrations

ERA-20CM (Hersbach et al., 2013) is an ensemble of ten atmospheric-model runs for 1899-2009 that sample uncertainty in prescribed SSTs from the first release of the HadISST2 analysis (HadISST2.0.0.0; Titchner and Rayner, 2014). It uses a 2012 version of the ECMWF model rather than the 2006 version used in ERA-Interim, with ~125km horizontal resolution and 91-level vertical resolution extending to 0.01hPa. Its prescribed time-varying concentrations of ozone and long-lived greenhouse gases, total solar irradiance and aerosols follow CMIP5 recommendations (Taylor et al., 2012).

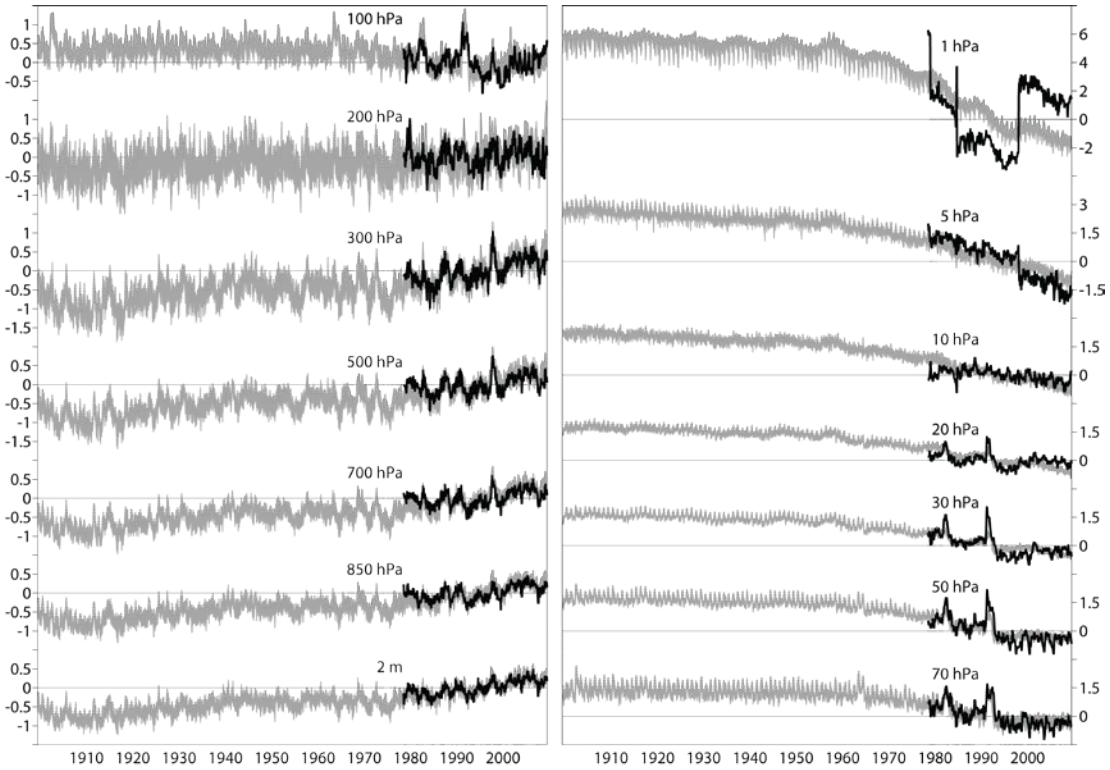

Figure 21: Anomalies in monthly and globally averaged temperatures (K) from ERA-Interim (black) and from ten model runs (grey) for 1900-2009 with prescribed distributions of SST (varying from run to run), sea-ice and trace constituents, at the indicated tropospheric and stratospheric levels. Anomalies are calculated relative to 1980-2009, and for each model run the anomaly is relative to the ensemble mean.

Figure 21 shows global-mean temperature anomalies for each ERA-20CM ensemble member, and for ERA-Interim from 1979 onwards. Results are for levels from 2m above the ground to 1hPa. Anomalies are with respect to 1980-2009, and to the ensemble-mean model climate for the ERA-20CM runs. Anomalies rather than twelve-month running means are presented as the aim here is to examine details of the variations over time, rather than document the mean errors that build up in the ERA-20CM model. ERA-20CM results are discussed more generally by Hersbach et al. (2013), but it is noted here that for the 1999-2009 period when ERA-Interim benefitted from assimilating AMSU-A data, the ERA-20CM ensemble mean is on average 2.3K warmer than ERA-Interim at 1hPa, 3.1K warmer at 5hPa, 1.6K warmer at 10hPa and 0.4K warmer at 20hPa. The bias switches to cold below 30hPa. For standard levels below 1hPa, the largest warm bias is 5.3K at 2hPa and the largest cold bias is 2K at 200hPa. Discontinuities in the ERA-Interim record are much larger at the 1hPa level illustrated than at 2hPa and the other standard upper-stratospheric levels not included in Figure 21.

ERA-20CM exhibits tropospheric warming that is quite coherent in the vertical. Temperature at 2m over sea is tightly constrained by the prescribed SSTs, and agreement over land with CRUTEM4 is good for the whole period, apart from somewhat lower temperatures over North America in ERA-20CM prior to the 1940s (Hersbach et al., 2013). There is likewise particularly good agreement with ERA-Interim at this level. ERA-Interim lies just within the ensemble in the free troposphere, but warms at a slightly lower rate than the ensemble mean. Ensemble spread increases with height in the

troposphere, consistent with the known amplification with height of low frequency variability and trends in temperature discussed for the tropics in section 9.

Ensemble spread is smaller relative to changes over time at stratospheric levels. Here there is general cooling in ERA-20CM, punctuated by warming at lower levels associated with the volcanic eruptions of Santa Maria in 1902 and Agung in 1963, as well as El Chichón and Pinatubo. ERA-20CM underestimates the vertical extent of the warming associated with the latter two events compared to ERA-Interim. The general cooling over time in ERA-20CM increases only slowly with height in the lower to middle stratosphere, but cooling is much more substantial at 5hPa and especially 1hPa. ERA-20CM includes prescribed ozone variations linked to the eleven-year solar cycle, and the associated variations in heating due to solar absorption explain the oscillation in temperature evident at 1hPa.

ERA-20CM agrees with ERA-Interim in showing little variation in lower-stratospheric temperature between the eruptions of El Chichón and Pinatubo and after the fall in temperature immediately following Pinatubo. It does however exhibit more cooling than ERA-Interim from 10 to 30hPa over the final few years of the period, and in this regard is closer to MERRA and JRA-55. Above 10hPa the temperature time series for ERA-Interim display near-discontinuities linked to the SSU and AMSU-A changes discussed earlier. The spike in early 1985 occurs for a short period when few SSU-3 data are available. In between the sharp shifts, the ERA-Interim temperatures vary similarly to those of ERA-20CM, at both 5 and 1hPa. Similar behaviour is found at the 2, 3 and 7hPa levels (not shown). There is good consistency between the height-resolved upper-stratospheric cooling extracted from SSU and AMSU-A data by ERA-Interim and the corresponding cooling simulated by ERA-20CM due to prescribed changes in carbon dioxide and ozone.

Figure 22 is an expanded plot for 1979-2009 comparing also with MERRA, for levels from 30 to 1hPa. ERA-20CM is more consistent with MERRA than ERA-Interim with regard to cooling over the later years of the period for the lower three of these levels. ERA-20CM also does not warm from the late 1990s to early 2000s, and is closer to MERRA, and indeed to JRA-55, than to ERA-Interim earlier in the period at 10 and 20hPa. Conversely, ERA-20CM is more consistent with ERA-Interim than MERRA is at 5 and 1hPa, discounting the discontinuities in ERA-Interim. The annual oscillation seen for MERRA at 5hPa is due to a shift in annual cycle with the change from SSU to AMSU assimilation (S. Pawson, personal communication), and MERRA does not cool over time at this level to the extent that ERA-20CM and ERA-Interim do. MERRA does not exhibit the near-discontinuities present in ERA-Interim, but otherwise does not match as closely the variations over time of ERA-20CM. Nevertheless, both reanalyses indicate cooling at 1hPa that is much more marked than at lower levels. It can be inferred from Figure 7 that the net cooling of JRA-55 from 1979 to 2009 is similar to that of ERA-20CM at both 5 and 1hPa.

8 The SSU data record

Thompson et al. (2012) drew attention to substantial differences between two processed SSU records, an earlier Met Office one subsequently discussed by Nash and Saunders (2013) and a more recent one reported by Wang et al. (2012). Differences are marked for SSU-1 and SSU-2, with Wang et al.

showing declining brightness temperatures in between the warmings associated with the eruptions of El Chichón and Pinatubo, and a much more substantial cooling at the end of the Pinatubo event. The fall in brightness temperature derived by Wang et al. for these channels is around 1K larger from 1985 to 1996 than seen in the Met Office record, and is about 2.5K overall from 1980 to 2005. Thompson et al. also showed that the Met Office record for SSU-1 is in much better agreement with corresponding simulations by coupled atmosphere-chemistry (CCMVal2; Eyring et al., 2010) and atmosphere-ocean (CMIP5) models. For SSU-2 both sets of models produce cooling between that of the two differing records. The Met Office and Wang et al. data are in quite close agreement for the higher-sounding SSU-3 channel, with both showing more cooling than the models. Thompson et al. used neither radiosonde nor HIRS-2 data, nor any reanalysis outputs, to shed light on the discrepancy in the SSU-1 and SSU-2 records.

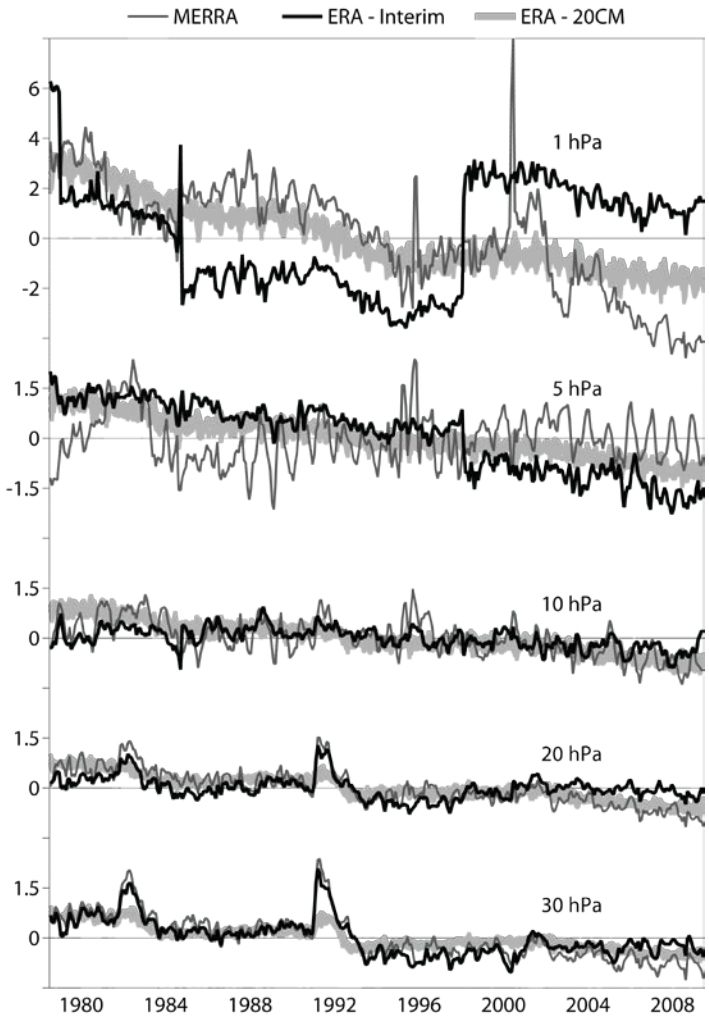

Figure 22: As Figure 21, adding the MERRA anomalies (thinner lines, darker grey) but showing only the uppermost five highest levels and the period 1979-2009.

The results presented here provide no evidence for cooling of the extent indicated by Wang et al. (2012) for SSU-1. None of ERA-Interim, MERRA and JRA-55 show net stratospheric cooling up to 10hPa between the El Chichón and Pinatubo events, and none cool as much immediately after

Pinatubo as indicated for SSU-1 by Wang et al.. Cooling of the order of 2.5K or larger from 1980 to 2005 occurs only at 5hPa and above in the results presented earlier. This is incompatible with the similar cooling shown by Wang et al. for SSU-1 and SSU-2, given that SSU-1 senses lower in the stratosphere than SSU-2. The fits of ERA-Interim to bias-adjusted SSU, HIRS-2, MSU-4 and relevant AMSU-A radiances, and to radiosonde data, show no evidence of a substantial discrepancy in net cooling from 1980 to 2005, and there is no indication from the time series of bias adjustments that a missing cooling has been absorbed by the adjustment process. Furthermore, the bias-adjusted radiances from SSU-1 have been shown in Figure 15 to be coherent with those from HIRS-2 and AMSU-A-11, and are much more similar in their variation over time to the Met Office record than to that of Wang et al.. Finally, the agreement between the reanalyses and ERA-20CM simulations is generally closer than in the comparisons presented by Thompson et al. (2012) for equivalent SSU radiances (discounting ERA-Interim's upper-stratospheric discontinuities), even for the CCMVal2 models that, like ERA-20CM, used prescribed SSTs.

It is beyond the scope of this report to delve into the various adjustments made by Wang et al. (2012) in constructing their SSU record. There is, however, a fundamental limitation to the approach of adjusting brightness temperatures from deep sounding channels to identify stratospheric temperature changes. Reanalysis and modelling both point to a large increase in cooling with increasing height in the upper stratosphere. This trend in lapse rate contributes to change in measured radiances in addition to the contribution from the trend in deep-layer-mean temperature. The adjustments documented by Wang et al. still result in a radiance record that is not easy to interpret in terms of temperature change. Further comment is given in section 10.

In contrast, there is good reason to suppose that reanalysis will be able in future to improve its height-resolved stratospheric temperature products by making better use of the SSU data record. Aside from the potential for algorithmic improvement of the analysis process and reduction of systematic error in the background model, which remains relatively large in the upper stratosphere in the case of ECMWF (Hersbach et al., 2013), improvements have already been incorporated in the fast radiative transfer model used in assimilating the SSU data for ERA and JRA. Version 11 of RTTOV already enables use of time-varying CO2 concentrations and cell pressures that can change from satellite to satellite, and over time for individual satellites (Saunders et al., 2013).

9 Temperature and humidity in the tropical upper troposphere

The extent to which temperature trends amplify with height in the tropical troposphere has been a longstanding issue due to the need to reconcile conflicting estimates from observations and models (Thorne et al., 2011). Recent evidence points to increasing agreement between the two (Haimberger et al., 2012; Lott et al. 2013; Mitchell et al., 2013) and reanalysis provides further reconciliation. Examination of the consistency of humidity fields is also required, as questions have been raised over the capability of reanalysis to provide reliable information on humidity changes in the tropical upper troposphere and their links with temperature changes. This is important for quantifying radiative feedbacks from water-vapour changes associated with warming from increased greenhouse-gas emissions. In this regard, Dessler and Davis (2010) have documented an improving situation, pointing

out that the newer ERA-Interim (based on the original data from 1989 onwards) and MERRA reanalyses are in better agreement than earlier reanalyses with theory and observation for both short- and longer-term variations.

Figure 23 shows time series of tropical-mean temperature anomalies at 2m and 300hPa, and corresponding specific- and relative-humidity anomalies at 300hPa. Results are shown for ERA-40 as well as ERA-Interim. The two reanalyses agree quite closely for temperature, apart from a short spell in the mid-1980s. Both vary similarly over time at the two levels, but the amplitude of the variations is much larger at 300hPa. An amplification factor of 2.2 from 2m to 300hPa is found for ERA-Interim by least-squares fitting the full variability of the two time series; least-squares linear trends differ by a factor 2.4. The amplification factors computed in these ways for levels from 200 to 400hPa fit within the range of model values reported by Lott et al. (2013) and Mitchell et al. (2013) for trends over the past 30 to 50 years, agreeing better than the values computed directly from homogenised radiosonde data. Agreement is marginal at 500 and 600hPa, however. Variability is a little larger at 500hPa than at 2m, but the linear trend is close to zero at this level (Haimberger et al., 2012). The decadal difference map for 500hPa in Figure 10 shows cooling over Africa that is likely linked with a change in observational coverage and the exaggerated decline in rainfall there that was noted by Dee et al. (2011a). No such behaviour occurs in ERA-20CM, whose tropical warming at 500hPa is larger than at 2m and smaller than at 300hPa.

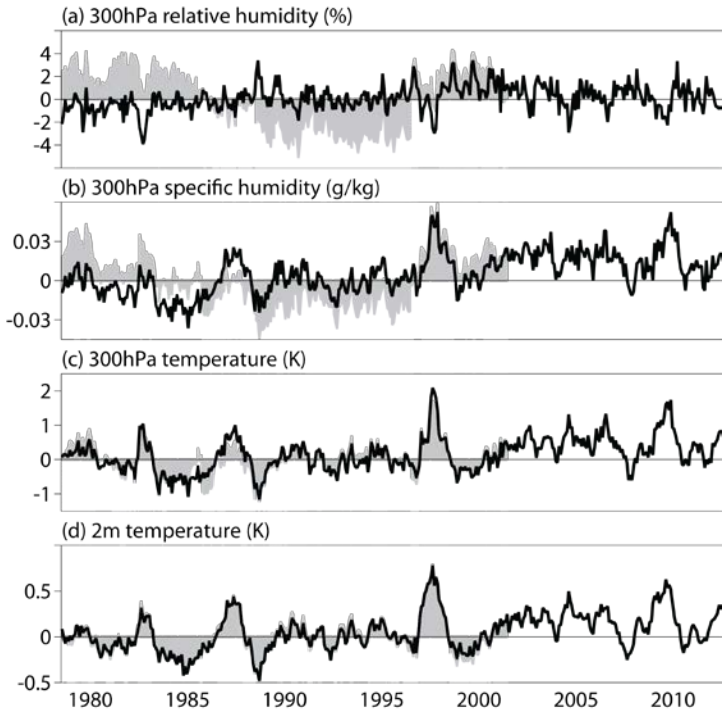

Figure 23: Anomalies in monthly and tropical averages relative to 1979-2001 from ERA-Interim (black lines) and ERA-40 (shading) at 300hPa for (a) relative humidity (%), (b) specific humidity (g/kg) and (c) temperature (K), and at 2m (d) for temperature (K).

The variations of 300hPa specific humidity shown in Figure 23 follow closely those of temperature for ERA-Interim, and the corresponding relative humidity does not shift substantially over time, in

contrast to the behaviour of ERA-40. Relative humidity does increase by around one percentage point in ERA-Interim during the second half of the 1980s and slightly higher values and short-term variability are seen from around 2000 onwards. The latter may reflect an actual change in short-term variability in SST over the last decade, but could be artefacts of supplementing use of clear-sky HIRS humidity data with use of all-sky data from the microwave sounder AMSU-B on NOAA-16 and -17, and the similar Microwave Humidity Sounder (MHS) on subsequent EUMETSAT and NOAA platforms. The dips in 300hPa relative humidity that occur in conjunction with El Niño events agree with a HIRS-based finding of McCarthy and Toumi (2004). Shi and Bates (2011) show a high degree of correlation between anomalies in SST and upper-tropospheric humidity derived directly from inter-calibrated HIRS data over the central tropical Pacific; their time series show clear qualitative agreement with the tropical means for temperature and specific humidity presented in Figure 23.

The consistency of variations in temperature and specific humidity is reassuring, as different types of observation constrain the analyses for the different variables and levels. Figure 24 shows observation-background differences for several sources of temperature and humidity data for the tropical upper troposphere, and shows also the bias estimates for the humidity-sensitive radiances. Considering temperature first, the background fits to radiosonde data show much larger biases and more variability over time for ERA-40 than ERA-Interim, notwithstanding the generally close agreement between anomalies in the analyses. The variations over time in ERA-Interim fits may be due in part to changes in the distribution of the radiosonde data rather than changes in character of the background forecasts,

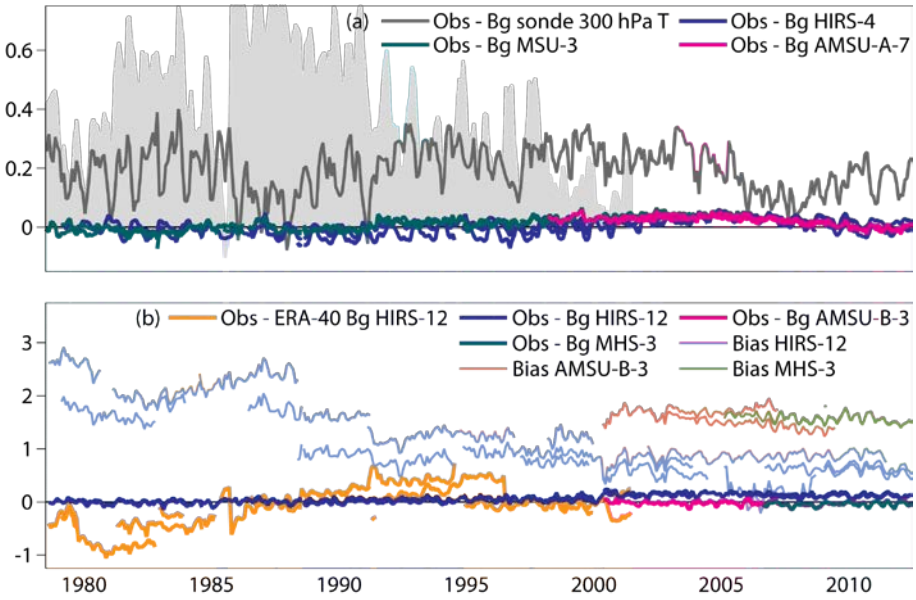

Figure 24: Monthly tropical means of: (a) observation-minus-background brightness temperature differences (K) for channels sensitive to upper-tropospheric temperature (HIRS-4, MSU-3 and AMSU-A-7) for ERA-Interim, and background fits to radiosonde temperatures at the 300hPa standard level and significant levels between 275 and 350hPa, for ERA-Interim (grey line) and ERA-40 (shading); (b) observation-minus-background differences in brightness temperature and bias estimates (K) for channels sensitive to upper-tropospheric humidity (HIRS-12, AMSU-B-3 and MHS-3) for ERA-Interim, and background fit to HIRS-12 for ERA-40. Observations are bias-adjusted.

but are in any case small compared with analysed variations in 300hPa temperature. Brightness temperatures are fitted to about the same accuracy as shown earlier for global data from lower-sounding channels. There is a similar bowing of the curves in the latter part of the period, with largest background deviation in 2005. A small shift in the fit for MSU-3 occurs at the beginning of 1992, which may be associated with the start of assimilation of a significant amount of rain-affected SSMI radiance data.

The background fits for the humidity sounding channels are much more stable over time for ERA-Interim than ERA-40. The close fit to HIRS-12 seen for the first twenty or so years shifts to a cold background bias in brightness temperature of about 0.15K from late 2000, when the microwave data begin to be assimilated and fitted closely. The bias estimates used to adjust the radiances vary little over time for individual instruments, especially later in the period, apart from small seasonal oscillations. The estimates for the microwave sounders are particularly close to each other, especially those for the MHS-3 data from NOAA-18 and Metop-A, which are virtually identical[2]. The levels of bias estimated for HIRS-12 are also quite similar from satellite to satellite later in the period. The reduction in HIRS-12 bias estimates from satellite to satellite during the late 1980s may be related to the small increase in relative humidity that occurs in ERA-Interim at this time. Incorrect specification of spectral response functions may play a role in this regard.

ERA-Interim also assimilated increasing amounts of clear-sky water-vapour radiance data from US and subsequently European and Japanese geostationary satellites, starting in late 2001. Background fits to these brightness temperatures in the tropics lie mostly in the range -0.1 to 0.2K, with bias estimates that range from 0 to 5K from satellite to satellite but are quite uniform over time for individual satellites, apart from two cases where there are abrupt changes in the calibration applied by data suppliers.

This increases confidence in the temporal variations of tropical upper-tropospheric humidity from ERA-Interim. It must be noted, however, that the bias estimates are almost entirely on the positive side, for data from both polar and geostationary orbit. This is indicative of moist bias in ERA-Interim's background forecasts to which the variational bias adjustment scheme responds by reducing brightness temperatures prior to their assimilation. Dessler and Davis (2010) showed the ERA-Interim analyses to be moister than AIRS retrievals, but drier than MERRA analyses.

Figure 25 compares the 300hPa tropical-mean temperature and humidity analyses from ERA-Interim and MERRA. Generally good agreement has already been noted for temperature at this level, although a shift of about 0.2K occurs in the tropics early in the 1990s, after which MERRA is slightly warmer than ERA-Interim. Relative humidity from ERA-Interim is distinctly below that of MERRA until the latter half of the 1980s. The general level for ERA-Interim then rises, as noted above, and the two reanalyses come into a quite close agreement. This lasts until 2002, when differences in specific humidity become smaller and relative humidity falls quite substantially in MERRA. A gradual rise follows. Dips in relative humidity occur in conjunction with several El Niño events in MERRA as well as ERA-Interim.

[2] MHS-3 data from NOAA-19 were assimilated for only two months due to excessive noise in the data.

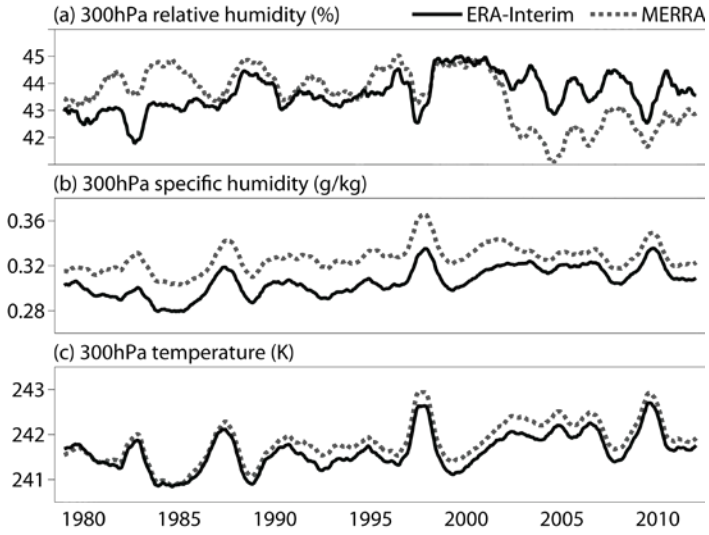

Figure 25Twelve-month running means of tropical averages from ERA-Interim (solid) and MERRA (dotted) at 300hPa for (a) relative humidity (%), (b) specific humidity (g/kg) and (c) temperature (K).

10 Summary and conclusions

Progress in estimating multi-annual variability and trends in atmospheric temperature through use of reanalysis has been documented in this report. Conclusions can be separated into those related to the evolution of the observing system and diagnosis and improvement of reanalysis, and those related to the nature and reliability of the estimated temperature changes.

10.1 Evolution of the observing system and the diagnosis and improvement of reanalysis

ERA-Interim has been shown to improve quite substantially over the earlier ERA-40 in several ways, and is for the most part closer to the newer MERRA and JRA-55 than to ERA-40, though only from the middle troposphere upwards in the case of MERRA. Improvement over time of the underlying observing system has been indicated by plots showing generally increasing observation counts, decreasing inferred observational biases and decreasing differences between ERA-Interim and MERRA. Moreover, several of the differences identified between ERA-Interim and its assimilated data, and some between ERA-Interim and MERRA or JRA-55, have been attributed to known deficiencies in ERA-Interim production for which solutions are to hand for use in its successor. They include:

- use of consistent SST and sea-ice concentration analyses for the full period of the reanalysis;
- better use of all-sky microwave imager radiances;
- incorporation of inter-annual and longer-term changes in radiatively active constituents in the assimilating model;
- better fast radiative transfer modelling of the measured infrared and microwave radiances;
- use of updated radiosonde homogenisation;

- adjustment of aircraft measurements of temperature for bias.

Reanalysis will also continue to benefit from overall improvements to the assimilating model and analysis technique driven largely by the need to improve the important applications of analysis and forecasting systems other than reanalysis. These include, for example, substantial improvements to the land-surface model and analysis of snow cover and soil moisture since the start of ERA-Interim, with consequent improvements in near-surface temperature analysis. Reduction of model biases generally poses a longer-term challenge however, and issues such as whether or not to adjust variationally the biases of the uppermost sounding channels and how to deal with shifts caused by introducing GPSRO data are likely to persist. If time series change abruptly due to observational changes, scope exists for application of homogenisation to the output products of reanalysis. Examples are the adjustments developed by McLandress et al. (2013) to remove the discontinuities in upper-stratospheric temperature in ERA-Interim, and the adjustments applied by Fueglistaler et al. (2013) to account for bias in temperature around the tropical tropopause.

The combined study of fits to observations, of the bias adjustments made to those observations and of the differences among reanalyses and between reanalyses and model simulations is seen as key to gaining understanding and building confidence in the outputs of reanalyses. To this can be added comparison with independent, unassimilated observations, even if available for only a limited period, something that has not been done in the present study. Time series of fits of the analysis and background to pre-homogenised radiosonde data have been shown to be directly informative. In contrast, ERA-Interim draws very closely to radiances that are fully adjusted for perceived biases, and growth of error in the background forecasts is generally slow, although informative on a few matters. Key diagnostic information for the radiance data is provided mainly by the temporal variation of the inferred bias adjustments. If they exhibit similar variations from one type of instrument to another, or from satellite to satellite for a single type of instrument, it is most likely that a problem in the data assimilation is the cause. Conversely, where behaviour differs between instruments of the same type, the problem may be of instrumental origin and a potential may exist for better characterising the instrument in computing model equivalents of what is observed. Examples of each have been presented. Addressing these types of bias at a more fundamental level offers the prospect of reducing dependence on the variational adjustment scheme, enabling parameters to be set for the scheme that reduce the risk of absorbing true signals in the bias adjustments.

Further insight may come from comparing time series of the bias-adjusted radiances, notwithstanding the need for careful interpretation, as has been illustrated for deep stratospheric sounding. In this case, the difficulties in estimating temperature changes solely from analysis of the SSU record is in contrast to what can be provided by reanalysis, which extracts a model-assisted, height-resolved signal from multiple observational data sources, albeit with scope for improvement. Progress is more likely to be achieved by focussing on the homogenising adjustments needed to improve observational input to reanalysis, and on the radiative transfer modelling needed to map model variables to measured radiances, than by seeking to retrieve temperature information directly from radiances measured by a single type of instrument.

It has been shown how reanalyses can vary in performance from region to region and from period to period. Spread among various reanalyses may not be a good measure of uncertainty in the better

(generally the newer) of them, particularly if attention is being paid to a particular region and period. This, plus the different uses to which a product may be put in climate studies, makes it difficult to assign simple general indicators of "climate quality" or product applicability (Thorne and Vose, 2010; Dee et al., 2011b). Ensemble data assimilation, as adopted for long-term reanalysis of surface-only observations (Compo et al., 2011; Poli et al., 2013), provides estimates of uncertainty associated with flow-dependent predictability, but does not account for uncertainty due to model errors and unknown biases in observations. It thus remains necessary for potential users of gridded products from a reanalysis to be provided also with access to the observations presented to the reanalysis, quality control information, the fits to assimilated observations and the bias adjustments, as well as guidance from individual studies such as this one or from broader assessments such as the incipient S-RIP reanalysis intercomparison project (Fujiwara and Jackson, 2013), in order that they can make the most informed choices as to which products to use and how to apply them in their specific studies (Dee et al., 2011b).

10.2 Nature and reliability of the estimated temperature changes

The Summary for Policymakers of the Fifth Assessment Report by Working Group I of the Intergovernmental Panel on Climate Change (IPCC, 2013) states there to be "medium confidence in the rate of warming and its vertical structure in the Northern Hemisphere extra-tropical troposphere and low confidence elsewhere." This is identified as a key uncertainty in the corresponding Technical Summary (Stocker et al., 2013). The Report provides quantitative observation-based evidence for temperature trends in the free atmosphere only for radiosonde and MSU datasets, not for reanalyses, and does not discuss the extent of the cooling observed in the upper stratosphere, citing the uncertainties in SSU data products discussed in the present report.

Haimberger et al. (2012) compare time series and linear trends of ERA temperatures with both MSU records and radiosonde equivalents, and with the profile information provided by radiosondes. They show that the ERA-Interim temperatures provide credible alternatives to the radiosonde and MSU datasets, notwithstanding the deficiencies identified in this report. Reanalysis also provides information at middle- and upper-stratospheric levels, linked wind and humidity information and more. Whilst the findings of the present study cannot be reduced to a few unequivocal general statements as to "climate quality", the summaries given below for different levels in the atmosphere give indications of how well key observations are fitted and of the consistency between variables, reanalyses and model simulations. The reanalyses considered here are limited to those examined earlier: ERA-Interim, MERRA and JRA-55, and for context, ERA-40. In brief, the summaries paint a quite confident picture for the upper troposphere and lower stratosphere, and identify consistent estimates of substantial cooling of the upper stratosphere.

10.2.1 Lower and middle troposphere

The reanalyses show similar low-frequency variability in global-mean temperature in the lower to middle troposphere, but their trends differ more than in the lower stratosphere. ERA-Interim suffers later in the period from problematic assimilation of SSMI radiances and a spurious shift in SST. This cools the analysis over sea and induces a slight warming degradation over land regions where

radiosonde data are fitted to very high accuracy and stability until the late 1990s. Global-mean fits to the homogenised radiosonde temperatures remain below about 0.1K, however. The SSMI and SST problems appear also to be responsible for a variation over time in the amount of precipitation over land compared to gauge data, and to contribute to a weaker overall rate of warming than JRA-55. MERRA appears to warm excessively over time due to observational changes that reduce an overall cold bias relative to JRA-55 and ERA-Interim. No simple regional picture emerges: relatively large differences between ERA-Interim and MERRA are found in the tropics and sub-tropics of both hemispheres, whereas the smaller global-mean differences between ERA-Interim and JRA-55 come mainly from the extratropical southern hemisphere at 850 and 700hPa but the tropics at 500hPa.

10.2.2 Tropical upper troposphere

Fluctuations in temperature and net warming in the tropical upper troposphere have been shown to correlate well with those near the surface in ERA-Interim, but to have more than double the amplitude, in line with model simulations. Upper-tropospheric specific humidity varies in harmony with temperature for ERA-Interim and MERRA, much more so than for ERA-40, and relative humidity in ERA-Interim appears to shift by no more than about one percentage point, aside from dips at the times of El Niño events. Similar dips occur in MERRA. Stable radiance bias adjustments and close fits to the assimilated observations indicate that ERA-Interim does not reproduce model behaviour by failing to draw to the observations. This marks a further step towards reconciling what once appeared to be a considerable discrepancy between observation and modelling.

10.2.3 Tropopause

A quite severe problem specific to ERA-Interim is a cold bias of more than 0.5K in temperature at the tropical tropopause prior to assimilation of GPSRO data, although multi-annual near-tropopause temperature variations are similar in all reanalyses. Local degradation occurs near the extratropical tropopause in regions where substantial amounts of warm-biased aircraft data are assimilated, in particular over North America later in the period.

10.2.4 Lower stratosphere

Time series of lower-stratospheric (50 and 70hPa) temperature are in particularly close agreement for all reanalyses, despite differences in the degree of homogenisation of their input radiosonde and MSU radiance data. The lower-stratospheric analysis is constrained by the assimilated radiance data, infrared as well as microwave, and by the model background to cool slightly more slowly than indicated by the pre-homogenised radiosonde data. This suggests that the homogenisation underestimates the net cooling correction that the earlier radiosonde data need, though only by around 0.1K. Radiance bias adjustments for channels sounding this region are generally smaller and more stable over time than for tropospheric-sounding channels. Perturbations to global-mean ERA-Interim temperatures from underestimating the warming following the El Chichón and Pinatubo volcanic eruptions and from assimilating recent GPSRO data are at most 0.2K, less than 10% of the analysed anomaly range and less than 20% of the net change since 1979 at 50hPa.

10.2.5 Middle stratosphere

Agreement among the reanalyses is generally less good in the middle stratosphere. Variations over time are quite similar until the late 1990s at 30hPa, and also at 20hPa for all but ERA-40. Behaviour differs thereafter, associated with differing treatments of the change from SSU to AMSU-A and the availability of increasing amounts of largely unadjusted radiosonde data. Differences are larger at 10hPa, with MERRA showing an overall cooling with time that is intermediate between ERA-Interim and JRA-55. MERRA and JRA-55 behave for the most part more similarly to the ERA-20CM model simulations than ERA-Interim does. Differences between the reanalyses at the 20 and 30hPa levels (and indeed down to 100hPa) are generally smaller in the northern extratropics than in the tropics and southern extratropics, but are particularly small in the extratropics of both hemispheres for ERA-Interim and JRA-55 after 2006, when both assimilate GPSRO data.

10.2.6 Upper stratosphere and stratopause

Cooling rates increase substantially with increasing height in the upper stratosphere. Here there is mostly good consistency between the ERA-20CM model simulations, ERA-Interim and JRA-55, although for ERA-Interim the net cooling since 1979 has to be either pieced together from the segments of the time series that lie either sides of the discontinuities in plots such as Figure 22, or recomputed from homogenised monthly means (McLandress et al., 2013). The net cooling from 1979 to 2009 is around 3K at 5hPa and 4.5K at 1hPa. ERA-Interim also exhibits fluctuations in cooling rate that align reasonably well with the fluctuations linked to the solar cycle in ERA-20CM. It is noteworthy that this pattern of cooling must have come from the assimilated SSU and AMSU-A data in ERA-Interim, but from the modelled radiative forcing due to the CMIP5-specified changes of carbon dioxide and ozone in ERA-20CM, as ERA-Interim did not include long-term variations of radiatively active constituents in its assimilating model. ERA-20CM points to a stratopause cooling of around 7K since 1950. A specific evaluation against the sparse available rocketsonde data, such as analysed for stratopause changes by Lysenko and Rusina (2003), would be worthwhile.

10.3 Final remarks

It has become common practice to continue reanalyses in close to real time once they have completed analysis of their past periods. Appreciated as this is by a large user base, it does not obviate the need for replacement reanalyses from time to time, and for users to adapt to them. ERA-Interim is a particular case in point as it was conceived originally as an intermediate contribution, so had some limitations from the outset, in addition to later restrictions on use of new types of observation and on bias adjustment of an increasing number of radiosonde data. At the time of writing there is opportunity to supplement the present work with a more thorough study of the recently available output from JRA-55. In turn, ECMWF and partners are preparing the reanalysis that will replace ERA-Interim, as part of a wider programme for reanalysis and the provision of climate services (Dee et al., 2014).

Acknowledgments

ERA-Interim is funded directly by ECMWF, supported by secondments of staff funded by the UK National Centre for Atmospheric Science and the Japan Meteorological Agency. Sakari Uppala served as project leader until October 2008. ERA-CLIM is funded by the European Commission under the EU Seventh Research Framework Programme, contract number 265229. Comments from Leo Haimberger, Roger Saunders, Ted Shepherd and two referees are gratefully acknowledged.

References

Albergel, C, Dorigo W, Reichle RH, Balsamo, G, de Rosnay, Muñoz-Sabater, J, Isaksen, L, de Jeu, R, and Wagner, W. 2013. Skill and global trend analysis of soil moisture from reanalyses and microwave remote sensing. J. Hydrometeor., **14**:1259-1277. doi: 10.1175/JHM-D-12-0161.1

Bauer, P, Lopez, P, Benedetti, A, Salmond, D, and Moreau, E. 2006. Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. Q.J.R. Meteor. Soc., **132**: 2277–2306. doi: 10.1256/qj.05.189

Becker, A, Finger, P, Meyer-Christoffer, A, Rudolf, B, Schamm, K, Schneider, U, and Ziese, M. 2013. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. Earth Syst. Sci. Data, **5**: 71-99. doi:10.5194/essd-5-71-2013

Blunden, J, and Arndt, DS, Eds. 2013. State of the Climate in 2012. Bull. Amer. Meteor. Soc., **94** (8): S1–S238. doi: 10.1175/2012BAMSStateoftheClimate

Bormann, N, and Thépaut, J-N. 2007. Assimilation of MIPAS limb radiances in the ECMWF system. I: Experiments with a 1-dimensional observation operator. Q.J.R. Meteor. Soc., **133**: 309–327. doi: 10.1002/qj.46

Brohan, P, Kennedy, JJ, Harris, I, Tett, SFB, and Jones PD. 2006. Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. J. Geophys. Res, **111**, D12106, doi:10.1029/2005JD006548.

Bromwich, DH, Nicolas, JP, Monaghan, AJ, Lazzara, MA, Keller, LM, Weidner, GA, and Wilson, AB. 2013. Central West Antarctica among the most rapidly warming regions on Earth. Nature Geosci, **6**: 139–145. doi:10.1038/ngeo1671

Cao, C, Goldberg, M, and Wang, L. 2009. Spectral Bias Estimation of Historical HIRS Using IASI observations for improved fundamental climate data records. J. Atmos. Ocean. Tech., **26**: 1378-1387. doi: 10.1175/2009JTECHA1235.1

Christy, JR, Spencer, RW, Norris, WB, Braswell, WD, and Parker, DE. 2003. Error estimates of Version 5.0 of MSU–AMSU bulk atmospheric temperatures. J. Atmos. Oceanic Technol., **20**: 613–629. doi: 10.1175/1520-0426(2003)20<613:EEOVOM>2.0.CO;2

Chung, E-S, and Soden, BJ. 2010. Investigating the influence of carbon dioxide and the stratosphere on the long-term tropospheric temperature monitoring from HIRS. J. Appl. Meteor. Climatol., **49**: 1927-1937. doi:10.1175/2010JAMC2486.1

Compo, GP, Whitaker, JS, Sardeshmukh, PD, Matsui, N, Allan, RJ, Yin, X, Gleason, BE, Vose, RS, Rutledge, G, Bessemoulin, P, Brönnimann, S, Brunet, M, Crouthamel, RI, Grant, AN, Groisman, PY, Jones, PD, Kruk, MC, Kruger, AC, Marshall, GJ, Maugeri, M, Mok, HY, Nordli, Ø, Ross, TF, Trigo, RM, Wang, XL, Woodruff, SD and Worley, SJ. 2011. The Twentieth Century Reanalysis Project. Q.J.R. Meteorol. Soc., **137**: 1–28. doi: 10.1002/qj.776

Compo, GP, Sardeshmukh, PD, Whitaker, JS, Brohan, P, Jones, PD, and McColl, C. 2013. Independent confirmation of global land warming without the use of station temperatures. Geophys. Res. Lett., **40**: 3170-3174. doi: 10.1002/grl.50425

Dee, DP, and Uppala, S. 2009. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. Q.J.R. Meteor. Soc., **135**: 1830–1841. doi: 10.1002/qj.493

Dee, DP, Uppala, SM, Simmons, AJ, Berrisford, P, Poli, P, Kobayashi, S, Andrae, U, Balmaseda, MA, Balsamo, G, Bauer, P, Bechtold, P, Beljaars, ACM, van de Berg, L, Bidlot, J, Bormann, N, Delsol, C, Dragani, R, Fuentes, M, Geer, AJ, Haimberger, L, Healy, SB, Hersbach, H, Hólm, EV, Isaksen, L, Kållberg, P, Köhler, M, Matricardi, M, McNally, AP, Monge-Sanz, BM, Morcrette, J-J, Park, B-K, Peubey, C, de Rosnay, P, Tavolato, C, Thépaut, J-N, and Vitart, F. 2011a. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteor. Soc., **137**: 553–597. doi: 10.1002/qj.828

Dee, DP, Källén, E, Simmons, AJ, and Haimberger, L. 2011b. Comments on "Reanalyses Suitable for Characterizing Long-Term Trends". Bull. Amer. Meteor. Soc., **92**: 65-70. doi: 10.1175/2010BAMS3070.1

Dee, DP, Balmaseda, M, Balsamo, G, Engelen, R, Simmons, AJ and Thépaut, J-N. 2014. Toward a consistent reanalysis of the climate system. To appear in Bull. Amer. Meteor. Soc..

Derber, JC, and Wu W-S. 1998. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon. Weather Rev. **126**: 2287–2299. 299. doi: 10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2

Dessler, AE, and Davis SM. 2010. Trends in tropospheric humidity from reanalysis systems. J. Geophys. Res., **115**, D19127. doi:10.1029/2010JD014192

Ebita, A, Kobayashi, S, Ota, Y, Moriya, M, Kumabe, R, Onogi, K, Harada, Y, Yasui, S, Miyaoka, K, Takahashi, K, Kamahori, H, Kobayashi, C, Endo, H, Soma, M, Oikawa, Y, and Ishimizu, T. 2011. The Japanese 55-year Reanalysis "JRA-55": An Interim Report. SOLA, **7**: 149−152. doi:10.2151/sola.2011-038

Eyring, V, Shepherd, TG, and Waugh, DW (eds.). 2010. Chemistry-Climate model validation. SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526. http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/

Fueglistaler, S, Legras, B, Beljaars, A, Morcrette, J-J, Simmons, A, Tompkins, AM, and Uppala, S. 2009. The diabatic heat budget of the upper troposphere and lower/mid stratosphere in ECMWF reanalyses. Quart. J. R. Meteor. Soc., **135**: 21-37. doi: 10.1002/qj.361

Fueglistaler, S, Liu, YS, Flannaghan, TJ, Haynes, PH, Dee, DP, Read, WJ, Remsberg, EE, Thomason, LW, Hurst, DF, Lanzante, JR, and Bernath, PF. 2013. The relation between atmospheric humidity and temperature trends for stratospheric water, J. Geophys. Res. Atmos., **118**: 1052–1074. doi:10.1002/jgrd.50157

Fujiwara, M, and Jackson, D. 2013. The SPARC [Stratosphere-troposphere Processes and their Role in Climate] Reanalysis Intercomparison Project (S-RIP) Planning Meeting, SPARC Newsletter, **41**: 52-55. Available from http://www.sparc-climate.org/

Geer, AJ, Bauer, P, and Lopez, P. 2008. Lessons learnt from the operational 1D + 4D-Var assimilation of rain- and cloud-affected SSM/I observations at ECMWF. Q.J.R. Meteor. Soc., **134**: 1513–1525. doi: 10.1002/qj.304

Haimberger, L. 2007. Homogenization of radiosonde temperature time series using innovation statistics. J. Climate, **20**: 1377-1403. doi: 10.1175/2007JCLI4050.1

Haimberger, L, Tavolato, C, and Sperka, S. 2008. Toward elimination of the warm bias in historic radiosonde temperature records: Some new results from a comprehensive intercomparison of upper-air data. J. Climate, **21**: 4587–4606. doi: 10.1175/2008JCLI1929.1

Haimberger, L, and Andrae, U. 2011. Radiosonde temperature bias correction in ERA-Interim. ERA Report Series, 8. ECMWF: Reading, UK. Available from www.ecmwf.int

Haimberger, L, Tavolato, C, and Sperka, S. 2012. Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations. J. Climate, **25**: 8108–8131. doi: 10.1175/JCLI-D-11-00668.1

Hersbach, H, Peubey, C, Simmons, A, Poli, P, Dee, D, and Berrisford, P. 2013. ERA-20CM: a twentieth century atmospheric model ensemble. ERA Report Series, 16. ECMWF: Reading, UK. Available from www.ecmwf.int

Houghton, J, Townshend, J, Dawson, K, Mason, P, Zillman, J, and Simmons, A. 2012. The GCOS at 20 years: the origin, achievement and future development of the Global Climate Observing System. Weather, **67**: 227-235. doi: 10.1002/wea.1964

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, TF, Qin, D, Plattner, G-K, Tignor, M, Allen, SK, Boschung, J,

Nauels, A, Xia, Y, Bex, V, and Midgley, PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Isaksen, L, Vasiljevic, D, Dee, DP, and Healy, S. 2012. Bias correction of aircraft data implemented in November 2011. ECMWF Newsletter , **131**:6. Available from www.ecmwf.int

Jones, PD, and Moberg, A. 2003. Hemispheric and large-scale surface air temperature variations: An extensive revision and update to 2001, J. Climate, **16**: 206 – 223, doi:10.1175/1520-0442(2003)016<0206:HALSSA>2.0.CO;2

Jones, PD, Lister, DH, Osborn, TJ, Harpham, C, Salmon, M, and Morice, CP. 2012. Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010, J. Geophys. Res., **117**, D05127, doi:10.1029/2011JD017139

Kobayashi, S, Matricardi, M, Dee, D, and Uppala, S. 2009. Toward a consistent reanalysis of the upper stratosphere based on radiance measurements from SSU and AMSU-A. Q. J. R. Meteor. Soc., **135**: 2086–2099. doi: 10.1002/qj

Lott, FC, Stott, PA, Mitchell, DM, Christidis, N, Gillett, NP, Haimberger, L, Perlwitz, J, and Thorne, PW. 2013. Models versus radiosondes in the free atmosphere: A new detection and attribution analysis of temperature. J. Geophys. Res. Atmos., **118**: 2609–2619. doi:10.1002/jgrd.50255

Lu, Q, and Bell, W. 2013. Characterising channel center frequencies in AMSU-A and MSU microwave sounding instruments. ECMWF Tech. Memo., 700. Available from www.ecmwf.int

Lysenko, EV. and Rusina, VY. 2003. Long-term changes in the stratopause height and temperature derived from rocket measurements at various latitudes. Int. J. Geomagn. Aeron., **4**: 67-81.

McCarthy, MP, and Toumi, R. 2004. Observed Interannual Variability of Tropical Troposphere Relative Humidity. J. Climate, **17**: 3181–3191.
doi: 10.1175/1520-0442(2004)017<3181:OIVOTT>2.0.CO;2

McLandress, C, Plummer, DA, and Shepherd, TG. 2013. Technical Note: A simple procedure for removing temporal discontinuities in ERA-Interim upper stratospheric temperatures for use in nudged chemistry-climate model simulations. Atmos. Chem. Phys. Discuss., **13**: 25801-25825. doi:10.5194/acpd-13-25801-2013

Mears, CA, and Wentz, FJ. 2009. Construction of the Remote Sensing Systems V3.2 Atmospheric Temperature Records from the MSU and AMSU Microwave Sounders. J. Atmos. Oceanic Technol., **26**: 1040–1056. doi: 10.1175/2008JTECHA1176.1

Miller, DE, Brownscombe, JL, Carruthers, GP, Pick, DR, and Stewart, KH. 1980. Operational temperature sounding of the stratosphere. Phil. Trans. R. Soc. Lond. **A 296**: 65-71.

Mitchell, DM, Thorne, PW, Stott, PA and Gray, LJ. 2013. Revisiting the controversial issue of tropical tropospheric temperature trends. Geophys. Res. Lett., **40**: 2801-2806. doi: 10.1002/grl.50465

Morice, CP, Kennedy, JJ, Rayner, NA, and Jones, PD. 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. J. Geophys. Res., **117**, D08101, doi:10.1029/2011JD017187.

Nash, J, and Saunders, RW. 2013. A review of satellite sounding radiance observations in support of climate trends investigations and reanalyses. Forecasting Research Technical Report No 586, Met Office. Available from www.metoffice.gov.uk

Onogi, K, Tsutsui, J, Koide, H, Sakamoto, M, Kobayashi, S, Hatsushika, H, Matsumoto, T, Yamazaki, N, Kamahori, H, Takahashi, K, Kadokura, S, Wada, K, Kato, K, Oyama, R, Ose, T, Mannoji, N, and Taira R. 2007. The JRA-25 Reanalysis. J. Meteor. Soc. Japan, **85**: 369-432. doi: 10.2151/jmsj.85.369

Poli, P, Healy, SB, and Dee, DP. 2010. Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis. Q.J.R. Meteor. Soc., **136**: 1972–1990. doi: 10.1002/qj.722

Poli, P, Hersbach, H, Tan, D, Dee, DP, Thépaut,J-N, Simmons, A., Peubey, C, Laloyaux, P, Komori, Y, Berrisford, P, Dragani, R, Trémolet, Y, Hólm, E, Bonavita, M, Isaksen, L, and Fisher, M. 2013. The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). ERA Report, 14. Available from www.ecmwf.int

Rienecker, MM, Suarez, MJ, Gelaro, R, Todling, R, Bacmeister, J, Liu, E, Bosilovich, MG, Schubert, SD, Takacs, L, Kim, G-J, Bloom, S, Chen, J, Collins, D, Conaty, A, da Silva, A, Gu, W, Joiner, J, Koster, RD, Lucchesi, R, Molod, A, Owens, T, Pawson, S, Pegion, P, Redder, CR, Reichle, R, Robertson, FR, Ruddick, AG, Sienkiewicz, M, and Woollen, J. 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. J. Climate, **24**: 3624-3648. doi: 10.1175/JCLI-D-11-00015.1

Saha, S, Moorthi, S, Pan, H-L, Wu, X, Wang, J, Nadiga, S, Tripp, P, Kistler, R, Woollen, J, Behringer, D, Liu, H, Stokes, D, Grumbine, R, Gayno, G, Wang, J, Hou, Y-T, Chuang, H-Y, Juang, H-MH, Sela, J, Iredell, M, Treadon, R, Kleist, D, Van Delst, P, Keyser, D, Derber, J Ek, M, Meng, J, Wei, H, Yang, R Lord, S, Van Den Dool, H, Kumar, A Wang, W, Long, C, Chelliah, M, Xue, Y, Huang, B, Schemm, J-K, Ebisuzaki, W, Lin, R, Xie, P, Chen, M, Zhou, S, Higgins, W, Zou, C-Z, Liu, Q, Chen, Y, Han, Y, Cucurull, L, Reynolds, RW, Rutledge, G, and Goldberg, M. 2010. The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., **91**: 1015–1057. doi: 10.1175/2010BAMS3001.1

Santer, BD, Wigley, TML, Simmons, AJ, Kållberg, PW, Kelly, GA, Uppala, SM, Ammann, C, Boyle, JS, Brüggemann, W, Doutriaux, C, Fiorino, M, Mears, C, Meehl, GA, Sausen, R, Taylor, KE, Washington, WM, Wehner, MF, and Wentz, FJ. 2004. Identification of anthropogenic climate

change using a second-generation reanalysis. J. Geophys. Res., **109**, D21104. doi:10.1029/2004JD005075

Saunders, R, Hocking, J, Rundle, D, Rayer, P, Matricardi, M, Geer, A, Lupu, C, Brunel, P, and Vidot, J. 2013. RTTOV v11 Science and Validation Report. Available from http://research.metoffice.gov.uk/research/interproj/nwpsaf/rtm/

Schoeberl, MR, Dessler, AE, and Wang, T. 2012. Simulation of stratospheric water vapor and trends using three reanalyses, Atmos. Chem. Phys., **12**: 6475-6487. doi:10.5194/acp-12-6475-2012

Shi, L, and Bates, JJ. 2011. Three decades of intersatellite-calibrated High-Resolution Infrared Radiation Sounder upper tropospheric water vapour. J. Geophys. Res., **116**, D04108. doi:10.1029/2010JD014847

Shine, KP, Barnett, JJ, and Randel, WJ. 2008. Temperature trends derived from stratospheric sounding unit radiances: The effect of increasing CO2 on the weighting function. Geophys. Res. Lett., **35**, L02710, doi:10.1029/2007GL032218

Simmons, AJ, Jones, PD, da Costa Bechtold, V, Beljaars, ACM, Kållberg, PW, Saarinen, S, Uppala, SM, Viterbo, P, and Wedi, N. 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

Simmons, AJ, Willett, KM, Jones, PD, Thorne, PW, and Dee, DP. 2010. Low-frequency variations in surface atmospheric humidity, temperature and precipitation: Inferences from reanalyses and monthly gridded observational datasets. J. Geophys. Res., **115**, D01110, doi: 10.1029/2009JD012442

Stocker, TF, Qin, D, Plattner, G-K, Alexander, LV, Allen, SK, Bindoff, NL, Bréon, F-M, Church, JA, Cubasch, U, Emori, S, Forster, P, Friedlingstein, P, Gillett, N, Gregory, JM, Hartmann, DL, Jansen, E, Kirtman, B, Knutti, R, Krishna Kumar, K, Lemke, P, Marotzke, J, Masson-Delmotte, V, Meehl, GA, Mokhov, II, Piao, S, Ramaswamy, V,Randall, D, Rhein, M, Rojas, M, Sabine, C, Shindell, D, Talley, LD, Vaughan, DG, and Xie, S-P. 2013: Technical Summary. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, TF, Qin, D, Plattner, G-K, Tignor, M, Allen, SK, Boschung, J, Nauels, A, Xia, Y, Bex, V, and Midgley, PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Taylor, KE, Stouffer, RJ, and Meehl, GA. 2012. An overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., **93**: 485–498. doi: 10.1175/BAMS-D-11-00094.1

Thompson, DWJ, Seidel, DJ, Randel, WJ, Zou, C-Z, Butler, AH, Mears, C, Osso, A, Long, C, and Lin, R. 2010. The mystery of recent stratospheric temperature trends. Nature, **491**: 692-697. doi: 10.1038/nature11579

Thorne, PW, and Vose, RS. 2010. Reanalyses suitable for characterizing long-term trends: Are they really achievable? Bull. Amer. Meteor. Soc., **91**: 353–361. doi: 10.1175/2009BAMS2858.1

Thorne, PW, Lanzante, JR, Peterson, TC, Seidel, DJ, and Shine, KP. 2011. Tropospheric temperature trends: History of an ongoing controversy. WIREs Climate Change, **2**: 66–88. doi:10.1002/wcc.80

Titchner, HA, and Rayner, NA. 2014. The Met Office Hadley Centre Sea Ice and Sea-Surface Temperature data set, version 2, part 1: sea-ice concentration. J. Geophys. Res., in press.

Trenberth, KE, Anthes, RA, Belward, A, Brown, O, Haberman, E, Karl, TR, Running, S, Ryan, B, Tanner, M, and Wielicki, B. 2012. Challenges of a sustained climate observing system. In Climate Science for Serving Society: Research, Modelling and Prediction Priorities, Hurrell, JW and Asrar, G, eds, Springer, 485p. ISBN 978-94-007-6691-4

Uppala, SM, Kållberg, PW, Simmons, AJ, Andrae, U, da Costa Bechtold, V, Fiorino, M, Gibson, JK, Haseler, J, Hernandez, A, Kelly, GA, Li, X, Onogi, K, Saarinen, S, Sokka, N, Allan, RP, Andersson, E, Arpe, K, Balmaseda, MA, Beljaars, ACM, van de Berg, L, Bidlot, J, Bormann, N, Caires, S, Chevallier, F, Dethof, A, Dragosavac, M, Fisher, M, Fuentes, M, Hagemann, S, Hólm, E, Hoskins, BJ, Isaksen, L, Janssen, PAEM, Jenne, R, McNally, AP, Mahfouf, J-F, Morcrette, J-J, Rayner, NA, Saunders, RW, Simon, P, Sterl, A, Trenberth, KE, Untch, A, Vasiljevic, D, Viterbo, P and Woollen, J. 2005. The ERA-40 re-analysis. Q.J.R. Meteor. Soc., **131**: 2961–3012. doi: 10.1256/qj.04.176

Vose, RS, Applequist, S, Menne, MJ, Williams Jr, CN, and P. Thorne P. 2012. An intercomparison of temperature trends in the U.S. Historical Climatology Network and recent atmospheric reanalyses. Geophys. Res. Lett., **39**, L10703. doi:10.1029/2012GL051387

Wang, L, Zou, C-Z, and Qian, H. 2012. Construction of stratospheric temperature data records from Stratospheric Sounding Units. J. Climate, **25**: 2931-2946. doi: 10.1175/JCLI-D-11-00350.1

Zou, C, Goldberg, MD, Cheng, Z, Grody, NC, Sullivan, JT, Cao, C, and Tarpley, D. 2006. Recalibration of microwave sounding unit for climate studies using simultaneous nadir overpasses. J. Geophys. Res., **111**, D19114, doi:10.1029/2005JD006798.