Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

881

Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets

Adrian Simmons, Hans Hersbach, Joaquín Muñoz-Sabater, Julien Nicolas, Freja Vamborg, Paul Berrisford (Copernicus Department) Patricia de Rosnay (Research Department), Kate Willett (Met Office) and Jack Woollen (National Centers for Environmental Prediction)

February 2021

Series: ECMWF Technical Memoranda

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

Contact: library@ecmwf.int

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

Abstract

The ERA5 reanalysis of atmospheric and surface observations provides new estimates of low frequency variability and trends from 1950 onwards. Near-surface atmospheric values of temperature and humidity are considered here. They are compared with corresponding values from two earlier reanalyses, ERA-Interim and JRA-55, and with other observationally based estimates. Topics covered include global trends and the fits of ERA5 and ERA-Interim to land-station data, sea-surface and marine-air temperatures, estimates of actual global temperatures, regional variations of actual temperatures over land and sea-ice, trends and variability of monthly temperature anomalies from multiple datasets, and some local issues for ERA5. The discussion of relative humidity is shorter, but covers many of the same lines.

The trend and low frequency variability of global-mean temperature from ERA5 are largely consistent with values provided by other datasets from 1950 onwards. ERA5 is biased cold over the majority of the land surface prior to 1967 because the cold bias of its background forecasts is less well constrained by analysed observations then than in later years. The effect is smaller, however, than the uncertainty in all datasets that arises from differences in sea-surface temperature (SST) analysis, and no larger than is estimated to arise in some datasets from using SST rather than air temperature over sea. ERA5's use of marine air temperature rather than SST is also one reason its global temperatures are a little higher than those from other datasets for the latest few years.

ERA5 performs relatively well for Europe from 1950 onwards, but uncertainty is somewhat larger in the mid-1960s, when there are significant gaps in observational data coverage. Its worst persistent temperature bias is over Australia prior to the 1970s. The issue in this case is not only lack of surface observations but also an unusually large warm bias of the background forecasts. Aside from this, the ERA5 surface analysis scheme does not cope well with the preponderance of observations from Australia that are for non-standard times. In addition, agreement over Australia between several reanalyses and monthly climatological datasets tends to be poorer during occasional wet spells. This stems in part at least from different definitions of daily average temperature. A number of issues elsewhere, of a more-local nature, have been identified. They relate to data gaps, questionable representation of fractional sea-ice cover, inconsistent coastal SSTs and erroneous temperatures of the Great Lakes.

ERA5 generally agrees best with GISTEMP and HadCRUT5 when its temperature anomalies are compared with those from monthly climatological datasets. Agreement is particularly close over the land masses of the extratropical northern hemisphere. Differences in the tropics and southern extratropics are more pronounced earlier in the period, when ERA5 suffers from an absence of synoptic data that is more acute than the absence of monthly averages of climatological observations. ERA5 and other reanalyses provide a more physically sound calculation of temperature over ice-covered sea than the extrapolation of land values used in the production of several of the monthly climatological datasets.

Reanalyses and the HadISDH datasets give similar depictions of interannual variability and longer term changes in moisture from the early 1970s onwards, including a net increase in specific humidity but decrease in relative humidity over land. Interannual variability is similar over sea and land, with changes over sea preceding those over land by a month or so. ERA5 is nevertheless moister over sea than HadISDH for recent years. It also appears from data fits and comparisons with later years to be biased dry as well as cold over land for the 1950s and 1960s. Long-term average values are drier over south-east Asia for ERA5 than for JRA-55 and HadISDH, but both reanalyses are drier than HadISDH over the Arabian Peninsula.

Much of the trend and low frequency variability in the reanalyses studied here is captured in the background fields of the data assimilation. Aside from helping to improve analyses of land-surface conditions, the surface air analyses of temperature and humidity act largely to reduce biases in the products provided for these variables. Issues arise where bias in the background values is large and observational coverage varies in time. A number of potential improvements have been identified.

Contents

1 Introduction

Several studies carried out at the European Centre for Medium-Range Weather Forecasts (ECMWF) have compared low-frequency variations and trends in surface air temperature from the ERA-40 (Uppala *et al.*, 2005) and ERA-Interim (Dee *et al.*, 2011) reanalyses with values from other reanalyses and from direct analyses of monthly climatological data. Simmons *et al*. (2004) showed that variations in temperature over land from ERA-40 compared favourably with variations from the CRUTEM2v dataset of Jones and Moberg (2003), especially from the late 1970s onwards. Deficiencies in the coverage of the observational data available to ERA-40 prior to the 1970s and a few deficiencies in CRUTEM2v due to erroneous observational data were also identified. Subsequent studies include an evaluation of temperatures over sea as well as land from the first two decades of production of ERA-Interim (Simmons *et al.*, 2010), an assessment that focussed on Arctic warming (Simmons and Poli, 2015) and a more extensive comparison of ERA-Interim with other datasets (Simmons *et al.*, 2017). These and other studies have shown how the various datasets have tended to come into closer agreement as new versions and products have become available. They have nevertheless identified spells of anomalous temperatures from individual datasets, such as the relatively high global temperatures from ERA-Interim in 2005 and 2006 illustrated in this report. They have also identified occasional shifts to persistently different values in the time series from reanalyses. Examples are a shift to lower marine air temperatures (MATs) from 2002 onwards due to different sources of sea-surface temperature (SST) analysis in ERA-Interim (also illustrated in this report) and a shift in 2008 to lower temperatures over land and sea-ice in MERRA-2 (Gelaro *et al*., 2017; Simmons *et al*., 2017).

The principal finding of Simmons *et al.* (2010) related not to the temperature but rather to the relative humidity of surface air. Comparison was made between ERA-Interim and a forerunner, HadCRUH, of the current HadISDH dataset (Willett *et al*., 2014), which is based on direct analysis of surface synoptic measurements of water vapour and temperature. Both datasets showed a sharp decline in relative humidity beginning in the late 1990s. This decline has since continued, though more slowly than from 2000 to 2008 (Willett *et al*., 2020a).

Based on these and other studies, the Copernicus Climate Change Service (C3S) adopted surface air temperature and relative humidity from ERA-Interim as two of the variables for which it provides monthly updates and brief commentaries (https://climate.copernicus.eu/climate-bulletins) and annual reports (https://climate.copernicus.eu/ESOTC). ERA-Interim data have also been used in the annual statements issued by the World Meteorological Organization on the State of the Global Climate, and in the annual State of the Climate reports published in the Bulletin of the American Meteorological Society (e.g. Blunden and Arndt, 2020).

ERA5 (Hersbach *et al.*, 2020) is ECMWF's latest comprehensive reanalysis, and the replacement for ERA-Interim. Produced as a contribution to C3S, it covers the period from 1950 to the present and is being updated in close to real time. Release of data from 1979 onwards (the period covered by ERA-Interim) for public use was completed early in 2019, and a first version of data for 1950-1978 was released in November 2020. This version is used here. ERA5 data have now superseded ERA-Interim data in the latest editions of the climate monitoring reports identified above.

The primary purpose of the present report is to document the quality of the surface air temperature and humidity analyses from ERA5. The evaluation of ERA5 includes comparison with the ERA-Interim and JRA-55 reanalyses and with the latest available versions of other established datasets. This serves also to identify several issues with datasets other than ERA5. Evaluation of ERA5 is also made in terms of how well observations are fitted by its background forecasts and analyses.

The various datasets are introduced in the following section. This is followed by an extensive consideration of temperature. It covers global trends and the fits of ERA5 and ERA-Interim to land-station data, sea-surface and marine-air temperatures, estimates of actual global temperatures, regional variations of actual temperatures over land and sea-ice, trends and variability of monthly temperature anomalies from multiple datasets, and some local issues for ERA5. A shorter discussion of relative humidity, though covering many of the same lines, is given in section 4. This is followed by some concluding remarks. The Table of Contents provides a more detailed listing of the topics covered.

2 ERA5 and other datasets

ERA5 is produced using a version of ECMWF's Integrated Forecast System (IFS), CY41r2, that was used for operational forecasting during part of 2016. ERA-Interim was produced for the period from January 1979 to August 2019 using an IFS version from 2006, CY31r2. ERA5's high-resolution (HRES) assimilating model has a horizontal grid resolution of about 31 km, compared with 78 km for ERA-Interim, and a 137-level vertical resolution rather than the 60 levels of ERA-Interim. The HRES ERA5 data assimilation uses background-error estimates that utilise the output from a ten-member ensemble data assimilation using a lower horizontal resolution of about 63 km. ERA5 also uses new externally-produced analyses of sea-surface temperature and sea-ice concentration, variations in radiative forcing derived from CMIP-5 specifications, and various new and reprocessed observational data records. It provides hourly output fields. A comprehensive account of ERA5 is provided by Hersbach *et al.* (2020) and further details can be found in supporting online documentation. In this report, unless stated otherwise, references to ERA5 relate to the analyses and other products from the HRES data assimilation system, using the ERA5.1 version (Simmons *et al*., 2020) for the period from 2000 to 2006. ERA5.1 provides better products for the upper troposphere and stratosphere for this period, and is similar to the original ERA5 in the lower troposphere.

The ERA5 analyses of surface air temperature and relative humidity are produced similarly to those from ERA-40 and ERA-Interim in that the fields are derived using an optimal interpolation (OI) analysis of synoptic observations that takes its background forecasts from the primary cycled variational assimilation used for upper-air data, the high-resolution four-dimensional variational (4D-Var) assimilation in the case of ERA5. The OI scheme used in ERA5 is basically as used in the earlier ERA reanalyses (Simmons *et al*., 2004; 2010), but an innovation for ERA5 is the application of the scheme hourly rather than six-hourly. Hersbach *et al*. (2020) describe how this was done. Moreover, as the analysis is carried out on the grid of the model used in the primary data assimilation, it is produced at finer resolution in ERA5. ERA5 also differs from ERA-Interim in that the surface air analysis operates only over land. Over sea, the analysis simply takes the background values. This avoids the problematic analysis of observations from ships noted for ERA-40 (Tett, personal communication; Simmons *et al*., 2004). It also avoids unrealistic changes to background values over coastal seas. These can arise via the structure functions used in the OI analysis of observations over those coastal lands for which background errors are unrepresentative of the background errors over nearby seas. This was found to be a particular issue around the coast of Norway during early production of ERA5.

The ERA5 scheme also differs from the ERA-Interim scheme in that it keeps a record during production of how observations were used in the OI scheme, though this is for only a subset of the data, as discussed in the following section. These *feedback* data contain for each stored observation detailed information such as the observed value, the equivalent of the observed value derived from the background forecast, the analysed value, a flag indicating whether the observation was used by the analysis scheme or rejected by quality control, and other relevant information such as the height of the model terrain and the nature of the underlying model surface, all at the location of the observation. Particular attention is paid in this

report to the *background departure*, the difference between (or fit of) the background and the observation, and the *analysis departure*, the difference between (or fit of) the analysis and the observation. Attention is also paid to the *analysis increment*, the difference between the analysis and the background. This is defined on the model grid, not at observation points, and describes the net change (or correction) to the background forecast due to all assimilated observations. In the OI scheme for surface air variables, up to 50 observations located within 1000km are used to calculate the analysis increment for each model grid-point. The horizontal structure function that determines the spreading of information away from the observation location has an e-folding distance of 420km. Background departures are defined as background minus observation, rather than observation minus background, to make the sign of the departure indicate the sign of the model error, if observations are accurate. Where observations are plentiful and accurate, the background departures thus defined are opposite in sign to the analysis increments.

The sources of the synoptic observational data used in ERA5 are of particular relevance for the present report. The observations of two-metre (dry-bulb) temperature and humidity (derived from additional reports of dew-point temperature) come from the following sources:

(i) 1950 to June 1959 – From data held by the US National Centers for Environmental Prediction (NCEP). These data are for dry-bulb temperature only, and data coverage is more limited than in later years. No humidity data was assimilated for this period. This applies to radiosonde data also. The sources are the TD13, TD14 and USSR datasets already processed for later years for ERA-40, and additional datasets for Canadian Meteorological Service and British Antarctic Survey stations. The original source datasets are available from https://rda.ucar.edu/, with identifiers ds467, ds470, ds475 and ds487, and https://data.bas.ac.uk, with identifier GB/NERC/BAS/PDC/00794.

(ii) September 1957 to 1978 – The various sources of historical data used for the earlier ERA-40 analysis, documented by Uppala *et al*. (2005). As illustrated later in this report, among the various gaps in this ERA-40 data collection is an absence of data over substantial regions in 1965 and 1966.

(iii) 1950 to 1978 – Data from version 2.5.1 of the International Comprehensive Ocean–Atmosphere Data Set (ICOADS, Woodruff *et al*., 2011).

(iv) 1979 – FGGE Final Level 2b data and ECMWF's holdings of data received operationally from the WMO Global Telecommunications System (GTS).

(v) 1980 to 1994 – A merge (carried out for ERA-40) of the holdings of data received operationally by ECMWF and NCEP.

(vi) 1995 to present – ECMWF's holding of data received operationally.

It is evident from the above and from what is presented in the following section that there is a need for recovery and rehabilitation of additional synoptic data for use in ERA6 and other future reanalyses.

ERA5 is compared here with the Japanese reanalysis JRA-55 (Kobayashi *et al*., 2015; available from 1958 to the present) as well as ERA-Interim. JRA-55 also provides surface air temperature and humidity analyses (and additionally wind analyses) based on an optimal interpolation of screen-level observational data. The JRA scheme potentially improves on ERA5 by using the so-called FGAT (first guess at the appropriate time) approach to adjust for misfits in timing between observation and analysis, using the differences between background values at observation and analysis time. Its ~55 km horizontal resolution is coarser than the ~31 km resolution of ERA5, however. The current comprehensive reanalyses CFSR/CFSv2 (Saha *et al*., 2010, 2014) and MERRA-2 (Gelaro *et al*., 2017) are not used

because they do not provide a separate analysis of surface air temperature and humidity, and there are known inhomogeneities in the time series of the products they provide for these variables.

Comparisons are also made with several datasets based on direct gridding of observational data. For temperature, these datasets provide anomalies in monthly climatological data relative to a climatological reference period. Four of the datasets used are updated versions of those for which comparisons with ERA-Interim were made by Simmons *et al*. (2017). The versions mainly used are NASA's GISTEMP v4 (Hansen *et al*., 2010; Lenssen *et al*., 2019), NOAAGlobalTemp v5 (Smith *et al*., 2008; Zhang *et al*., 2019) and HadCRUT5 (Morice *et al*., 2020). HadCRUT5 comes in two variants; the "HadCRUT5 analysis" that is spatially extended to provide more-complete coverage is used here. A further dataset, Berkeley Earth (Rohde and Hausfather, 2020), is included in the present comparison. Past values may change slightly with each new monthly release of some of these datasets; the results shown here were produced using the versions available in mid-January 2021, all of which provided data to the end of 2020.

Comparison has also been made with version 6.0.0 of HadCRUT4 (Morice et al., 2012) and version 2 of the krigged spatial extension of HadCRUT4 developed by Cowtan and Way (2014), referred to as Had4krig. Although these datasets have recently been superseded by HadCRUT5, a few results are included here. This is to illustrate how the change to HadCRUT5 brings for the most part improved agreement with ERA5 and other datasets, and to raise a question concerning the SST analyses used by ERA5 and JRA-55.

It is important to note that there are degrees of dependence between the various datasets. GISTEMP and NOAAGlobalTemp use the same set of quality-controlled monthly climatological data from land-station observations and the same set of SST analyses (ERSSTv5; Huang *et al*., 2017). Berkeley Earth and HadCRUT4 are likewise both based on the same SST analysis (HadSST3; Kennedy *et al*., 2011). Had4krig differs from HadCRUT4 only in regions where HadCRUT4 provides no data. The ERA5 data assimilation system is an evolution of that used for ERA-Interim, and although the JRA-55 system is more different, not least in SST analysis, it has features in common with the ERA systems. The SST analyses used by ERA5 are not the same as used in either HadCRUT4 or HadCRUT5, but all three sets of analyses are produced by the Met Office. As shown in section 3.4.1 the temperature increase from the 1950s and 1960s to the 1980s from the HadISST2 dataset used by ERA5 is much closer to that from HadSST3 than to that from the HadSST4 dataset (Kennedy *et al.*, 2019) used in HadCRUT5.

The HadCRUT4, HadCRUT5, HadSST3 and HadSST4 datasets each comprise ensembles of either 100 or 200 members. Except where stated otherwise, reference to these datasets refers to the ensemble means that are provided by the Met Office Hadley Centre for HadCRUT5 and the medians that are provided for the other three datasets.

The various datasets require some processing, particularly to relate data to a common climatological reference period, to handle differences in spatial resolution and to construct regional and global averages. This processing is carried out as described by Simmons *et al*. (2017) and references therein. It includes an adjustment over ice-free sea for ERA-Interim to account for the principal effect of temporal inhomogeneity introduced by changes in source of SST analyses, as discussed further in section 3.2. It also includes use of background rather than analysed temperatures over sea for ERA-Interim and JRA-55, which is consistent with ERA5's explicit suppression of the OI surface analysis over sea. An exception for ERA-Interim is that the use of the background rather than the analysis is applied only over ice-free sea. The use of Arctic ice-station and ice-buoy data in ERA-Interim has been discussed by Simmons and Poli (2015). The resulting analysis does not inherit all of the warm bias of background temperatures in cold conditions. The suppression of the surface analysis over sea in ERA5 was applied over all sea, ice-covered as well as ice-free.

The reanalysis datasets used are for air temperature at two-metre height everywhere. The other datasets merge anomalies in SST analyses over ice-free sea with anomalies in surface air temperature over land. Their treatments over sea-ice differ. Here these datasets are regarded as representing anomalies in two-metre air temperature everywhere, consistent with the view expressed by Hansen *et al*. (2010) for GISTEMP. As discussed by Simmons *et al*. (2017) and returned to briefly in section 3.4, the indication from the reanalyses is that the error introduced by this is small for the type of time series presented here, albeit systematic for global trends, as shown first by Cowtan *et al*. (2015) for climate models. It appears, however, to be rather larger in the 1950s and 1960s than subsequently.

For humidity, comparison is made with the HadISDH dataset based on synoptic observations over land (Willett *et al*., 2014) and sea (Willett *et al*., 2020b), using for the most part the latest available versions (4.2.0.2019f for land, 1.0.0.2019f for sea), although version 4.1.0.2018f of the land dataset was used to fill in some large unexplained data gaps that occur for April 2015 in version 4.2.0.2019f.

Many results presented here cover the whole of the period from 1950 onwards, but the period from 1979 to 2018 receives additional attention as it is the period for which detailed comparison can be made between ERA5 and ERA-Interim using complete annual sampling.

3 Surface air temperature

3.1 Global trends and the fits to station data

3.1.1 Global averages

Figure 1(a) presents twelve-month running mean times series of global average two-metre temperature anomalies relative to 1981-2010 from ERA5, ERA-Interim, and the five other datasets mainly considered in this report. Each dataset shows considerable warming over the period shown, with a peak early in 2016 associated with a strong El Niño event, generally high values over the past six years, and values for 2020 that peak close to the largest values reached in 2016. ERA5, ERA-Interim and GISTEMP are the three datasets with the largest anomalies since 2016, but not the datasets with the largest trends over the period since 1979 covered by ERA-Interim. Trends from ordinary least squares fits to monthly values for the forty years 1979 to 2018 are 0.182ºC/decade for ERA5, 0.184ºC/decade for ERA-Interim, 0.178ºC/decade for JRA-55 and otherwise range from 0.170ºC/decade for NOAAGlobalTemp to 0.188ºC/decade for HadCRUT5. The trend from HadCRUT4 is 0.176ºC/decade. ERA-Interim was known to be relatively warm in 2005 and 2006 compared with other datasets (Simmons *et al.*, 2017); ERA5 is closer to the other datasets for this period. Conversely, ERA5 has a positive anomaly in 1990/91 that is larger than that of all other datasets.

The spread among datasets would be expected to be larger outside the common 1981-2010 reference period used for computing anomalies, but signals emerge nevertheless. Going back in time, the spread tends to increase prior to around 1970, arising especially from differences in SST analyses and differences over the Antarctic, as is shown later. ERA5 is biased cold over land in the early years due to limited observational coverage and a cold bias of its background model, but this is countered by the contribution of SSTs that are warmer than used in several other datasets. It will also be seen later that part of the reason ERA5 appears on the cold side for the early years and on the warm side for recent years is because over sea the contribution to its global-average temperature is based on air temperature (as over land), whereas the datasets based directly on monthly climatological data use anomalies in SST.

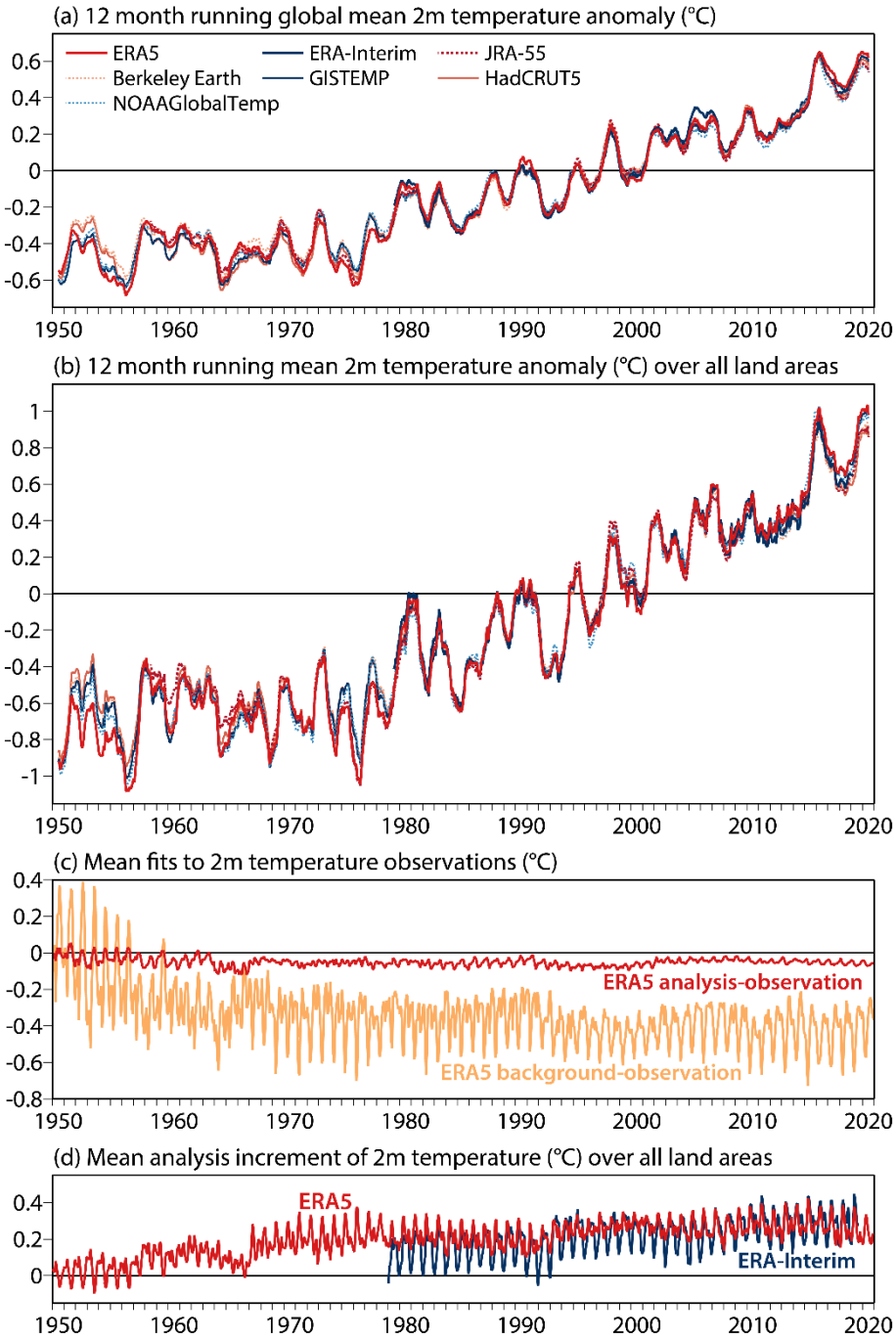

Figure 1 Time series of monthly values from 1950 to 2020. (a) Twelve-month running averages of the global average two-metre temperature (°C) anomaly with respect to 1981-2010 from ERA5, ERA-Interim, JRA-55, Berkeley Earth, GISTEMP, HadCRUT5 and NOAAGlobalTemp. (b) As (a), but for averages over land only. (c) ERA5 analysis - observation (red) and background - observation (orange) departures averaged over all available surface air temperature data from selected land stations. (d) The average analysis increment (analysis – background) in two-metre temperature over land for ERA5 (red) and ERA-Interim (dark blue).

Figure 1(b) shows corresponding time series for averages over all land. Its larger temperature range indicates that warming over the past forty years has been larger over land than sea. The linear trend of about 0.3°C/decade in surface air temperature over land for the period 1979-2018 is more than 50%

higher than the trend in global-mean temperature. The discrepancies between ERA5 and ERA-Interim in global averages for 1990/91 and 2005/06 are not seen in the averages over land; the implied differences over sea are discussed further in the following two sections. Also evident is a separation between ERA5 and ERA-Interim later in the period, with ERA5 having the slightly larger anomalies. The absolute values presented in section 3.3 show that the slightly larger trend of ERA5 compared with ERA-Interim is such as to bring the two reanalyses into closer agreement: ERA5 is colder in the global mean than ERA-Interim, more so in the earlier part of their common period. ERA5 has the coldest temperatures relative to 1981-2010 for most of the period from 1950 to 1957.

More comprehensive comparisons, including plots of temperature differences between ERA5 and the other datasets, and more extensive discussion of these and additional datasets, are presented in section 3.4.

Evidence relating to the reliability of the temperature trend over land in ERA5 is shown in Figure 1(c). This panel presents mean differences between background forecasts and observations (the background departures, or the fit of the background to observations) and between analyses and observations (the analysis departures), calculated for each month of the period. The average is taken over most of the subset of observations for which this type of feedback data was generated during production of the ERA5 surface air analyses; no such data was generated for ERA-Interim. The only archived feedback data from these ERA5 analyses that are omitted in this report, with one exception, are from coastal and small-island stations for which the model's fractional land-sea mask evaluated at the station location is less than 0.5. This is needed for temporal consistency because the archived feedback data from some of these stations are known to be unreliable because of the way the surface analysis was suppressed over sea during an initial phase of the production of ERA5. Its impact is generally small, but detectable in time series for Antarctica, for instance. The exception relates to two island stations for which the time series of feedback data shown later are known to be reliable.

No allowance has been made in Figure 1(c) for variations in the density of observations in space and time, but the variation over time in these plots is generally small compared with the changes in analysed values, although the statistics on background fits are rather different for 1950-1957 than later. The variations over time are particularly small in the case of the mean analysis departures, which lie in the range from 0.05 to -0.12°C with average -0.05°C. The OI surface air analysis evidently corrects for much but not all of the overall cold bias (averaging -0.34°C) in model background temperatures at observation locations.

The final panel of Figure 1 shows the average over all land of the monthly analysis increment, the monthly average difference between the analysis and background fields. These are true area averages, and are available and shown here for ERA-Interim as well as ERA5. The average increment for ERA5 can be seen to follow quite closely the increment averaged over all observation points inferred from the difference between the two sets of values plotted in Figure 1(c). This is reassuring as the data on observation fits that form the basis of Figure 1(c) have very limited local sampling of the diurnal cycle (illustrated later), in contrast to the hourly resolution of the increments that are averaged to produce Figure 1(d).

The ERA5 increment is smaller prior to 1967 than afterwards, and particularly small from 1950 to 1957. This reflects the relatively low numbers of observations supplied to ERA5 for these years, illustrated and discussed below. The ERA5 increment tends to increase a little over time thereafter, though not by as much as the ERA-Interim increment increases over the common period from 1979 onwards. The ERA-Interim increase shows that various factors such as the climate of the background model and the forcing from the SST analysis and assimilated upper-air data tend in combination to underestimate the trend implicit in the surface air temperature observations. Simmons *et al*. (2014) noted this, and gave

reasons to suppose that the mismatch in trends would be reduced in the reanalysis that succeeded ERA-Interim. The ordinary least squares trends of global-average temperatures from 1979 to 2018 are 0.169°C/decade for the ERA-Interim background and 0.175°C/decade for the ERA5 background. The ERA5 background trend is thus about 4% smaller than the trend given by ERA5's analysis of surface air temperature observations.

Feedback information relating to all surface air temperature observations is available from the ERA-Interim and the ERA5 4D-Var data assimilations, even though these data are not assimilated directly by the 4D-Var scheme. The data have been passed passively through the system, and statistics of how well the background forecasts and analyses fit them have been recorded. Here only the background departures are discussed, as the background forecasts are integral to the OI surface air analysis as well as the 4D-Var analysis. Fits of the 4D-Var analysis to the observations are generally similar to the fits of the background, as the other observations assimilated in the 4D-Var do little to correct the background biases in two metre temperature. The 4D-Var background departures provide a complete sampling of the observations made available to the ERA5 OI scheme for which only partial feedback is available, as well as enabling a comparison between ERA5 and ERA-Interim. Feedback is however provided for some observations that are quite clearly erroneous, either in reported measurement or in reported location, as no quality control is provided in 4D-Var for these unassimilated observations. The simple expedient of removing from consideration all observations that differ from the background forecast by more than 22.5°C is applied here. This is three times the limit used for rejection of observations by the OI scheme, but allows into consideration observations from reliable high-latitude stations where departures can be very large in wintertime, particularly for ERA-Interim, as illustrated later. The OI scheme also does not use observations from stations whose elevation differs by more than 300m from the orographic height of the model background; no such representivity check is applied to the 4D-Var feedback data. Only observations in traditional alphanumeric codes are considered for ERA5, as these are the only observations processed in ERA-Interim and the only observations used in the ERA5 OI analysis for the time period considered here, which runs up to the end of October 2019 for ERA5. The ERA5 4D-Var also provides feedback on data received in the BUFR code format that has gradually been introduced in recent years. Although many of these data have been duplicates (albeit with marginally different values and locations) of traditionally coded synoptic (SYNOP) data that were still reported, they include additional METAR data reported by automatic stations at airports. BUFR SYNOP data (with redundancy checks) have been used in the ERA5 OI analyses since 3 November 2019.

The upper two panels of Figure 2 show time series of the means and standard deviations of the 4D-Var background fits to observations, calculated month by month for all departures that pass the simple quality check specified in the preceding paragraph. Panel (c) shows the corresponding numbers of observations, expressed as average daily values for each month. It also shows the number of observations for which feedback data are diagnosed from the ERA5 OI analysis. Mean ERA5 departures are generally larger than shown in Figure 1 for the feedback from the OI scheme, due to the more-lax quality control, the absence of a check on elevation differences and the complete sampling in time. The net cold bias of the ERA-Interim background is smaller than that of ERA5, but the standard deviation of the background departures is smaller for ERA5. Mean absolute values of the departures (not shown) are also generally smaller for ERA5. The ERA5 background is thus closer overall to observations, but more systematically biased cold from one station or region to another. It should be noted that some of the larger bias of ERA5 may be due to larger differences in elevation between station and model, as illustrated in section 3.1.6.

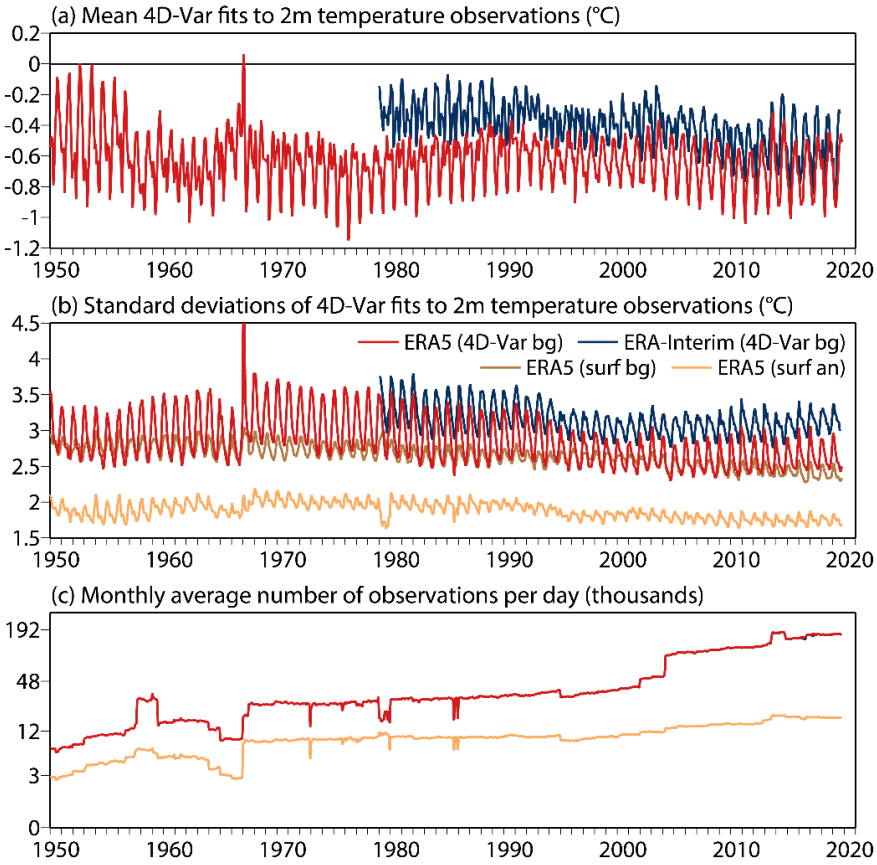

Figure 2 Time series of monthly values from 1950 to October 2019 of (a) mean background-observation differences in two-metre temperature (°C) from ERA5 and ERA-Interim for all selected observations processed by 4D-Var, (b) corresponding standard deviations, plus the standard deviations of the available background and analysis departures for the ERA5 surface analysis, and (c) the average daily number of observations, for all selected observations processed by 4D-Var, and for the observations used in the surface analysis for which departure statistics are available.

3.1.2 Monthly and longer-term variations in the number of observations

The number of temperature observations sampled from the 4D-Var feedback is similar for ERA5 and ERA-Interim, increasing over time due to more frequent reporting and a rise in the number of observing stations (GCOS, 2015). Other variations seen in Figure 2(c) relate to issues in data collection and processing, rather than fundamental characteristics of the observing system. These include:

(i) The peak from September 1957 to June 1959 in the numbers processed passively by 4D-Var, which is an artefact of supplying observations from both ERA-40 and NCEP holdings. Many but not all observations in the NCEP holdings had already been collected for ERA-40 for this period. The used observations for which the surface analysis provides values show a much more modest increase in numbers for the period.

(ii) The low numbers of data in 1965 and 1966, which will be seen later to be due to missing observations from a number of countries.

(iii) The lower numbers of data in 1979, which is due to ECMWF's failure in the 1980s to archive observations at 03, 09, 15 and 21UTC from its FGGE data holdings.

(iv) Missing data in December 1979 and in August and December 1985, for reasons that are not yet known.

(v) The drop in data numbers at the end of 1994, when addition of NCEP's data holdings to those of ECMWF ceased. NCEP held more data than ECMWF was receiving operationally from the GTS at the time.

(vi) The jump in numbers in 2004, which is due to ECMWF starting to collect METAR data then rather than earlier.

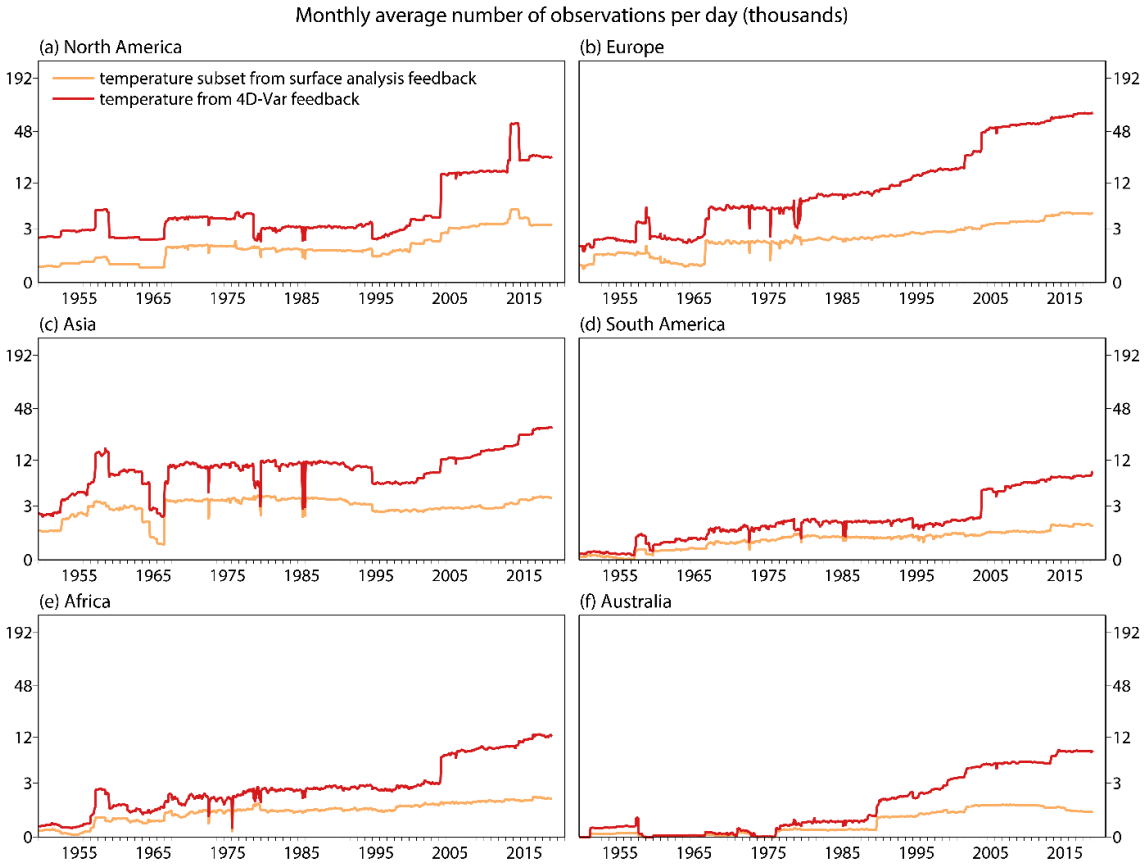

Figure 3 Time series of monthly averages of the daily number of observations, as in Figure 2 but with averages computed over land for (a) North America (170°W-50°W, 15°N-85°N), (b) Europe (20°W-40°E, 35°N-80°N), (c) Asia (60°E-180°E, 0°-85°N), (d) South America (90°W-25°W, 65°S-15°N), (e) Africa (25°W-55°E, 40°S-35°N) and (f) Australia (110°E-160°E, 50°S-10°S).

A breakdown of the observation numbers by region is presented in Figure 3. Again, it should be noted that some of the differences between regions are not fundamental characteristics of the observing system, but a consequence instead of limitations in the international exchange of synoptic data on the GTS that affect receipt of data within Europe, as reflected in ECMWF's holdings of data received on the GTS since 1979. For example, the drop in data numbers at the end of 1994 is seen to relate to additional data held by NCEP from North and South America, and Asia, but not from Europe, Africa and Australia. Otherwise, the low data numbers in 1965 and 1966 relate mainly to observations from Asia, and the low numbers prior to 1957 relate primarily to South America, Africa and Australia; maps showing data coverage for these periods are presented later, in section 3.1.4. The increase in the number of available observations since 1995 can be seen to have occurred for all regions, although the increase from starting to use METAR data is proportionally larger for North America, South America and Africa

than for the other regions. Numbers peak over North America from August 2013 to October 2014 because exceptional amounts of METAR data were available for this period.

ERA-40 had access to very little surface data from Australia prior to July 1976, and this is reflected in low ERA5 numbers from 1958 to 1976. This is one of several problems in ERA5's analysis of Australian temperatures that are discussed in various places in this report, starting in the next section. Antarctic data coverage is discussed in section 3.1.5.

At least some of the issues identified here are likely to be remedied by data recovery and consolidation of datasets prior to future reanalyses. In most cases they do not appear to have significant effects on global statistics of the type presented here, but some results prior to 1967, in particular from 1950 to 1957, and in 1965 and 1966 appear to have been influenced by reduced data coverage.

3.1.3 Sub-daily variations in the number of observations

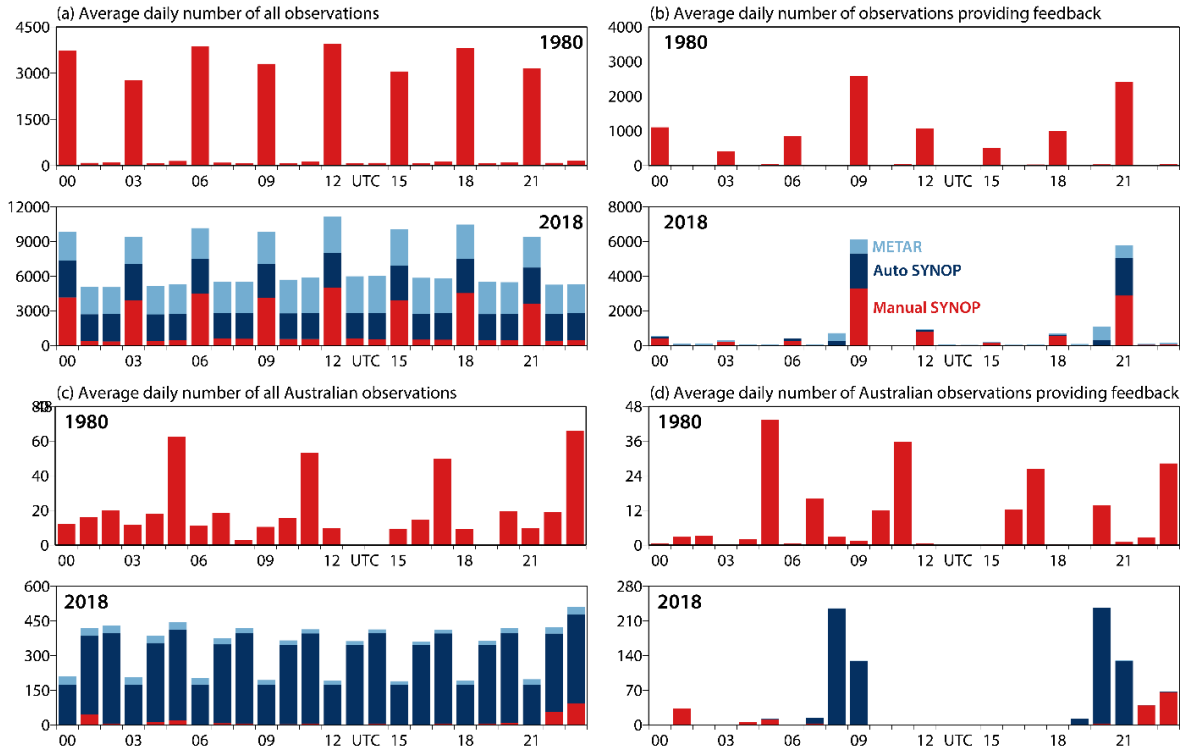

Figure 4 Average number of two-metre temperature observations for each hour of the day for 1980 and 2018. (a) All selected observations processed by ERA5; (b) All observations used in the surface analysis for which feedback data on analysis and background departures are available for selected stations; (c) and (d) as (a) and (b) respectively, but only for those observations in the Australian region (110°E-160°E, 50°S-10°S). Results are shown separately for observations coded as Manual Land SYNOPs (red), Automatic Land SYNOPs (dark blue) and METARs (light blues).

Feedback data from the OI analysis of surface air temperature observations has been archived for far fewer observations than were presented to it, as indicated by the difference between the numbers for 4D-Var and the surface analysis shown in Figure 2(c) and Figure 3. This is because pre-existing software for archiving feedback data was used when hourly surface analysis was introduced for ERA5. This restricted the data stored to one observation per 12-hour 4D-Var cycle for each observing station. This is commonly the feedback for the final hour of the cycle, 09 or 21 UTC. This is illustrated in the upper pairs of panels of Figure 4, which show the average number of observations for each hour of the day,

for the years 1980 and 2018, for all observations processed (from the 4D-Var feedback) and for all observations for which there is feedback from the surface analysis (aside from the coastal and small-island station data discussed earlier). For 1980 the majority of observations are available three-hourly, and although feedback availability peaks at 09 and 21 UTC, the main hours of 00, 06, 12 and 18 UTC are quite prominent also. Many more observations for intermediate hours are available in 2018, and the 09 and 21 UTC peaks in feedback availability from the surface analysis are even more prominent.

The lower pairs of panels of Figure 4 show corresponding results, but restricted to Australian observations. Australia differs from the world at large in that its primary observation times are not the usual three- or six-hourly times starting each day at 00 UTC. Instead, the traditional observation times are 05, 11, 17 and 23 UTC, as shown clearly in the complete values for 1980. The observation times switch to a predominantly three-hourly frequency by 2018, but with a greater number of observations at all hours except those that are the regular three-hourly synoptic hours for most countries. The surface analysis feedback peaks at 05 and 11 UTC in 1980, but at 08 and 20 UTC in 2018.

This feature of Australian observations has implications not only for the temporal consistency of the feedback data, but more importantly also for the analysis itself. Concern has already been expressed regarding analysed temperature trends over Australia from both ERA-40 and ERA-Interim (Simmons *et al*., 2004; 2014). The 00, 06, 12 and 18 UTC analysis times of ERA-40 and ERA-Interim are the times with relatively few Australian observations, and many of the observations used in these analyses were made at times that differed from the analysis time by one or two hours, with no adjustment for the time difference and a changing mix of observation times over the periods of the reanalyses. This is still an issue for ERA5, though lessened as its monthly means involve averaging over analyses for each hour. Some of these analyses use observations at the appropriate time; others do not. Use of the FGAT approach already employed for JRA-55 could be used to address the issue in future versions of ERA.

Separate contributions to the observation numbers presented in Figure 4 are shown for data reported to be from manual and automatic observations and transmitted in SYNOP code, and for METAR data. All data for 1980 are reported to be from manual observation. Globally, the number of manual observations is much the same in 2018 as 1980. The growth in observation numbers between these years comes approximately equally from automatic SYNOPs and METARs, and approximately equally for each hour of the day. For Australia, most of the observations in 2018 are reported to be automatic, but there are nevertheless fewer of them for the standard synoptic hours (00UTC, 03UTC, 06UTC, …) than for other hours. It should again be noted that this refers to observations received and processed by ECMWF from the GTS, so may indicate an issue in data transmission rather than an issue in making the actual observations.

3.1.4 Regional variations in increments and data fits

Figure 5 presents global maps showing averages for 1980-1984 and 2014-2018 of the analysis increments over land and sea-ice, for ERA-Interim and ERA5. Corresponding net values have been shown month by month in Figure 1. Local values tend to be larger for ERA-Interim than for ERA5 for both periods; net increments over land are smaller for ERA-Interim in the earlier period because of a larger cancellation of contributions from regions of positive and negative increment. The geographical patterns of the increments are similar for the earlier and later periods, and of larger magnitude in places for the more recent period due to increased observational coverage, notably over Antarctica.

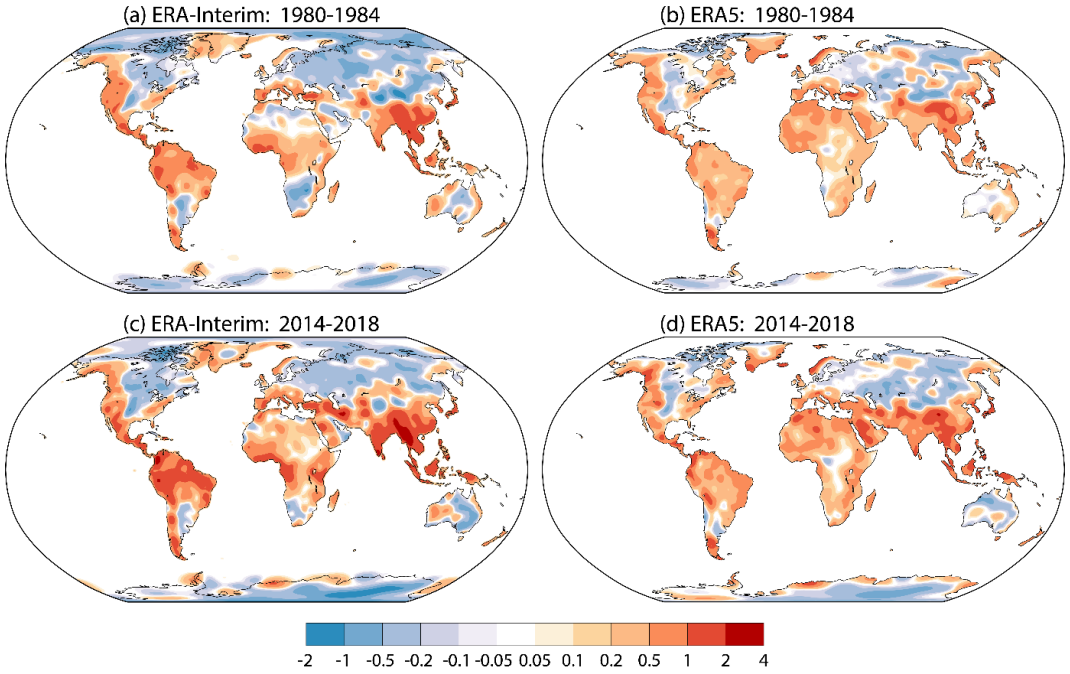

Figure 5 Five-year mean analysis increments in two-metre temperature (ºC) over land and sea-ice from the surface air analyses for: (a) ERA-Interim 1980-1984; (b) ERA5 1980-1984; (c) ERA-Interim 2014-2018; (d) ERA5 2014-2018.

The patterns of the increments over land are also similar, though to a lesser extent, between ERA-Interim and ERA5, and indeed between these two reanalyses and ERA-40, for which increments were presented by Simmons *et al*. (2004). The largest temperature increase from background to analysis is over south-east Asia; warming increments occur almost everywhere in the tropics and sub-tropics. Relatively large cooling occurs in recent years over the Antarctic plateau, for which Fréville *et al*. (2014) and Jones and Lister (2015) demonstrate a warm bias of ERA-Interim. One region in which there is a large difference between the two periods is south of the Caspian Sea. This is one of the local issues discussed later.

ERA-Interim's increments over sea-ice are as discussed by Simmons and Poli (2015). They cool a warm-biased background over most of the Arctic Ocean and coastal Arctic seas, but warm over Baffin Bay and the Greenland Sea. They exhibit reasonable continuity with the increments over neighbouring land. Increments from coastal land stations are spread over sea-ice in ERA-Interim, but only to a quite limited extent, as can be seen around Antarctica.

Maps based on related information from observational feedback data are presented in Figure 6. It shows average departures for 1979-2018 plotted for all observations within 2°x2°grid boxes. The ERA5 4D-Var departures show the background forecasts to be largely biased cold, but with little bias over Europe, western Asia and central/eastern USA and Canada. ERA-Interim gives a more mixed picture. Although it has a smaller net bias when averaged over all observations, as shown in Figure 2, ERA-Interim has generally larger absolute biases, less extensive regions with low bias, and a greater occurrence of warm biases. The feedback archived from the ERA5 OI analysis provides a pattern of background departures that is similar to that from the ERA5 4D-Var, indicating little general sensitivity of this diagnostic to the limited sampling in the case of the OI feedback. The OI analysis corrects for much of the bias of the background forecasts: analysis departures are close to zero in many locations.

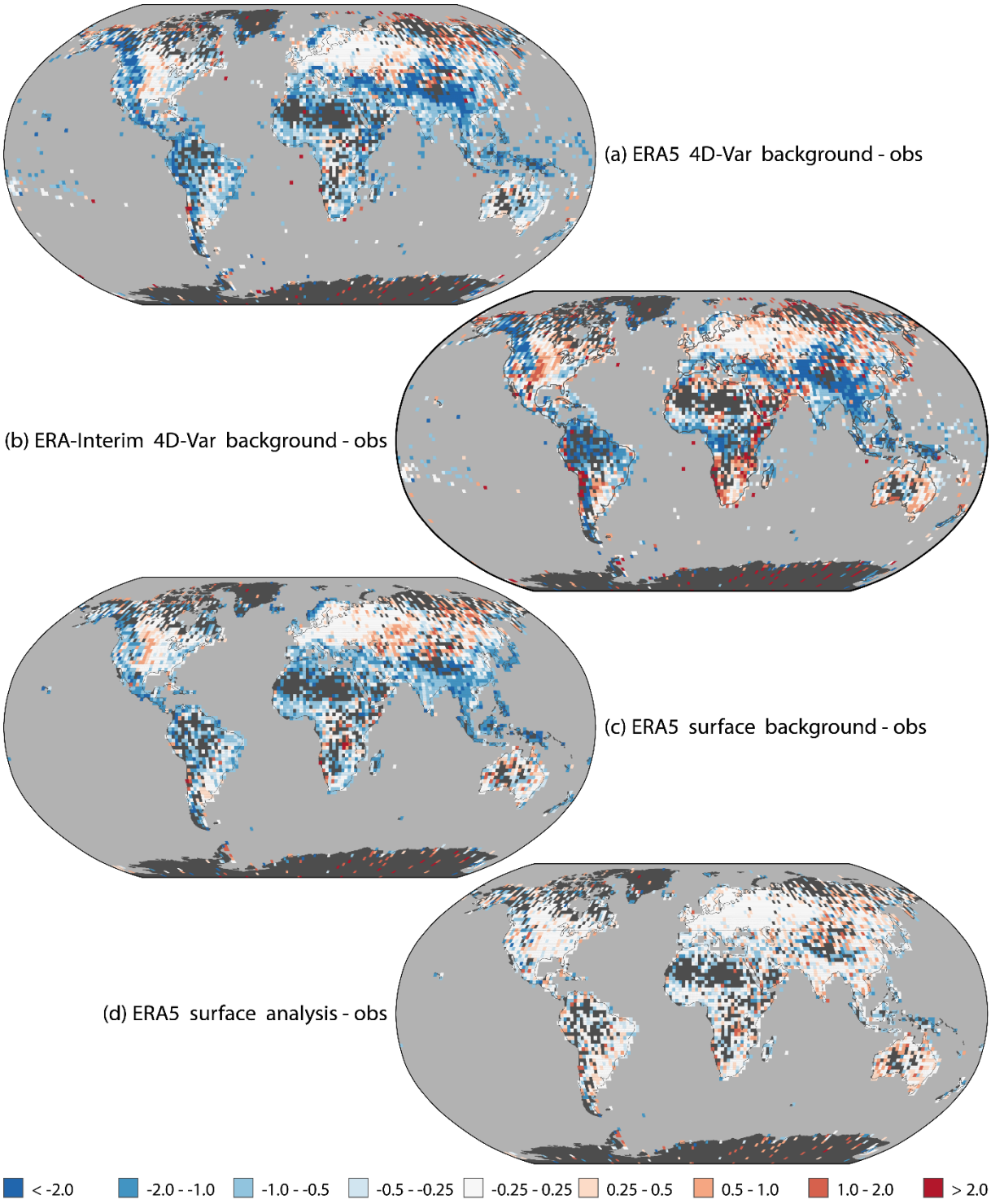

Figure 6 Mean (background - observation) and (analysis - observation) fits to surface air temperature observations (°C) from 1979-2018 for 2°x2° grid boxes containing at least one observing station and 100 observations. (a) Background fit for all selected observations processed by the ERA5 4D-Var, (b) corresponding fit of the ERA-Interim 4D-Var background, (c) background fit for the processed sample of observations used in the ERA5 surface analysis, and (d) corresponding surface analysis fit.

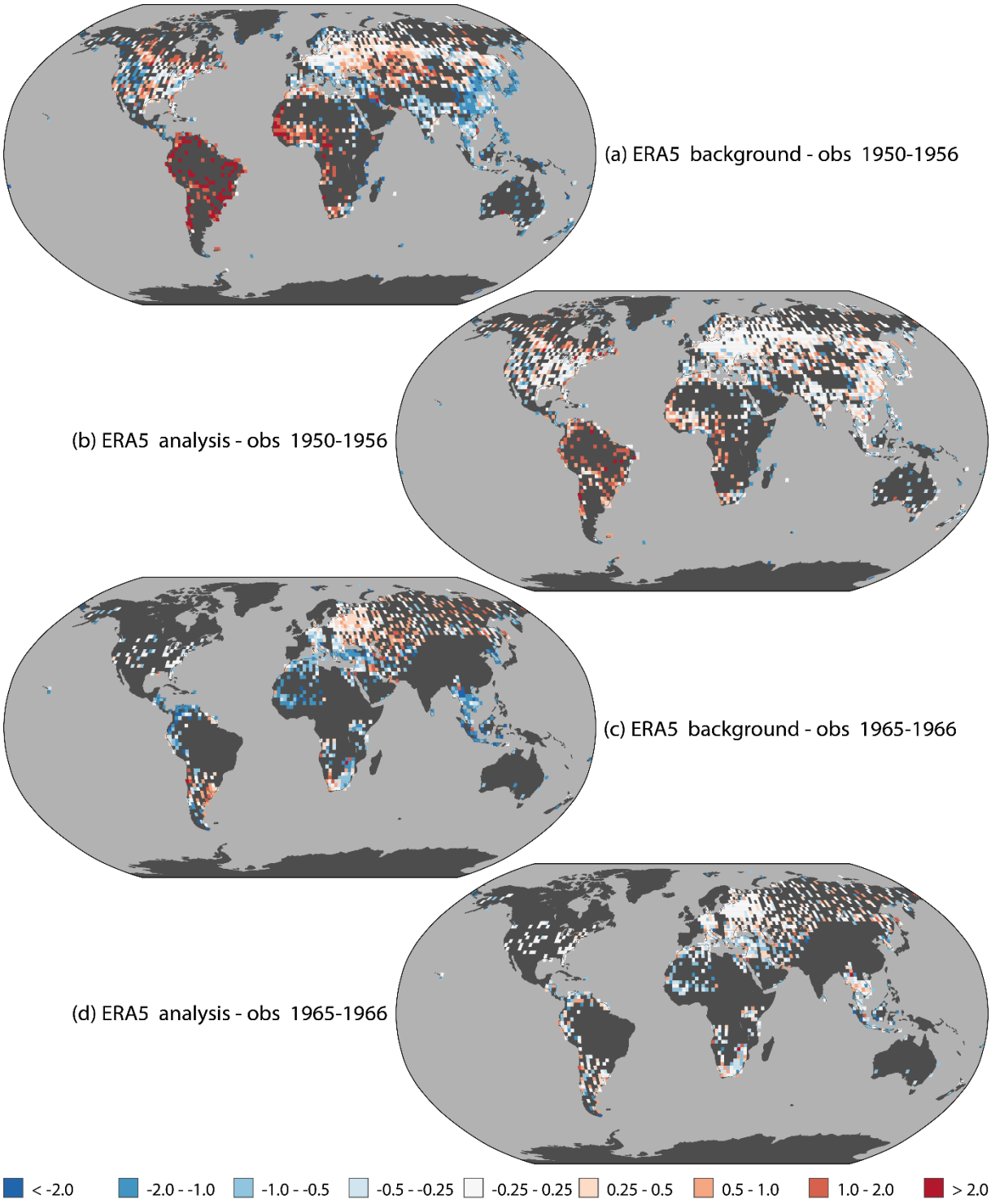

Figure 7 Mean (background - observation) and (analysis - observation) fits to surface air temperature observations (ºC) for 2°x2° grid boxes containing at least one observing station and 100 observations from the ERA5 surface analysis scheme: (a) background fit 1950-1956, (b) analysis fit for 1950-1956, (c) background fit for 1965-1966, and (d) analysis fit for 1965-1966.

Two examples from the pre-1979 period of ERA5 are presented in Figure 7. Average background-observation and analysis-observation differences from the surface-analysis scheme are shown for 1950-1956 and 1965-1966. Overall data counts are low for the first of these two periods, but the situation for

the extratropical northern hemisphere is much as illustrated in Figure 6 for 1979-2018, although background departures are locally larger and data coverage not quite as good. Conversely, data coverage is quite a lot poorer for much of the tropics and southern hemisphere, and background departures are opposite in sign to those for 1979-2018 in many places, most conspicuously over South America and Africa. There are in fact considerable temporal variations in data coverage in these regions for the 1950-56 period, and overall counts are much lower than for the extratropical northern hemisphere. Accordingly, the analysis increments shown in the left-hand panel of Figure 8 are small over most of South America and Africa. This is true also of much of China. Elsewhere the increments are broadly similar to those shown in Figure 5 from ERA5 for 1980-84 and 2014-2018.

The second period shown in Figure 7, 1965-66, was selected to illustrate the dearth of observations from many countries for these two years. The plots speak for themselves. The corresponding increment map presented as the right-hand panel of Figure 8 unsurprisingly shows much less widespread increments than seen for earlier and later periods.

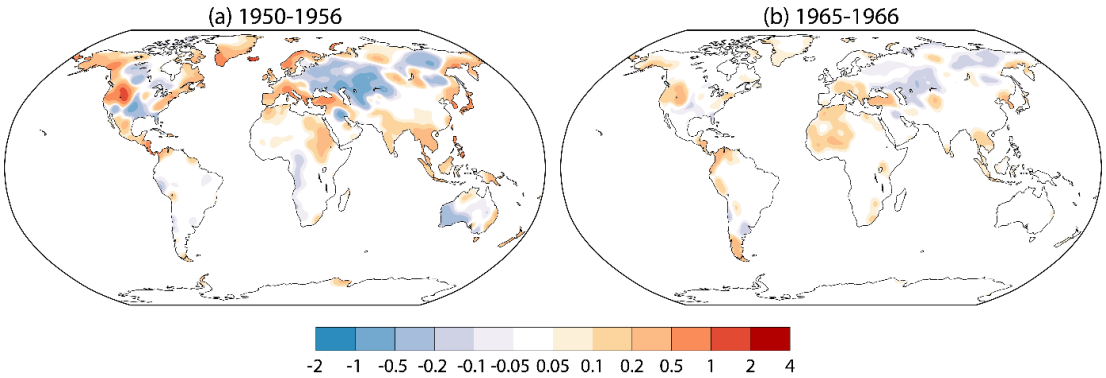

Figure 8 Mean analysis increments in two-metre temperature (°C) over land from the ERA5 surface air analyses for: (a) 1950-1956; (b) 1965-1966.

3.1.5 Three regions

3.1.5.1 Australia

Australia provides one exception to the general picture summarised above for Figure 6. The mean analysis departures shown in Figure 6 are larger for Australia than for many other well-observed regions of the world, with a predominantly warm bias in the archived sample that is not evident in the corresponding background departures. Time series of monthly departures from the 4D-Var and surface-analysis feedback averaged from 110°E to 160°E and 10°S to 50°S are presented in Figure 9. Here the quality check of the observations passed through 4D-Var was adjusted to remove evidently erroneous reports of 0°C temperatures in mid-1979 that otherwise are enough to cause misleading spikes in the 4D-Var departures. Time series are plotted only from July 1951 to December 1957 and from 1977 onwards because data counts are particularly low for other years, as shown in Figure 3.

The departures archived by the surface-analysis scheme can be seen in Figure 9 to shift markedly between the 1980s and the 1990s, and to change further over the final twenty years of the period. The corresponding 4D-Var background departures vary more smoothly over time. The shifts in the archived departures of the surface analysis coincide with increases in the total number of available (mainly manual) observations, first during December 1988 and then during February 1990. Figure 10 shows that the change from 1987 to 1991 comprises an increase in the number of observations at times other than the main synoptic hours, with the largest changes at 05, 20 and 23 UTC. Changes at the standard synoptic

hours are small, with both increases and decreases. This implies an increase in the number of stations for which observations are used in the surface analyses for times that differ from the observation time by one or two hours. The changes between the 1980s and 1990s in the means and standard deviations of the analysis scheme's departures are likely indicative of changes in the match between model and observation due to these differences between analysis and observation time.

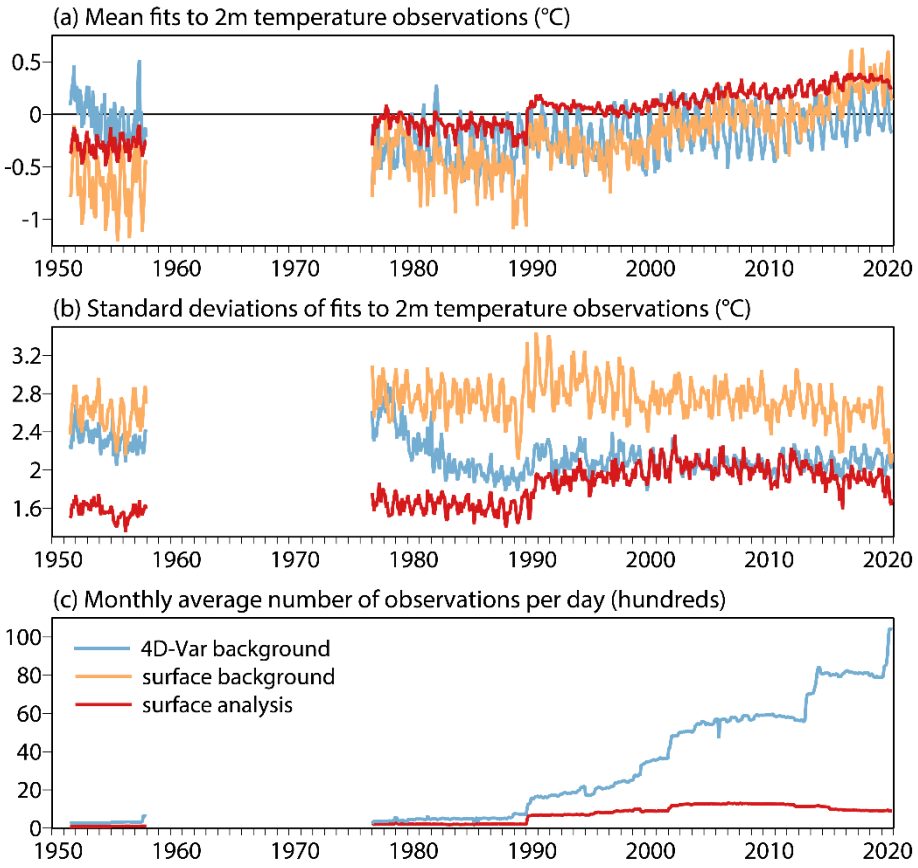

Figure 9 Time series of monthly values from July 1951 to December 1957 and from 1977to 2020 of (a) mean background - observation differences from 4D-Var feedback (blue) and surface-analysis feedback (orange) and mean analysis - observation differences from surface-analysis feedback (red), for two-metre temperature (°C) and all selected observations for the Australian region (110°E-160°E, 50°S-10°S). (b) Corresponding standard deviations. (c) The average daily number of observations for the Australian region from 4D-Var and surface-analysis feedback. The 4D-Var statistics are shown only for data transmitted in traditional alphanumeric code.

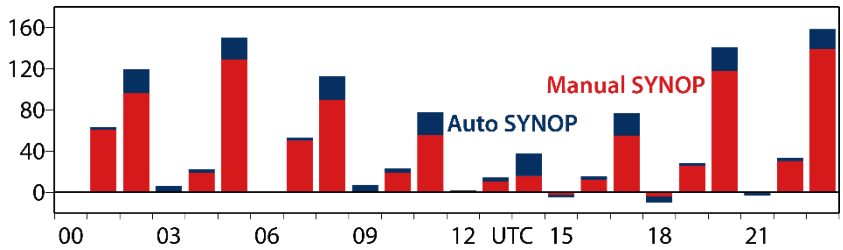

Figure 10 Change from 1987 to 1991 in the annual average number of surface air temperature observations reported from manual (red) and automatic (blue) measurements from the Australian region, for each hour of the day.

Differences in behaviour between the 1950s and 1980s can also be seen in Figure 9. The feedback from the surface analysis shows a larger cold bias of both background and analyses for the 1950s. This is consistent with what is shown in Figure 7, but at odds with what is shown later, in section 3.4.2, namely that the ERA5 analyses are substantially warmer than the monthly climatological datasets for the 1950s and 1960s, relative to the 1981-2010 norm. There is also quite a large discrepancy between the background departures from the 4D-Var and surface analysis schemes for the 1950s. These results have yet to be understood.

3.1.5.2 Antarctica

Antarctica provides a second exception to the general picture summarised for Figure 6. Figure 11 shows maps of departures for Antarctica. In addition to presenting more clearly than in Figure 6 the 1979-2018 data for the continent (due not only to the map projection but also to plotting feedback averaged for 1°x1° rather than 2°x2° grid boxes), surface-analysis departures are also shown for averages over the periods 1980-1984, 1997-2001 and 2014-2018, and for separate averages for June and December over the whole period. Although much of Antarctica has been observed with a reasonable density of measurement at some time or other within the last forty years, observational coverage of inland regions of the continent has varied quite considerably over the period. Coverage over the plateau was very sparse early in the period, and coverage in recent years, when observation numbers have been higher than in previous years, has gaps in regions that had previously been observed for a while.

The overall bias over the Antarctic plateau is a warm one, as found for ERA-Interim by Fréville *et al*. (2014) and Jones and Lister (2015), and discussed by Dutra *et al*. (2015). The bias is somewhat smaller for the ERA5 background than the ERA-Interim background, and somewhat smaller for the analysis than the background in the archived feedback data from the ERA5 surface-analysis scheme. This masks a significant difference between ERA5 and ERA-Interim, however. The ERA5 feedback shows a marked annual cycle in values over the plateau. The analysis is biased cold in summer months, as illustrated for December in Figure 11, and warm for the rest of the year. This contrasts with ERA-Interim, for which the 4D-Var departures show the background to be biased warm over the plateau in every month.

Stations at the periphery of Antarctica exhibit a mixture of biases. These biases do not in general exhibit a pronounced annual cycle such as seen for ERA5 in the interior of the continent. Biases also do not vary substantially from year to year. Evaluated over all observations from 1979 to 2018, the ERA products are generally biased cold, as found for ERA-Interim analyses from 1979 to 2013 by Jones and Lister (2015) for a set of 40 stations. The bias is only 0.1°C for the coastal feedback data from the ERA5 surface-analysis scheme, however. Simmons *et al*. (2017) examined a set of six stations with long reporting records, showing mostly reasonable agreement between the annual-mean anomalies in ERA-Interim background forecasts and observations for individual stations. Repeating the calculation for ERA5 gives a similar conclusion, with ERA5 performing somewhat better at two of the stations (both on the Antarctic Peninsula) but worse at a third.

Figure 12 shows mean background and analysis departures based on all observations south of 60°S for which there is feedback from the ERA5 surface air analysis. The annual cycle discussed above for stations located in the continental interior is evident in these area averages. The analysis scheme reduces both warm and cold biases as expected; the larger background biases and greater reduction of bias by the analysis scheme seen from 2008 onwards is presumably due to a greater number and changed geographical distribution of observations. Overall, there is little long-term variation in the analysis fits, despite the changes in the numbers and locations of observations. There is thus no evidence from these feedback statistics to indicate that the analyses do not reproduce the large-scale trend and low-frequency

variability information implied by the observations. Caution must nevertheless be exercised when examining local changes.

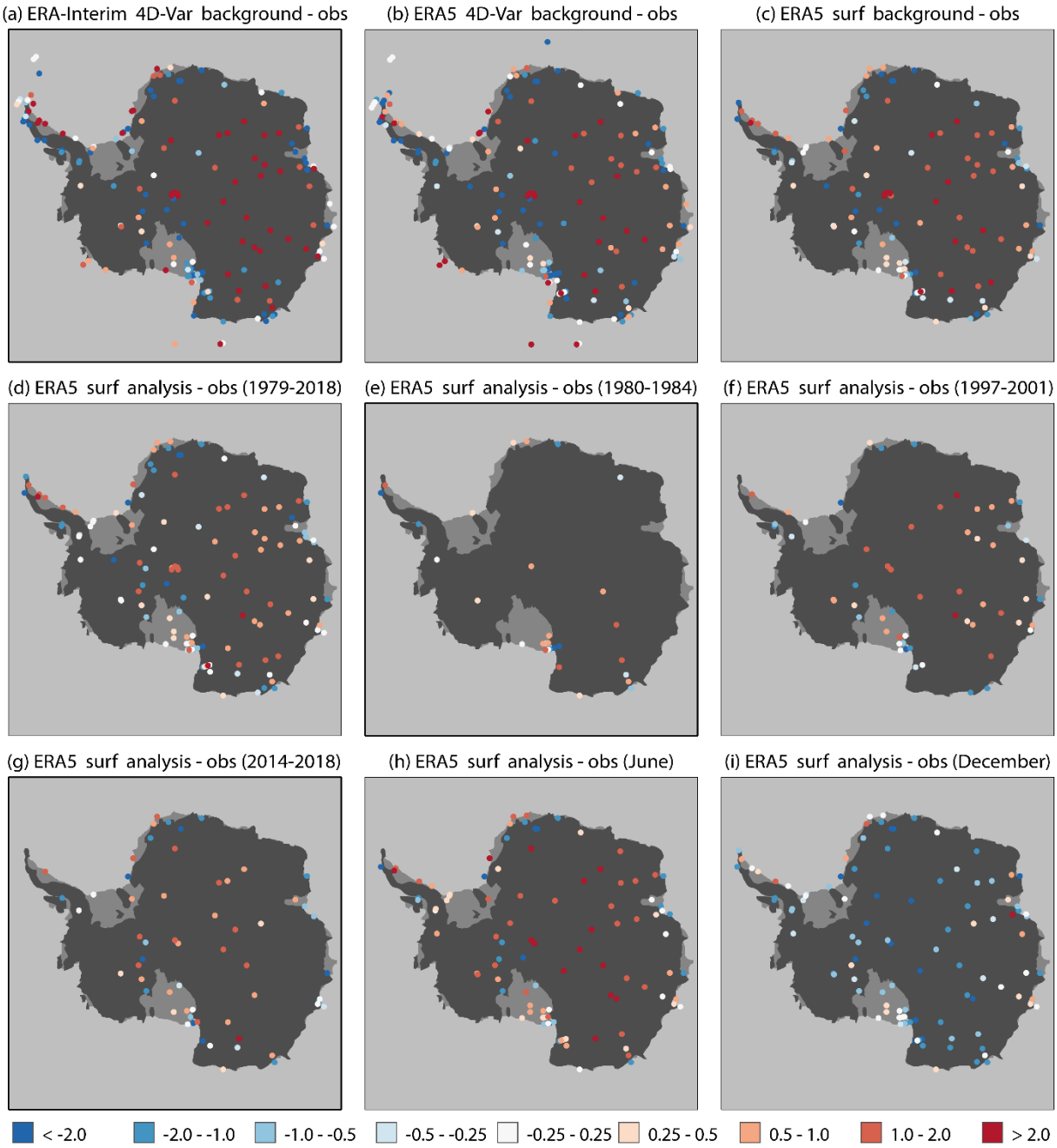

Figure 11 Mean (analysis - observation) and (background - observation) fits to surface air temperature observations (ºC) for 1°x1° grid boxes containing at least one observing station and 100 observations. (a) Background departures for all selected observations from 1979-2018 processed by the ERA-Interim 4D-Var, (b) corresponding ERA5 4D-Var background departures, (c) background departures for sample of 1979-2018 observations used in the ERA5 surface analysis, (d) corresponding surface analysis departures, (e) as (d) but for 1980-1984, (f) as (d) but for 1997-2001, (g) as (d) but for 2014-2018, (h) as (d) but for June only and (i) as (d) but for December only.

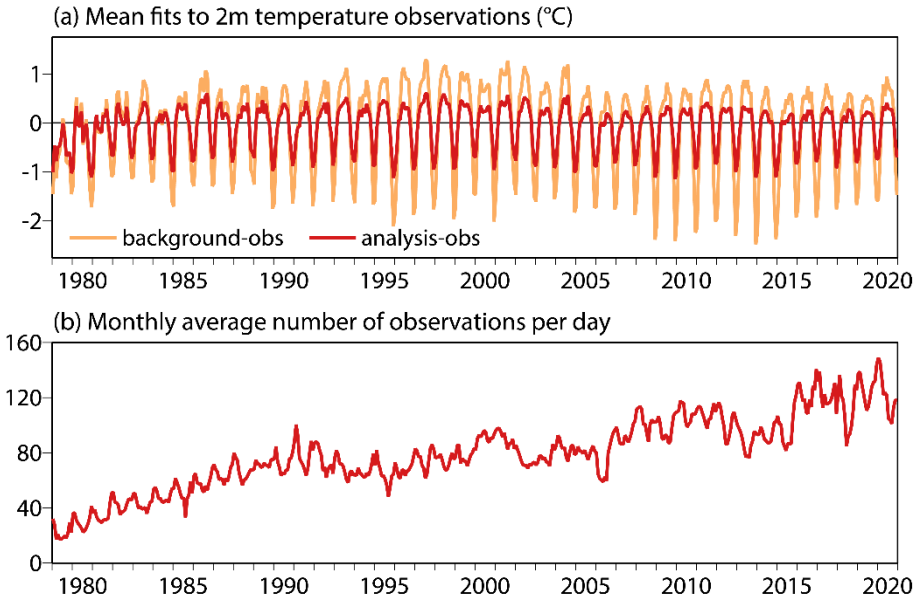

Figure 12 (a) Monthly (analysis – observation) (red) and (background – observation) (orange) departures, and (b) monthly averages of number of observations per day, from 1979 to 2020 averaged over all surface air temperature data (°C) from stations south of 60°S for which statistics from the ERA5 surface analysis have been processed.

3.1.5.3 Arctic sea-ice

A set of observing stations located on the drifting Arctic sea-ice was operated by the USSR between 1937 and 1991. The second such station (NP-2) ran from April 1950 to April 1951, and subsequent stations (NP-3 to NP-31) ran between 1954 and 1991. The surface pressure and 10m wind observations from these stations were assimilated in ERA5's 4D-Var, as were the data from the radiosonde ascents made from the stations. The surface air temperature data from the stations were not analysed, but were among the data processed passively by the 4D-Var. The data from 1979 onwards were known from ERA-Interim to be of high quality (Simmons and Poli, 2015), and the same appears true of the data back to 1950. The ice stations thus provide important data for evaluating ERA5 temperatures in a region that is otherwise almost devoid of data for much of the period. As the surface air temperature analysis is suppressed over sea-ice, the "analysis" values there from ERA5 are simply the same as the background values from its 4D-Var data assimilation.

Figure 13 presents time series of the observed temperatures and ERA5 background values. The fall in data numbers at the end of 1978 is due to the incomplete holdings of ice-station observations in the data sources used by ERA5 (and ERA-Interim) from 1979 onwards. The ICOADS dataset used for 1950 to 1978 contains substantially more ice-station data per month than ECMWF held from 1979 onwards. It has been confirmed from a two-year overlap assimilation (Bell *et al.*, 2021) that the ICOADS dataset also contains substantially more ice-station data for 1979 and 1980 than were processed in the original ERA5 production stream for these years.

A quite clear picture of the annual cycle emerges from Figure 13. Temperatures range from close to 0°C during the summer melt period to winter values that drop below -40°C on several occasions each year, with values lower than -50°C reported in three of the years. Several milder spells, with temperatures in the range from -20°C to a few degrees below 0°C, also occur each winter. These can be seen more clearly in Figure 14, which presents expanded plots for the three 12-month (July to June) periods that include temperatures below -50°C.

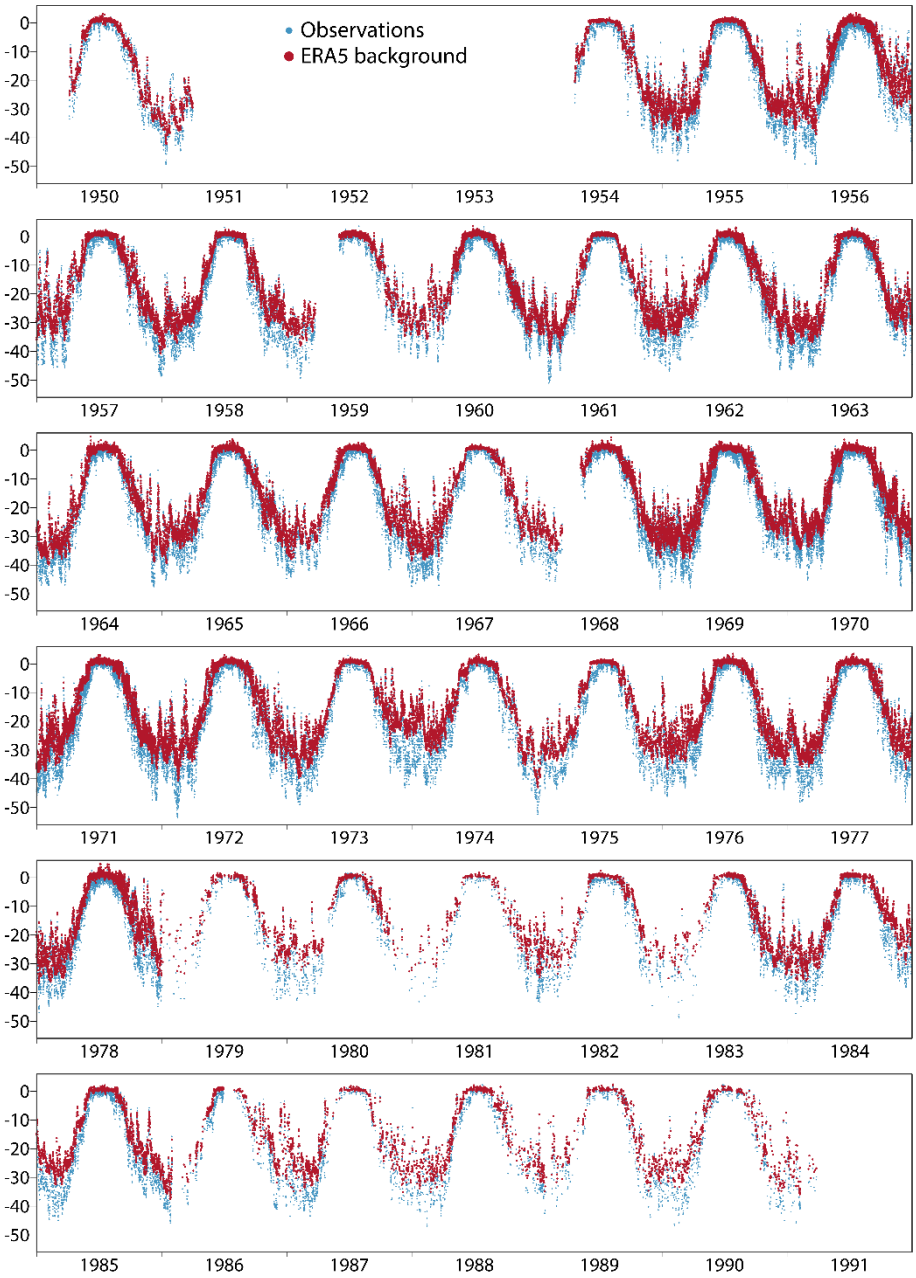

Figure 13 Observations of surface air temperature (ºC, blue) from drifting ice stations (NP-2 to NP-31) operated by the USSR between 1950 and 1991, and corresponding ERA5 background values (red).

The locations of the ice-station observations within the Arctic Ocean changes due to drift of the ice, and the number of stations providing data in any one month varies from none to four, the latter being the case from May 1970 to October 1971. The two distinct sets of values around 10ºC apart seen in Figure 14 for the middle of May 1972 come from two quite widely separated stations: NP-19 close to the North Pole and NP-21 close to 75ºN, 176ºE. These characteristics make it difficult to draw conclusions as to interannual variability and longer-term trends from these data alone.

ERA5 reproduces the general characteristics of the annual temperature cycle and the higher frequency variations in cold-season temperatures. Compared with the ice-station data, ERA5 temperatures are biased warm throughout the year, but with a considerable seasonal variation in magnitude. Calculated for each month of the year over the whole period, the warm bias of the ERA5 background is smallest in

July, at 1.1ºC, and largest in March, at 5.0ºC. The standard deviation of the background departures is also smallest in July, at 0.9ºC, and largest from October to February, with values in the range from 4.1 to 4.4ºC. The correlation of the two time series, with the mean annual cycle removed, is 83.2%.

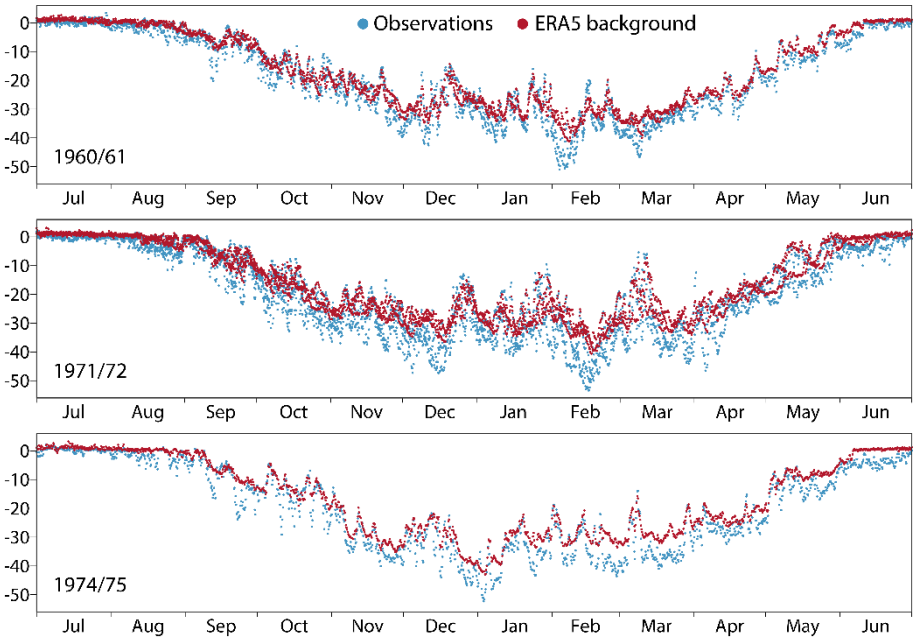

Figure 14 As Figure 13, but showing twelve-month periods from July to June for 1960/61, 1971/72 and 1974/75.

Surface air temperature measurements over Arctic sea-ice are also provided by buoys that drift with the ice. These observations have been processed passively by the ERA5 4D-Var in the same way as the ice-station data, for the period 1982-2018. This was likewise the case for ERA-Interim, for which analysis of the departure statistics for ice-station data from 1981 and buoy data up to 2010 was reported by Simmons and Poli (2015), including comparison with the statistics for data from ships and land stations north of 70ºN. In contrast to the ice-station data, a number of quality issues were evident for the buoy observations, although the majority of observations were of good quality and gave results compatible with those obtained for the ice-station data. Among the issues was a warm summer bias, with a number of reports of temperature well above 0ºC, something seen in neither the ice-station data nor the reanalyses. Rigor *et al*. (2000) had noted this bias in an earlier comparison of buoy, ice-station and reanalysis data (including ERA-15), and ascribed it to solar heating, an issue partly addressed for newer buoys, though (as will be seen) apparently still present for recent years.

The variations over the twelve months from July to June of buoy observations and the ERA5 background are shown in Figure 15 for the sample years 1985/86, 2000/01 and 2015/16. No quality control is applied to these data other than to suppress data close to land that may not be representative and may come from moored coastal buoys. Several observational problems can be inferred from the various stray lines that appear in these plots. Nevertheless, the basic picture presented earlier for the ice-station data from 1950 onwards can again be seen, with ERA5 evidently too warm in the coldest months of the year for both 1985/86 and 2000/01, a result found for most other years. This is not the case in 2015/2016, however. It is also not the case in 2016/17. Applying a further quality control step of rejecting suspect data for which the background departure is greater than 15ºC, ERA5 is found to have a net cold bias in the calendar years 2016 and 2017. All earlier years have a warm bias, but it is only 0.3ºC for 2015. The same is true of the ERA-Interim background. The largest annual warm bias, 5ºC, occurs for 1985, for both reanalyses.

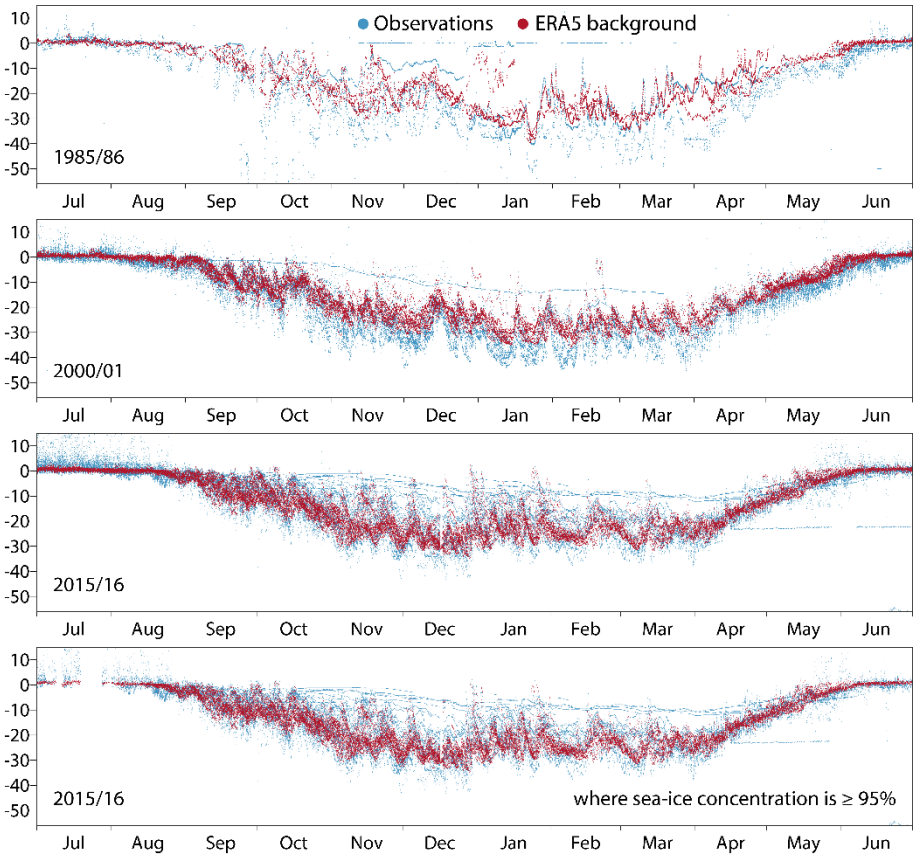

Figure 15 Observations of surface air temperature (ºC, blue) from drifting buoys, and corresponding ERA5 background values (red), for 12-month periods from July to June, for 1985/86, 2000/01 and 2015/16. Results for 2015/16 are also shown for only those observations and background values for which the sea-ice concentration used in ERA5 is at least 95%.

Figure 15 also illustrates how the buoy data include some temperatures that are several degrees and occasionally (in 2015) as much as 15ºC above 0ºC in summer. The fourth panel of the figure shows only those data for 2015/16 that are located where the ERA5 sea-ice concentration is at least 95%. This removes a considerable amount of data in the summer, but reported values of up to 15ºC are again seen. Most of the summer data are close to 0ºC, however. The 1982-2017 July average temperature is 1.0ºC for the buoy data. The July average for all the ice-station data from 1950 to 1990 is -0.2ºC.

Wang *et al.* (2019) compared ERA5 and ERA-Interim temperatures with buoy data for 2010-2016. Both analyses had warm biases of several degrees in winter and spring, but the bias was larger in the case of ERA5. Biases were smaller in summer and autumn, when ERA5 was slightly colder than ERA-Interim. Figure 16 shows this to be the case for both ice-station data averaged monthly for 1979-1990 and buoy data averaged monthly for 1982-2017, applying the 15ºC quality-control limit in the case of the buoy data.

ERA-Interim also captures variability better than ERA5. For the ice-station data from 1979 to 1990 the correlations of the time series of observed and background temperatures, with the mean annual cycle removed, are 80% for ERA5 and 85% for ERA-Interim. For the buoy data from 1982 to 2017 the correlations are 76% for ERA5 and 78% for ERA-Interim.

Aside from any differences due to the different versions of the assimilating model used by ERA-Interim and ERA5, background differences between ERA-Interim and ERA5 arise due to differences in the analyses of sea-ice concentration they use. In addition, ERA-Interim, unlike ERA5, applied its surface

analysis scheme over sea-ice, which resulted in analyses that did not inherit all the bias of its background values (Simmons and Poli, 2015). Further illustrations and discussion are given later in this report, in sections 3.3 and 3.5.4.

The warmer winter temperatures and smaller model biases shown in Figure 16 for the buoy data compared with the ice-station data cannot be simply interpreted. This is because of differences in the locations sampled by the two types of data. The buoy data are primarily for more recent years than the ice-station data, however. The warmer winter temperatures are thus to some extent indicative of the recent warming of the Arctic (see section 3.4.4), which is suppressed in surface air over sea-ice in summer due to melting.

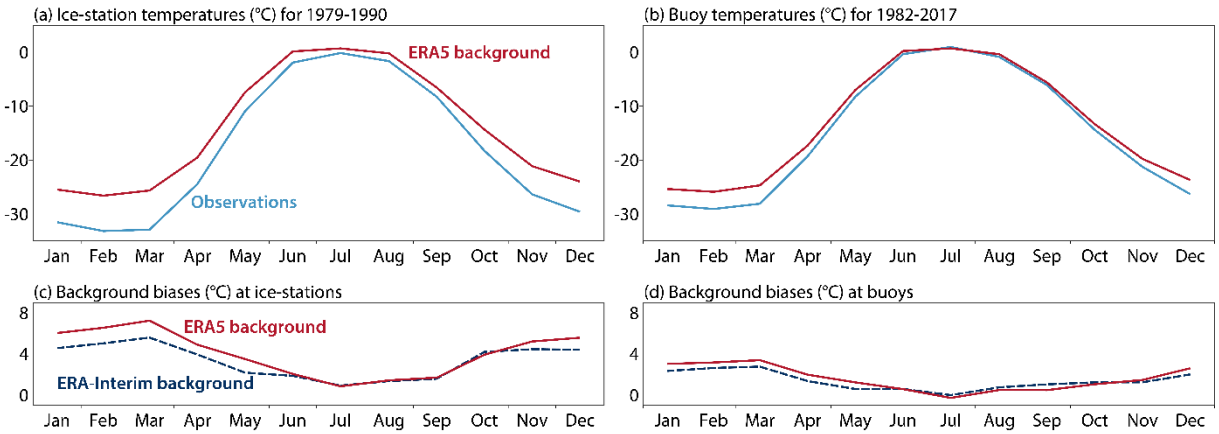

Figure 16 Observed (light blue) and co-located ERA5 background (red) Arctic temperatures (°C) averaged for each month of the year for (a) ice-stations (1979-1990) and (b) buoys (1982-2017). Panels (c) and (d) show corresponding (background - observation) differences for ERA5 (red) and ERA-Interim (dark blue).

3.1.6 Local behaviour

Examples of local behaviour since 1979 for four high-latitude stations with long and reasonably complete reporting records are presented in Figure 17. Three are in the Arctic. Two of them are stations whose data were compared with background forecasts from ERA-Interim by Simmons and Poli (2015): Alert on Ellesmere Island, close to north-western Greenland, and Ostrov Vize, in the Kara sea north-east of Novaya Zemlya. The third is Longyearbyen, Svalbard, chosen both because of the large changes observed there and its challenging topography, which is reflected in substantial differences in the background departures of ERA-Interim and ERA5. The fourth station, Marambio, is located on an island close to and east of the northern limit of the Antarctic Peninsula. Out of the six Antarctic stations for which comparisons of observations with ERA-Interim were presented by Simmons *et al*. (2017), Marambio was the station for which ERA-Interim performed most poorly. Differences between the station elevation and the orographic heights of the ERA-Interim and ERA5 background models are less than 300m for each of these stations, so their data are used in the surface analysis if background departures are not too large.

The left-hand panels of Figure 17 show, for each of the four stations, the monthly mean 4D-Var background departures for ERA-Interim and ERA5, and the mean fits of the ERA5 surface analysis. The ERA5 background evidently fits the station data much more closely than ERA-Interim does at Alert and Longyearbyen, for most of the year at least. This is particularly the case in mid-winter, when in most years the monthly mean ERA-Interim background departure exceeds the 7.5°C threshold beyond which observations are rejected for use in the surface analysis. The (austral) mid-winter biases at Marambio

are also larger for ERA-Interim than ERA5 in almost all years. Biases are smaller at Ostrov Vize, especially in summer, and similar for ERA-Interim and ERA5, particularly in recent years. Vize island lies in a region where there has been a substantial decline in sea-ice cover, and better reanalysis performance is to be expected as air temperatures over the island become more strongly influenced by the relatively well analysed temperatures of ice-free sea.

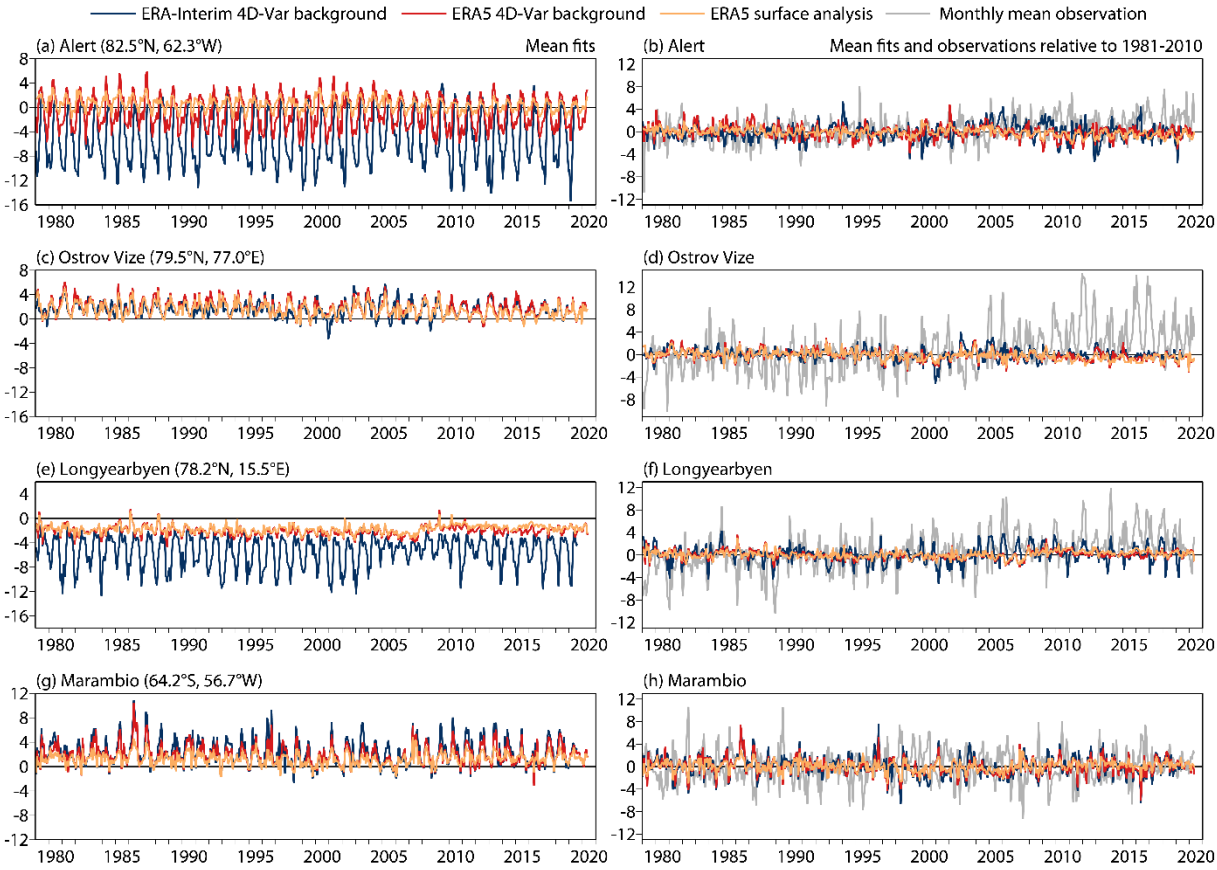

Figure 17 Monthly mean 4D-Var (background-observation) departures of 2m temperature (ºC) at (a) Alert (82.5°N, 62.3°W) , (c) Ostrov Vize (79.5°N, 77.0°E, (e) Longyearbyen (78.2°N, 15.5°E)) and (g) Marambio (64.2°S, 56.7°W) for ERA5 (red) and ERA-Interim (blue), and corresponding (analysis-observation) departures for the ERA5 surface analysis (orange), for observations transmitted in traditional alphanumeric codes from January 1979 to May 2020. Panels (b), (d), (f) and (h) show corresponding values plus the monthly mean observations (grey) expressed as anomalies relative to 1981-2010.

The feedback from the surface analysis available for ERA5 shows predominantly smaller departures for the analysis than for the background. This is especially the case for the larger errors that occur in wintertime, and is more generally the case for Alert, for which background departures are much larger and more seasonally variable than for the other two Arctic stations. The ERA5 analysis also fits the wintertime observations from Marambio distinctly more closely than the ERA5 background does.

The right-hand panels of Figure 17 show corresponding results plotted as anomalies relative to 1981-2010, and adds the anomalies in observed temperatures. The poor ERA-Interim background is much less prominent in these plots. Warming over the period since 1979 is particularly large at Ostrov Vize. It is also larger at Longyearbyen than Alert. Marambio is characterised more by variability than trend, as will be illustrated later for the analysis over the Antarctic as a whole. There is in contrast little net trend and relatively small variability in the anomalies in departures, implying that the reanalyses,

particularly ERA5, provide a faithful representation of long-term variability and trends for these four stations.

More generally, cases of large monthly average 4D-Var background fit tend to be more pronounced for ERA5 than ERA-Interim. This happens in particular when the observing station is located in a valley or is otherwise relatively close to a high mountain ridge. In these circumstances the mismatch between elevation of the station and the background model's orographic height tends to be larger for the higher-resolution ERA5 model than for the ERA-Interim model. Examples are presented in Figure 18 for Madrushkat, a station located in a mountain valley in Tajikistan, and Anta/Huaraz, a station located close to the highest point of the Peruvian Andes. Model and station elevations differ by more than 300m for both stations and both reanalyses (see figure caption for elevations), so no data from these stations is used in either surface analysis. ERA5's orography is more than 1000m higher than that of ERA-Interim at the location of both stations and this is reflected in much lower temperatures and larger deviations from the values measured at the lower elevations of the stations. Despite this, the intra-monthly standard deviations of the temperature departures are for the most part lower for ERA5 than for ERA-Interim at these stations, as indeed they are found to be when calculated over all stations for which the monthly average ERA5 and ERA-Interim background temperatures differ by more than 5°C.

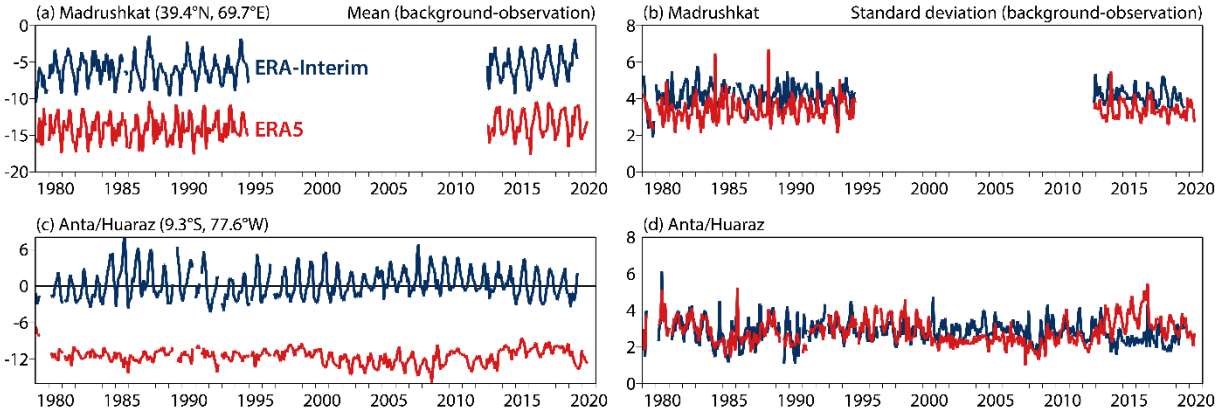

Figure 18 Monthly mean 4D-Var (background-observation) departures of 2m temperature (°C) at (a) Madrushkat (39.4°N, 69.7°E) and (c) Anta/Huaraz (9.3°S, 77.6°W) for ERA5 (red) and ERA-Interim (blue) for observations transmitted in traditional alphanumeric codes from January 1979 to May 2020. Corresponding monthly standard deviations of the departures are shown in panels (b) and (d). The heights of the surface elevations for Madrushkat are: Station 2234m, ERA-Interim 2598m and ERA5 3550m. For Anta/Huaraz they are: Station 2759m, ERA-Interim 3149m and ERA5 4189m.

3.2 Sea-surface and marine-air temperatures from the ERA and JRA-55 reanalyses

Comparison is made here between the SSTs from ERA-Interim, ERA5 and JRA-55, and MATs (two-metre temperatures over sea) from ERA5 and JRA-55. Anomalies in MAT and SST are discussed for a wider range of datasets in section 3.4.

Figure 19 shows averages taken over all ice-free sea. Interannual variations in SST are generally similar for all three analyses, and the same for ERA5 and ERA-Interim after January 2009, apart from negligible differences arising from interpolation to different model grids. The COBE (Ishii *et al.*, 2005) SST analyses used by JRA-55 are nevertheless some 0.2°C warmer on average than the combination of the HadISST2 and OSTIA (Donlon *et al.*, 2012) analyses used by ERA5 (Hirahara *et al.*, 2016). The SST analyses used by ERA-Interim have previously been documented as shifting by around 0.1°C at the end of 2001 due to a change in source of the analyses, and has led to SST and 2m temperature over ice-free sea being reduced by 0.1°C prior to 2002 for ERA-Interim data when comparing with other datasets

(Simmons *et al*., 2017). As noted in section 2, this is done here also. Figure 19(a) shows that the 0.1°C adjustment brings the ERA-Interim SST quite closely into line with that of ERA5.

Notwithstanding the general agreement on variability, some differences are apparent in Figure 19(a). ERA5 exhibits more pronounced maxima than ERA-Interim and to a lesser degree JRA-55 in 1981/82, 1990/91 and 1997/98, for example. The figure also shows the spread of the SST analyses used in ERA5's ensemble data assimilation. This provides no indication that the maxima in ERA5 are due to extreme behaviour of the ensemble member chosen for use in the HRES analysis. It can also be seen that the average difference between the JRA-55 and ERA5 SSTs declines slightly over time from the early 1970s, such that the ERA5 SST has slightly the larger warming trend. Differences are also smaller prior to the 1970s. Further discussion is given in section 3.4, in the light of results from other datasets.

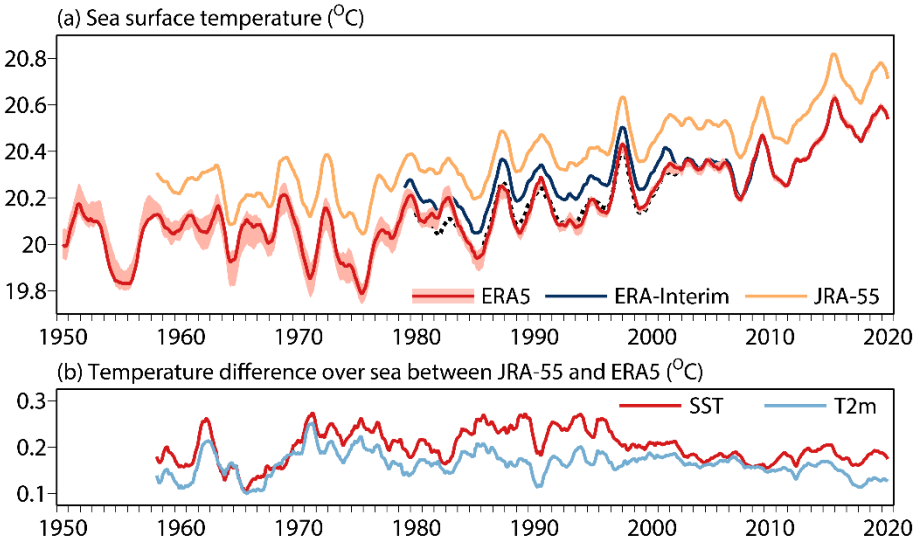

Figure 19 (a) Twelve-month running averages of sea surface temperature (SST; °C) taken over all ice-free sea between 60.625°N and 60.625°S for ERA5 (red), ERA-Interim (dark blue) and JRA-55 (orange) from 1950 to 2020. The dotted curve shows adjusted values for ERA-Interim used prior to 2002 in estimating trends. Pink shading denotes the spread of the ERA5 ensemble. (b) Corresponding differences between JRA-55 and ERA5 in SST (red) and two-metre temperature (blue; °C).

Differences between the absolute values provided by different SST datasets may occur due to differences in the mix of observations used and the bias corrections applied to the observations. Differences may also be due to differences in the depths in the uppermost ten or so metres of the ocean to which the datasets apply. The surface-layer parametrizations of assimilating models may however compensate to a greater or lesser degree for differences in the SST analyses that the models use. It may thus be asked whether the MATs from ERA5 and JRA-55 are more similar than their SST analyses. Figure 19(b) shows this to be the case, to a limited degree. The differences in two-metre temperature between JRA-55 and ERA5 are seen to be a little more uniform over time and generally smaller than the differences in SST. SST and MAT both increase at a slightly faster rate from 1970 in ERA5 than in JRA-55.

3.3 Estimates of absolute global and regional temperatures

The systematic difference in MAT between ERA5 and JRA-55 shown in Figure 19 prompts the question as to how well the global-average air temperature is known in absolute terms. Time series of surface-air temperatures are often presented as anomalies with respect to a climatological reference period, as in Figure 1 and later figures in this report. Indeed, although absolute temperatures are provided routinely by the reanalyses, the well-established datasets based on monthly climatological station data and SST

analyses generally provide anomalies rather than absolute values. This is partly because merging SST anomalies with surface air temperature anomalies over land provides a reasonable approximation to global anomalies in surface air temperature (discussed further in the following section), avoiding the need to analyse MAT or infer it from SST using modelling. Moreover, the analysis methods used over land produce grid-square average monthly values by combining the anomalies calculated for each observing station, avoiding the need to make adjustments for variations in station heights and other observational characteristics within the grid square.

Jones *et al.* (1999) did derive a globally complete surface air temperature climatology for 1961-1990, however. Their estimate of the annual and global average temperature was 14.0ºC, and they stated that comparison with earlier climatologies suggested that this average was within 0.5ºC of the true value. Hansen *et al*. (2010) subsequently produced an estimate of 14ºC, but for 1951-1980, with an estimated uncertainty of several tenths of a degree Celsius. The GISTEMP average for 1961-1990 is 0.1ºC above its 1951-1980 reference level.

The estimate of Jones *et al*. (1999) has been brought forward to recent years using the anomalies in global-mean temperature provided by GISTEMP and HadCRUT5, and compared with the values computed directly from ERA5, ERA-Interim and JRA-55. The annual averages for each of the past six years are shown in Table 1. Using the Hansen *et al*. (2010) estimate with GISTEMP gives values 0.1ºC higher than those tabulated for the Jones *et al*. (1999) estimate updated with GISTEMP. Rounded to 0.1ºC the different estimates are within 0.2 or 0.3ºC of each other, depending on year. ERA5 is a few hundredths of a degree Celsius colder than ERA-Interim for two of the years compared, and around 0.1ºC colder than JRA-55. The Jones *et al*. (1999) estimate for 1961-1990 updated using either GISTEMP or HadCRUT5 gives values close to those from JRA-55.

	2015	2016	2017	2018	2019	2020
ERA5	14.64	14.81	14.72	14.64	14.78	14.81
ERA-Interim	14.66	14.85	14.76	14.66	-	-
JRA-55	14.75	14.93	14.80	14.72	14.88	14.86
Jones *et al*. updated with GISTEMP	14.80	14.92	14.83	14.75	14.89	14.92
Jones *et al*. updated with HadCRUT5	14.83	14.93	14.85	14.76	14.89	14.92

Table 1 Global-mean temperature (ºC) averaged for the years from 2015 to 2020 from ERA5, ERA-Interim and JRA-55, and as estimated by Jones et al. (1999) for 1961-1990, updated using anomalies from HadCRUT5 and GISTEMP.

The global-mean temperature is larger in boreal summer than winter, associated with the larger seasonal cycle over land and the greater land mass of the northern hemisphere. Global-mean temperature was estimated by Jones *et al*. (1999) to vary from about 13ºC in January to a maximum close to 16ºC in July. The annual range from ERA5 is rather larger. Figure 20 shows the annual variation of global-mean temperature from 1950 to present. Over this period the annual temperature range of the 31-day running averages shown varies from 3.6 to 4.3ºC, with an average of 3.9ºC.

JRA-55 also has a larger annual range in global-mean temperature than estimated by Jones *et al.* (1999). The annual variations of monthly values averaged for 1981-2010 are presented in Figure 21 for ERA5 and JRA-55. The temperature difference between July and January is 3.6ºC for JRA-55 and 3.8ºC for ERA5. The two have similar temperatures for July August and September, but JRA-55 is warmer for the rest of the year. Corresponding results for the climatological reference period 1961-1990 are similar.

The temperature differences between July and January are 0.1ºC larger for both datasets for this earlier reference period. This is consistent with an indication from ERA5 in Figure 20 that the rate of temperature increase is a little lower in boreal summer than at other times of the year. The picture is complicated by the interannual variability of boreal winter temperatures, however.

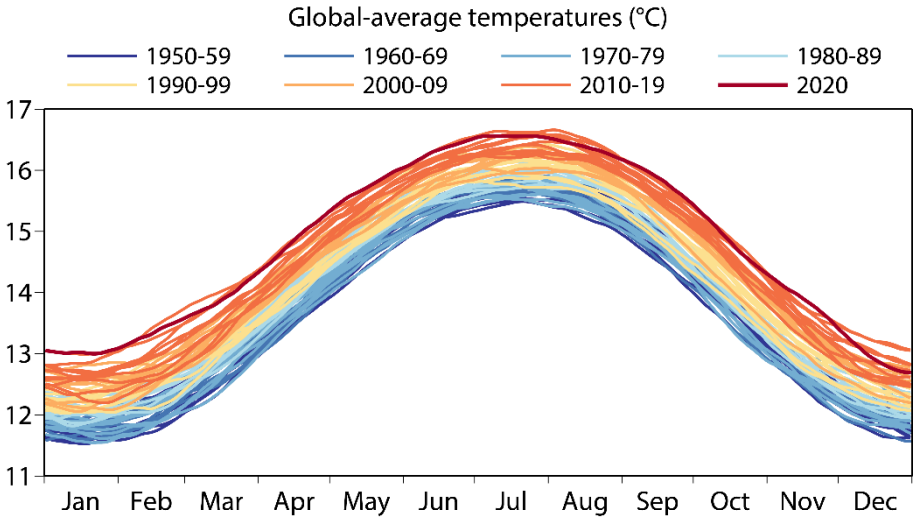

Figure 20 Annual variation of the global-mean ERA5 temperature (ºC) for each year since 1950, coloured according to decade. 31-day running averages of daily values are plotted.

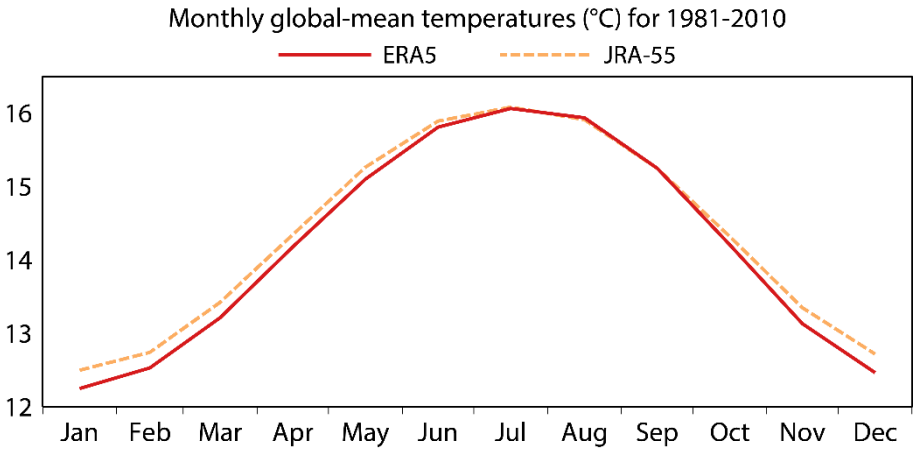

Figure 21 Annual variation of monthly global-mean temperatures (ºC) for the climatological reference period 1981-2010, from ERA5 (solid) and JRA-55 (dashed).

The differences in absolute temperatures between ERA5, ERA-Interim and JRA-55 have been examined further. It has already been seen in Figure 19 that air temperatures over ice-free sea are lower in ERA5 than JRA-55, by an amount that varies over time but is mostly in the range from 0.1 and 0.2ºC. ERA5 temperatures are also lower than JRA-55 temperatures over land. This is also the case for ERA-Interim, though to a lesser extent. This is shown by the monthly values averaged over all land presented in the upper panel of Figure 22. As noted earlier, ERA-Interim has a lower temperature trend over land than ERA5 (or indeed JRA-55). ERA-Interim can be seen in Figure 22 to have an absolute mean temperature over land that is closer to JRA-55 than ERA5 in its early years and closer to ERA5 than JRA-55 for recent years.

Panel (b) of Figure 22 includes the CRU TS dataset (Harris *et al*., 2020) in the comparison. This is a dataset based on direct analysis of observations over all land other than Antarctica, and the reanalyses accordingly are averaged over all land northward of 60°S in this case, which results in values that are some 4.5°C higher than those in panel (a). Differences among datasets are smaller in this case. JRA-55 is particularly close to CRU TS in these averages, and the differences in trends are such that all datasets are close to each other in absolute terms for the most recent years.

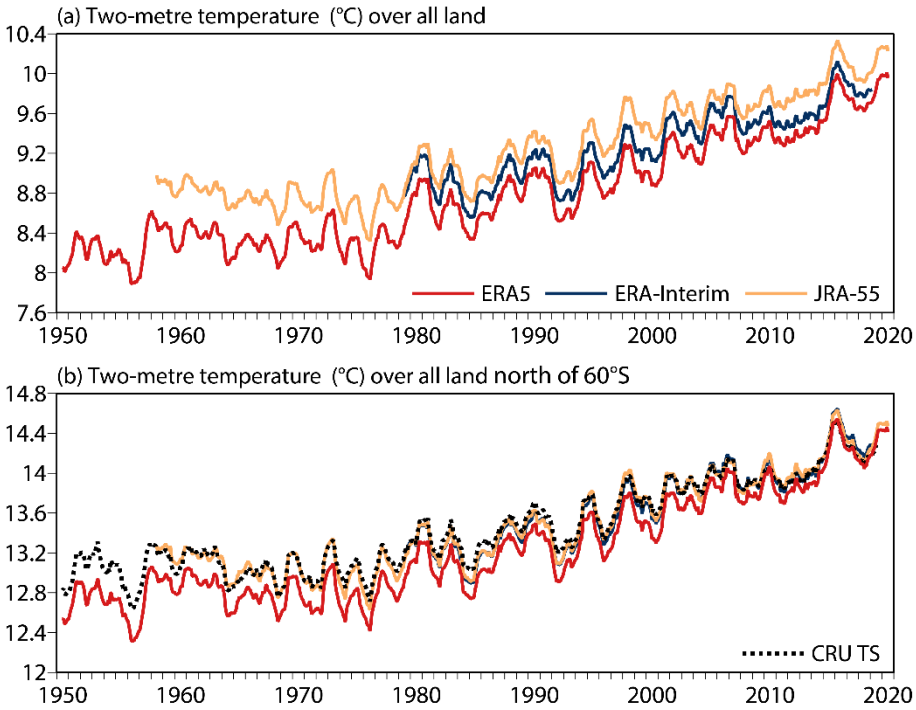

Figure 22 (a) Twelve-month running averages from 1950 to 2020 of two-metre temperature over all land (°C) for ERA5 (red), ERA-Interim (dark blue) and JRA-55 (orange). (b) Corresponding values for land north of 60°S, including from version 4.04 of the CRU TS dataset (black, dotted).

Maps of the average differences for the period 1979-2018 between ERA5 and respectively CRU TS, JRA-55 and ERA-Interim are presented in Figure 23. Differences between the ERA5 and JRA-55 background forecasts are also shown.

Although JRA-55 matches CRU TS more closely than ERA5 does in the average over all land north of 60°S, it can be inferred from the difference maps in panels (a) and (b) of Figure 23 that there are regions where this is not the case. In particular, ERA5 has much higher temperatures than CRU TS over north-eastern Siberia and the southern Arabian Peninsula, regions where JRA-55 is warmer still. There are other such regions, though with smaller differences between ERA5 and CRU TS, but on the whole JRA-55 is the closer to CRUTS over much of the land surface. ERA5 is much colder than CRU TS over Greenland, and here JRA-55 is closer to CRU TS, though colder. ERA5 is also colder than CRU TS over the Tibetan Plateau, where it very much overestimates snow depth. This is mainly due to an overestimation of precipitation, the effect of which cannot be corrected by the snow-analysis scheme due to absence of *in situ* observations (Orsolini *et al.*, 2019). The problem is less marked for ERA-Interim as it assimilated NOAA IMS snow-cover data (https://nsidc.org/data/G02156) for the region from 2004. ERA5 also assimilates IMS data, but only in non-mountainous areas.

The most striking differences seen in Figure 23 are between ERA5 and JRA-55 over Antarctica, and over the sea-ice regions around it and over the Arctic Ocean. JRA-55 is warmer than ERA5 over the

Antarctic Plateau and (to a lesser degree) over Greenland and other Arctic land areas. As other evidence, discussed in section 3.1.5.2, points to ERA5 being biased warm over the interior of Antarctica, the indication is that JRA-55 has an even larger bias in this region. ERA-Interim is also warmer than ERA5 over Antarctica, though to a smaller degree. This is consistent with what was shown earlier for the background fits to observations. ERA5 is thus seen to be a modest overall improvement on ERA-Interim for inland Antarctic temperature.

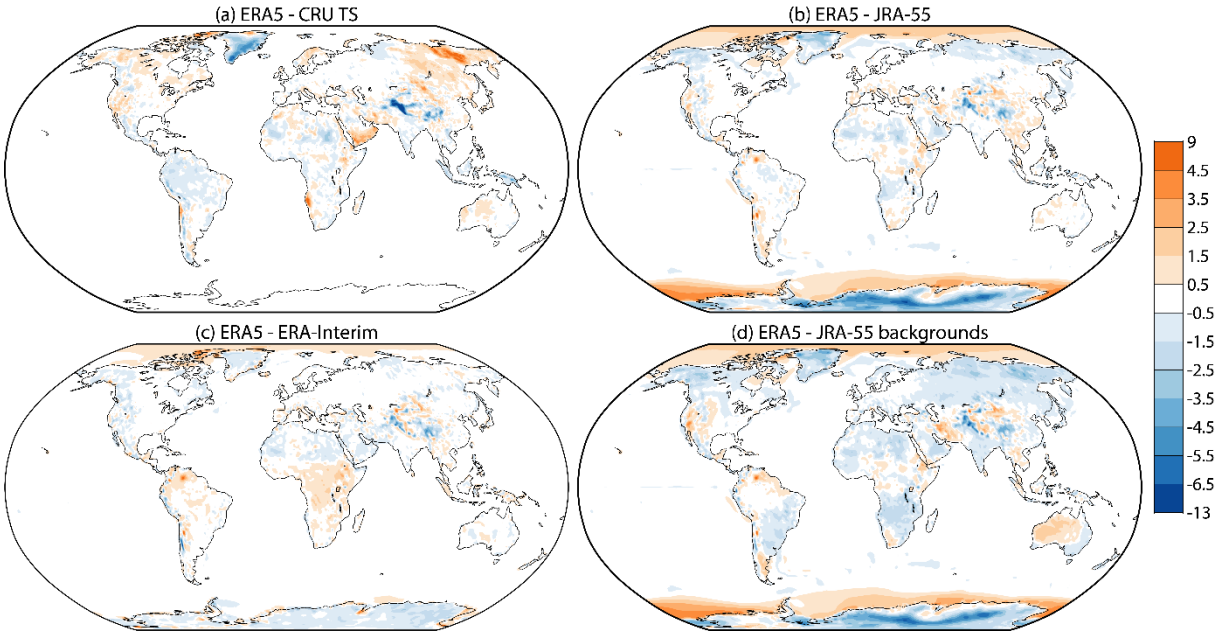

Figure 23 Difference in two-metre temperature (°C) between (a) ERA5 and CRU TS, (b) ERA5 and JRA-55, (c) ERA5 and ERA-Interim, and (d) the ERA5 and JRA-55 background forecasts, averaged from 1979 to 2018. The differences in (a) are computed on the 0.5° grid of the CRU TS dataset, and apply the ERA5 land-sea mask to suppress island values that are resolved in one dataset but not the other. The differences in other panels cover land and sea, and were computed on the 1.25° resolution at which the JRA-55 data were downloaded.

ERA-Interim has a warm wintertime bias over Arctic sea-ice, and ERA5 is slightly warmer still (section 3.1.5.3). JRA-55 is colder than ERA5 over both Arctic and Antarctic sea-ice, but it is unclear whether it is generally superior in this regard. For example, Batrak and Müller (2019) compared clear-sky ice-skin temperatures from these (and other) reanalyses with measurements from satellite for 2015-2017. ERA5 displayed too-high temperatures almost everywhere, whereas the lower temperatures from JRA-55 were in better agreement with the satellite data over a quite large portion of the central and Canadian sector of the Arctic Ocean, but were considerably too low and in poorer agreement with the satellite data over the northern Barents Sea, and generally too low in regions away from the North Pole.

The colder temperatures of ERA-Interim compared with ERA5 over Arctic sea-ice are associated with higher sea-ice concentrations and the cooling from application of the surface analysis over sea-ice. These are discussed further in section 3.5.4. One reason the JRA-55 temperatures are colder still over sea-ice is that the ERA assimilation systems use fractional sea-ice cover when it is analysed to be greater than 20% in the case of ERA-Interim and 15% in the case of ERA5, and no cover otherwise, whereas the JRA system uses either 100% sea-ice cover or none, according to whether the analysed fractional sea-ice cover is greater or less than 55% (Kobayashi *et al*., 2015). ERA5's use of fractional sea-ice cover gives higher air temperatures at those times of the year when the ice-covered fraction of a model grid

square has a surface temperature substantially lower than that of the ice-free fraction, which cannot be lower than the freezing point of sea water, -1.7°C.

Comparing panels (b) and (d) of Figure 23 shows that there are few areas over land where the ERA5 background is warmer than the JRA-55 background. It also shows how the respective surface analysis schemes are successful in producing analyses that are closer than the background forecasts are at non-polar latitudes. Panels (b) and (d) are identical over sea by construction, due to this report's suppression of the surface analysis over sea.

Corresponding differences between ERA5 and JRA-55 are shown for the months of January and July in Figure 24. ERA5 is for the most part colder than JRA-55 over land in January, particularly in the Arctic. Differences over sea-ice are larger over Antarctica in July than they are over the Arctic in January. Sea-ice extent is also larger over Antarctica in July than over the Arctic in January. Conversely, Arctic sea-ice extent is larger in July than Antarctic sea-ice extent in January, but as discussed earlier, temperatures over melting summer Arctic sea-ice are constrained to be close to 0°C, and the differences between ERA5 and JRA-55 temperatures over sea-ice are relatively small then. JRA-55 is generally a little warmer over open sea than ERA5, as seen in Figure 19. The differences for July average out to give the similar global-mean temperatures for ERA5 and JRA-55 illustrated in Figure 21, but ERA5 is colder overall than JRA-55 in January, by a little under 0.25°C.

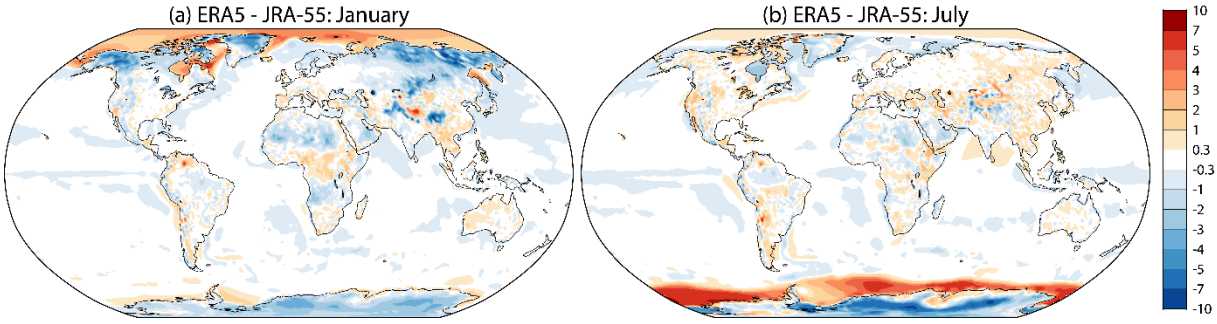

Figure 24 Difference in two-metre temperature analysis (°C) between ERA5 and JRA-55, averaged from 1979 to 2018 for (a) January and (b) July.

3.4 Temperature anomalies from multiple datasets

This section expands on the comparison of the monthly average surface air temperature anomalies from ERA5 with the values from other datasets, touched on already in the discussion of the upper two panels of Figure 1 in section 3.1.1.

3.4.1 Comparisons over land and sea

Figure 25 presents a breakdown of the contribution from four parts of the globe to the global-mean temperature time series shown in Figure 1(a). The split is into the land and (mostly ice-free) sea areas between latitudes 60°S and 60°N, and the polar regions south of 60°S and north of 60°N, where the averages are taken over both land and a mix of ice-covered and open sea. The breakdown is shown explicitly for ERA5, and the range of values provided by the six other datasets is also shown. Panel (b) includes a corresponding plot for the ERA5 SST between latitudes 60°S and 60°N.

It was shown in Figure 1 that temperature has risen faster on average over land than sea in recent decades. Sea covers a greater fraction of the Earth's surface, however. Figure 25 shows that about one half of the contribution to net global warming since the mid-1970s comes from the temperature rise over the tropical and mid-latitude oceans, and about one third comes from the rise over tropical and mid-

latitude land. Most of the remainder, around 15%, comes from the Arctic and near-Arctic latitudes north of 60°N, a region that comprises under 7% of the Earth's surface. There has been little net temperature change over the Antarctic during the past forty years, although a small net warming is seen there over the full seventy-year period covered by ERA5.

Panel (b) of Figure 25 shows that the trend in MAT is larger than the trend in SST for ERA5. This was first found for climate models by Cowtan *et al*. (2015) and identified as a source of discrepancy when the surface air temperatures from these models were compared with products such as GISTEMP and HadCRUT4 that blended anomalies in SST with anomalies in surface air temperature over land. It was discussed by Simmons *et al*. (2017) for ERA-Interim and JRA-55, for which the 1979-2018 least squares linear trends in global-mean temperature are respectively 2 and 4% smaller when SST rather than MAT is used. The discrepancy is only 1% for this period in the case of ERA5. Mean differences between SST and MAT anomalies relative to 1981-2010 are quite pronounced for ERA5 in the 1950s and 1960s, however, and the use of MAT rather than SST also contributes to ERA5 values being the highest for 2016-2020. The global-mean temperature trend for 1958-2020 is 7% smaller for ERA5 when SST rather than MAT is used. For JRA-55 the figure is 6%. Further assessment is needed to determine to what extent these differences represent a real climatological change, rather than a consequence of changes in observational quality and quantity.

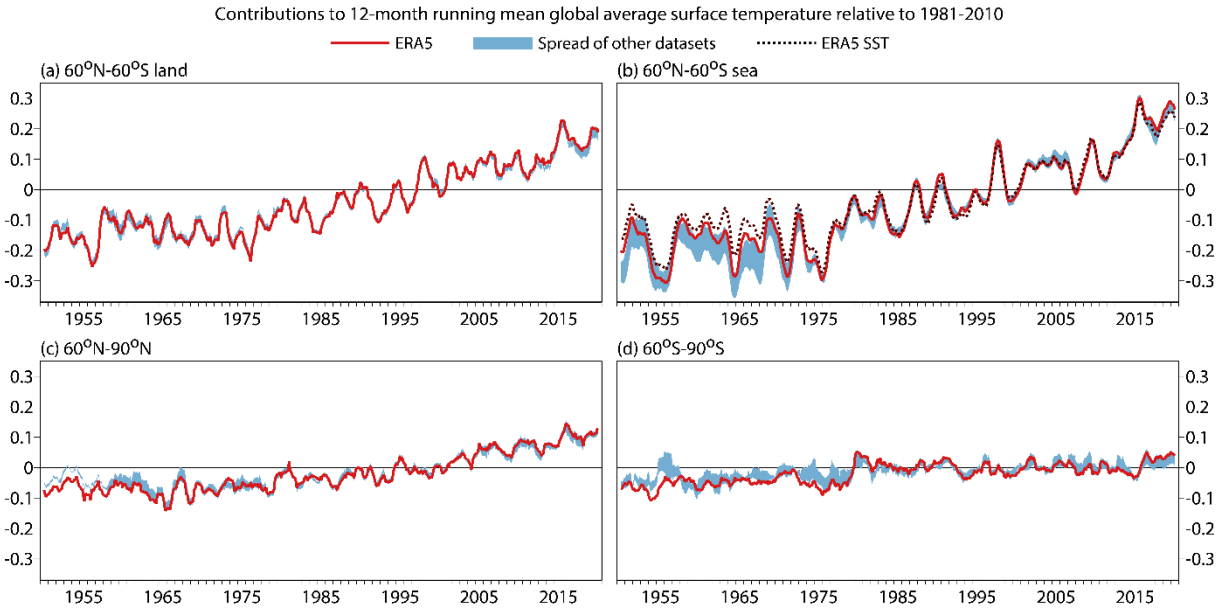

Figure 25 Contributions to twelve-month running mean global average surface air temperature anomalies (°C) relative to 1981-2010 for ERA5 (red, solid) from (a) land located between 60°S and 60°N, (b) sea located between 60°S and 60°N, (c) 60°N to 90°N, and (d) 60°S to 90°S, from 1950 to 2020. Blue shading denotes the range of corresponding values available from the six other datasets listed in the caption of Figure 1. The black dotted curve in (b) is for the ERA5 SST, using air temperature only where the surface is ice-covered.

The spread in data values is an indication of uncertainty. It is largest over sea in the 1950s and 1960s, but has also been surprisingly large over land for the past few years. More detail is provided in Figure 26, which shows, with an expanded scale, the differences between the various other datasets and ERA5. Differences are also shown for HadCRUT4 for regions and periods for which there are noteworthy differences with HadCRUT5.

It is evident from Figure 26(a) that the main contribution to the increased spread over tropical and mid-latitude land in recent years came from just one dataset, HadCRUT4, which is distinctly colder relative to 1981-2010 than all other datasets from 2015 onwards, and which has been fully superseded by HadCRUT5 as of January 2021. ERA5 has the highest values over land in this period, but its values are little different to those from ERA-Interim, GISTEMP and NOAAGlobalTemp. Berkeley Earth, HadCRUT5 and JRA-55 are colder, but broadly as close to ERA5 as they are to HadCRUT4 from 2017.

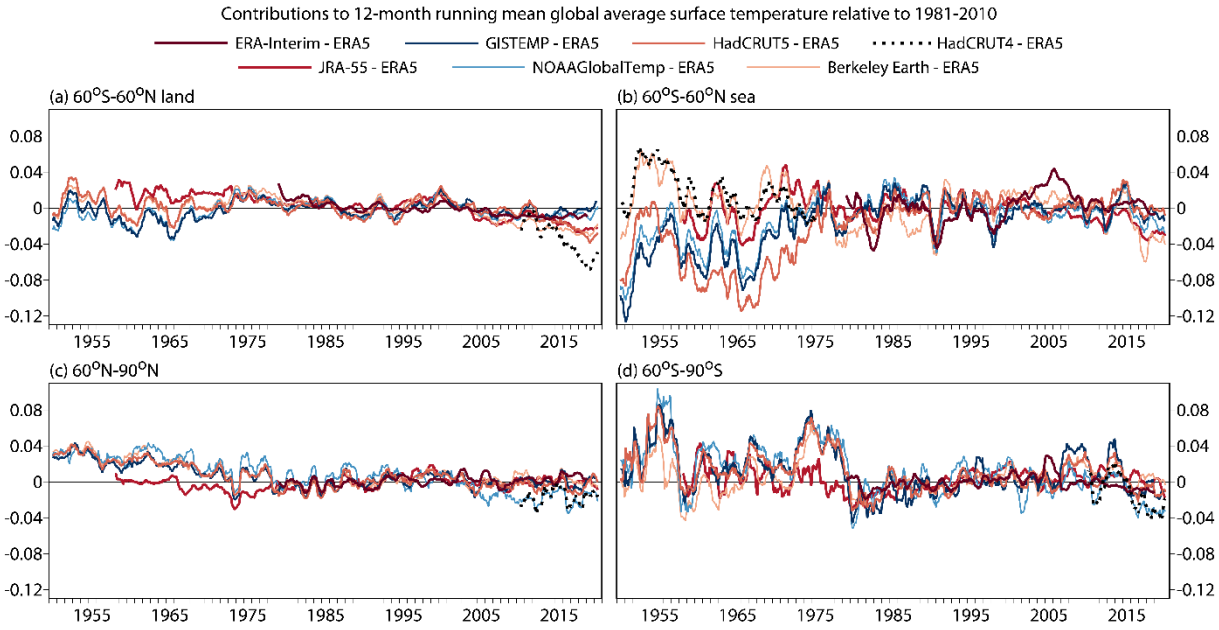

Figure 26 Differences between six datasets and ERA5 in contributions to twelve-month running mean global average temperature anomalies (°C) relative to 1981-2010, from 1950 to 2020. The six datasets are as in Figure 1 and indicated in the legend. Contributions are from the same four regions as in Figure 25. In addition, black dots denote values from HadCRUT4, plotted from 1950 to 1975 in (b) and 2010-2020 otherwise. These are the regions and periods where HadCRUT4 and HadCRUT5 differ most.

Relatively large differences over tropical and mid-latitude sea in the 1950s and 1960s are seen in Figure 26(b) to stem partly from the differences between Berkeley Earth and HadCRUT4 on one hand and GISTEMP and NOAAGlobalTemp on the other, reflecting the use of a common SST analysis by the first pair of datasets and a common but different SST analysis by the second pair. HadCRUT5 is the coldest of all, relative to ERA5, from the mid-1950s to the mid-1970s. Although MAT differences between ERA5 and JRA-55 fluctuate over time, they tend to lie within the range defined by the other pairs of datasets in the late 1950s and the 1960s. The JRA-55 values drift a little colder than ERA5 values after the mid-2000s, as illustrated also in Figure 19.

Further time series relating to differences in some of the SST analyses are presented in Figure 27. Averages from 60°S to 60°N of the analyses used in ERA5 (HadISST2/OSTIA) and JRA-55 (COBE) are compared in the upper two panels respectively with corresponding averages of the HadSST3 analyses used in Berkeley Earth and HadCRUT4 and the HadSST4 analyses used in HadCRUT5. The third panel compares the spreads of the ERA5 and HadSST4 ensembles.

Differences between the ERA5 and JRA-55 SSTs are in general smaller than the differences between HadSST3 and HadSST4, and closer to HadSST3 than to HadSST4. This is especially the case in the 1950s, 1960s and first half of the 1970s, when HadSST4 is distinctly colder (relative to 1981-2010) than the other datasets shown in Figure 27, though closer to the ERSSTv5 analyses used in GISTEMP and

NOAAGlobalTemp (Kennedy *et al*., 2019). During this period the spread of the HadSST4 ensemble barely overlaps, if at all, with the spread of the ERA5 ensemble. Moreover the ERA5 ensemble member selected to provide the SST for the primary ERA5 HRES data assimilation happens to be one of the warmest few ensemble members for much of the 1960s, which increases the difference between ERA5 and both HadSST4 and JRA-55 for these years.

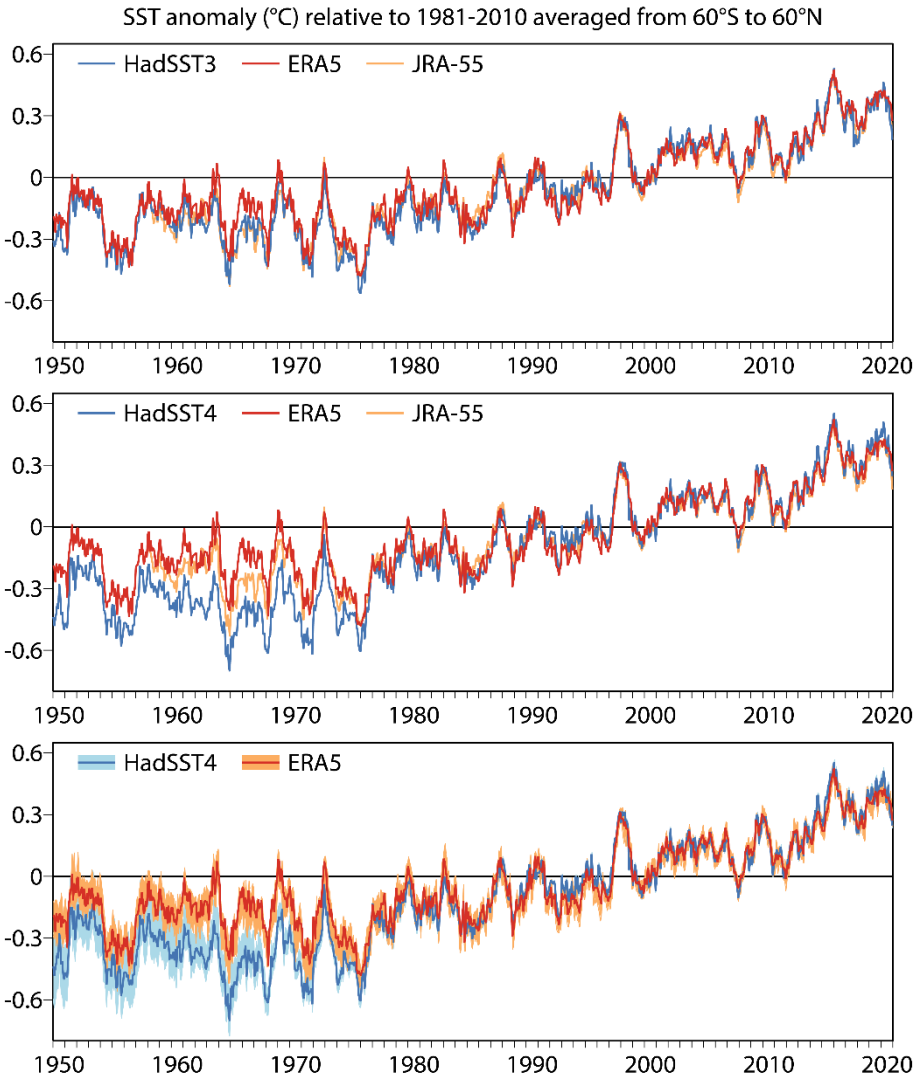

Figure 27 Sea-surface temperature anomalies relative to 1981-2010 averaged from 60°S to 60°N for (a) ERA5, JRA-55 and HadSST3, (b) ERA5, JRA-55 and HadSST4, (c) ERA5 and HadSST4 including ensemble spread, from 1950 to 2020. ERA5 and JRA-55 are sampled to have the same coverage as HadSST3 and HadSST4 respectively.

Another feature of Figure 26(b) is the relatively large difference between ERA-Interim and ERA5 (but not between JRA-55 and ERA5) in 2005 and 2006. It has already been shown in Figure 1 that the global-mean temperature anomaly is larger for ERA-Interim than for all other datasets for these years. There is, however, little difference between the SST analyses used by ERA-Interim and ERA5 for this period, as shown in Figure 19. It thus appears that the relationship between SST and MAT is anomalous for ERA-Interim for these years, although the reason for this has not been pinpointed.

Panel (c) of Figure 26 shows that ERA5 and JRA-55 are in good agreement back to the 1950s as regards temperature change at Arctic and near-Arctic latitudes. All datasets based on monthly climatological

data are relatively warm prior to the late 1970s compared to ERA5 and JRA-55, but Arctic warming in the 2000s is similar in all datasets other than NOAAGlobalTemp and HadCRUT4. This is understood to be associated with the limited representation of the Arctic in the latter two datasets (Cowtan and Way, 2014; Simmons and Poli, 2015). Simmons *et al*. (2017) noted that global and regional temperature averages from the ERA-Interim and JRA-55 reanalyses tend to agree better with averages from those monthly climatological datasets (now including HadCRUT5) that infill where they lack direct observations than with the averages from those datasets (including HadCRUT4) that do not use infilling.

Spread in values for the Antarctic is no surprise given the sparsity of long-term observational data coverage there, and the consequent differences in sampling of the region between the reanalyses and those datasets with little or no infilling. ERA5 and JRA-55 are in quite good agreement throughout, with all or most other datasets relatively warm for spells in the 1950s and the 1970s. Recent discrepancies include relatively high values from ERA-Interim in 2005 (as found also over tropical and mid-latitude sea), relatively high values from GISTEMP in two spells between 2007 and 2015, and (as for the Arctic) relatively low values from NOAAGlobalTemp and HadCRUT4 for the latest few years.

3.4.2 Comparisons over regions

Comparisons of averages over continental land regions have also been made. Figure 28 presents temperature time series for ERA5 and the spread of values from other datasets, as in Figure 25. The definitions of the regions are as before. Figure 28 also shows the average ERA5 analysis increment over each region. The corresponding differences between each of the other datasets and ERA5 are presented in Figure 29. These plots are complemented by the geographical maps of temperature trends that are presented in section 3.4.4.

Figure 28 shows that temperatures have risen over each region since the mid-1970s. The rise is largest over Europe, which is the second smallest region (after Australia) and the region centred the furthest north. ERA5 differs from all datasets other than JRA-55 in having relatively high temperatures over Australia in the 1950s and 1960s, as discussed further below. Agreement among the datasets is best for Europe from the late 1960s onwards, but there is a period in the mid-1960s when ERA5 is colder than all other datasets, by up to around 0.3ºC. There is a reduction in the analysis increment at the time, and it is likely that the substantial gap in data coverage over the west of the continent illustrated for 1965-1966 in Figure 7, combined with a cold bias of the background forecasts there, is responsible for this behaviour. More generally, the analysis increments are small for the observation-poor period from 1950 to 1957, especially for Asia, South America and Africa. Increments are reasonably uniform over time from the 1970s onwards, and generally such as to increase the background temperature, with the exception of Australia.

It can be seen from Figure 29 that ERA5 is mainly colder than the other datasets for Europe and North America in the 1950s and 1960s (relative to 1981-2010 levels), with the exception of GISTEMP for Europe. Aside from Australia, the largest differences are found for South America, where JRA-55 is briefly a warm outlier around the start of 1960, and datasets briefly separate again in 1961 with ERA5 the warmest and GISTEMP the coldest. JRA-55 and Berkeley Earth are similar to HadCRUT5 in giving colder recent temperatures than the other datasets for Africa and South America.

Another feature for Africa is a period in 1965 and 1966 in which ERA5 is distinctly warmer than all other datasets, by some 0.4ºC. This is a period with some pronounced country-specific gaps in observational coverage (Figure 7), but this spell is unusual in that ERA5 tends to be biased cold in such circumstances. In this case, however, there is also abnormally low precipitation in ERA5, something not seen in other datasets, as shown by Bell *et al*. (2021). Further discussion of this and two other instances of such behaviour, including South America in 1961, is given in section 4.7.1.

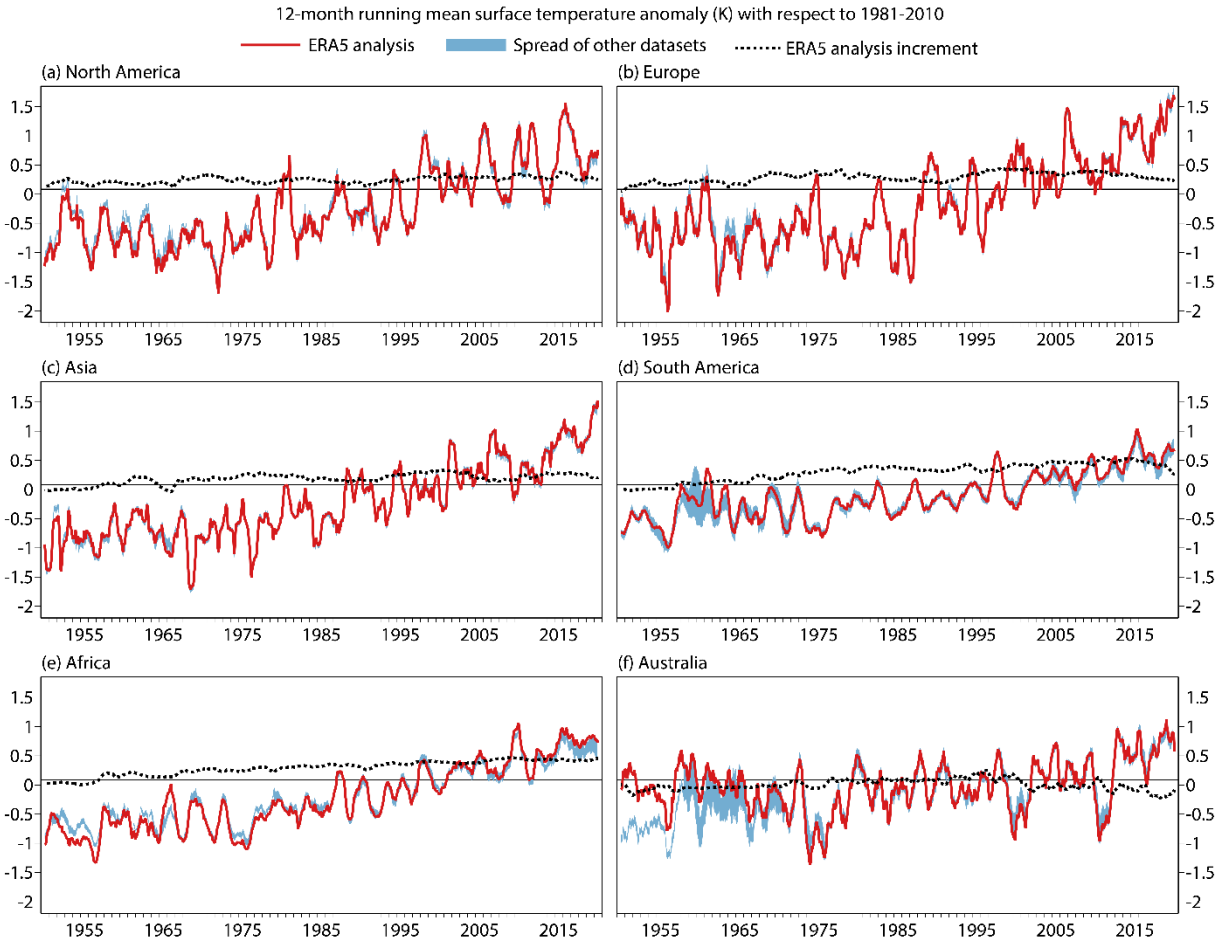

Figure 28 Time series of twelve-month running means of the ERA5 analysis (red, solid) expressed as anomalies relative to 1981-2010, and analysis increment (black, dotted) for two-metre temperature (ºC) averaged over land for (a) North America, (b) Europe, (c) Asia, (d) South America, (e) Africa and (f) Australia, for 1950 to 2020. The corresponding range of the temperature estimates available from the other six datasets listed in the caption of Figure 1 are shown by blue shading. Regions are as specified in the caption of Figure 3.

The results shown in Figure 29 are particularly interesting for Australia, as there is a clear difference in behaviour between the reanalyses and the datasets based directly on monthly climatological data. This is most evident in the 1950s and 60s, when ERA5 and (from 1958) JRA-55 are both quite substantially warmer (relative to 1981-2010) than the other datasets. The difference is over 1ºC for ERA5 in the early 1950s. In addition, ERA5, JRA-55 and ERA-Interim are each colder than the monthly datasets for periods in 2000/1, 2011/2 and 2016/7, and the same is true for ERA5 and JRA-55 in 1973/4 and 1976. Reference to Figure 28 shows that these are periods of relatively low temperatures for which the reanalyses tend to give more anomalous values than the other datasets.

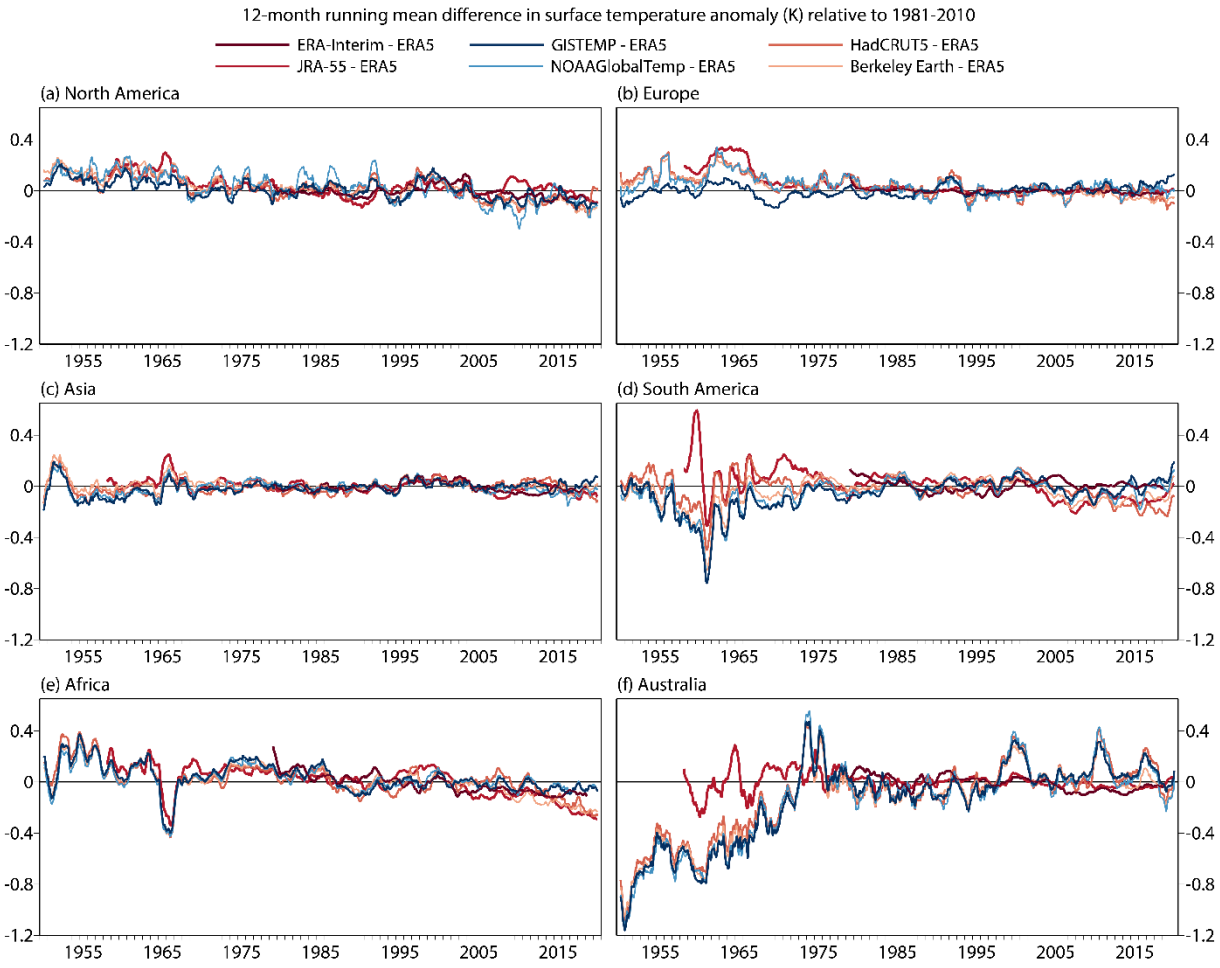

Figure 29 Differences between six datasets and ERA5 in twelve-month running mean temperature anomalies (ºC) relative to 1981-2010, averaged over land for (a) North America, (b) Europe, (c) Asia, (d) South America, (e) Africa and (f) Australia, for 1950 to 2020. The six datasets are those listed in the caption of Figure 1, as indicated in the legend. Regions are as specified in the caption of Figure 3.

Figure 30 provides further information for Europe and Australia. Temperature comparisons are extended to include regional datasets: E-OBS for Europe (Cornes *et al*.,2018; here version 21 at 0.25º resolution) and ACORN-SAT for Australia (Trewin, 2013; here country-average version-2 values, based on homogenised observational data from 112 stations). These datasets provide absolute values, not the anomalies relative to (differing) reference periods that are provided by the global datasets based on monthly climatological data.

ERA5 is in close absolute agreement with E-OBS for these area-averaged temperatures; this is demonstrated in Figure 30(a) by using the ERA5 climate for 1981-2010 as the reference for defining both the ERA5 and the E-OBS anomalies. There is little sensitivity of the anomalies to missing values in some of the datasets. The ERA5 values are sampled to match the data coverage of E-OBS in Figure 30(a), but these values are evidently close to those derived from the various other datasets, the spread of which is included in the plot. The 1981-2010 average temperature is reduced from 8.53 to 8.09ºC by the sampling.

It is clear from Figure 30(b) that GISTEMP is in very good agreement with the ACORN-SAT temperature record over the whole period from 1950 onwards. Reference back to Figure 29(f) shows somewhat poorer agreement prior to the 1970s with the other datasets based on monthly climatological

station data, with HadCRUT5 a little warmer than GISTEMP, but the main discrepancy is again with ERA5 and JRA-55. This appears not to be directly due to the sparse observations of surface air temperature analysed by ERA5 and JRA-55, nor to the way the analysis is carried out, as the background forecasts from the two reanalyses are in poorer agreement with the other datasets than the analyses are: such surface air observations as are analysed bring the reanalyses somewhat closer to the other datasets, though by no means close enough. This can be seen for ERA5 from the analysis increments shown in Figure 28(f), which are slightly negative in the early years.

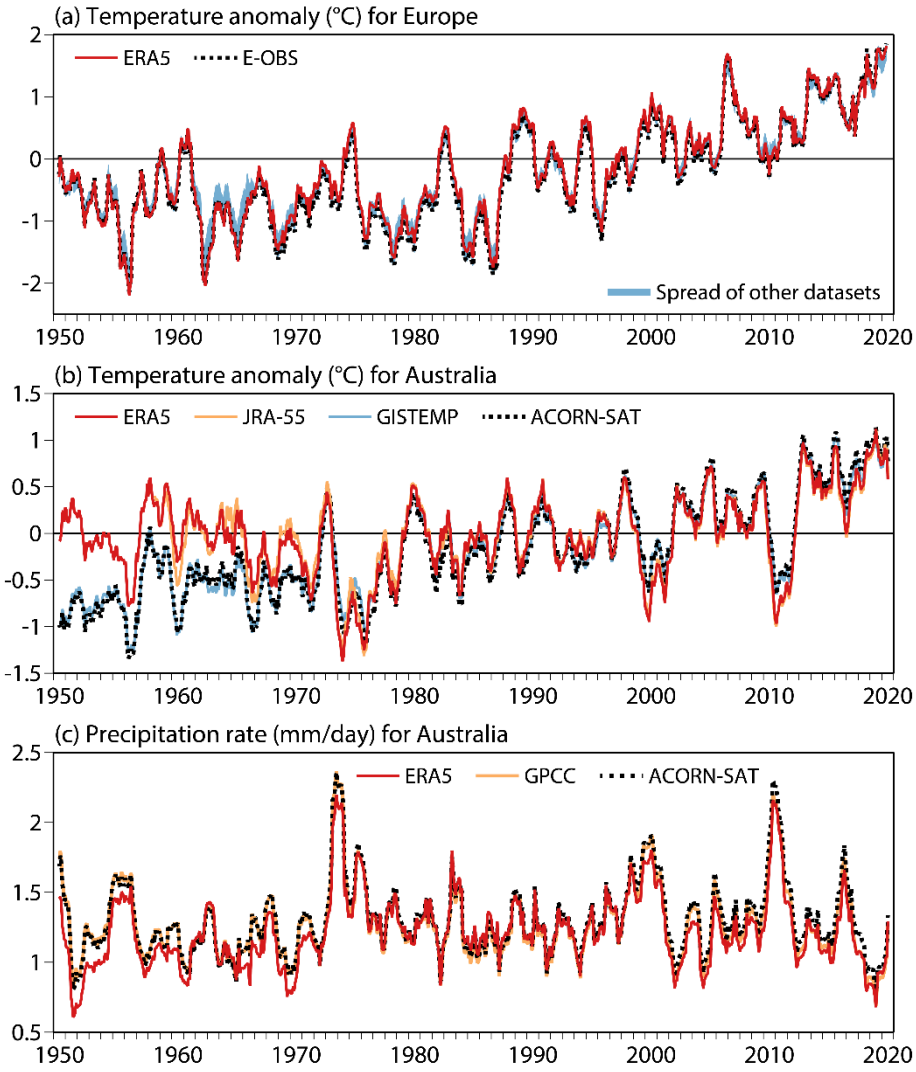

Figure 30 Time series from 1950 to 2020 of twelve-month running means: two-metre temperature anomalies (°C) relative to 1981-2010 for (a) Europe from ERA5 (red, solid) and E-OBS (version 21.0e; black dotted), and the spread (blue shading) of these and other datasets, and (b) Australia from ERA5 (red, solid), JRA-55 (orange, solid), GISTEMP (blue, solid) and ACORN-SAT (version 2; black, dotted); (c) precipitation rate (mm/day) for Australia from ERA5 (red solid), GPCC (orange solid) and ACORN-SAT (black dotted). In (a), ERA5 data are sampled with the same coverage as E-OBS, and the ERA5 climate for 1981-2010 is used to calculate the E-OBS as well as the ERA5 anomalies.

Panel (c) of Figure 30 shows time series of precipitation rate averaged over Australia, comparing ERA5 with the independent gauge-based estimates from ACORN-SAT and the Global Precipitation Climatology Centre (GPCC; Becker *et al.*, 2013). ERA5 precipitation agrees better with GPCC values for Australia than for any of the other continental regions studied here (Bell *et al.*, 2021). GPCC and

ACORN-SAT are in close agreement from 1950 onwards, differing only in the most recent decade when GPCC gives slightly lower values, closer to those from ERA5. ERA5 agrees with both other datasets for earlier decades back to the 1970s, but gives systematically lower precipitation in the 1950s and 1960s. This suggests a change over time in the performance of the upper-air analysis scheme. Analysis increments over Australia tend to cool the lower troposphere, but less so in the 1950s and 1960s than in later years. They also tend to increase the specific humidity of the lower troposphere from the mid-1960s to the 1990s, but not earlier (and less so later). This is consistent with lower precipitation and too-high surface air temperatures in the 1950s and 1960s compared with at least the next three decades. Soil moisture and cloud cover are also found to be lower in the 1950s and 1960s than in the following decades.

3.4.3 Calculation of daily-mean temperature, and the diurnal temperature range

It has already been noted that the reanalyses are similar in temperature to each other but colder than the other datasets over Australia for 1973/4, 1976, 2000/1, 2011/2 and 2016/7, periods when temperatures are relatively low in all datasets. Figure 30 shows that these are also periods of above-average precipitation, which tend to occur in conjunction with La Niña events (http://www.bom.gov.au/climate/about). One reason for the temperature differences between datasets for these periods is that the values from the reanalyses are averages over synoptic times, whereas the other datasets are based primarily on monthly records of the average of daily maximum and minimum temperatures. As discussed below, the difference between the two types of average increases at times of plentiful rainfall. Biases in the background forecasts of the reanalyses may also change in such circumstances, although there is little evidence for this from the time series of analysis increments in Figure 28(f).

Rennie *et al*. (2014) present an example of different ways of calculating daily mean temperature using the observations from a single European station, for a winter month. Results differed by up to around 1ºC. The monthly averages of the reanalyses used in this report are calculated by averaging over a set of fixed hours and all days of the month. Analyses for every hour of the day are used for ERA5. ERA-Interim and JRA-55 use their six-hourly analyses for 00, 06, 12 and 18UTC. Conversely, the monthly average ACORN-SAT values are arithmetic means of the monthly averages of daily maximum and minimum temperatures. The monthly global datasets are also based primarily on monthly averages of the means of daily maximum and minimum temperatures, with a small number of days of missing data allowed. E-OBS uses data produced using a mix of national practices, with preference given to mean temperature derived using multiple (hourly) observations if available, as specified in online metadata (Else van den Beselaar, personal communication). For example, the daily values for Swedish stations are calculated as a blend of the maximum and minimum temperatures for the 24 hours from 18UTC the previous day, and the temperatures observed at 06, 12 and 18UTC on the day in question (Ma and Guttorp, 2013).

The 1981-2010 averages over Europe are 8.09ºC for ERA5 (sampled as E-OBS) and 7.98 ºC for E-OBS. The complete averages for Europe are 8.53ºC for ERA5, 8.71ºC for ERA-Interim and 8.56ºC for JRA-55. The averages over Australia are 22.01ºC for ERA5, 22.14ºC for ERA-Interim, 21.99ºC for JRA-55 and 22.15 ºC for ACORN-SAT. Using ERA5 values only for 00, 06, 12 and 18h rather than all 24 hourly values makes a small difference: the averages are 8.49ºC for Europe (without sampling) and 22.13ºC for Australia.

Calculations for ERA5 have been repeated using the mean of its maximum and minimum hourly temperature analyses each day, rather than the mean of each of the 24 hourly values. This reduces the 1981-2010 temperature average for Europe by 0.09ºC. Conversely, it increases the average for Australia by 0.21ºC. Using the mean of the maximum and minimum hourly temperatures as an approximation of

the daily average temperature thus improves the agreement between ERA5 and both E-OBS and ACORN-SAT.

The amplitude and shape of the diurnal cycle change depending on factors that include soil wetness. Accordingly, there are variations over time in the extent to which the true daily average temperature differs from the average of the maximum and minimum daily temperatures. This is seen for Australia in particular. For ERA5, the average of its 24 hourly temperatures is colder than the average of its maximum and minimum hourly temperatures by a 1981-2010 average value of 0.21ºC, as noted above, but the annual-mean differences vary from 0.30ºC and 0.31ºC respectively for the years 2000 and 2011, when precipitation was high and temperatures low, to 0.11ºC for the calendar year 2019, a year with exceptionally low precipitation and high temperature.

The diurnal temperature range (DTR) over the continental areas examined in section 3.4.2 is largest for Australia and smallest for Europe, averaging 12.0ºC over Australia and 7.9ºC over Europe for 1981-2010, based again on ERA5's hourly analyses. The corresponding averages are 13.5ºC for ACORN-SAT and 8.8ºC for E-OBS. DTR is expected to be higher for the observed values, as the hourly resolution of ERA5 limits the range. There may also be issues relating to the temperature analysis for Australia (as discussed earlier) and to the fidelity of the background model's representation of the diurnal cycle.

Figure 31 shows twelve-month running averages of the anomalies in DTR relative to 1981-2010, for Europe from ERA5 and E-OBS and Australia from ERA5 and ACORN-SAT. Notwithstanding the long-term differences in values, ERA5 evidently captures the interannual variations in DTR in good agreement with the datasets derived more directly from the station data. This is seen most clearly for Europe over the 1981-2010 reference period, and for the much larger variations over Australia from the late 1980s onwards. Notable in the latter case are the relatively low values of DTR at the times of low mean daily temperature and high precipitation discussed earlier.

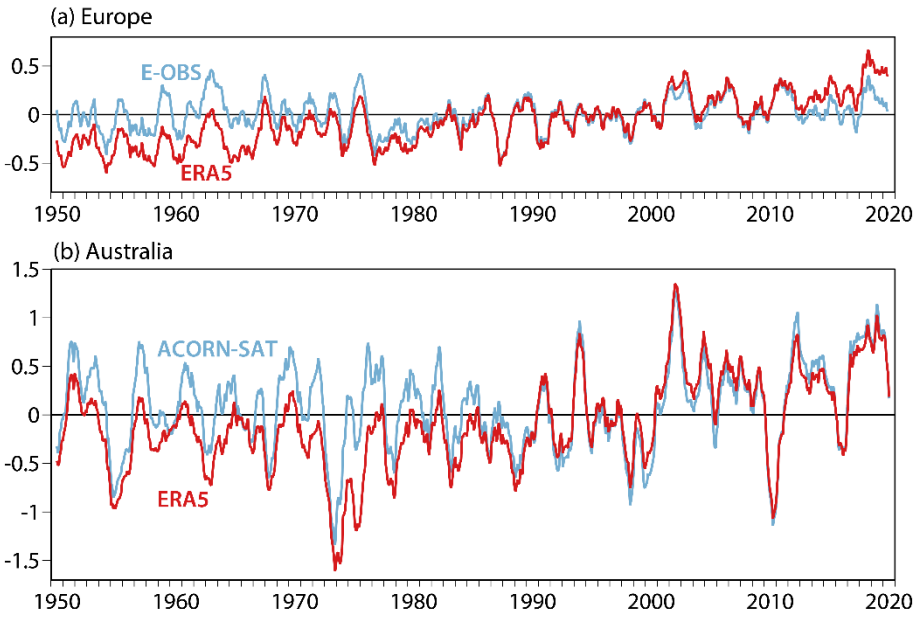

Figure 31 Time series from 1950 to 2020 of twelve-month running means of anomalies in the diurnal range of two-metre temperature (ºC) relative to 1981-2010 for (a) Europe from ERA5 (red) and E-OBS (blue), and (b) Australia from ERA5 (red) and ACORN-SAT (blue). In (a), ERA5 data are sampled with the same coverage as E-OBS.

The larger long-term increase in DTR for ERA5 compared with both E-OBS and ACORN-SAT may well stem from changes in the times and numbers of observations available to ERA5. The available synoptic observations for Europe were predominantly six-hourly in the 1950s and three-hourly in the 1960s and 1970s. The next forty years saw an increasing proportion of hourly observations, while the number of observations for each of the main synoptic hours increased by an order of magnitude. The situation for Australia has been discussed in section 3.1.5, where low data counts prior to 1977 are noted and a change in observation characteristics from the late 1980s to the early 1990s is discussed.

The discrepancy between ERA5 and E-OBS in DTR for Europe from 2013 onwards has yet to be well understood. ERA5 and E-OBS agree well as regards the recent trend in daily maximum temperature, but ERA5 has a smaller recent trend in daily minimum temperature. Relative to 1981-2010, the minimum temperature is higher in ERA5 than in E-OBS for the 1950s and 1960s. The converse is the case for maximum temperature. The ERA5 background forecasts, like the E-OBS analyses, do not exhibit the long-term trend in DTR over Europe shown for the ERA5 analyses, but do not match the variations in DTR from E-OBS as well as the analyses do from the late 1960s to the early 2010s.

3.4.4 Regional variations in warming trends

A more detailed depiction of geographical variations is provided by maps of the local least squares fit linear trends in temperature. These are presented in Figure 32 for the period from 1979 to 2018 that is common to all datasets studied. Results are shown for the background forecasts and analyses for ERA-Interim and ERA5, for the JRA-55 analyses and for the Berkeley Earth, GISTEMP, HadCRUT5 and NOAAGlobalTemp datasets. As a strict requirement of computing trends only for locations with complete data records is adopted, even small amounts of missing input data results in missing data on trends, which affects NOAAGlobalTemp and to a lesser extent HadCRUT5 in particular. Accordingly, the trends from the infilled Had4krig dataset based on HadCRUT4 are also presented. Specific issues affecting both the ERA-Interim and the ERA5 trends south of the Caspian Sea, and the ERA5 trend north of Greenland, are discussed separately in sections 3.5.1.1 and 3.5.4 respectively.

The datasets presented in Figure 32 are in good overall agreement. All have a maximum warming rate over the Arctic, and of the datasets that provide complete or near-complete coverage of the region, all indicate maximum warming in a band eastward from Svalbard over the Barents and Kara Seas, even though this is not evident for GISTEMP with the contour intervals used. All datasets also indicate relatively large warming over central and eastern Europe, the Middle East, north-eastern Africa and the southwestern USA and northern Mexico, albeit with local differences that in part reflect the differences in resolution of the datasets.

Another common feature that can be inferred from the maps is the lower average warming rate over sea than over land discussed earlier. All datasets show a cooling trend over the eastern South Pacific and either cooling or only weak warming trends over the Southern Ocean. Here Berkeley Earth and GISTEMP show a fair degree of similarity with the reanalyses. Had4krig is in poorer agreement. Although there are some commonalties over Antarctica itself, there are also many differences.

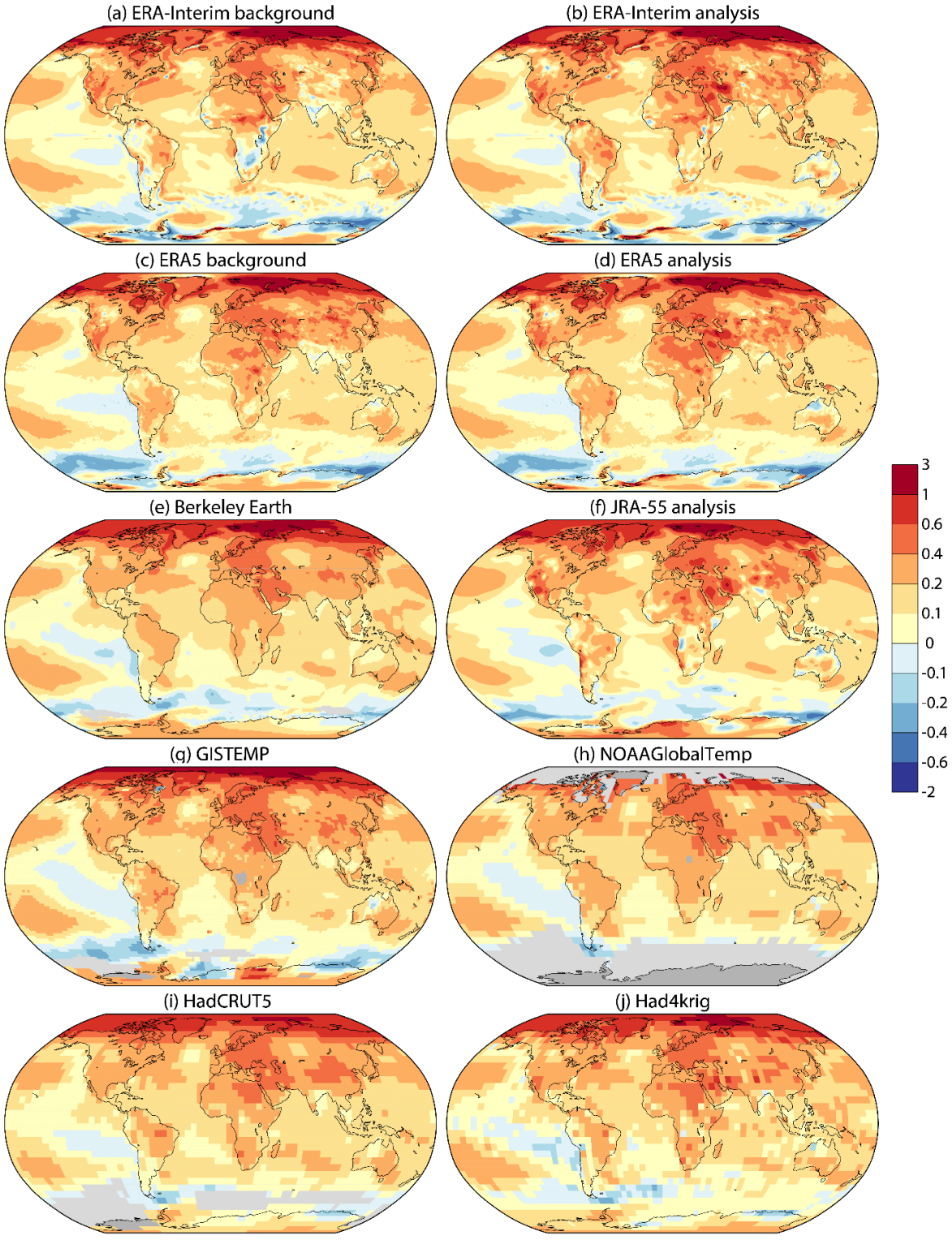

Figure 32 Linear trend in two-metre temperature (ºC/decade) from 1979 to 2018 for: (a) and (c) the ERA-Interim and ERA5 background forecasts; (b), (d) and (f) the ERA-Interim, ERA5 and JRA-55 analyses; (e) and (g) to (j) Berkeley Earth, GISTEMP, NOAAGlobalTemp, HadCRUT5 and Had4krig. All calculations use monthly-mean data with the mean annual cycle for 1979-2018 removed.

3.4.5 Comparisons of monthly values

Attention up to now has been concentrated on comparing behaviour on timescales from a year or so upwards. The relatively large variations that occur from month to month have been largely filtered out by use of twelve-month averaging.

Scatter plots comparing ERA5 analyses with the ERA5 background and the various other datasets are presented in Figure 33 for monthly global-mean temperature anomalies and in Figure 34 for monthly European-mean anomalies. The period covered is from January 1979 to August 2019, for which all datasets provide values. The clustering of points close to the diagonal indicates good general agreement between the datasets. The 488-month period is divided into eight consecutive periods of 54-month duration and a final one of 56 months duration, and the colour assigned to a particular month depends on into which of the nine periods the month falls. The tendency for deeper blue colours lower on the diagonal and deeper red colours higher on the diagonal indicates the warming of the atmosphere over time, and the spread of points with a particular colour indicates the amount by which values spread over a period of four to five years. The root-mean-square (rms) differences over the full period between ERA5 and each other dataset are shown.

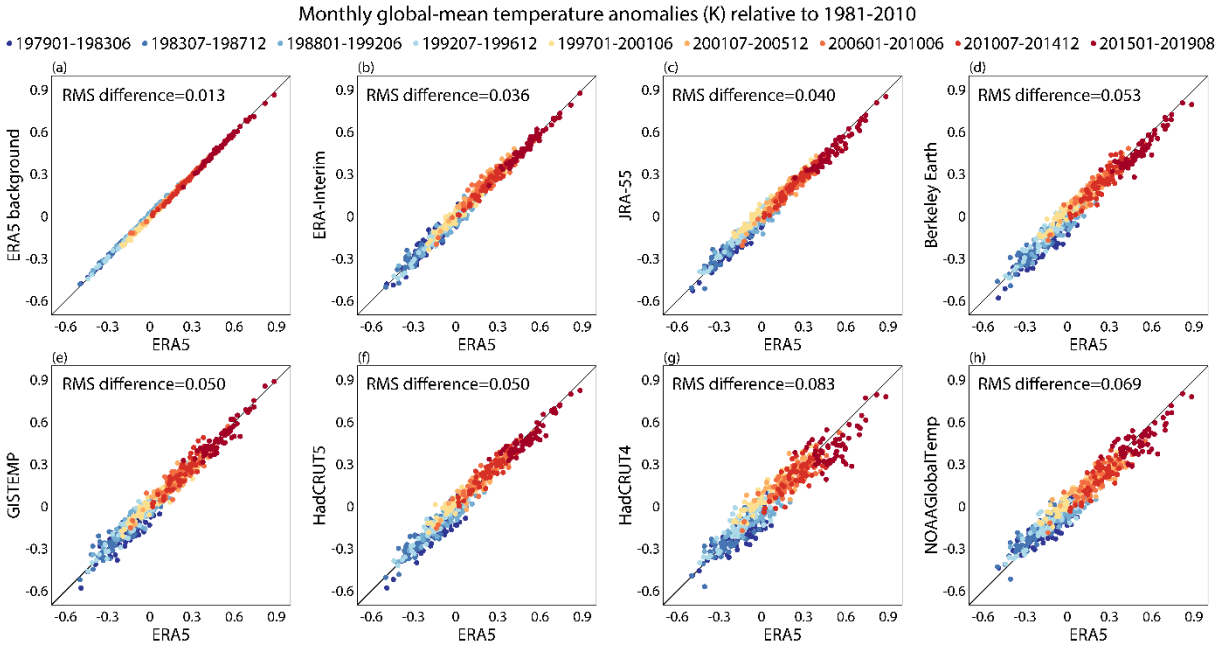

Figure 33 Monthly global-average two-metre temperature (ºC) anomalies relative to 1981-2010 for the period from January 1979 to August 2019, for the ERA5 analysis compared with values from (a) the ERA5 background, (b) the ERA-Interim analysis, (c) the JRA-55 analysis, (d) Berkeley Earth, (e) GISTEMP, (f) HadCRUT5, (g) HadCRUT4 and (h) NOAAGlobalTemp. Each colour represents values from one of the nine sub-periods shown in the legend, with the deepest blue indicating the earliest period and the deepest red the latest period. Early periods are plotted before later periods, so redder dots are seen where there is overlap.

It is not surprising that the ERA5 analysis is closest to its own background values globally, because analysis and background are by design the same over sea. This remains so, however, if comparisons are made for values averaged over all land (not illustrated), although in this case the rms difference of 0.043ºC is not much smaller than the rms difference of 0.054ºC between ERA5 and ERA-Interim.

Otherwise, for the full global averages the monthly ERA5 analyses are only a little closer to ERA-Interim than they are to JRA-55. Of the other datasets, GISTEMP and HadCRUT5 are the closest to

ERA5 and HadCRUT4 the furthest. As found by Simmons *et al*. (2017) for ERA-Interim, ERA5 agrees better with those datasets that provide near-global or global coverage than with those that use little or no infilling, but its agreement with the Had4krig spatial extension of HadCRUT4 (not included in Figure 33) is not as close as its agreement with HadCRUT5; root-mean-square differences are 0.05ºC for HadCRUT5, 0.061ºC for Had4krig and 0.083 ºC for HadCRUT4. The only change to the ordering when averages are restricted to land only is that ERA5 is then slightly closer to Berkeley Earth than it is to JRA-55 and GISTEMP.

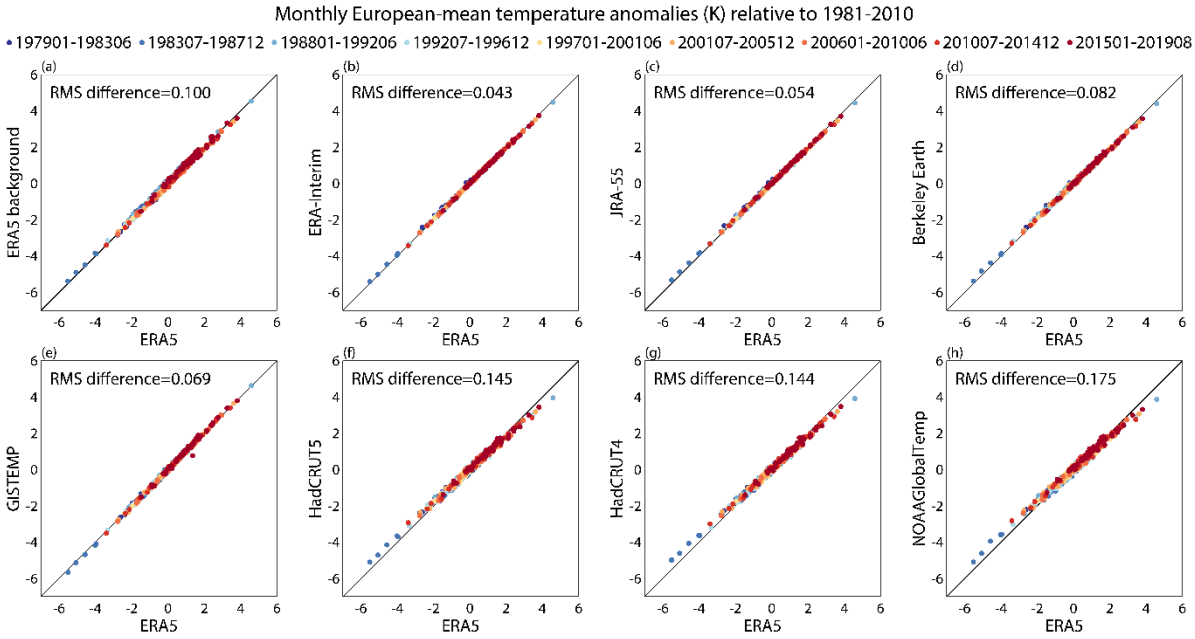

Figure 34 As Figure 33, but for European-average temperatures.

The scatter plots for European averages can be seen in Figure 34 to be more tightly clustered along the diagonal than the global values in Figure 25. However, the month-to-month variability of values is much higher for the relatively small European area, and rms differences for Europe are generally a little larger than global values in absolute terms. In this case the ERA5 analyses are more different from the ERA5 background values than they are from the other reanalyses and from Berkeley Earth and GISTEMP. HadCRUT5, HadCRUT4 and NOAAGlobalTemp stand out as more different, both in terms of their larger rms differences from ERA5 and their tendency for slightly smaller extremes, both warm and cold. This may be a consequence of their lower 5ºx5º geographical resolution. Here HadCRUT5 is marginally further from ERA5 than HadCRUT4 is.

Comparisons for 1950-2019 are shown for both global and European averages in Figure 35. Here ERA5 analyses are compared with Berkeley Earth, GISTEMP, HadCRUT5 and NOAAGlobalTemp. There is no change in the rankings of these datasets compared with ERA5 from those for the 1979-2019 period, with smallest differences for GISTEMP and largest differences for NOAAGlobalTemp. Globally, spread for the earlier years is larger for HadCRUT5 and Berkeley Earth than it is for GISTEMP and NOAAGlobalTemp. For Europe, the fit of ERA5 to E-OBS (not illustrated) is comparable with the fit of ERA5 to GISTEMP, provided grid squares with missing data in E-OBS are taken into account by sampling ERA5 so that it has the same spatial coverage each month as E-OBS.

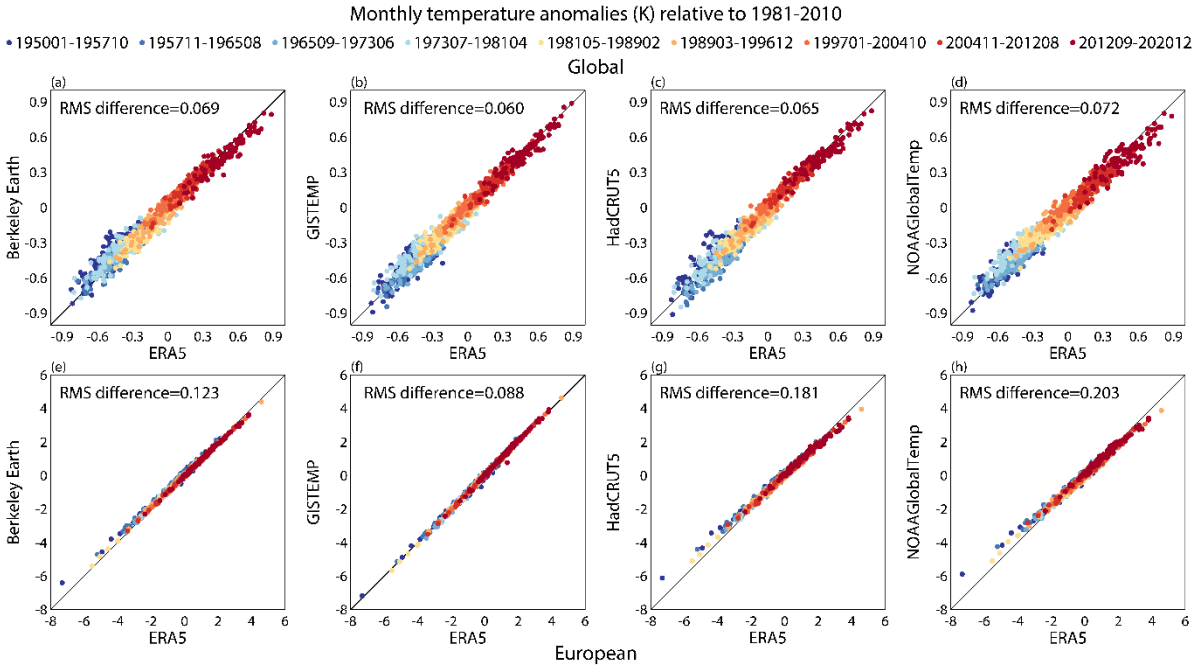

Figure 35 As Figure 33, but for the period from 1950 to 2020 for global-average (upper row) and European-average (lower row) anomalies, comparing ERA5 with (a, e) Berkeley Earth, (b, f) GISTEMP, (c, g) HadCRUT5 and (d, h) NOAAGlobalTemp.

Geographical variations in the agreement between monthly values are illustrated in Figure 36, which shows maps of correlations for 1979-2018 between the reanalyses and the HadCRUT5 and GISTEMP datasets, and in Figure 37, which shows maps of the correlations of ERA5 with HadCRUT5 and GISTEMP for 1950-2018. Correlations are computed for each of the (5ºx5º) HadCRUT5 and (2ºx2º) GISTEMP grid boxes for which a complete time series of values is provided, using reanalysis values averaged over the grid-box concerned. Corresponding correlations derived for the continental land regions specified earlier are presented in Table 2, which also includes correlations between ERA5 and Had4krig.

Considering values over land first, correlations are highest over Europe and western Asia, China and central and eastern North America. They are lowest over the tropical zone, especially over central Africa. Correlations over South America, Africa and Australia are higher overall for ERA5 than ERA-Interim and JRA-55. Correlations between ERA5 and HadCRUT5 for these regions are higher than those between ERA5 and GISTEMP, and higher still than those between ERA5 and Had4krig. Differences are smaller and more mixed for the higher-correlation regions of the extratropical northern hemisphere. Relatively high correlations are also found over the Antarctic Peninsula and around the coast of East Antarctica, where all datasets are constrained by the observations made there over the past few decades.

Correlations for 1950-2018 are generally lower than those for 1979-2018. This is particularly so for South America, Africa and Australia, and to a lesser extent for Asia, a result consistent with ERA5's poorer coverage of observational data over these regions in the 1950s in particular, the 1960s and to some extent the 1970s.

Correlations are generally lower over sea than land. Here there are additional differences between HadCRUT5 and GISTEMP on the one hand and the reanalyses on the other that arise from differences between SST and MAT anomalies. There is reasonably good agreement between the correlations of the reanalyses with HadCRUT5 and GISTEMP over northern extratropical seas, and the reanalyses correlate quite highly with both GISTEMP and HadCRUT5 over the tropical eastern Pacific Ocean and southern

subtropics. Correlations are generally low over the tropical western Pacific and over and immediately to the north of the Southern Ocean.

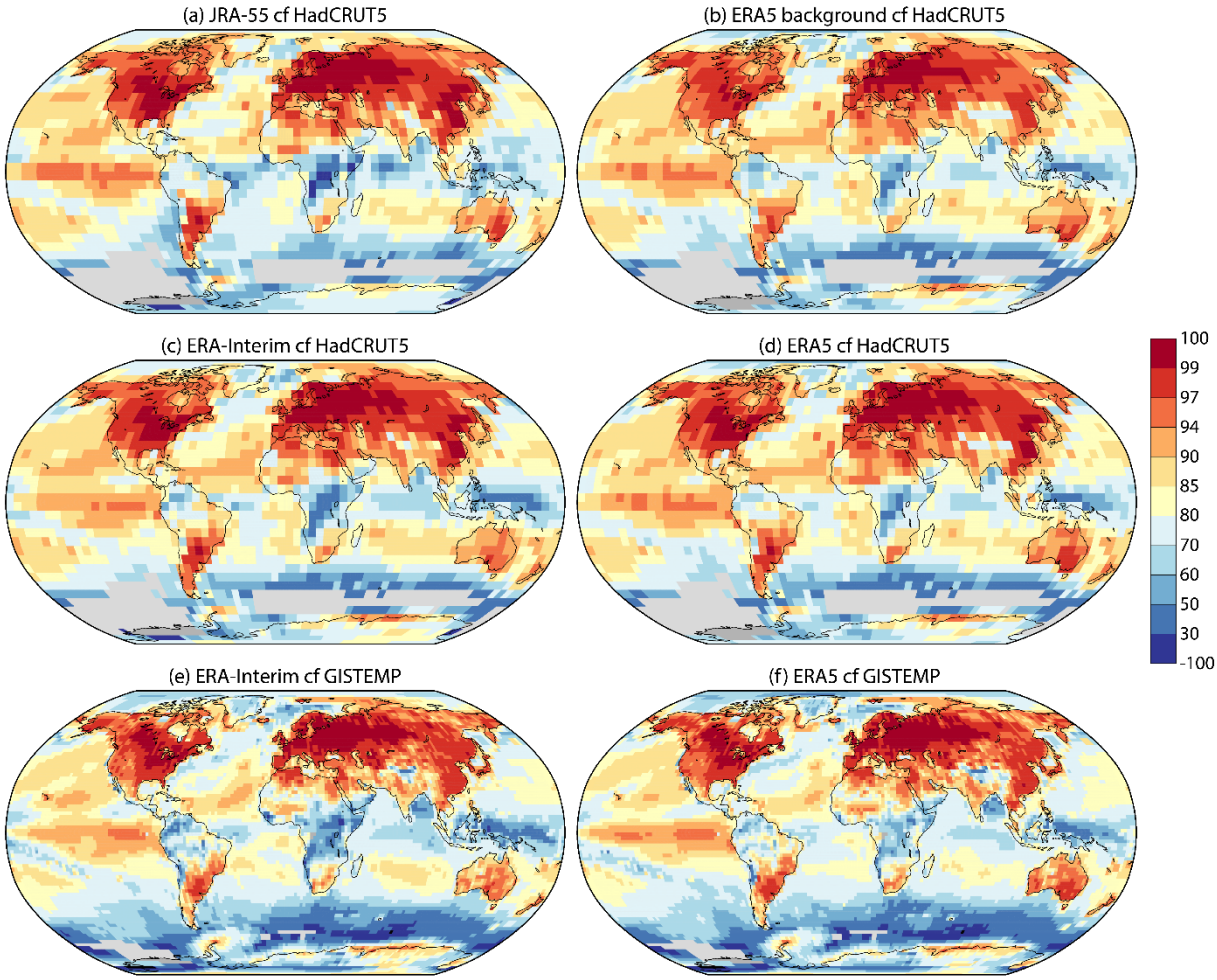

Figure 36 Correlation (%) of monthly average temperatures for 1979-2018 between HadCRUT5 and (a) JRA-55, (b) the ERA5 background, (c) ERA-Interim and (d) ERA5, and between GISTEMP and (e) ERA-Interim and (f) ERA5. The mean annual cycle is removed from each dataset prior to calculating the correlations.

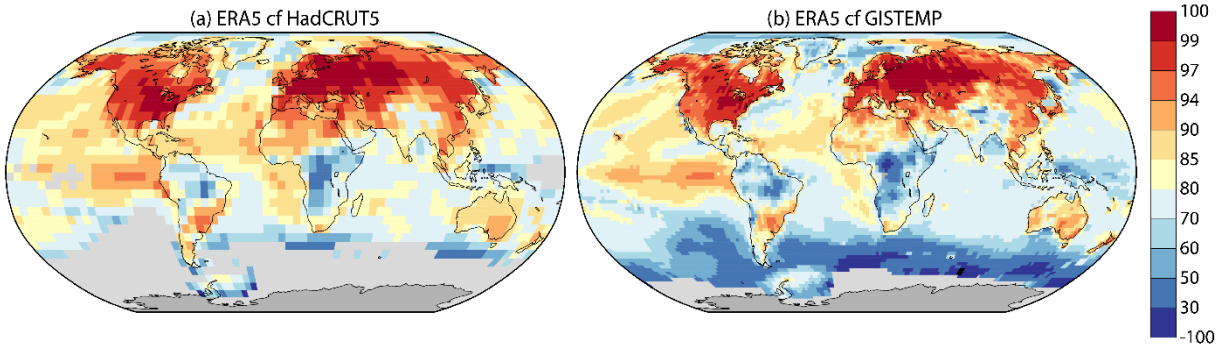

Figure 37 As Figure 36, but for the correlations between ERA5 and (a) HadCRUT5 and (b) GISTEMP for 1950-2018

	Datasets compared		North America	Europe	Asia	South America	Africa	Australia
1979-2018	HadCRUT5	ERA-Int	95.9	98.3	96.8	88.1	83.2	92.7
		JRA-55	96.4	98.1	96.7	87.1	83.3	92.8
		ERA5 bg	95.6	98.0	95.6	86.8	85.5	92.5
		ERA5	96.1	98.3	96.6	89.6	86.7	94.1
	Had4krig	ERA5	94.9	98.1	95.5	81.9	82.5	92.0
	GISTEMP	ERA-Int	94.5	98.1	95.6	81.7	79.5	91.4
	GISTEMP	ERA5	94.9	98.1	95.2	83.9	83.4	93.1
1950-2018	HadCRUT5	ERA5	95.7	98.0	95.6	79.5	81.0	88.4
	Had4krig	ERA5	94.5	97.7	94.4	72.7	76.3	87.3
	GISTEMP	ERA5	94.3	97.4	94.0	75.0	77.6	86.9

Table 2 Correlations (%) over the continental regions specified in the caption to Figure 28 between the datasets and for the periods for which correlation maps are presented in Figure 36 *and Figure* 37*, plus correlations between ERA5 and Had4krig.*

3.4.6 Examples of monthly extremes

Monthly temperature anomalies typically have maxima of several ºC, and maxima can reach ten or more ºC at high latitudes. The representation of such extremes in reanalyses is little affected by the much smaller biases that can mar estimates of trends. It can nevertheless still be sensitive to observational coverage. Figure 38 presents the first of three examples. It shows in panels (a) to (d) maps of anomalies in surface temperature for February 1956, chosen as it is from the early part of the reanalysis period and is the month with record cold average conditions over Europe, relative to the climatological normal for the month; it is represented by the lowest point on the diagonal in Figure 35. Maps are for GISTEMP, HadCRUT5 and the ERA5 background and analysis. Panels (e) and (f) show the mean background and analysis departures for ERA5.

The maps of anomalies are in good agreement where each analysis has access to observations. All depict the cold conditions over Europe, peaking over south-eastern Russia. Although not illustrated here, ERA5 shows this to be associated with anomalous easterly flow, with anticyclonic anomalies centred near Iceland and Novaya Zemlya, and relatively low pressure centred near Italy. Conditions are colder than normal to the east over Siberia, and it is also relatively cold over western North America. Consistent with the anomalous flow pattern, temperatures are higher than normal for February over the far north of Siberia and over the seas to the north and west. Positive anomalies also occur over eastern North America, Greenland, north-eastern Africa and the Middle East. Australia is relatively warm in the south and west, and cool in the north and east. The analyses differ in the magnitude of the warm conditions over the North Pacific and North Atlantic. The ERA5 background forecasts give a similar overall picture.

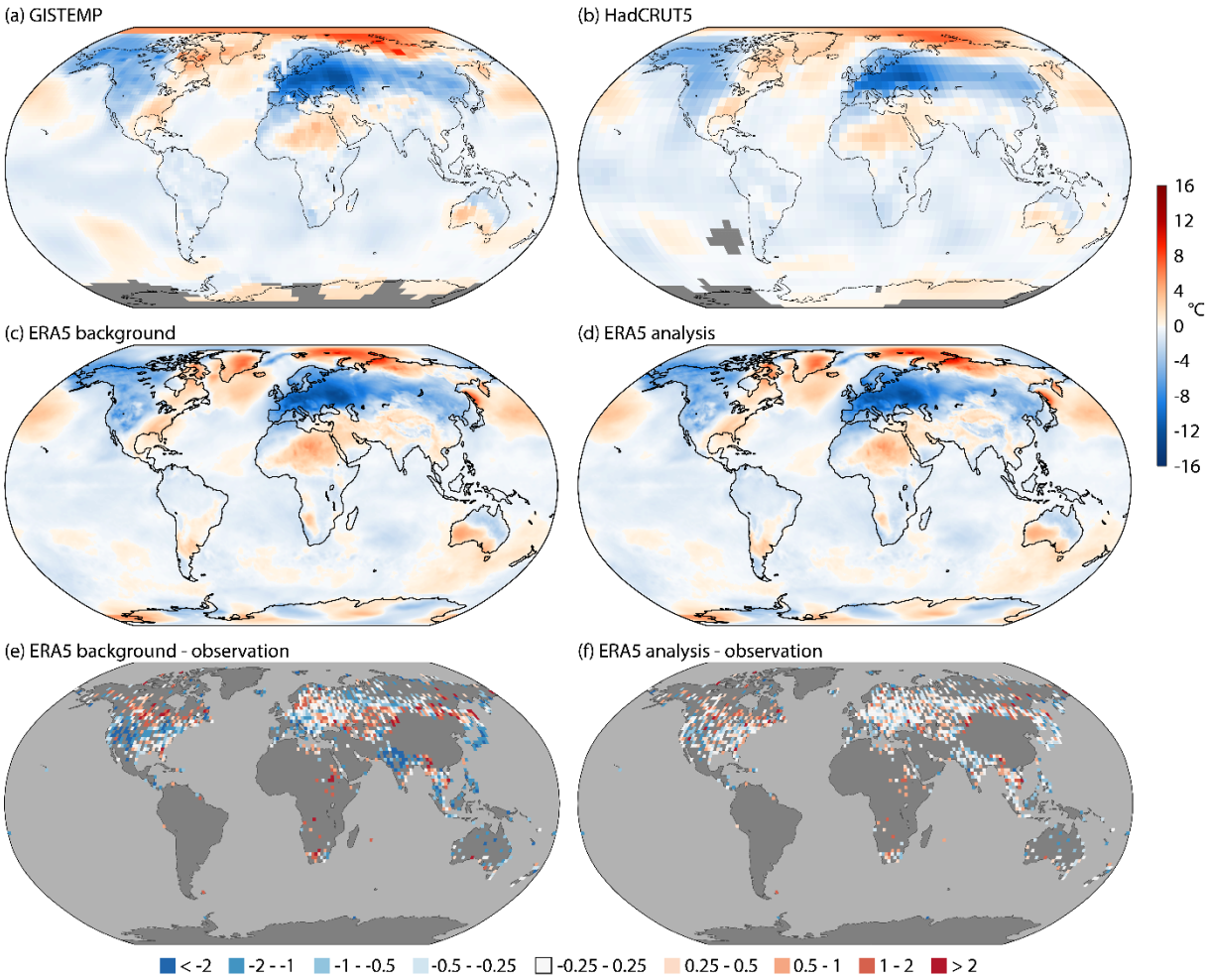

Figure 38 Average surface temperature anomalies (ºC) relative to 1981-2010 for February 1956: (a) GISTEMP; (b) HadCRUT5; (c) ERA5 background; (d) ERA5 analysis. Corresponding average (e) background-observation and (f) analysis-observation differences are shown for ERA5, based on values averaged over 2ºx2º grid boxes that contain at least ten observations during the month. Values for GISTEMP and HadCRUT5 are plotted only for grid squares for which no more than two months are missing in the calculation of the 1981-2010 reference.

There are nevertheless differences between ERA5 and the other datasets with regard to the weak anomalies over Argentina and south-western Africa. Panels (e) and (f) of Figure 38 show that few or no observations in these regions were supplied to ERA5, and the ERA5 analysis carries over positive anomalies from its background forecasts that are indicated by neither GISTEMP nor HadCRUT5. GISTEMP and HadCRUT5 have access to monthly climatological data for some regions where ERA5 does not have access to historical synoptic observations, over South America in particular for this month. ERA5 also lacks observations over China at this time, as noted above, but anomalies there are relatively weak. The spatial extrapolation used by GISTEMP and HadCRUT5 produces a temperature anomaly in the vicinity of the North Pole that agrees poorly with the anomaly from ERA5. ERA5 is also alone in producing relatively low temperatures near the ice-edge between Greenland and Svalbard.

Although the maps for the ERA5 background forecasts and analyses look similar at first sight, the analysis scheme for surface air temperature works as intended where observations are plentiful, significantly reducing biases in the background forecasts that reach up to 2ºC or higher in some

locations. This can be seen by comparing panels (e) and (f) over the western USA, northern India, and Japan, for example. The background bias is relatively small over Europe, however.

The second example is presented in Figure 39, for February 1990. This is the month with record warm average conditions over Europe, relative to the climatological normal for the month. It is represented by the highest point on the diagonal in Figure 35. The month is particularly warm over the north and far east of Europe, under the influence of a strongly anomalous cyclonic circulation pattern centred between Iceland and Scotland, with southerly flow over the Barents Sea and Arctic Ocean to the north, consistent with ERA5's depiction of above-normal temperatures in the high Arctic. The same is the case for JRA-55 (not shown), but this anomaly is less marked in GISTEMP and HadCRUT5. Elsewhere, temperatures are well below normal in a high-latitude belt from far eastern Siberia to Greenland.

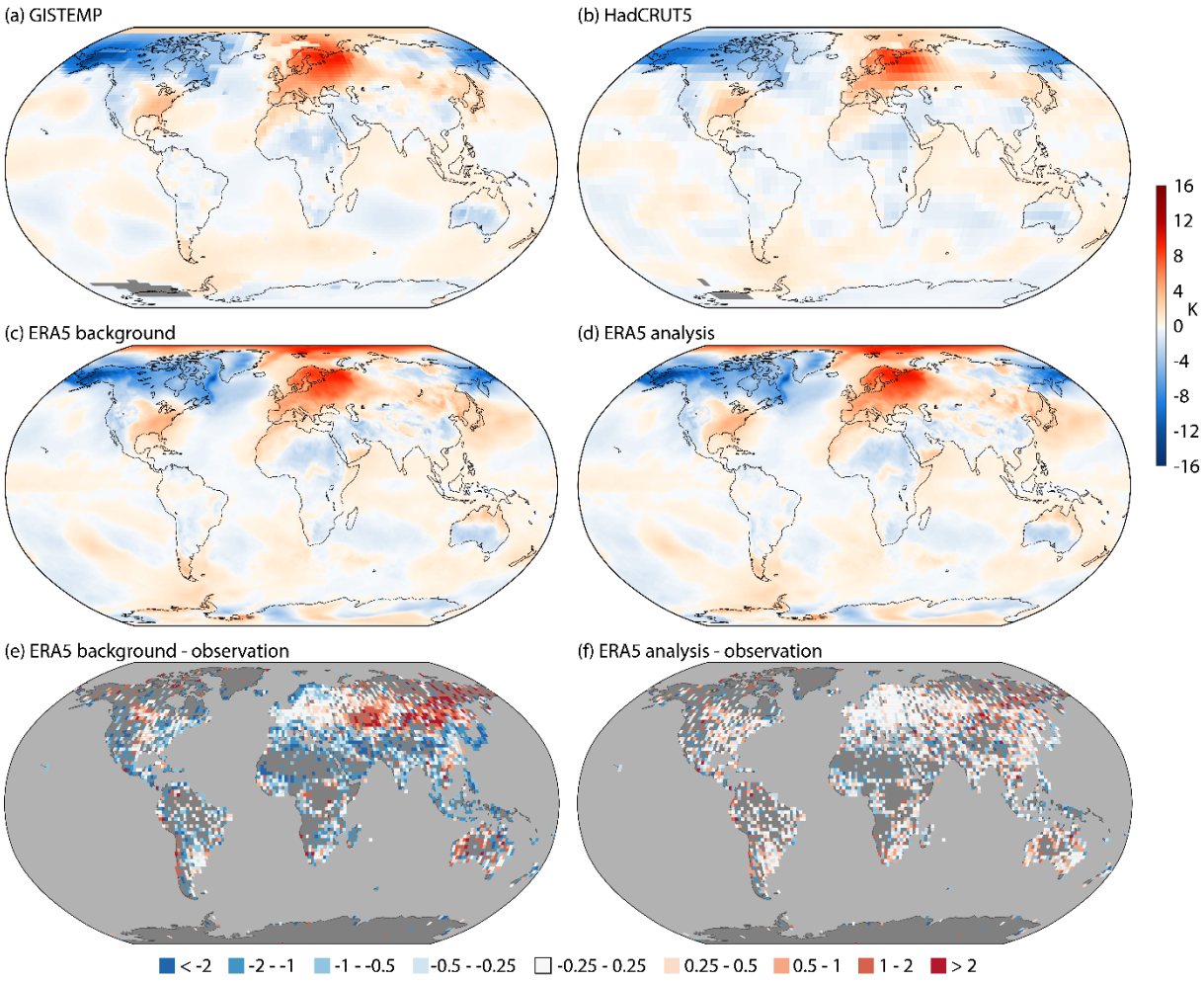

Figure 39 As Figure 38, but for February 1990.

In this case there is much better data coverage than in the 1956 case, and the weaker anomalies over much of the rest of the globe are similar in all analyses. The exception is Antarctica, where ERA5 has a pattern of temperature anomalies that is absent in GISTEMP and HadCRUT5. Cold bias in the ERA5 background is largely removed by the OI analysis scheme at many locations; a relatively large warm background bias over central latitudes of Asia and to the north-east is likewise largely removed by the analysis.

The final example is a recent case for boreal summer, June 2019. This is the warmest summer month in the ERA5 record for Europe, with an anticyclonic anomaly centred over the east of the continent, and anomalous southerly flow over the west of the continent. Temperature anomalies were larger still in Arctic Siberia, though not as extreme as in 2020. The last three Junes are the warmest three on record for this region (https://climate.copernicus.eu/index.php/arctic-siberias-unusual-warm-spell-continues).

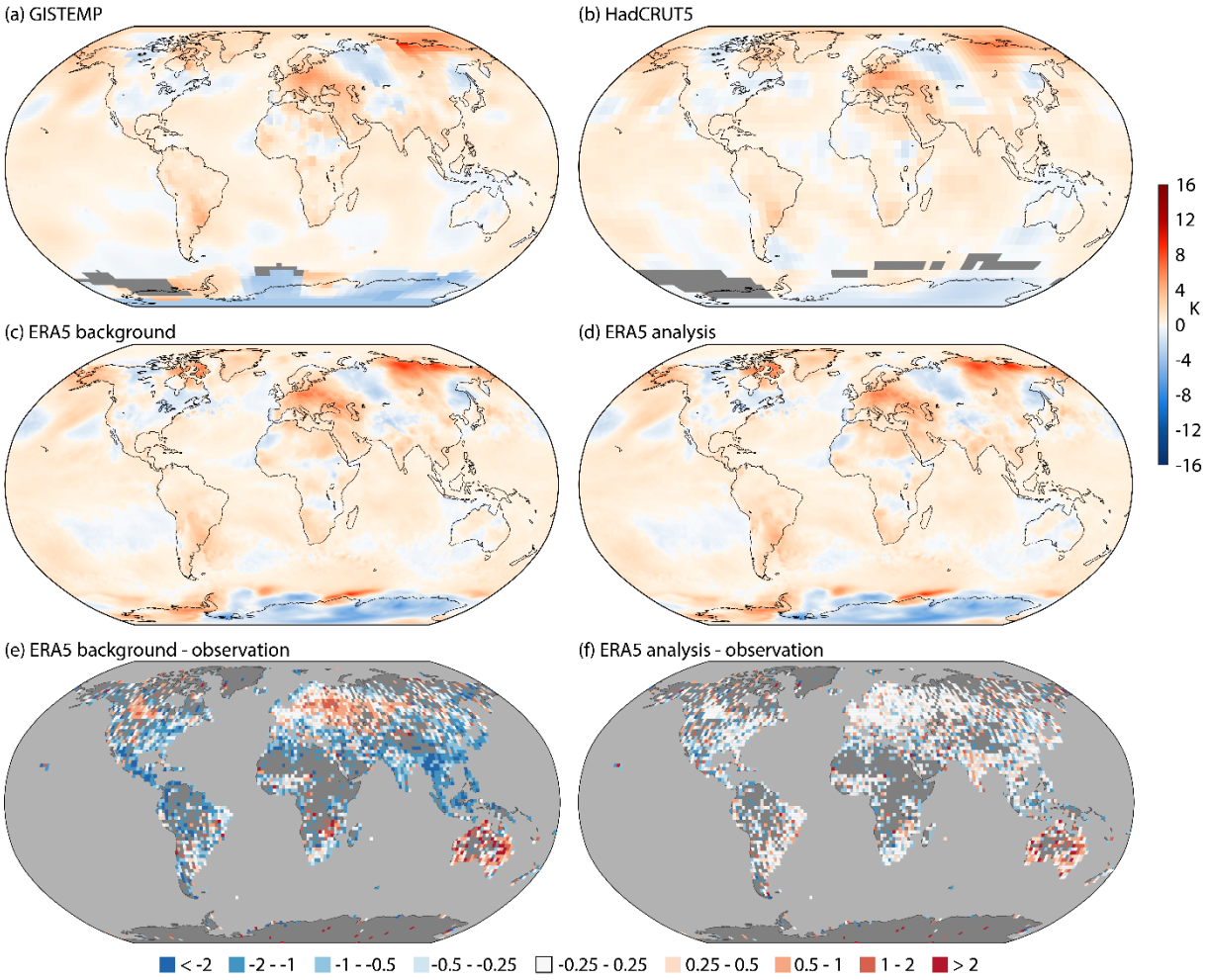

Figure 40 As Figure 38, but for June 2019.

There is again good general agreement between the various datasets in this case. More structure in the fields from ERA5 is again evident over and around Antarctica, where it is winter and there are accordingly local areas of anomalously warm conditions that are linked to anomalously low sea-ice concentration, corresponding maps of which may be found at https://climate.copernicus.eu/sea-ice-cover-june-2019.

Another region where sea-ice conditions come into play in determining differences between ERA5 and the datasets based only on climatological temperature observations is the Arctic Frontal Zone along the northern coastline of Siberia. Here ERA5 exhibits a sharp temperature gradient between air over land to the south that is anomalously warm in June 2019 and air over the coastal seas and Arctic Ocean, where near-surface temperatures are constrained to be close to 0°C by the temperature of the underlying melting ice or newly melted seawater. Anomalies are accordingly small. This is in contrast with the behaviour of GISTEMP and HadCRUT5, and also Berkeley Earth and Had4krig (not illustrated), each

of which extends the large anomalies observed over Arctic Siberian land out over the sea-ice to the north in an unphysical manner.

As is the case for the other two examples, the ERA5 background forecasts have a predominant cold bias that is significantly reduced by the surface analysis. In this example warm bias over the Canadian Prairies and western Russia is effectively removed also, but the problematic behaviour over Australia discussed earlier is evident, with only a limited improvement of the analysis over the background in the monthly-mean fit to observations.

3.5 Some local issues for ERA5

3.5.1 Impact of missing data on trends

3.5.1.1 Iran and Iraq

An issue that affects both ERA-Interim and ERA5 stems from the paucity of data from Iran and Iraq held in the ECMWF archives for the period from late 1980 to early 1989, during most of which the two countries were in conflict. Data numbers dropped suddenly in September 1980, and recovered slowly between November 1988 and April 1989. The region is characterised in both ERA-Interim and ERA5 by background temperatures that are biased cold, and by analysis increments that average around 0.6ºC when observations are present in sufficient numbers. The absence of observations for most of the 1980s results in too-cold temperatures during this period and too large a trend in temperature when calculated from 1979 or 1980 onwards, as is common. A local maximum south of the Caspian Sea can be seen for both ERA-Interim and ERA5 in the global maps of 1979-2018 trends shown in Figure 23.

Figure 41 presents time series of averages over a land region spanning much of Iran and Iraq, illustrated in Figure 42. The ERA-Interim and ERA5 analyses (panel (a)) show that the region warms over time, and that ERA-Interim is a little warmer than ERA5. Inter-annual variability is quite large, however, making it difficult to spot from the analyses alone that there is an issue. The issue is nevertheless clear from the time series of analysis increments presented in panel (b). When observations are available, the surface analysis scheme generally warms a background that the observations indicate is biased cold. The analysis increment increases a little over time, more so for ERA-Interim than ERA5 for recent years.

Panel (c) shows the total number of observation available to ERA5, taken from the archived 4D-Var feedback, and the numbers of observations for which ERA-Interim and ERA5 respectively had background departures larger in magnitude than 7.5ºC, the threshold above which data are rejected by the OI surface analysis scheme. The small number of observations available for much of the 1980s and a pronounced increase in observation numbers starting in early 2004 are evident. Also evident, especially from 2004 onwards, is an increase over time in the number of observations that exceed the rejection threshold. This number is generally higher for ERA-Interim than ERA5, highest in winter, and lowest in summer for ERA5 and autumn for ERA-Interim. Although a smaller number of observations are used by the surface scheme in the case of ERA-Interim, the background has a larger bias, resulting in the larger increments seen in panel (b). Analysis departures (panel (d)) are particularly small for ERA5 beyond 2004, showing only a small residual effect of the cold bias of its background. The small number of observations available for 1981-1988 are unrepresentative in that they show background and analysis departures opposite in sign to those of the earlier and later periods when many more data are available.

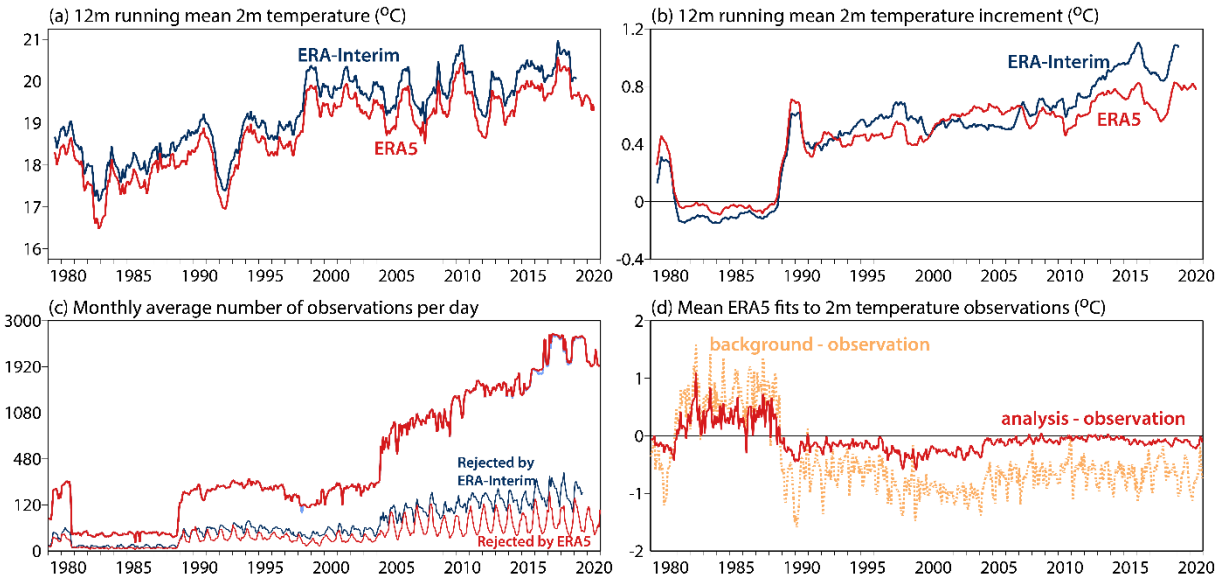

Figure 41 Twelve-month running averages taken over a land region spanning much of Iran and Iraq of (a) two-metre temperature (°C) and (b) the analysis increment in two-metre temperature for ERA5 (red) and ERA-Interim (dark blue) from 1979 to 2020. (c) Monthly average number of observations over the region available each day to ERA5 and ERA-Interim (heavy lines) and the number that exceed the surface analysis scheme's rejection threshold for background departures (light lines). (d) The monthly average background (orange) and analysis (red) departures from the ERA5 surface air analysis. The region for which averages are computed is shown in Figure 42.

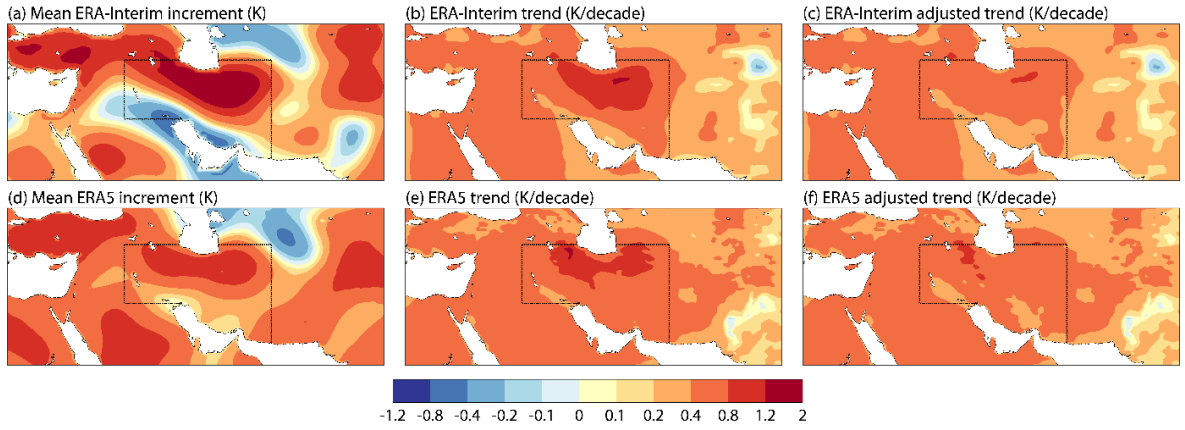

Figure 42 Average analysis increment in two-metre temperature (°C) for the years 1979 and 1990-2018 from (a) ERA-Interim and (d) ERA5. The linear trend for 1979-2018 in two-metre temperature over land (°C/decade) from (b) ERA-Interim and (e) ERA5. Corresponding trends adjusted for the 1980s data gap are shown in (c) and (f). The land region for which averages are shown in Figure 41 and adjustments are applied is depicted by black dotted lines on the maps.

The left-hand panels of Figure 42 are maps of the analysis increments for ERA-Interim and ERA5 over the region in question, averaged over years with good data coverage, 1979 and 1990-2018. ERA-Interim has a particularly large warming increment over Iran, south of the Caspian Sea, and a cooling increment around the Persian Gulf. The ERA5 increment has smaller spatial variation, and a weaker maximum south of the Caspian Sea.

The maps of unadjusted 1979-2018 temperature trends presented in the central panels of Figure 42 show maxima in the same region south of the Caspian Sea. These maxima are reduced in the trends shown in

the right-hand panels, which were derived using temperatures that were adjusted in the delineated region by applying the average increment calculated for the years 1979 and 1990-2018 to the temperatures analysed from October 1980 to March 1989.

3.5.1.2 Southern and eastern China

Similar considerations apply to some of the data gaps earlier in the period covered by ERA5. An example for a region covering most of southern and eastern China (105-122ºE; 22-40ºN) is presented in Figure 43. Time series of analysed and background two-metre temperature, and the corresponding analysis increments, are shown for ERA5. The OI temperature analysis primarily removes about 0.5ºC of cold bias in the background temperature; the background and analysis in general show a similar warming from the mid-1980s onwards.

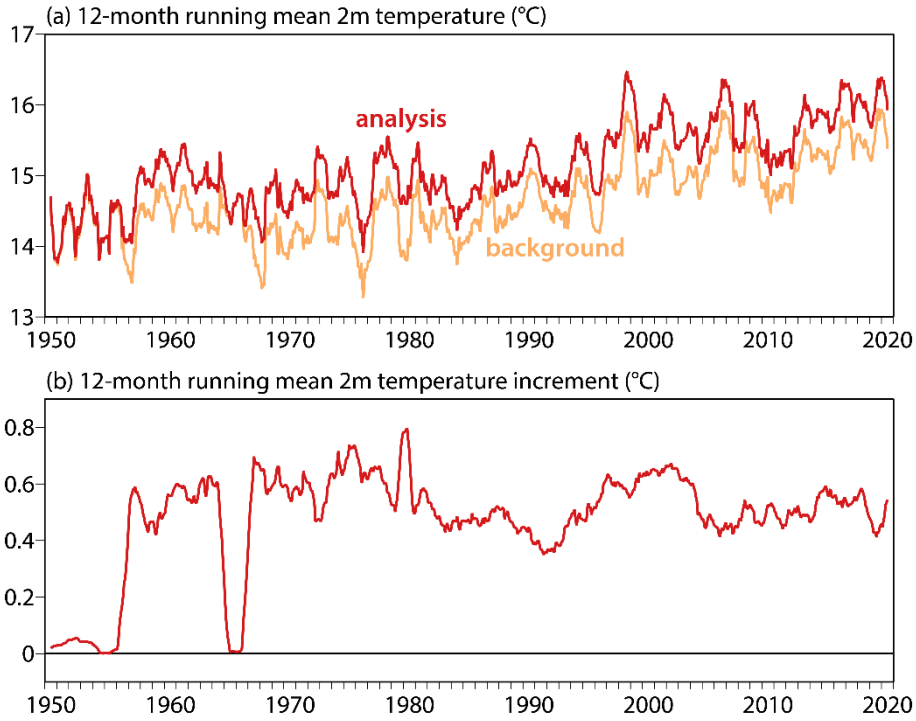

Figure 43 Twelve-month running mean time series of ERA5 two-metre temperature (ºC) averaged over southern and eastern China (105-122ºE; 22-40ºN), from 1950 to 2020, for (a) analyses and background forecasts, and (b) analysis increments.

The small increments from 1950 to 1955 and in 1965 and 1966 shown in Figure 43(b) are a consequence of ERA5 having access to little or no data from this region for these years. In this case too, the impact on the mean analysis appears to be straightforward; the time series for the analysed temperature could be "homogenised" by adding the known model bias of around 0.5-0.6ºC for periods when observations are absent, in order to derive a more reliable estimate of the long-term temperature trend.

3.5.2 Erroneous temperatures of the Great Lakes prior to 2014

In ERA-Interim, and in ERA5 from 2014 onwards, the surface temperatures of the North American Great Lakes were taken from SST analyses. They are not found to be especially problematic. It was also the intention to use SST analyses prior to 2014 in ERA5, but temperatures for these years were inadvertently taken from the assimilating model's simulation of lake temperatures rather than the SST analyses. As the lake modelling is inappropriate for lakes of the depth and area of the Great Lakes, the pre-2014 ERA5 temperatures of the Great Lakes are lower in winter and higher in summer than the

corresponding ERA5 temperatures from 2014 onwards. The same applies to the surface air temperatures over the Great Lakes. These temperatures are not corrected by the surface analysis scheme, which operates only over land in ERA5. The problem is most acute for Lake Superior, the Great Lake with the largest depth and surface area. It shows clearly in anomaly maps for recent years for monthly temperatures relative to the 1981-2020 climatology. It has little effect on twelve-month mean data as the winter and summer biases tend to cancel.

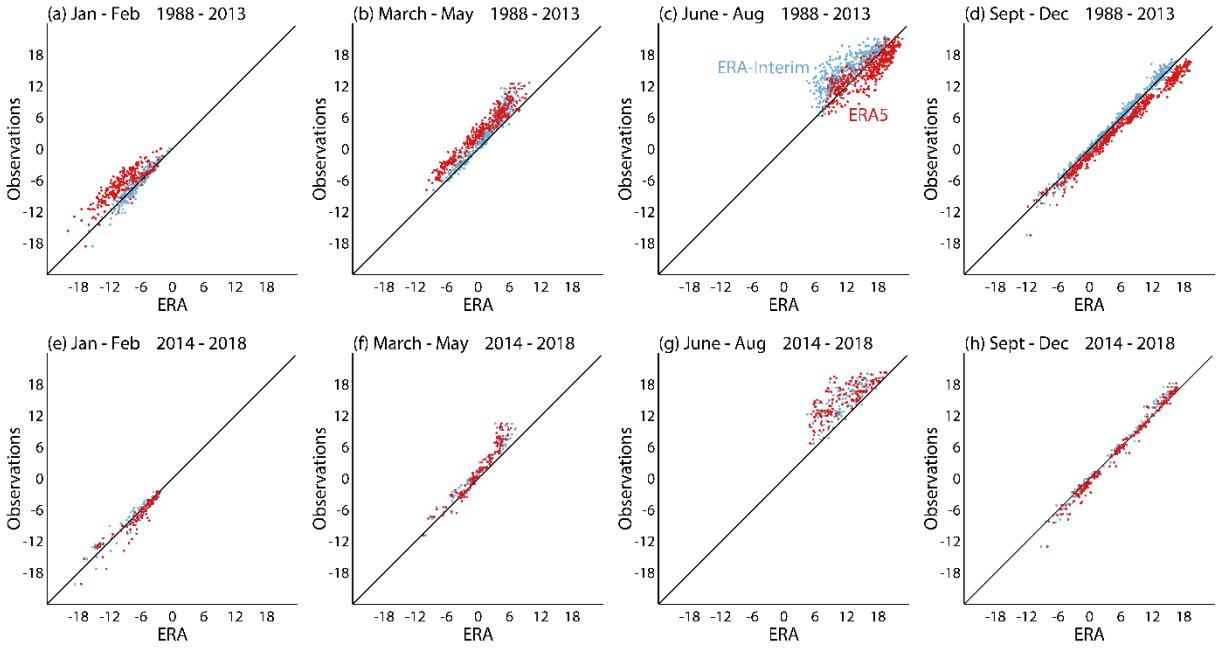

Figure 44 Comparison of monthly average ERA two-metre temperatures (ºC) with observed temperatures from the NOAA Great Lakes Environmental Research Laboratory coastal and marine station network around and over Lake Superior for ((a),(e)) January and February, ((b),(f)) March to May, ((c),(g)) June to August and ((d),(h)) September to December, from ((a) to (d)) 1988 to 2013 and ((e) to (h)) 2014 to 2018. Red dots denote values from ERA5 and pale blue dots denote values from ERA-Interim. The ERA5 values here are from the original production stream for 2000-2006, not from the ERA5.1 stream.

Figure 44 presents a comparison of ERA5 and ERA-Interim air temperatures with corresponding measurements made around and over Lake Superior. Data are separated into the periods 1988-2013 and 2014-2018, for the months January-February, March-May, June-August and September-December. For 1988-2013 and each set of months other than June-August, ERA-Interim fits the observations more closely than ERA5 does, with ERA5 biased cold for the first two sets and warm for the fourth set. ERA5 temperatures are also higher than ERA-Interim temperatures for June-August, although in this case ERA-Interim is biased warm and ERA5 biased cold compared with the observations. Conversely, ERA5 and ERA-Interim give similar temperatures for 2014-2018, which generally fit the observations well apart from a cold bias in June-August (and the latter part of May).

An adjustment of ERA5 two-metre temperatures has been developed in order to reduce spurious values over the Great Lakes in post-2013 monthly anomaly maps such as published routinely by C3S. The monthly ERA5 climate for 1981-2010 used to compute anomalies is adjusted over water in the region (92ºW-75ºW, 40ºN-50ºN) by adding the mean 1981-2010 difference between ERA-Interim and ERA5. Figure 45 shows the impact of the adjustment in scatter plots similar to those shown in Figure 44, in this case comparing ERA-Interim temperatures over Lake Superior with raw and adjusted ERA5 values. The

adjustment works as expected, given the differences between the ERA-Interim and raw ERA5 fits to observed temperatures shown in Figure 44.

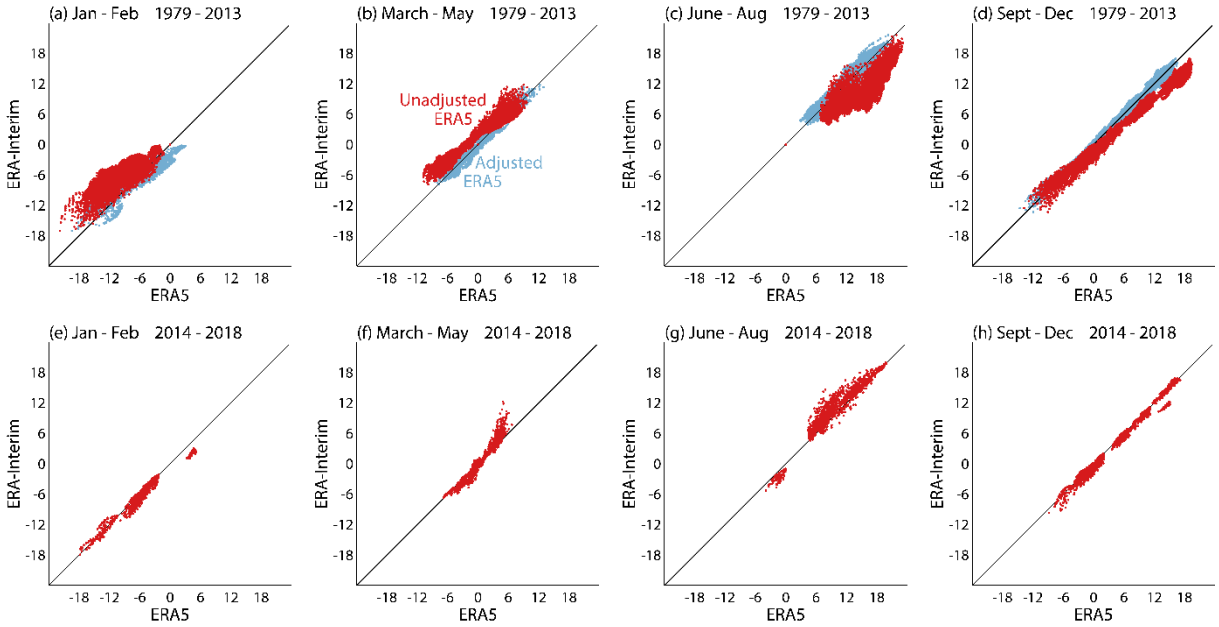

Figure 45 Comparisons of monthly average ERA5 and ERA-Interim two-metre temperatures (°C) interpolated onto a common regular 0.25° grid over Lake Superior for ((a),(e)) January and February, ((b),(f)) March to May, ((c),(g)) June to August and ((d),(h)) September to December, from ((a) to (d)) 1979 to 2013 and ((e) to (h)) 2014 to 2018. Red dots denote standard ERA5 values and pale blue dots denote ERA5 values for 1979-2013 that have been adjusted by adding the monthly climatological mean difference between ERA-Interim and ERA5 for 1981-2010.

Examples are presented in map form in Figure 46. The upper panels show the warming adjustment applied over the Great Lakes in February and the cooling adjustment applied in July. Without the adjustment the spurious anomalies over the Great Lakes, Lake Superior in particular, are evident in the maps of anomalies for February and July 2018, relative to 1981-2010. They are largely removed by application of the adjustment, though not entirely when judged by comparison with ERA-Interim.

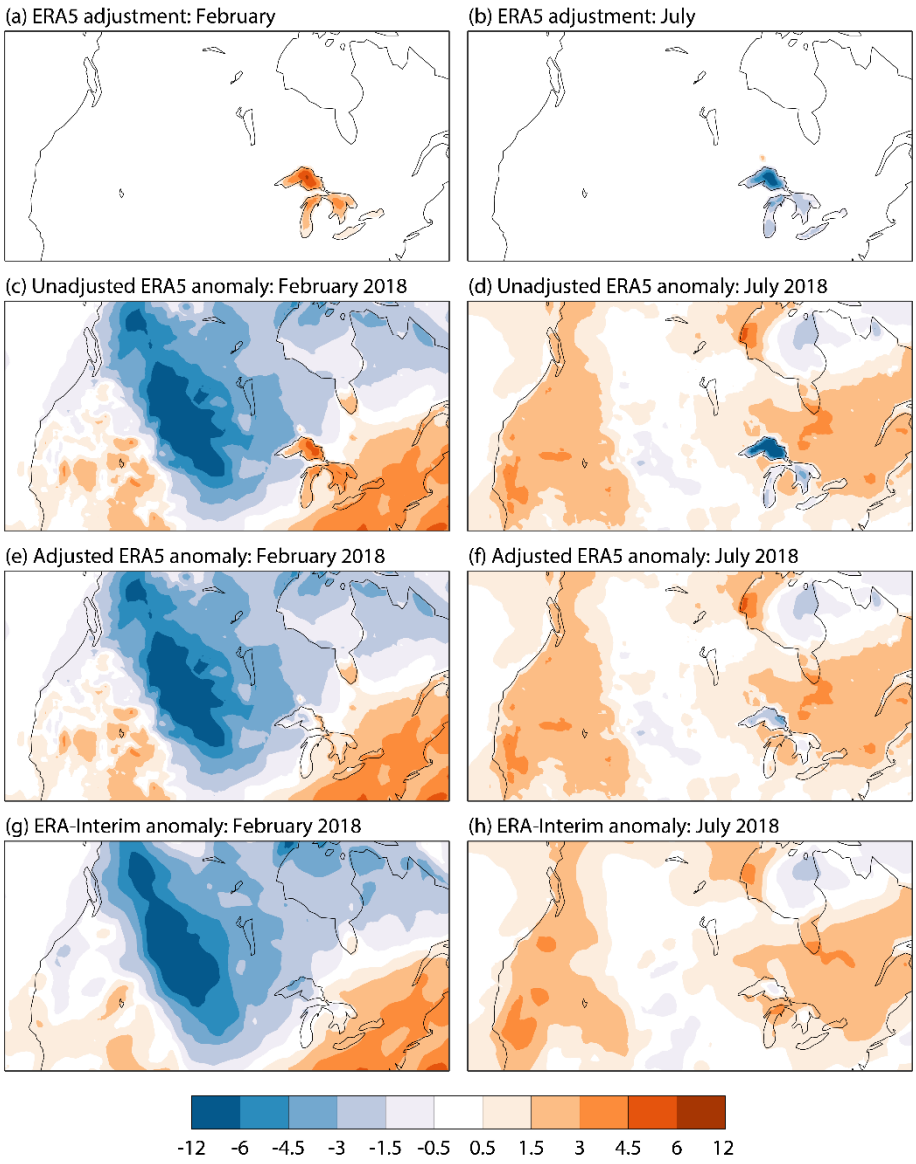

Figure 46 Adjustments applied in (a) February and (b) July to two-metre temperatures (ºC) over the Great Lakes, ERA5 anomalies in two metre temperature (ºC) relative to raw 1981-2010 climatologies for (c) February 2018 and (d) July 2018 and relative to adjusted 1981-2010 climatologies for the same months ((e) and (f)), and corresponding ERA-Interim anomalies ((g) and (h)).

3.5.3 Inconsistencies in sea-surface temperature analyses

Although care was taken in preparing for ERA5 to make choices of SST analysis that gave a generally consistent transition from one to the other, this was not achieved everywhere.

In preparing to move from using ERA-Interim to using ERA5 for the monthly temperature bulletins published by C3S, it was spotted that ERA5 had a larger temperature anomaly than ERA-Interim over the northern Caspian Sea in April 2019, as illustrated in Figure 47. This was quickly found to be an issue specific to the time of year, linked to a relatively cool sea surface in the 1981-2010 ERA5 climate.

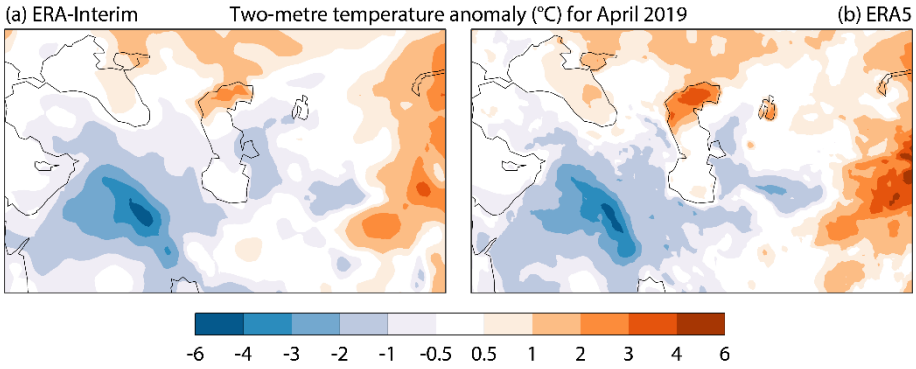

Figure 47 (a) ERA-Interim and (b)ERA5 anomalies in two metre temperature (°C) for April 2019 relative to 1981-2010 over the Caspian Sea and neighbouring regions.

Figure 48 shows SSTs for the months of March, April and May, averaged over the Caspian Sea north of 43°N, from ERA-Interim and ERA5. The two analyses are very similar for May from 1984 onwards, but a distinct change of behaviour can be seen for ERA5 for the other two months following the change made from September 2007 to use the OSTIA product available in near-real-time rather than the HadISST2 analysis used for earlier years. The values given by OSTIA for April from 2008 onwards are more consistent with the pre-2008 ERA-Interim values than they are with the pre-2008 HadISST2 values used in ERA5, and the 1981-2010 average is larger for ERA-Interim than for ERA5. The March value from HadISST2 is also more uniform than that from OSTIA, ranging only from 2.6 to 3.0°C in the period 1979-2007.

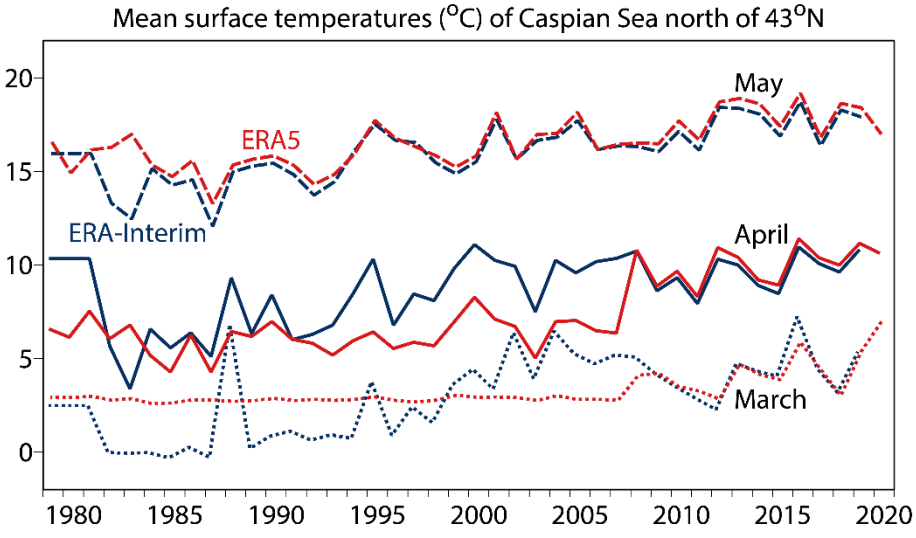

Figure 48 Time series from 1979 to 2020 of average surface temperatures (°C) of the Caspian Sea north of 43°N for the months of March, April and May, for ERA5 (red) and ERA-Interim (dark blue).

As this problem is not as severe in either magnitude or duration as that of the temperatures of the Great Lakes, no adjustment is applied in producing the C3S bulletins. The problem can thus be seen to reappear in April 2020. An adjustment similar to that applied for the Great Lakes based on average 1981-2020 differences between ERA-Interim and ERA5 could be applied for April if a specific need arises.

3.5.4 Inconsistencies in sea-ice analyses and temperatures north of Greenland

The maps of temperature trends presented in Figure 32 show warming over most of the Arctic in the case of ERA5, but little or no warming, or slight cooling to the north of Greenland. This is seen in none of the other datasets that provide values over this region, including ERA-Interim. It appears to be due to the concentrations of sea-ice in the analyses used by ERA5, which are lower prior to around 1990 than in subsequent years, leading to higher temperatures except in the melt season. A factor in this is likely to be the change in 1987 from the SSMR to the SSMI satellite-borne instruments from which sea-ice concentration is derived, although the interannual variability in estimates of sea-ice concentration cloud the picture. A further complicating factor is the setting of the sea-ice concentration to 100% north of 82.5ºN in the ERA-Interim assimilation from 1989 to early 2009. This was evidently wrong for summer and early autumn in the latter years of this period, when the analysed ice limit moved north of 82.5ºN at some longitudes.

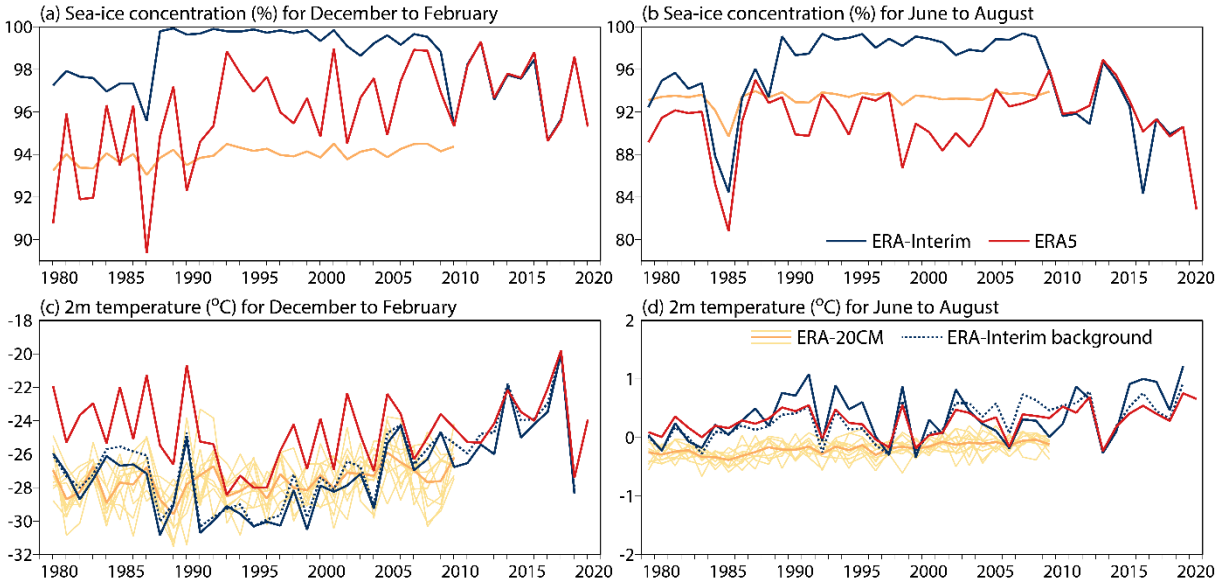

Figure 49 December-February ((a), (c)) and June-August ((b), (d)) means of ((a), (b)) sea-ice concentration (%) and ((c), (d) two-metre temperature (ºC) from ERA5 (red), ERA-Interim (dark blue, solid), the ERA-Interim background (dark blue, dotted) and from the ERA-20CM ensemble members and mean (orange) for averages north and north-east of Greenland (75ºW-0ºW, 82ºN-90ºN) from the period June 1979 to August 2020.

Evidence is presented in Figure 49, which shows time series of sea-ice concentration and two-metre temperature averaged from 75ºW to 0ºW and 82ºN to 90ºN for December-February and June-August. Values are shown for ERA5, ERA-Interim and the ERA-20CM ensemble of model simulations, which used a third sea-ice analysis. Details of sources are provided by Hersbach *et al.* (2020) for ERA5, Simmons and Poli (2015) for ERA-Interim and Hersbach *et al.* (2015) for ERA-20CM.

ERA-Interim has a higher sea-ice concentration than ERA5 in this region prior to 2009, including the 1979-1988 period when its concentration was not set to 100% north of 82.5ºN. It has largely similar concentrations to ERA5 from early 2009 onwards, when both reanalyses are based on the same OSTIA product, although differences are evident in some months. The higher sea-ice concentrations are expected from the discussion in section 3.3 to result in lower surface air temperatures in winter for ERA-Interim, and temperatures are indeed lower in ERA-Interim than ERA5, by several ºC, especially prior to 2005. This does not come solely from the sea-ice differences, however. A further difference between ERA-Interim and ERA5, already noted, is that the surface analysis was suppressed over all sea surfaces

in ERA5, but used over ice-covered sea in ERA-Interim, reducing a warm bias in its background values (Simmons and Poli, 2015). This can be seen from the differences between the background and analysis values shown for ERA-Interim in Figure 49.

The sea-ice concentration (from HadISST2) used in ERA-20CM is also generally lower than that used in ERA-Interim, but it has a smaller variation over time, as do the corresponding air temperatures, notwithstanding the variability of the ensemble members.

There are also large differences in sea-ice concentration in summer, but in this case the surface temperatures of melting ice and sea are similar, at close to 1°C. There is variability in the sign of the ERA-Interim analysis increment for this season.

	JJA sea		JJA land		DJF sea		DJF land	
	Number of obs	Mean obs - bg	Number of obs	Mean obs - bg	Number of obs	Mean obs - bg	Number of obs	Mean obs - bg
ERA-Interim	7266	-0.1°C	22290	2.4°C	6847	-6.7°C	14019	3.8°C
ERA5	7268	-0.1°C	22361	1.5°C	5656	-8.7°C	17103	0.6°C

Table 3 Observation feedback for the ERA-Interim and ERA5 background forecasts over the region (75°W-0°W, 82°N-90°N) from the period June 1979 to August 1989, averaged for June, July and August (JJA) and December, January and February (DJF). The number of observations with (observation – background) departures smaller than 15°C in magnitude and the corresponding average departures are shown for data reported in SHIP and BUOY codes (including ice-station data) and for data reported in Land SYNOP codes. The latter come from three stations: Alert on Ellesmere Island (see also Figure 17) and Cape Morris Jesup (83.6°N, 33.4°W) and Cape Harald Moltke (82.2°N, 22.9°W) on Greenland.

Table 3 shows mean fits for 1979-1989 of the ERA-Interim and ERA5 background forecasts of surface air temperature to observations in the region from 75°W to 0°W and 82°N to 90°N, again separated into the December-February and June-August periods. Both reanalyses have very low summer biases over (mostly ice-covered) sea, but large cold biases in winter. The cold winter bias is 2°C larger for ERA5 than for ERA-Interim. The region concerned includes just three land stations (see caption). The background forecasts have a net warm bias for these stations. This bias is smaller for ERA5 than for ERA-Interim.

4 Surface air humidity

4.1 Observations

Figure 50 is a repeat of Figure 3, but shows the numbers of observations of humidity as well as those of (dry-bulb) temperature. ERA5 used no observations of humidity prior to September 1957, and fewer observations of humidity than temperature in the period from September 1957 to June 1959 when the ERA-40 data holdings were supplemented by additional temperature but not humidity data from NCEP. Subsequently, for several decades, the number of observations of relative humidity is almost the same as the number of observations of temperature, the latter being barely visible as the graphs for humidity observations overlay the graphs for temperature observations. There are, however, small but more-evident differences from 2004 onwards in the number of observations used by the surface analysis, and these differences have a pronounced annual cycle for North America and Europe. This stems from an issue with METAR data that is discussed in section 4.7.2. Otherwise, apart from the absence of humidity

data prior to September 1957, the previous discussion of the coverage of temperature data applies also to humidity data.

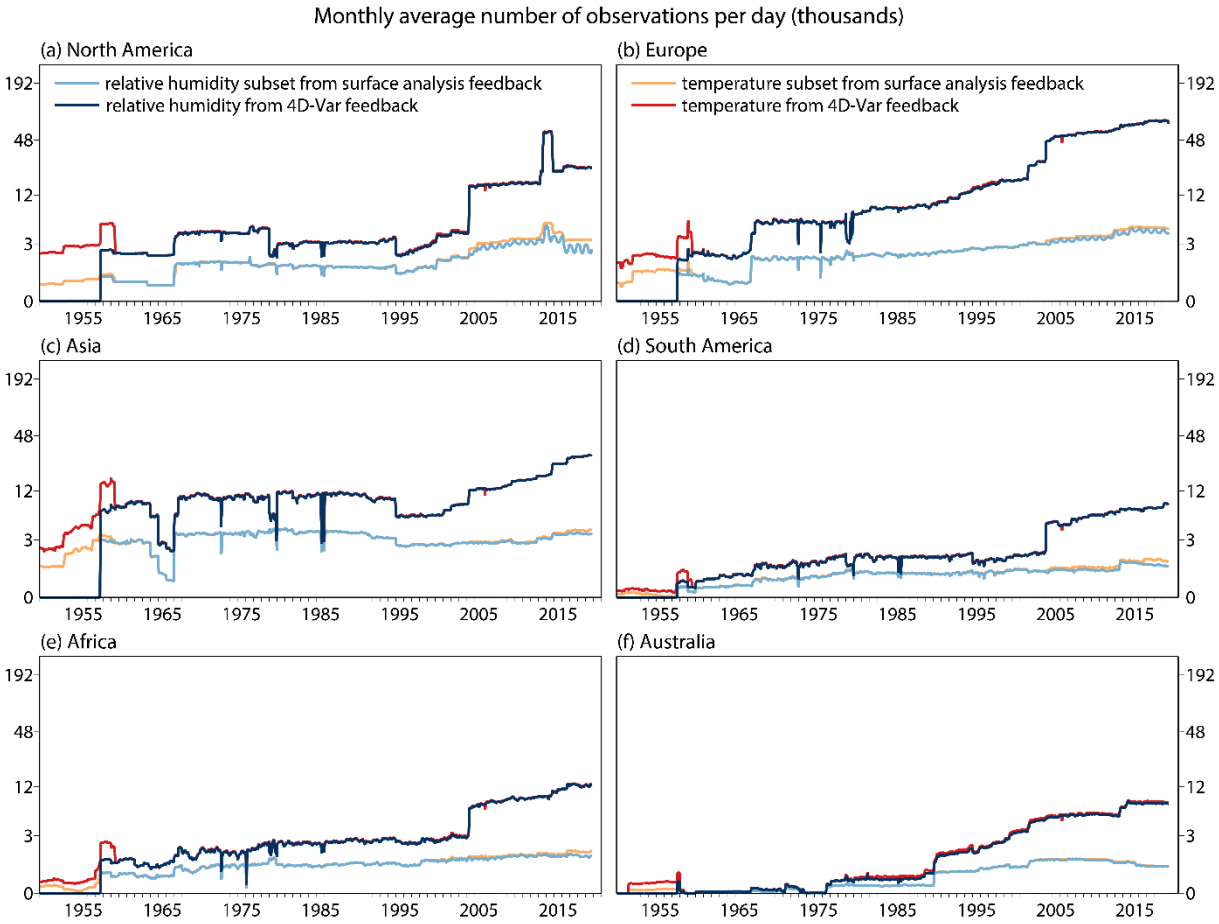

Figure 50 Time series of monthly averages of the daily number of observations by region, as in Figure 3 but showing the number of relative humidity observations (dark and light blue) as well as the number of temperature observations (red and orange).

4.2 Comparisons of datasets over land and sea

Figure 51 presents twelve-month running mean times series of specific humidity over land and sea. Averages over all land of actual values are shown in panel (a) for the ERA5, ERA-Interim and JRA-55 reanalyses. Also shown are the averages over all land grid boxes where HadISDH provides actual values, and corresponding averages of the ERA5 values taken over these grid boxes. The generally larger values of the HadISDH and sampled ERA5 averages reflects the fact that HadISDH under-samples the drier regions of the world, particularly over Antarctica, Greenland and the far north of Russia and North America, but also over deserts. This is a consequence of the limited observational coverage over these regions. If all other factors were to be equal, the values from reanalysis would be expected to be less reliable for these less-well-observed regions than for the regions where HadISDH provides plentiful values. The rise in HadISDH and sampled ERA5 values from 1973 to 1978 relative to the complete averages from the reanalyses indicates a change in observational coverage of HadISDH in this period.

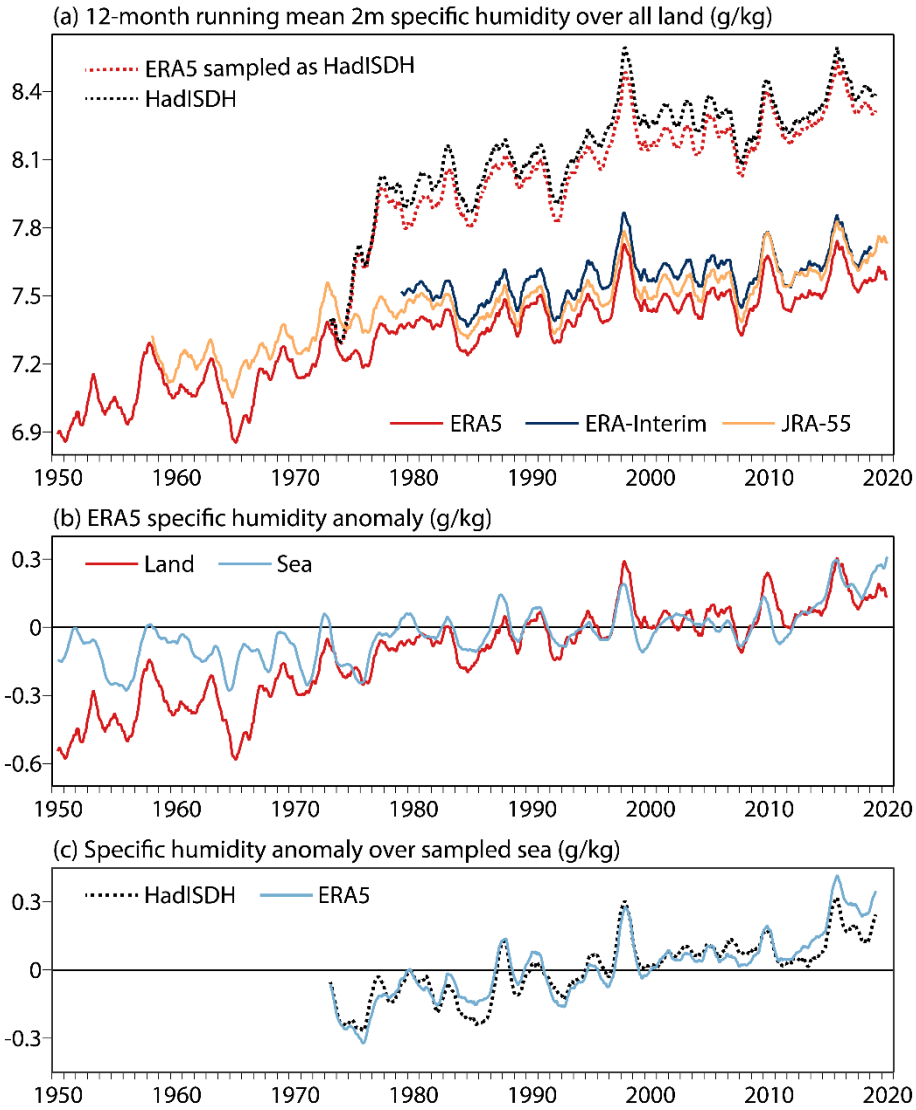

Figure 51 Twelve-month running means from 1950 to 2020. (a) Average over all land of two-metre specific humidity (g/kg) from ERA5 (red, solid), ERA-Interim (dark blue), JRA-55 (orange) and HadISDH (black, dotted), and from ERA5 sampled spatially to match the coverage of HadISDH (red, dotted). (b) ERA5 anomalies relative to 1981-2010, for averages over all land (red) and all ice-free sea (blue). (c) ERA5 and HadISDH anomalies over sea, with ERA5 sampled as HadISDH. HadISDH data are available only for 1973-2018.

Aside from the differences related to differences in data coverage, all datasets describe a similar long-term moistening, and shorter-term variability over time that tends to mirror the short-term fluctuations in temperature shown in Figure 1. ERA5 does drop to a deeper minimum in the mid-1960s than JRA-55; values at and prior to this time are likely biased dry due to a combination of a dry background-model bias and relatively few or no observations of humidity at these times. ERA-Interim has a slightly smaller moistening trend than ERA5 and JRA-55, reflecting its slightly lower warming trend.

Simmons *et al*. (2010) presented evidence from the HadCRUH forerunner of HadISDH and from ERA-Interim and ERA-40 showing the similarity of variations in surface specific humidity over sea and land, both for long-term trends and for the shorter term variations. Variations over land tended to lag those over sea by a month or so. This was most evident for the peaks associated with El Niño events. The

comparison was rather unsatisfactory in the case of ERA as the marine values were taken to be the saturation specific humidity derived from the SST analysis, due to issues with the humidity analyses over sea.

Panel (b) of Figure 51 compares time series of ERA5 specific humidity averaged over all land and over all ice-free sea, expressed as anomalies relative to 1981-2010. The marine values in this case are the background values from the 4D-Var data assimilation, as increments from the surface analysis scheme are suppressed over sea. They vary in a similar way to the land values from the late 1960s onwards, but changes over sea are again seen to lead the changes over land by a month or so. This includes the two most recent El Niño maxima in 2010 and 2016. The marine values in the 1950s and 1960s are consistent with variations in SST, but the corresponding land values are lower than would be expected from the results for later decades, providing a further indication that the ERA5 analyses are biased dry over land in this period when there are insufficient observations to counter the dry-biased background.

The marine values from ERA5 are compared with those from HadISDH in panel (c). The ERA5 values are averaged here only over those 5ºx5º grid boxes where HadISDH provides values; comparison with panel (b) shows that this makes only a small change to the time series. ERA5 and HadISDH are evidently in a fair degree of agreement as to the increase in specific humidity over sea and the interannual variations in values. ERA5 is nevertheless moister over sea (relative to 1981-2010) than HadISDH from 2012 onwards. This may be due in part to a small change in bias of the SST analysis used by ERA5, as suggested by the temperature comparisons shown in Figure 19 and Figure 26, although the possible change in temperature bias appears to be at most a few hundredths of a ºC, which is not large enough to explain easily the extent of the difference in specific humidity between ERA5 and HadISDH for the last few years of the time series.

Corresponding results for relative humidity are shown in Figure 52. All datasets show a decline in relative humidity over land since the 1970s. They also show similar interannual variability once allowance is made for the effect of incomplete coverage by HadISDH. JRA-55 does not show as steep a fall in the 2000s, but its overall fall since the mid-1970s is similar to that of the other datasets. Moreover, it does have as steep a fall if sampled as HadISDH (not illustrated). ERA5 and JRA-55 are relatively dry in the 1960s, particularly in 1965 and 1966 in the case of ERA5. ERA5's relative humidity is also low in the 1950s.

Time series of the analysis increments in specific and relative humidity averaged over all land are presented in Figure 53, for ERA5 and ERA-Interim. The ERA5 increment is zero in relative humidity until September 1957, as relative humidity is the analysed variable and no humidity data is analysed for earlier months. There is however a small increment in specific humidity, due to increments in temperature and surface pressure, as the relationship between specific and relative humidity depends on theses variables. The specific humidity increment is nevertheless small before September 1957. It is also lower prior to 1967 than in any subsequent year.

For ERA5, the moistening increments in specific humidity increase only slowly over time from 1967 to around 1990. The relative humidity increments, in the annual average, are small prior to the 1990s, although the monthly increments indicate a small negative bias for much of the year, and a larger positive bias in boreal springtime. Small increments can be the result either of having few observations to assimilate or of having only a small bias in the background humidity. ERA5 is biased dry in terms of specific humidity in its early years, but it is also biased cold. It is thus difficult to judge the extent to which the lower relative humidities prior to the early 1970s are a true feature. The very low values in 1965 and 1966, barely seen in JRA-55, are almost certainly spurious, however, coming from a too-dry and too-warm background that is uncorrected by observations. Evidence for this comes from the

unusually low rainfall and high temperatures over Africa at the time, noted earlier in section 3.4.2. Further discussion is given in section 4.7.1.

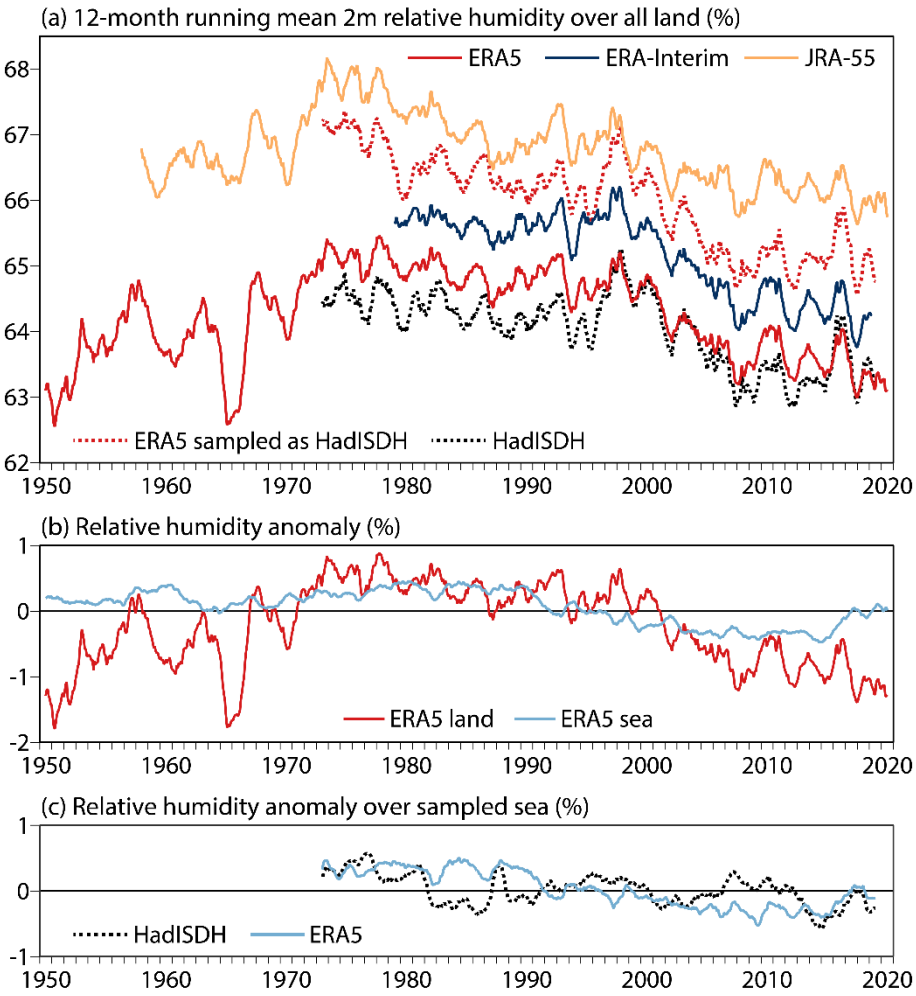

Figure 52 Twelve-month running means from 1950 to 2020. (a) Average over all land of two-metre relative humidity (%) from ERA5 (red, solid), ERA-Interim (dark blue), JRA-55 (orange) and HadISDH (black, dotted), and from ERA5 sampled spatially to match the coverage of HadISDH (red, dotted). (b) ERA5 anomalies relative to 1981-2010, for averages over all land (red, solid) and all ice-free sea (light blue, solid). (c) ERA5 and HadISDH anomalies over sea, with ERA5 sampled as HadISDH. HadISDH data are available only for 1973-2018.

Beyond 1990, ERA5's analysis increments in both specific and relative humidity rise until about the year 2000, and then fall back to return close to their 1990 level by 2020. This is primarily a response to a shift in the bias of background values. Further discussion is given in the following section, where comparisons of background and analysis values with observations are shown. ERA-Interim has a largely similar long-term variation of increments over time. It has slightly larger moistening increments for specific humidity, but its relative humidity increments are systematically below those of ERA5.

Relative humidity varies less over sea than over land. Figure 52 shows a small net decrease over the period of comparison for both ERA5 and HadISDH, but there is little agreement as to shorter-term variability.

The differences in absolute values of relative humidity over land shown in Figure 52 appear disconcerting, especially those between ERA and JRA-55 and between ERA5 and HadISDH when

ERA5 is sampled as HadISDH. They come, however, from specific geographical regions. This can be seen from the maps of average differences for 1979-2018 in specific and relative humidity between ERA5 and respectively JRA-55 and HadISDH presented in Figure 54.

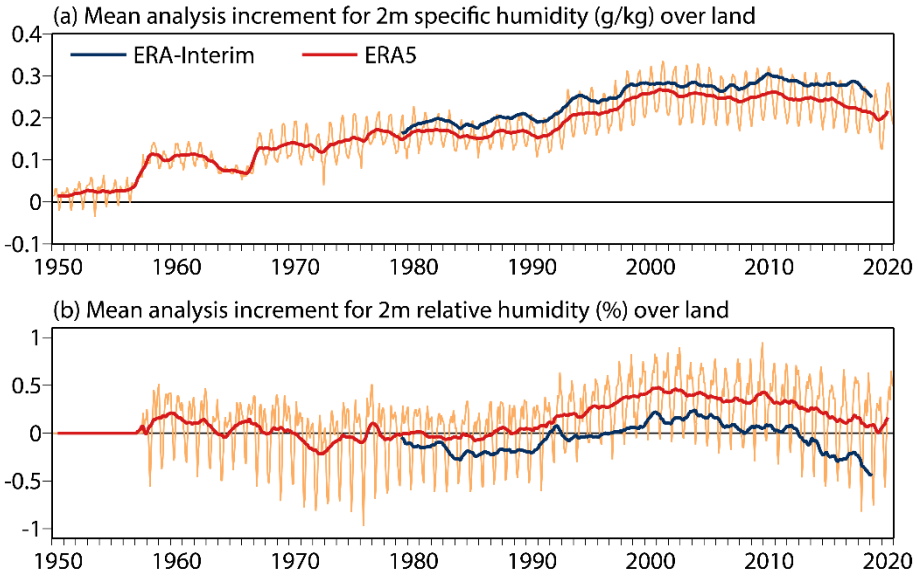

Figure 53 Average analysis increments over land for two-metre (a) specific humidity (g/kg) and (b) relative humidity (%), for ERA5 monthly (orange) and 12-month running (red) means, and from ERA-Interim 12-month running means (dark blue).

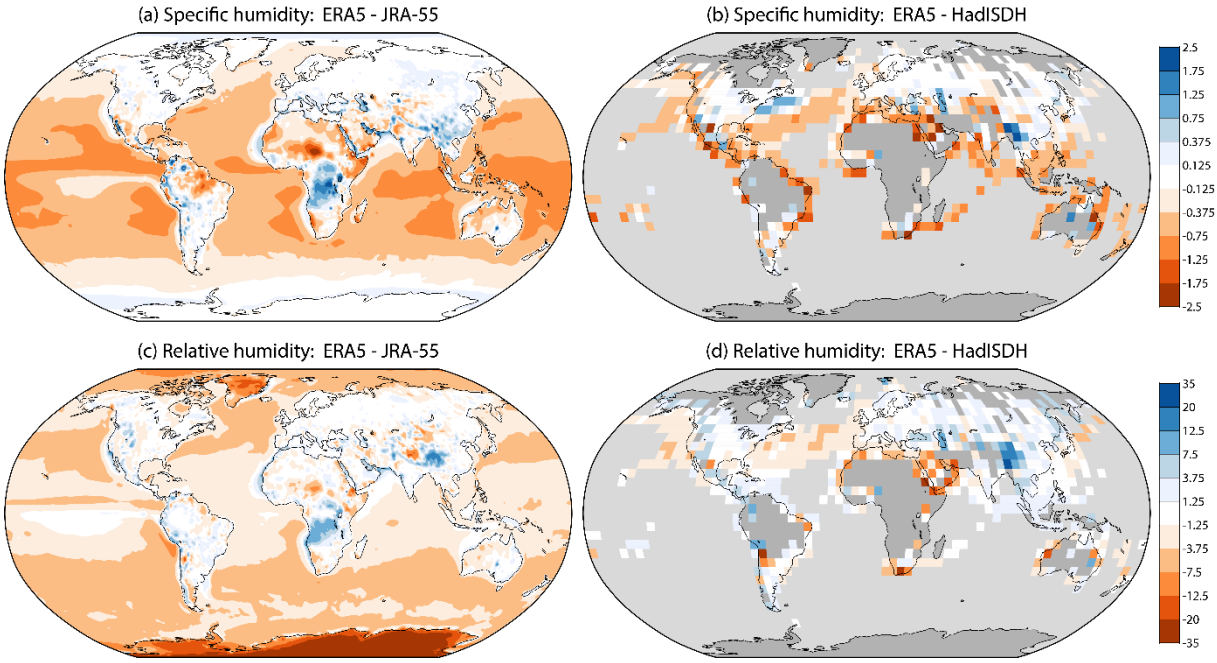

Figure 54 Averages for 1979-2019 of differences in surface air humidity: (a) specific humidity (g/kg), ERA5 – JRA-55, (b) specific humidity (g/kg), ERA5 -HadISDH, (c) relative humidity (%), ERA5 – JRA-55, and (d) relative humidity (%), ERA5 - HadISDH. ERA5 data are evaluated on the respective grids at which HadISDH and JRA-55 data are available. Values for the ERA5 - HadISDH differences are plotted only for grid squares where HadISDH provides values for at least 490 of the 492 months over which averages are made.

The absolute differences between ERA5 and JRA-55 are widespread over sea and, in the case of relative humidity, large over Antarctica. Over sea, the lower specific humidity of ERA5 is consistent with its lower SST (shown in section 3.2), and its lower relative humidity is consistent with its greater difference between SST and MAT (also shown in section 3.2). JRA-55 has been shown to be warmer than ERA5 over the Antarctic Plateau and (to a lesser extent) Greenland, so its higher relative humidity there implies a higher specific humidity. This is indeed the case, although as the specific humidities are low the absolute differences are small, and do not show in Figure 54.

Differences between ERA5 and HadISDH are quite large for a number of coastal grid boxes, where there may be issues as to how representative the 5°x5° grid values are. This has been noted previously for the comparison of HadISDH with ERA-Interim (Willett *et al.*, 2020b). Otherwise, HadISDH has predominantly higher humidities than ERA5 over sea, also as noted by Willett *et al.* for the comparison with ERA-Interim. HadISDH's specific humidities appear to be closer to those of JRA-55, and its relative humidities appear to be in between those of the two reanalyses in the northern extratropics, and closer to ERA5 in the tropics. HadISDH has higher humidities than either ERA5 or JRA-55 over and around the Arabian Peninsula (as singled out also by Willett *et al*. for the HadISDH/ERA-Interim comparison), but ERA5 has higher humidities than the other two datasets over south-eastern Asia, particularly over land to the north-east of the Bay of Bengal. The three datasets in general agree quite well over the mid-latitude land areas of the extratropical northern hemisphere.

4.3 Fits to land-station data

Maps based on observational feedback data for relative humidity are presented in Figure 55, showing average departures for 1979-2018 plotted for all observations within 2°x2°grid boxes as presented earlier for temperature in Figure 6. In the annual average the ERA5 background has a bias of too-low relative humidity at most locations when averaged over all observations processed by 4D-Var. The ERA-Interim background has a similar bias at extratropical latitudes, but conversely has higher relative humidity than is observed at most locations in the tropics. Averaged over all observations, the net bias of ERA5 is thus larger than that of ERA-Interim, even though there is little to choose between the magnitudes of the local biases of the two reanalyses.

The mean background departures for the surface analysis shown in Figure 55 differ more from those from the 4D-Var assimilation than is the case for temperature. The feedback statistics archived for the surface analysis are evidently limited in their representivity in the case of relative humidity. The analysis is seen nevertheless to remove much of the bias at extratropical latitudes of the northern hemisphere, although some residual bias is evident over eastern Asia. Rather more residual bias is evident in the analysis departures in the tropics and extratropical southern hemisphere. In particular, the analysis fails completely to reduce the bias over Australia.

A seasonal variation to the bias was noted in section 4.2, with the departures averaged over all observations showing too high rather than too low relative humidity in boreal springtime. This is illustrated in Figure 56 by maps showing the 1979-2018 average background and analysis departures for April and November, from the surface scheme. The background departures for the two months are largely similar in the tropics and extratropical southern hemisphere, but opposite in sign over much of the extratropical northern hemisphere, eastern Asia being the main exception. A trace of the differences between the months over the extratropical northern hemisphere can be seen in the analysis departures, but the analysis again can be seen to remove most of the bias for this region.

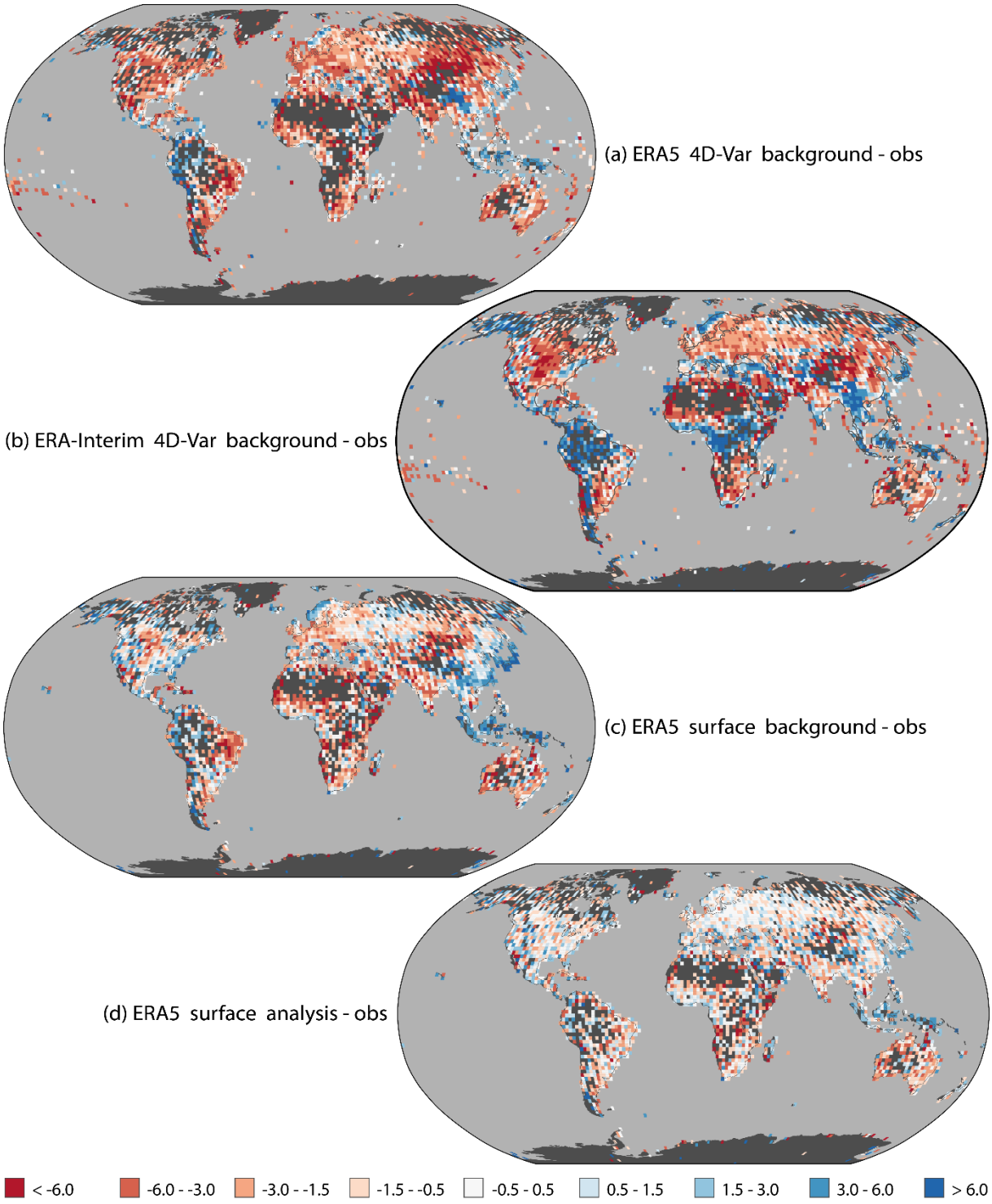

Figure 55 Mean (background - observation) and (analysis - observation) fits to surface air relative-humidity observations (%) from 1979-2018 for 2°x2° grid boxes. (a) Background departures for all selected observations processed by the ERA5 4D-Var, (b) corresponding ERA-Interim 4D-Var background departures, (c) background departures for the selected sample of observations used in the ERA5 surface analysis, and (d) corresponding surface analysis departures.

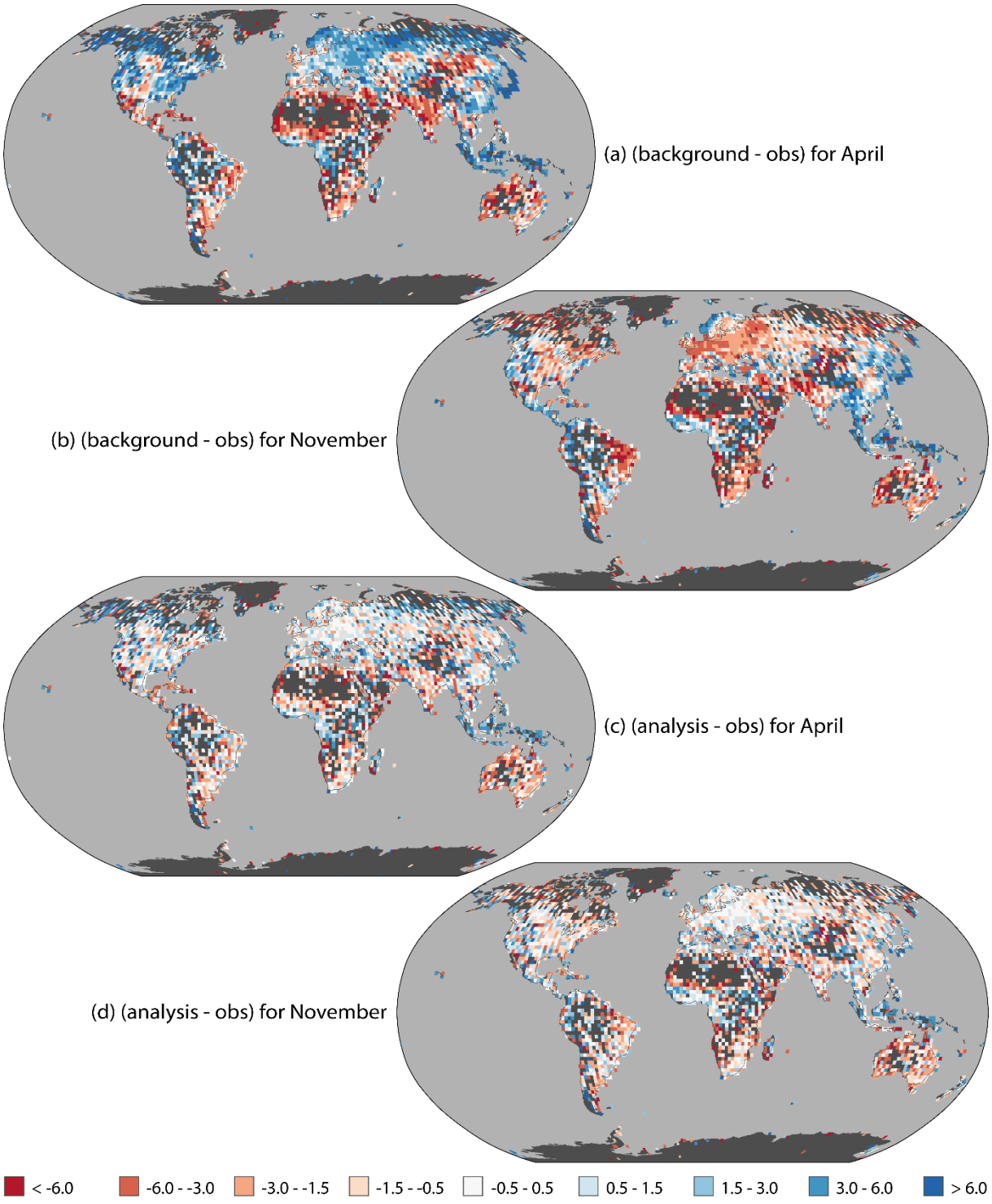

Figure 56 Mean (background - observation) and analysis - observation) fits to surface air relative-humidity observations (%) from 1979-2018 for 2°x2° grid boxes for the selected sample of observations used in the ERA5 surface analysis. (a) Background departures for April, (b) background departures for November, (c) analysis departures for April, and (d) analysis departures for November.

Time series of the monthly means and standard deviations of the background and analysis departures are shown in Figure 57 for the six continental regions defined earlier. All regions show increasingly negative mean background departures in the 1990s, and some reversal of this trend thereafter. Consistent

with the rise and fall in analysis increments (in specific as well relative humidity) shown in Figure 53, where observations are plentiful the surface analysis scheme removes most of the slowly varying bias in background relative humidities, providing analysed fields that match the observations closely, showing only a small dip in values in the 1990s and 2000s. This is especially the case for Europe, for which the monthly mean analysis departures are small throughout, and is largely true for North America and Asia. These latter two regions both exhibit shifts in their analysis departures in the mid-1960s and slightly larger dips around the year 2000 than seen for Europe. The analysis departures also show little net variation over time for South America, but there is some variability in the departures in the 1960s and a slightly more pronounced dip later in the period.

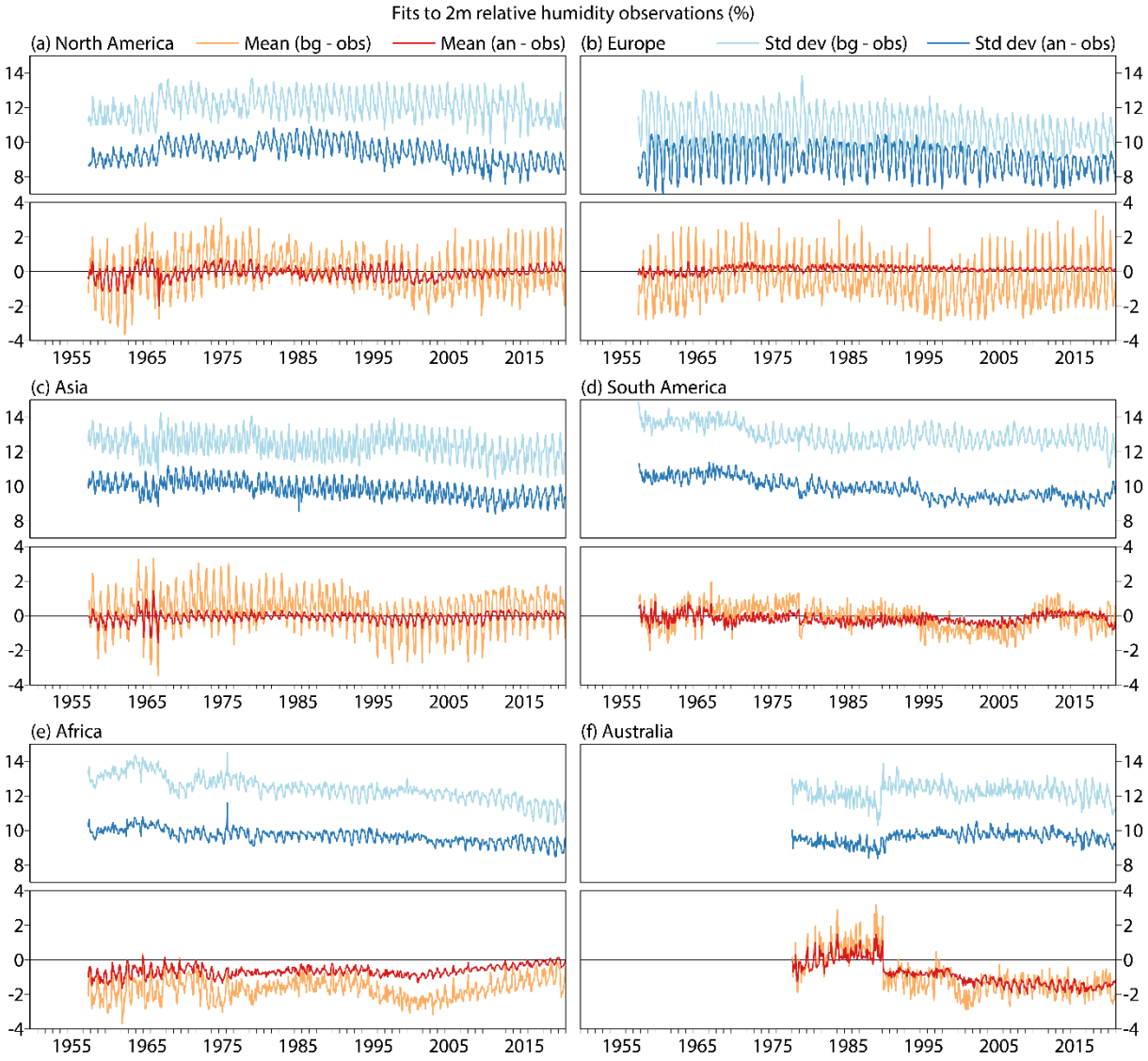

Figure 57 Monthly means and standard deviations of (analysis - observation) and (background - observation) differences averaged over all surface air relative humidity data (%) from land stations for which ERA5 surface-analysis statistics are processed, averaged over (a) North America, (b) Europe, (c) Asia, (d) South America, (e) Africa and (f) Australia. Regions are as specified in the caption of Figure 3.

The background relative humidity over Africa has a negative bias for all months of the year, and this bias increases in magnitude quite sharply between 1990 and 2000, before reducing to a relatively low

magnitude for the latest few years. In this case the analysis departure has a predominantly negative bias, though this is small at the end of the period. It also has a much smaller increase in magnitude between 1990 and 2000 than the background departure.

The performance of the surface analysis scheme again stands out as poorest over Australia. There is a marked change in character of the departure statistics following the changes in observational coverage in 1988 and 1990 discussed in section 3.1.5.1. From 1990 onwards the background departures in relative humidity have a relatively large negative bias and show little or no improvement in recent years. Moreover, the analysis departures have much the same bias as the background departures.

Australia also shows up relatively poorly in time series of the monthly standard deviations of the departures, which are included in Figure 57. The standard deviations are much larger in magnitude than the biases, and the surface analysis scheme ensures that the analysis departures have smaller standard deviations than the background departures. The standard deviations of the analysis departures tend to decrease over time for all regions other than Australia, for which values are larger after 1990 and do not decrease enough in recent years to fall below the values for the mid-1980s.

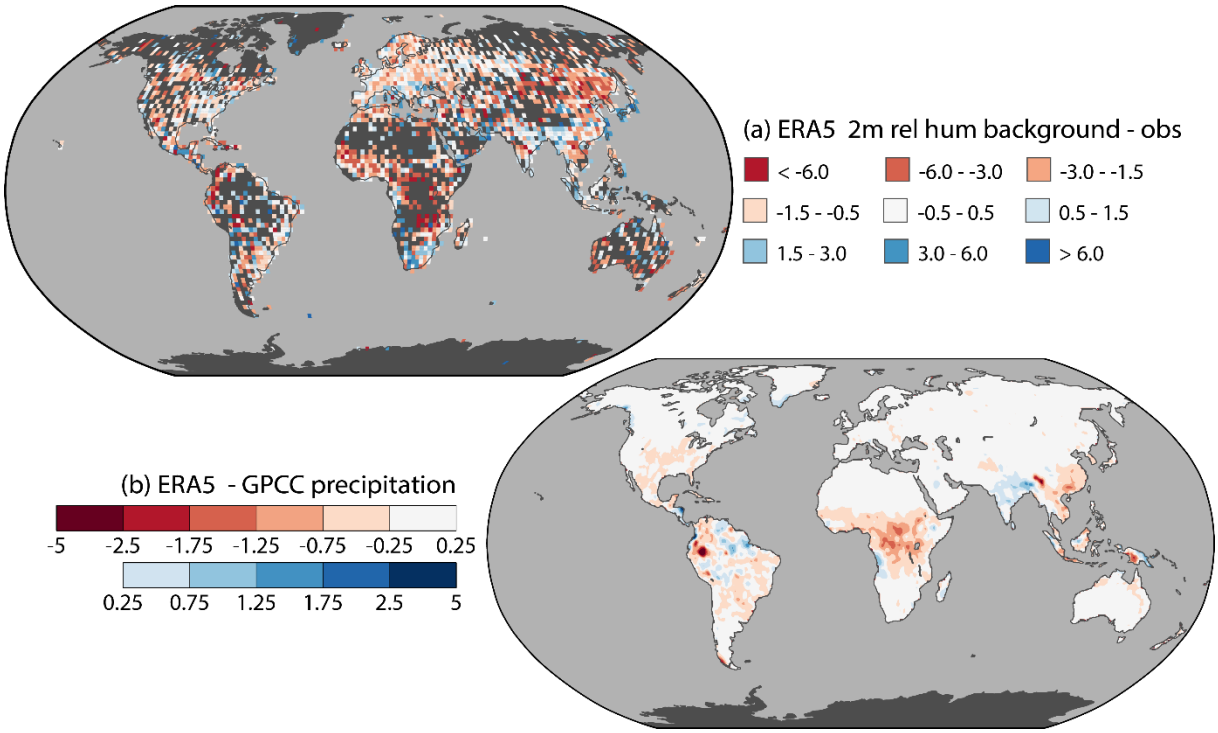

Figure 58 Decadal-mean differences between 2000-2009 and 1980-1989 for (a) ERA5 background departures (background - observation differences) in two-metre relative humidity (%), and (b) ERA5 - GPCC differences in precipitation rate (mm/day). The latter differences are calculated at the 1°x1° spatial resolution of the GPCC data, which are available for all land except Antarctica.

The increase in underestimation of relative humidity over Africa by the ERA5 background forecasts in the 1990s occurs during a period of rising temperatures (Figure 28) when there is good agreement among the various temperature datasets examined here (Figure 29) and time series of the background and analysis departures for ERA5 temperatures show little change over time (not illustrated). The implied increase over time in the dryness of the African-average background forecast can be linked with the decline in African precipitation over this period discussed by Hersbach *et al*. (2020). The decline was pronounced over the Congo Basin, and it is here that the background departures are seen to change most. This is illustrated in Figure 58, which presents maps of the differences between decadal averages for

2000-2009 and 1980-1989 for the ERA5 background departures in relative humidity and for the differences between ERA5 and GPCC precipitation fields. Figure 58 also shows declines in precipitation of ERA5 relative to GPCC over most other areas, parts of tropical South America and India being the main exceptions. It also shows a widespread predominance of more-negative departures in background relative humidity. Other factors may come into play, however, including changes related to the analysis of soil moisture.

4.4 Regional variations in increments

Figure 59 presents global maps showing averages for 1980-1984 and 2014-2018 of the analysis increments in specific humidity over land and sea-ice, for ERA-Interim and ERA5. Corresponding maps for relative humidity are presented in Figure 60.

The increments in specific humidity show the overall moistening illustrated for the average over all land in Figure 53. Although there are local differences, the increments are generally similar between ERA-Interim and ERA5, and between 1980-1984 and 2014-2018, although they are slightly larger for the later period, when observation counts are generally higher. Moistening is largest over tropical South America, West Africa, India and Southeast Asia. Drying occurs in places at middle and high latitudes.

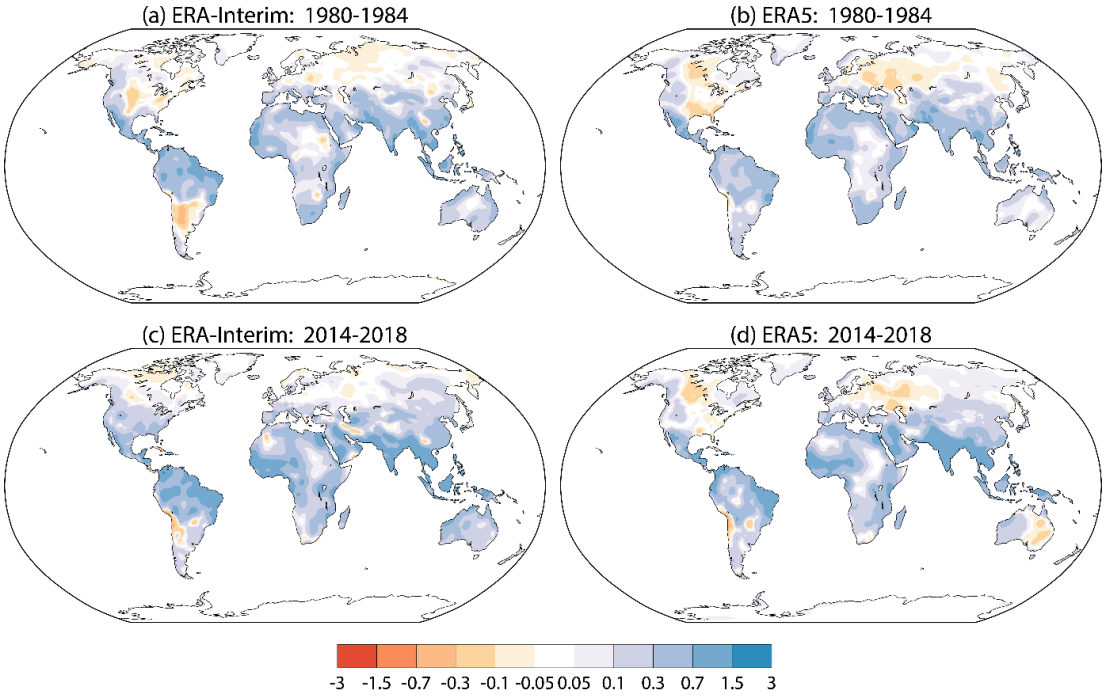

Figure 59 Five-year mean analysis increments in two-metre specific humidity (g/kg) over land and sea-ice from the surface air analyses for: (a) ERA-Interim 1980-1984; (b) ERA5 1980-1984; (c) ERA-Interim 2014-2018; (d) ERA5 2014-2018.

The increments in relative humidity, like those for temperature shown in Figure 5, are generally smaller for ERA5 than for ERA-Interim. The analysis scheme predominantly increases both specific humidity and temperature in the tropics; the associated changes in relative humidity are more variable spatially, and differ more between ERA-Interim and ERA5.

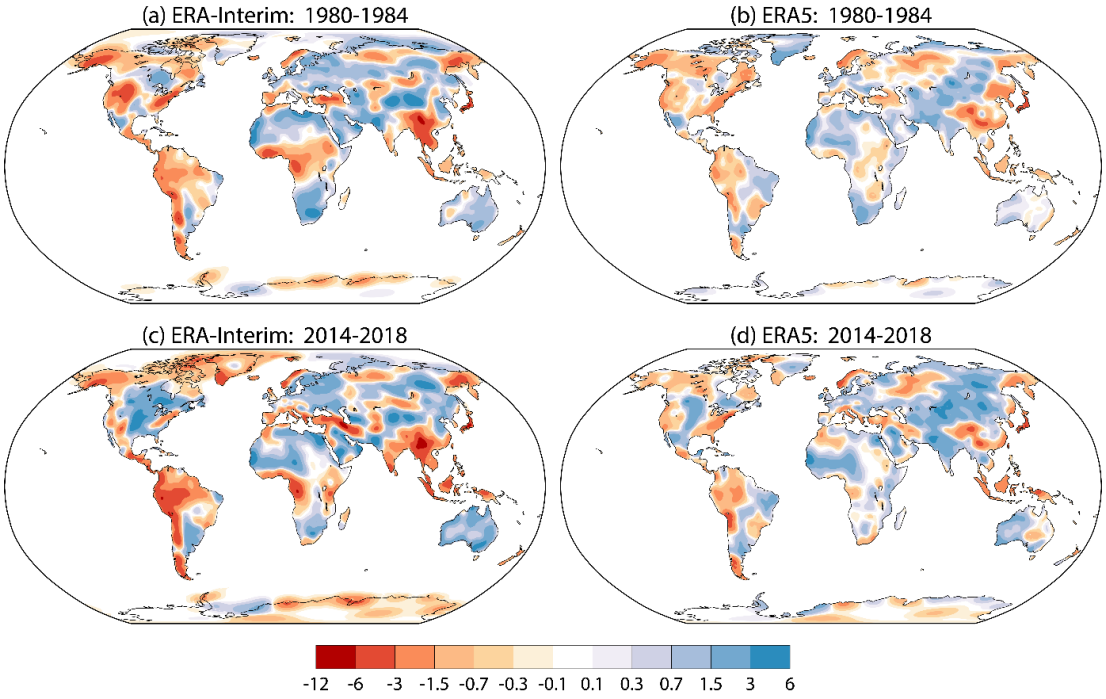

Figure 60 Five-year mean analysis increments in two-metre relative humidity (%) over land and sea-ice from the surface air analyses for: (a) ERA-Interim 1980-1984; (b) ERA5 1980-1984; (c) ERA-Interim 2014-2018; (d) ERA5 2014-2018.

The ERA5 increments in specific and relative humidity for the period 1958-1962 are presented in Figure 61. The very small increments over Brazil and Australia are due to the absence of significant numbers of observations from these countries for this period in the datasets used by ERA5. Elsewhere, the increments are similar to those shown for 1980-1984 and 2014-2018.

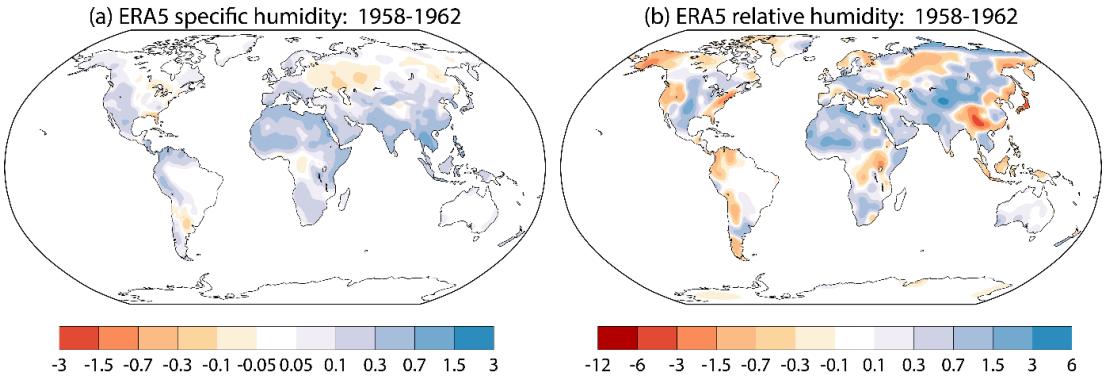

Figure 61 1958-1962 mean analysis increments in two-metre (a) specific humidity (g/kg) and (c) relative humidity (%) over land and sea-ice for ERA5.

4.5 Regional variations in trends

Maps of the local least squares fit linear trends in specific and relative humidity over the period from 1979 to 2018 are presented in Figure 62 for the ERA5 background forecasts and analyses. Evident are the widespread moistening over time in the case of specific humidity, and the decrease over time in relative humidity, as shown respectively in Figure 51 and Figure 52 for averages over all land.

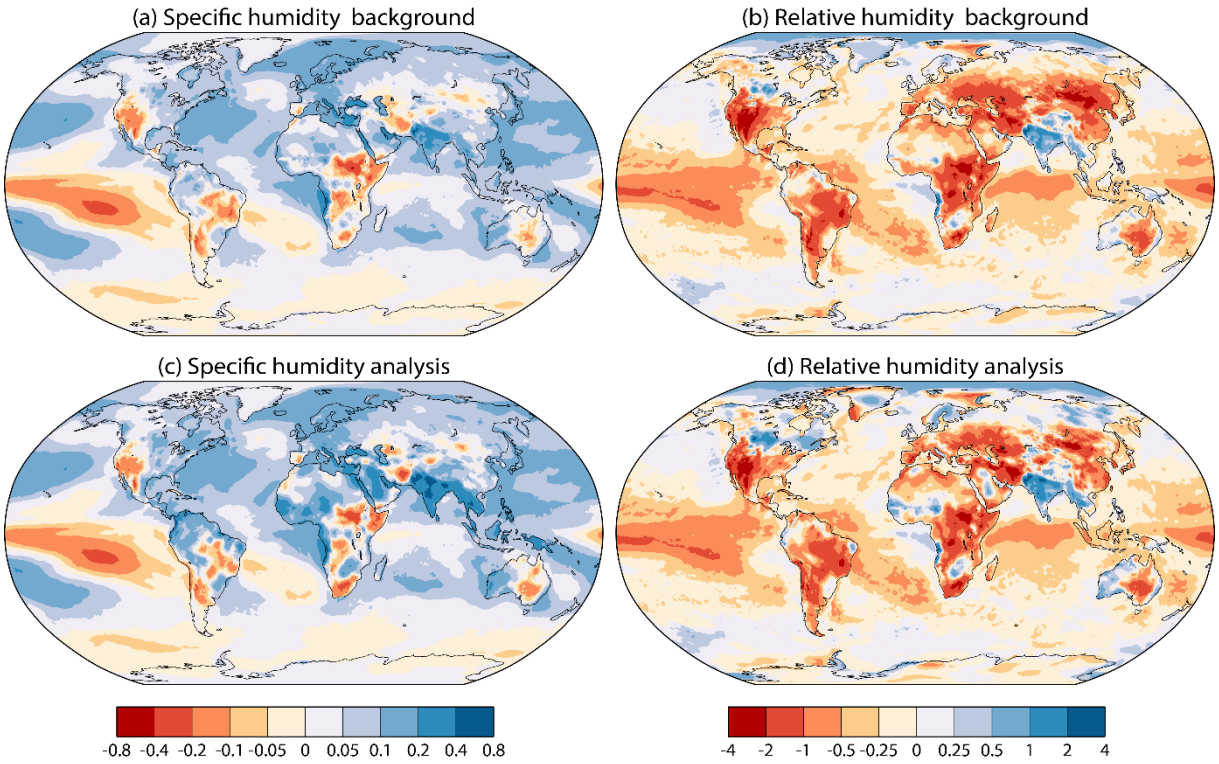

Figure 62 Linear trend in ((a), (c)) two-metre specific humidity ((g/kg)/decade) and ((b), (d)) relative humidity (%/decade) from 1979 to 2018 for: ((a), (b)) the ERA5 background forecasts and ((c), (d)) the ERA5 analyses. All calculations use monthly-mean data with the mean annual cycle for 1979-2018 removed.

Over sea, the drying over the south-eastern Pacific Ocean and over the oceans around Antarctica is consistent with the tendency for cooling in these regions illustrated in Figure 32, although the extent and intensity of the drying over the south-eastern Pacific Ocean is larger than might be inferred from the temperature tendency alone, as indicated by a high rate of reduction of relative as well as specific humidity in this region. Relative humidity decreases over time over other oceans in the tropics and southern subtropics, but it will be seen shortly that this behaviour is less pronounced in other datasets. There is little change over time in marine relative humidity at higher latitudes. One exception is a marked reduction over the Barents Sea, where ice cover has decreased and temperature has risen most rapidly, with lower relative humidity despite increased evaporation and a rise in specific humidity. A second exception is an increase in relative humidity over the Arctic Ocean.

Over land, the trend for decreasing relative humidity is pronounced over tropical South America and central Africa, and widely over the subtropical and middle latitudes of both hemispheres. The specific humidity trend in these regions is for drying or only weak moistening in many places.

Most of these changes over land in specific and relative humidity are features of the ERA5 background forecasts, rather than direct consequences of changes made by the surface analysis. The overall trend for decreasing relative humidity is nevertheless a little stronger in the background forecasts than in the analyses, as seen for example in the smaller rate of decrease over east Asia and the larger rates of increase over Canada and India. This overall result is consistent with what is shown in Figure 53 for the average increment over land, although that figure shows a reduction in increment after the year 2000 that suggests that this finding may not hold for future years. Trends over sea are the same for the background forecasts and analyses, as the surface analysis scheme is not applied over sea.

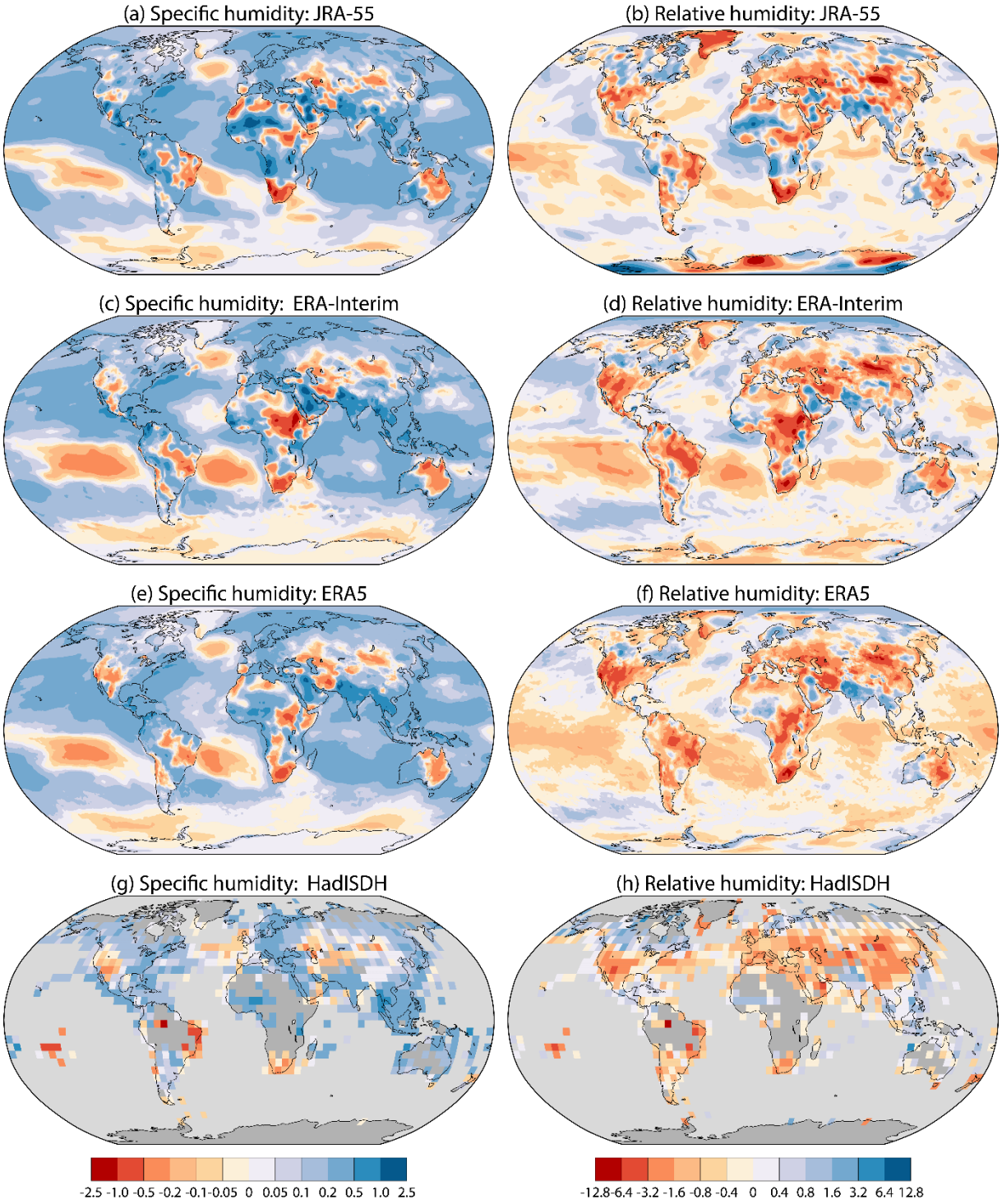

Figure 63 Averages for 2011-2018 of two-metre specific humidity (g/kg) for (a) JRA-55, (c) ERA-Interim, (e) ERA5 and (g) HadISDH, expressed as anomalies with respect to 1981-2010. Corresponding anomalies in two-metre relative humidity (%) are shown in panels (b), (d), (f) and (h) respectively.

As HadISDH provides data in the form of anomalies relative to 1981-2010, changes in surface air humidity from ERA5 are compared with those from other datasets in terms of the seven-year average anomalies for 2011-2018 and 1973-1980. Missing data are treated strictly: values for HadISDH are presented only for grid squares for which there is no missing data for the two periods. Comparison of

the HadISDH panels in Figure 63 and Figure 64 shows that HadISDH has better marine coverage for 1973-1980 than for 2011-2018. The converse is the case over land.

Figure 63 presents maps of the mean anomalies in specific and relative humidity for 2011-2018 from the JRA-55, ERA-Interim, ERA5 and HadISDH analyses. Comparison with Figure 62 in the case of ERA5 shows that the mean anomaly for 2011-2018 has much the same pattern as the linear trend for 1979-2018. Moreover, the three reanalyses are in general agreement, as is HadISDH where data are available.

There are nevertheless some substantial regional differences between the datasets. JRA-55 is quite different from the more similar ERA-Interim and ERA5 with regard to relative humidity in both the Arctic and the Antarctic, and there are differences between all three reanalyses over Africa. HadISDH provides few data values for these regions. The reanalyses are in quite reasonable agreement with each other over Australia, but not with HadISDH. HadISDH provides few data over sea, but those provided indicate drier conditions than indicated by the reanalyses, as illustrated in Figure 51 for the average over all its values compared with the corresponding average from ERA5.

Figure 64 compares specific and relative humidity anomalies averaged for 1973-1980 from ERA5 and HadISDH. Consistent with the trends over later decades, specific humidity is mostly lower for 1973-1980 than it is for 1981-2010, and relative humidity is mostly higher. In this case there is better agreement between ERA5 and HadISDH as regards humidity over sea. Agreement between the two datasets is also quite reasonable over land in the northern hemisphere, but there are discrepancies over the southern hemisphere, not only over Australia as before, but also over southern South America.

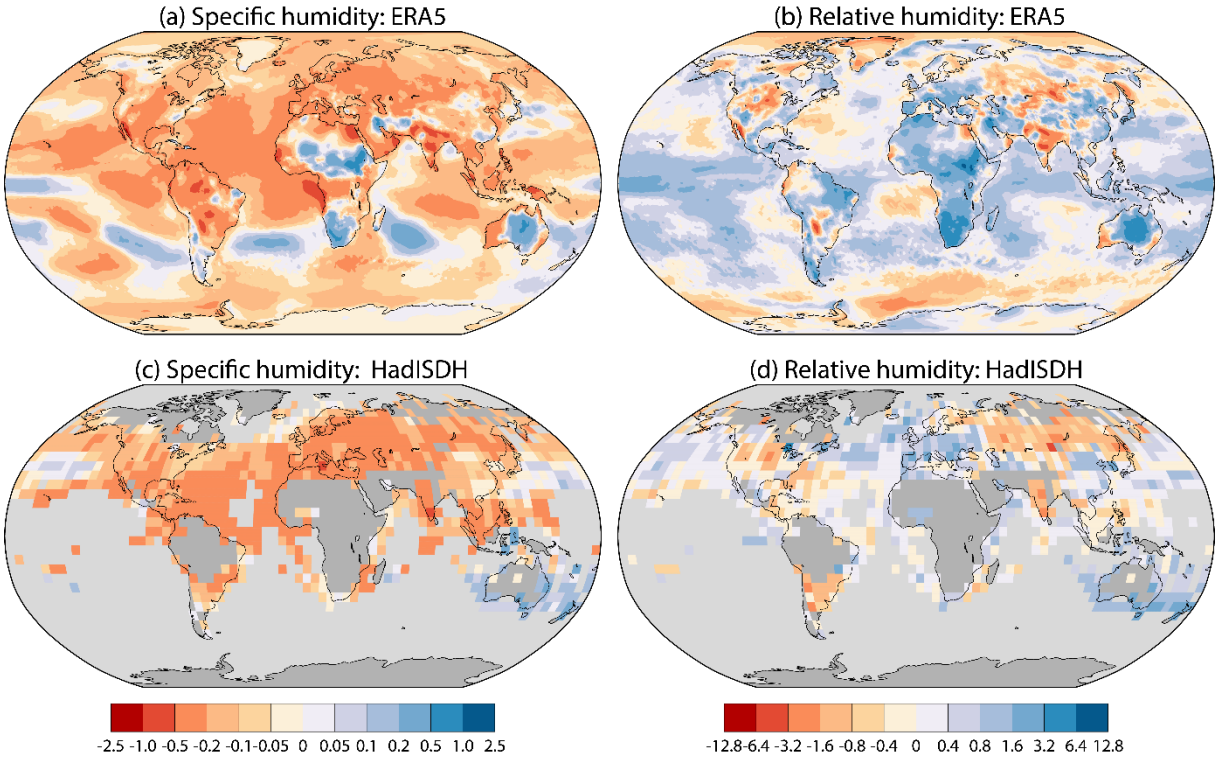

Figure 64 Averages for 1973-1980 of two-metre specific humidity (g/kg) for (a) ERA5 and (c) HadISDH, expressed as anomalies with respect to 1981-2010. Corresponding anomalies in two-metre relative humidity (%) are shown in panels (b) and (d) respectively.

4.6 Use of surface air humidity observations in the ERA5 4D-Var analysis

The relative humidity data derived from observations of dry-bulb and dew-point temperatures are used in two ways in ERA5. The discussion in preceding sections has been concerned with the OI analysis of the value at two-metre height that provides the analysis product disseminated for ERA5 and earlier ECMWF reanalyses. In addition, the observations over land for daytime hours are assimilated in 4D-Var, and thereby influence the analysis at model levels, which start at ten-metre height and extend upward into the boundary layer and then the free atmosphere. These data are assimilated in the conventional sense of the word in that the values used during one analysis cycle directly influence the atmospheric state used to start the background forecast for the next cycle. This is in contrast to the OI analyses of relative humidity and temperature, which are not used to adjust the initial atmospheric state for the background forecast, but which do influence the background forecasts indirectly, through their use to adjust the soil temperature and moisture of the initial state.

Figure 65 presents time series of monthly means and standard deviations of the background and analysis departures for all the daytime observations used in the ERA-Interim and ERA5 4D-Var analyses, taken from the feedback statistics archived for 4D-Var. In this case the analysis is not the OI surface analysis for two-metre relative humidity, but rather the relative humidity at two metres derived from the model-level 4D-Var analysis of relative humidity. The fit of this 4D-Var analysis to observations is accordingly much less close than provided by the OI analysis. Although the temporal sampling of the data presented in Figure 65 is very different from that for the departure statistics presented earlier for the OI analysis, the background and analysis fits again show the ERA5 relative humidities to have their most pronounced low biases in the 1990s and 2000s.

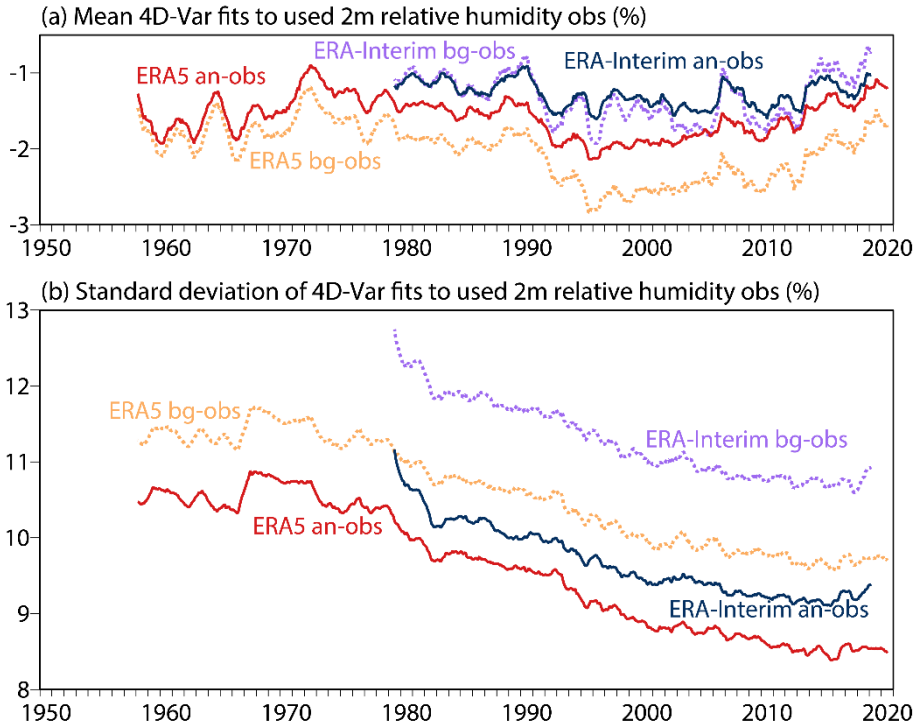

Figure 65 12-month running averages from September 1957 to 2020 of monthly (a) means and (b) standard deviations of (analysis – observation) (solid) and (background – observation) (dotted) differences averaged over all surface air relative humidity data (%) from land stations used in the ERA5 and ERA-Interim 4D-Var analyses.

The 4D-Var background has a larger negative bias for ERA5 than for ERA-Interim in the average over all observations used in the upper-air analysis, as discussed in section 4.3 in the case of the whole set of relative humidity data. Panel (a) of Figure 65 shows that the 4D-Var assimilation changes the bias only to a small extent for ERA-Interim, but reduces its magnitude quite substantially for ERA5, especially in the period when the background is most biased. The mean 4D-Var analysis departures are nevertheless larger for ERA5 than ERA-interim.

A different picture is seen for the standard deviations of the departures presented in panel (b) of Figure 65, which are substantially lower for ERA5 than ERA-Interim. Standard deviations of the ERA5 background between 1957 and 1979 are around the same level as those of the ERA-Interim background for the 1990s. Differences between the two reanalyses are smaller in the case of analysis departures. Standard deviations reduce over time for both reanalyses. Observation error must be kept in mind, as the prescribed values limit the extent to which the analyses are constrained to fit the observations. The standard deviation of observation error for relative humidity (in %) is specified in 4D-Var to be a function of temperature, limited to a range from 6% to 18%. Monthly averages vary from just below 10% in boreal summer to around 12% in boreal winter for ERA5. The OI surface analysis of relative humidity uses a fixed value of 10% for this error.

4.7 Local surface-analysis issues for ERA5

4.7.1 Excessive warm, dry anomalies in data-sparse regions

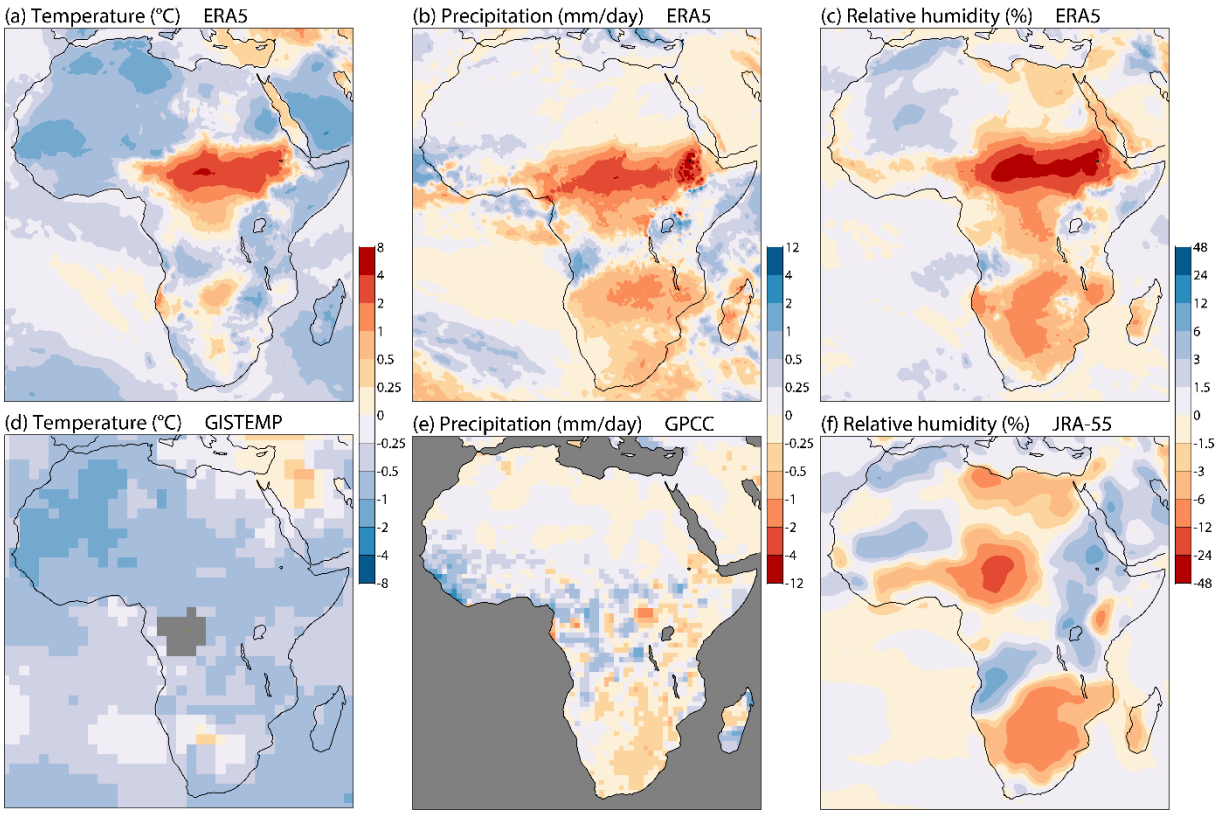

Figure 66 Anomalies for 1965-66 relative to 1981-2010 of surface temperature (°C) from (a) ERA5 and (d) GISTEMP, of precipitation (mm/day) from (b) ERA5 and (e) GPCC (at 1°x1° spatial resolution), and of two-metre relative humidity (%) from (c) ERA5 and (f) JRA-55.

Particularly high temperatures of ERA5 compared with all other datasets have been noted in section 3.4.2 for averages over Africa for 1965-66 and South America for 1961. They are accompanied by low precipitation and dry near-surface air. Although less evident in continental averages, a similar feature is found for part of Africa in the early 1950s. The three cases are illustrated here, comparing values from ERA5 with temperature from GISTEMP, precipitation from GPCC and (for two of the cases) relative humidity from JRA-55.

Figure 66 shows averages for 1965-1966 over Africa. In this case there is a zone extending from Nigeria to Ethiopia where ERA5 temperatures are up to 4ºC higher than the 1981-2010 average, in stark contrast to the below-average temperatures for the region shown by GISTEMP. ERA5 precipitation anomalies of several mm/day over the zone are not supported by the values from GPCC. Relative humidity from ERA5 is substantially lower than average over the zone; JRA-55 has below-average relative humidity over the west of the zone, but not over the east.

Agreement among datasets is better over other parts of the continent, in particular as regards the relatively cold temperatures in the north west, above-average precipitation over the western Sahel and dry conditions over southern Africa.

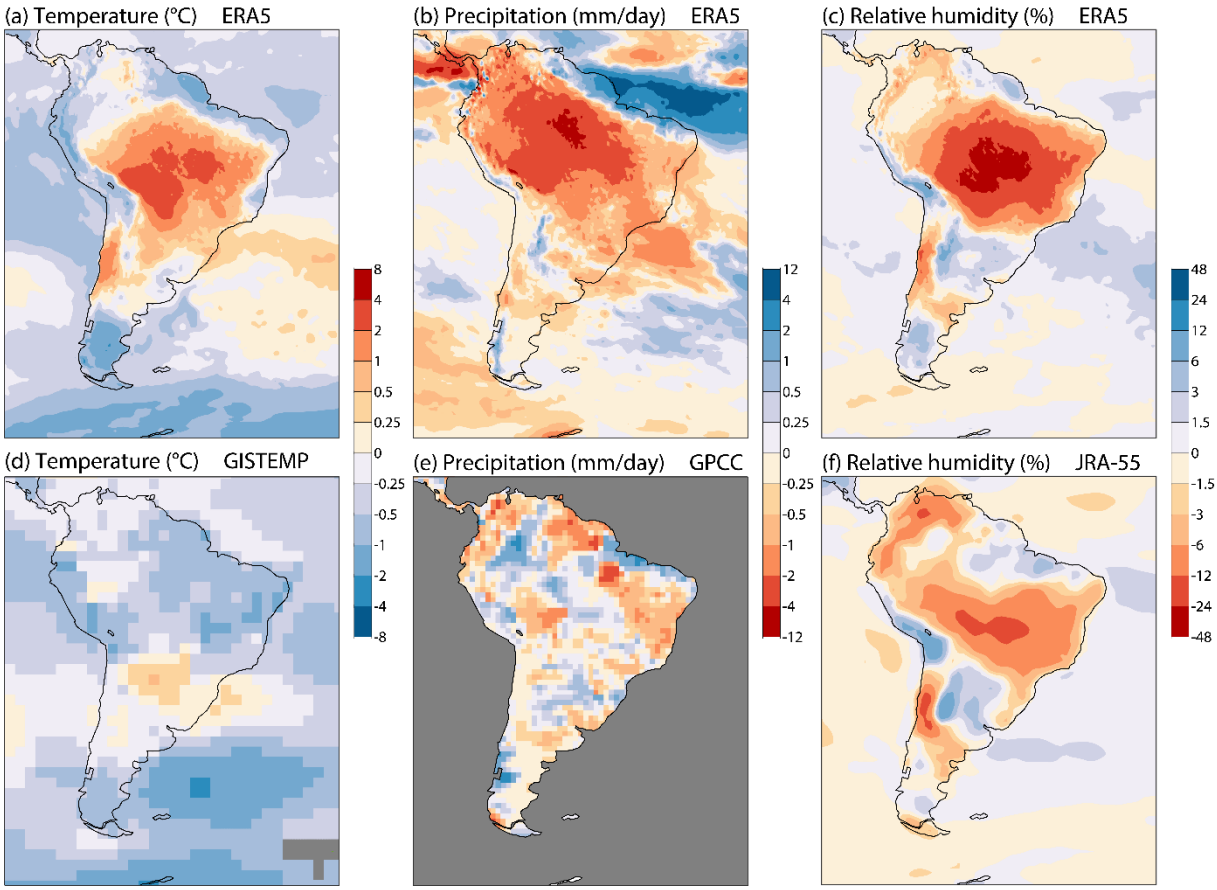

Figure 67 As Figure 66, but for the year 1961 over South America.

Figure 67 presents the corresponding maps for 1961 averages over South America. In this case the problematic region covers much of Brazil, and Peru and Bolivia east of the Andes. Here the ERA5 temperature is again much higher (relative to 1981-2010) than that from GISTEMP. Although there is some support from GPCC and JRA-55 for drier than average conditions over the region, the ERA5

anomalies in precipitation and relative humidity appear to be too widespread and too intense. ERA5 is more consistent with the other datasets as regards conditions over the south of the continent.

Maps for the third example are presented in Figure 68. They are for Africa again, but the averages are taken over the first three years of ERA5, from 1950 to 1952. In this case the excessively warm and dry regions of ERA5 are located over the Congo Basin and to the south. Compensating and equally dubious cold, wet conditions are indicated by ERA5 to the east and north-east, with the largest anomalies over the Sudan.

Although observational coverage is lacking over the problematic region of Africa in the 1965/66 case and over Brazil in 1961, this is less evidently so for this 1950-52 case. Background temperatures are biased high compared with observations over the Congo Basin in the early 1950s, where the relatively high analysed temperatures occur despite cooling analysis increments. Too-warm and too-dry conditions, with cooling analysis increments are also characteristic of the averages for Australia in the 1950s and 1960s discussed in section 3.4.2.

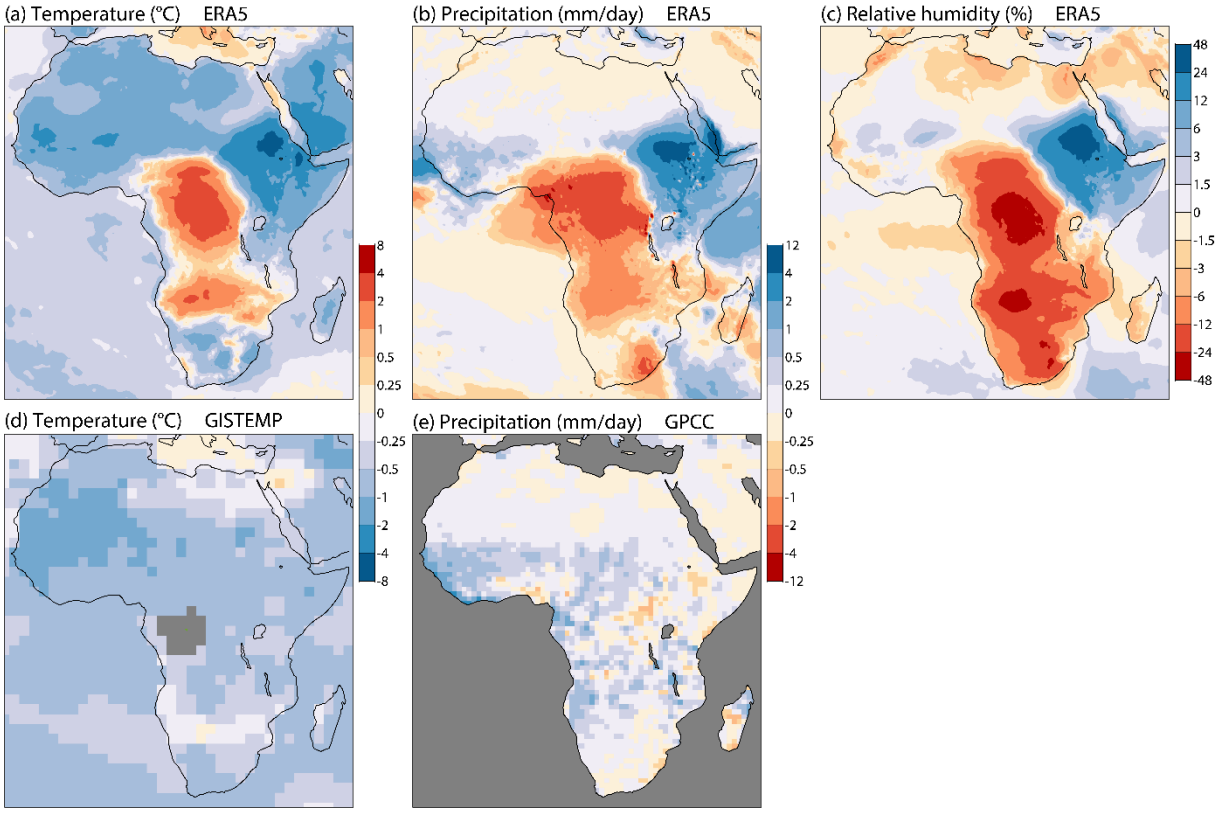

Figure 68 As Figure 66, but for 1950-52. JRA-55 data are not available for this period.

An issue that relates to the periods and locations of excessive warmth and dryness seen in ERA5 concerns the initialization of soil moisture, particularly for the lower layers of the ERA5 land-surface model. Figure 69 presents the volumetric water content anomalies of each of the four model layers, averaged over continental regions. It shows that the pre-1979 ERA5 production streams were generally started with water contents that were much lower than those typical of the period from 1979 onwards. The response is a moistening that is generally slower for the deeper layers, on a timescale that varies from continent to continent. Sharp changes occur when the production streams change, something that also happened in ERA5 production at the beginning of 2015.

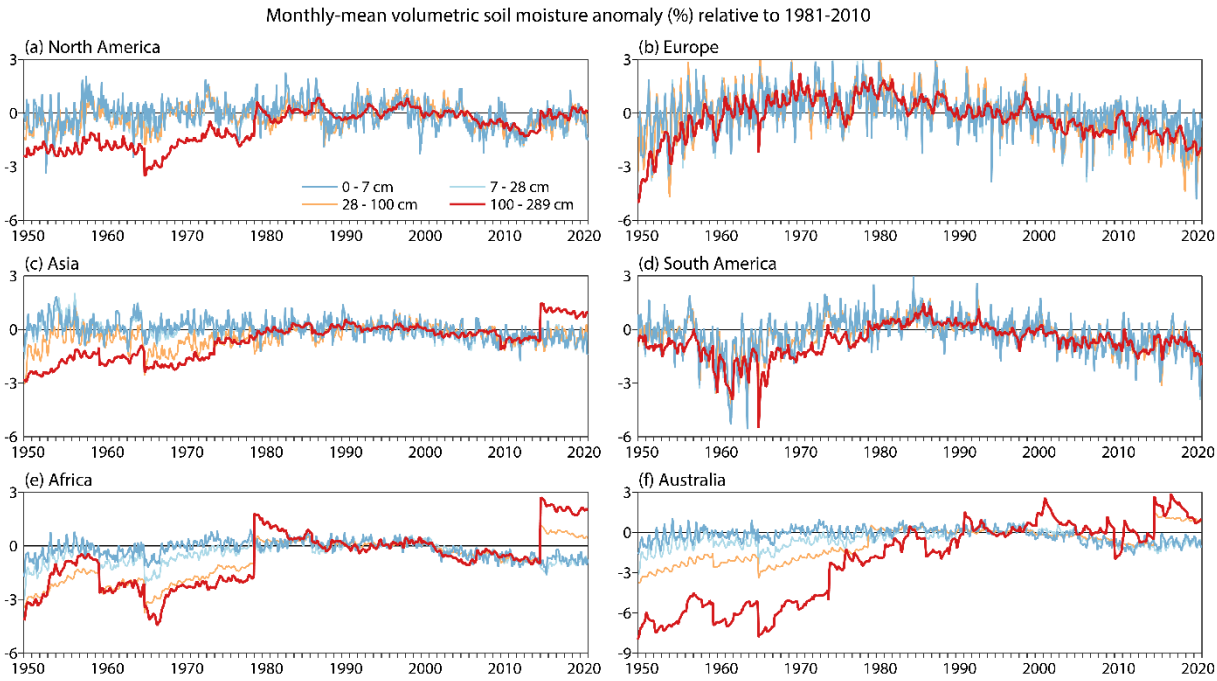

Figure 69 Monthly-mean volumetric soil moisture anomalies (%) relative to 1981-2020 for the four layers of the ERA5 land-surface model, averaged over the continental areas defined in the caption of Figure 3, from 1950 to 2020. The depths of the layers are shown in the legend.

This problem is least marked for Europe, which appears affected mainly in the early to mid-1950s. There are some later dips in European soil moisture that appear to be associated with production-stream boundaries, but there are also fluctuations in the 1960s and 1970s that are likely associated with sharp changes in precipitation over the continent. These precipitation changes are shown by Bell *et al*. (2021) to be common to ERA5, JRA-55 and GPCC.

The initialisation of soil moisture is most problematic for the deepest soil layer for Australia, and the deepest two layers for Africa. The uppermost layer shows little problem in general, but its soil moisture is relatively low for Africa in 1965 and 1966, when the start of a new production stream coincides with deficient observational cover. Near-surface soil moisture is also lower than the 1981-2010 average in the 1950s and 1960s over Australia, and from 1950 to 1952 over Africa.

Although the too-warm periods over Africa and Australia might be due to or exacerbated by the dryness of the deep soil, the situation is different for South America. Here there are pronounced minima in the moisture of the deepest soil layer associated with the changes in production stream in 1959 and 1965, but they are short-lived, and the minimum in 1961 is larger for the topmost layer than the deep layer, suggesting that the behaviour of soil moisture is more a response than a cause in this case.

Soil moisture was initialised at the start of the pre-1979 ERA5 production streams using scaled values derived from earlier comprehensive reanalyses. The related ERA5-Land product (Hersbach *et al*., 2020) improved on this by using climatological soil moisture derived for the post-1979 period. It provides the basis for better initialisation of soil moisture for future comprehensive reanalyses, and may be updated for this purpose should the land-surface model change significantly.

4.7.2 Assimilation of METAR data

The observation counts presented in Figure 50 show an annual cycle in the number of humidity data used in the surface analysis from 2004, after data from airports transmitted in METAR code began to

be analysed. No such cycle is seen for temperature data. This behaviour is most marked in averages over Europe and North America. Further detail for these regions is given in Figure 70, which shows time series related to relative humidity observations from 2004 to the present. Mean background and analysis departures and counts of used data are shown separately for manual and automatic observations transmitted in SYNOP code and for observations transmitted in METAR code. This information is again drawn from the archived feedback data from the surface analysis, so relates mainly to morning and evening data over Europe, and night-time and daytime data for North America.

For Europe, mean background departures for both types of SYNOP data exhibit an annual cycle, with higher background relative humidity than observed in spring, and lower values at other times of the year, as shown earlier in Figure 56. The analysis is close to both automatic and manual observations, with a slight separation between the two that could be due to a geographical difference in the locations of the two types of observation rather than a systematic, if small, difference in the measurements themselves. There is barely any drift in the departure statistics, despite an increase over time in the number of automatic observations and a decrease in manual observations.

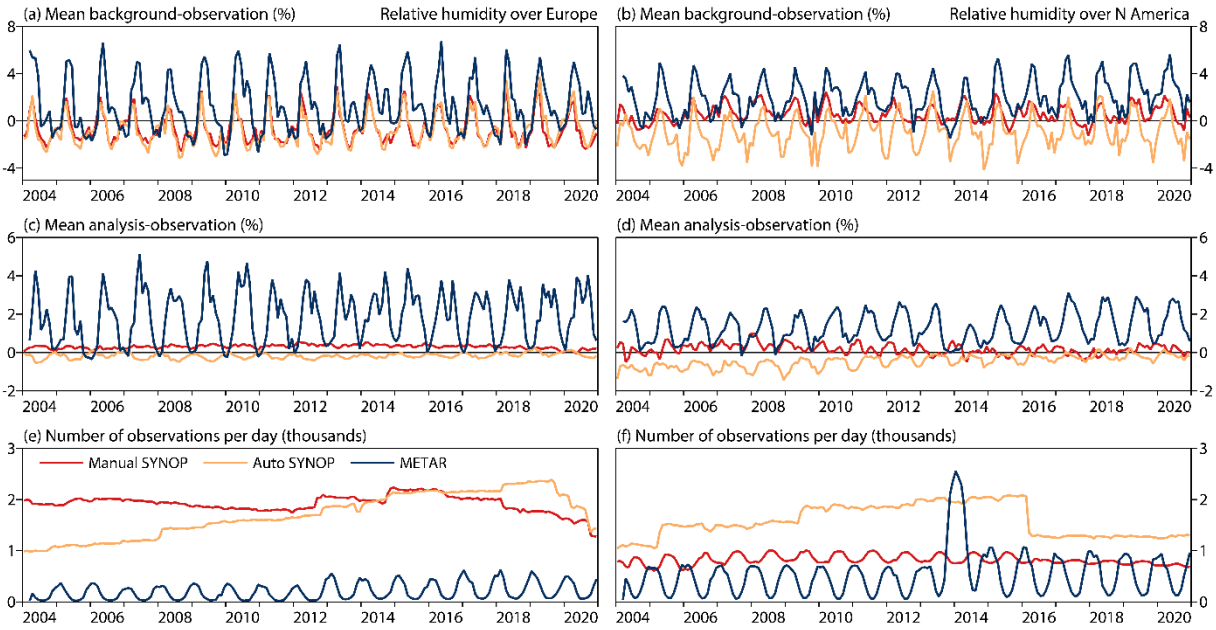

Figure 70 Time series from 2004 to 2020 of mean (a, b) background departures, (c, d) analysis departures and (e, f) numbers of observations of relative humidity over Europe (left) and North America (right), for manual SYNOPs (red), automatic SYNOPs (orange) and METARs (blue) used in the surface analysis.

There is a greater difference between the background departures for automatic and manual SYNOPs over North America, where there is a less homogeneous distribution of the two types of observation than in Europe, due to differences in the preponderance of automatic rather than manual observation between Mexico, the USA and Canada. The analysis fits show more variation of over time than over Europe, and larger bias in the case of automatic observations prior to the latest few years.

The statistics for METAR data are comparable in mid-winter with those for SYNOP data, for both regions. This is despite the reporting of temperature and dew point in whole numbers of degrees Celsius in METAR code, which limits the accuracy with which relative humidity can be derived. Few METAR data are used in mid-summer, but those that do pass data selection and quality-control checks in spring and summer have, on average, large mean background departures. These observations influence the

analysis locally, as the analysis departures are smaller than the background departures, but as relatively few observations are used, they are unlikely to cause a general degradation of the analysis, and no such degradation is suggested by the fits of the analysis to SYNOP data. Standard deviations of the departures are also larger for METAR data than for SYNOP data over Europe, especially in summer, but they are similar for METAR and SYNOP data for North America.

An explanation for this behaviour, which also occurred for some time in ECMWF operations, was found shortly before publication of this report. The quality control of relative humidity observations in the land surface analysis is at fault. It is affected in the version of the analysis used for ERA5 by a bug involving poor handling of alphanumeric METAR station identifiers that results in a spurious setting of a 275K dry-bulb temperature for the quality control of humidity data. This setting is intended to be used only for non-SYNOP snow reports. It in turn leads to widespread occurrence of larger dew-point than dry-bulb temperatures in summer, a criterion for data rejection. Among the changes made to ECMWF operations with the implementation of version CY46r1 of the IFS in June 2019 was one that solved the problem. This change will be considered for implementation in the near-real-time production stream of ERA5.

Figure 71 presents the corresponding picture for temperature. In this case the number of METAR data used does not vary with season. Background and analysis departures for the most part have larger biases for the METAR data than for either manual or automatic SYNOP data, but not to an extent that causes usage of the data to be questioned. Standard deviations of the departures are similar for METAR and SYNOP data for Europe, and a little smaller for METAR data than for SYNOP data for North America.

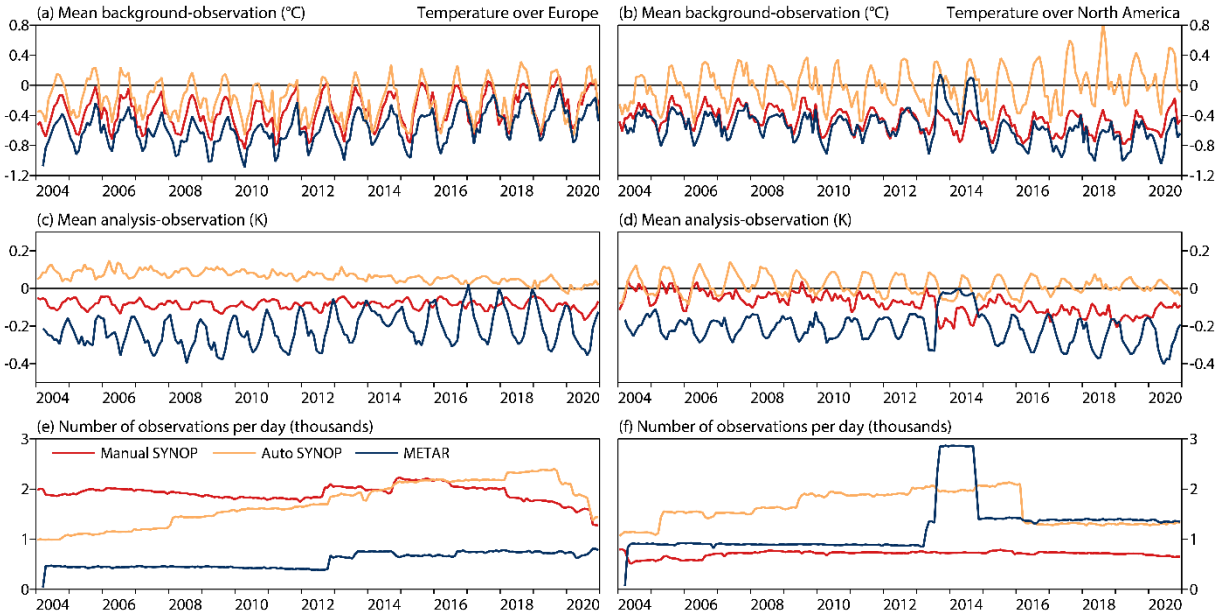

Figure 71 As Figure 70 but for temperature.

4.7.3 An uncertainty over China

Li *et al*. (2020) (see also Freychet *et al*., 2020) investigated biases in surface air humidity observations over China. They argued that ERA5 (and ERA-Interim) have a spurious trend in humidity due to assimilating observations whose bias changed due to a shift from manual to automatic synoptic observation in the early 2000s. The authors worked with a dataset of daily-mean observations from a dense network of 756 stations, made available by the Chinese Meteorological Administration (CMA). Time series from 643 of the stations were assessed to be inhomogeneous due to the change to automatic

observation. Some of the inhomogeneity may have come from changes to the daily average resulting from the more frequent (hourly) sampling that came in with automatic observation rather than directly from the change in instrumentation. It was nevertheless concluded that a decrease in two-metre relative humidity in ERA5 in the early 2000s was a result of analysing data whose bias changed.

Evaluation of the data used by ERA5 and the performance of its analysis scheme provides evidence that contradicts the above conclusion.

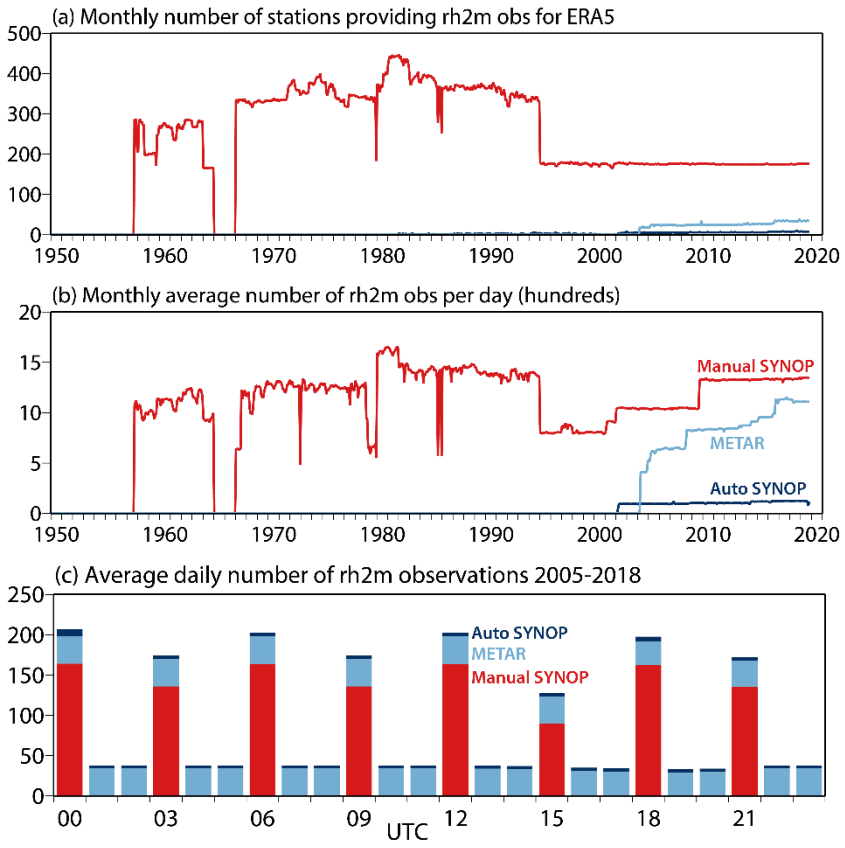

Figure 72 Aspects of ERA5's two-metre relative humidity analysis averaged over southern and eastern China (105-122ºE; 22-40ºN): (a) Monthly number of stations providing data; (b) Monthly average number of data per day (hundreds); (c) Average number of observations per hour for 2005-2018. Results are shown separately for observations coded as Manual Land SYNOPs (red), Automatic Land SYNOPs (dark blue) and METARs (light blues).

Figure 72 presents monthly statistics of the numbers of observing stations and two-metre relative humidity observations provided operationally by CMA and processed by ERA5. Averages are taken over southern and eastern China (105-122ºE; 22-40ºN), where observational density is highest in the dataset studied by Li *et al*. (2020). Time series are plotted only to October 2019, after which date the picture is complicated by the transition to BUFR coding for synoptic observations. Only a simple quality control that rejects relative humidities outside the range 0-120% is applied; neither the 4D-Var data selection nor the more stringent quality control applied in the OI surface analysis is activated. Values are shown separately for SYNOPs that are coded as being made from manual and automatic stations, and for METARs. In general, data from the stations coded as automatic are provided hourly, as are the METARs. The SYNOPs coded as from manual stations generally come three-hourly in recent years, as seen in panel (c) for the average from 2005 to 2018, with a slightly higher number of observations at the

main hours of 00, 06, 12 and 18UTC and fewest observations at the late-evening hour of 15UTC. A higher proportion of data were reported six-hourly in earlier years.

The absence of observations before September 1957 and in 1965 and 1966 is evident in Figure 72. The drop in the number of observations (but not stations) for most of 1979 reflects ECMWF's failure in the 1980s to archive observations at 03, 09, 15 and 21UTC from its FGGE data holdings. Data from some stations are missing in December 1979, however, and also in August and December 1985. This has been shown in Figure 50 to be the case for other regions also. There is also a sharp drop in the number of stations and data counts at the end of 1994. This is larger than seen for some other regions, and indicates that NCEP had access to considerably more data from southern and eastern China around this time than ECMWF received operationally on the GTS. Data from around 200 stations have been analysed by ERA5 for the period from 1995, far fewer than in the dataset used by Li *et al*. (2020).

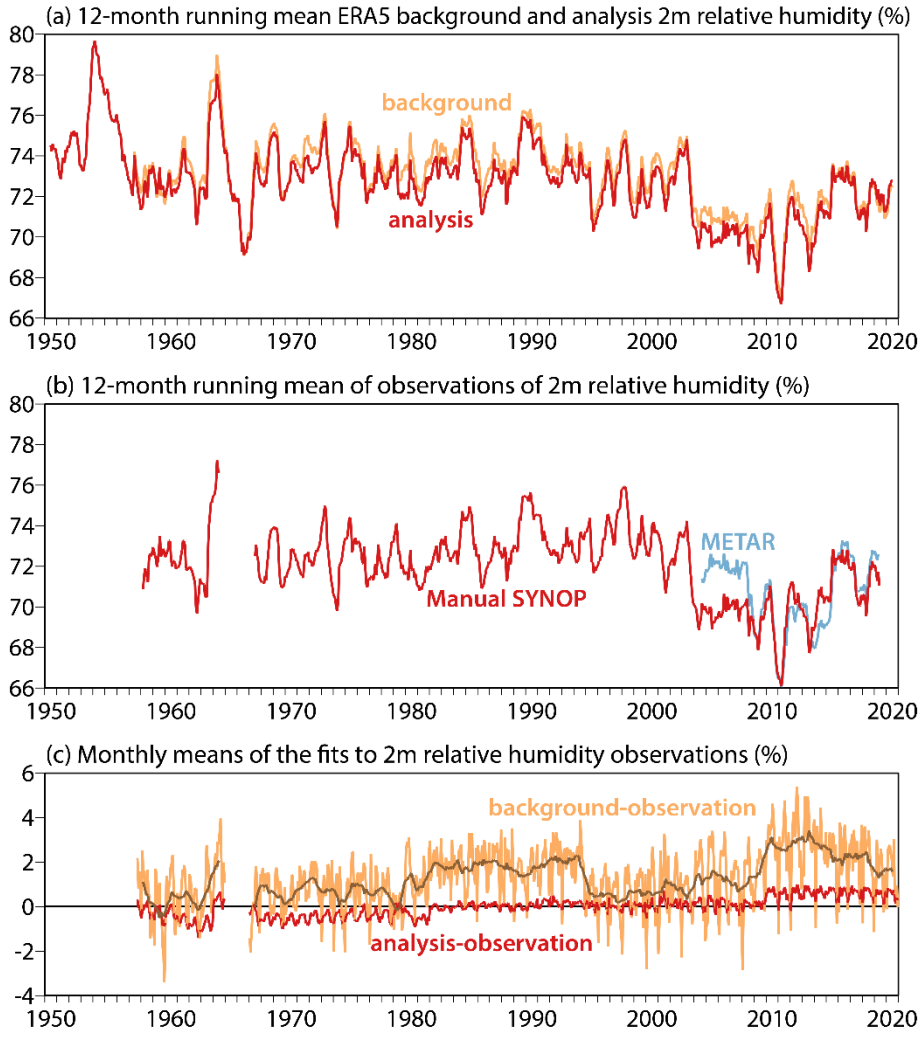

Figure 73 Aspects of ERA5's two-metre relative humidity analysis averaged over southern and eastern China (105-122°E; 22-40°N): (a) Twelve-month running means of the background (orange) and analysis (red); (b) As (a), but average over all reported manual SYNOP (red) and METAR (blue) observations; (c) Monthly mean fits of the background (orange) and analysis (red) to the archived subset of observations, and the twelve-month running mean of the background fit.

Panel (a) of Figure 73 presents values of the ERA5 background and analysis, averaged over the same region of southern and eastern China. The analysis values indeed fall quite suddenly near the end of

2003, just before the analysis scheme starts to be presented with METAR data. The monthly average reaches a minimum in March 2011, and rises thereafter to reach a value for the last five or so years that is not much below that for the second half of the 1990s. The background values exhibit similar variations. The drop near the end of 2003 is a fraction smaller: the analysis increment reduces relative humidity a little more between 2004 and 2013 than previously, but the difference is only around 0.02%. The increment is particularly small for the most recent years.

Panel (b) shows the corresponding time series of the averages over all observations available to ERA5 that were reported as Manual Land SYNOPs or METARs. The number and quality of the SYNOP observations and their distribution in space, and the closeness of the fit of the analysis to the observations, are sufficient for the direct average of the observations to be very similar to the area-average of the analyses. In particular, the observations reported to be manual show the same decrease starting in late 2003 as the analysis and background fields. The averages over the METAR data are quite similar after the increase in the numbers of these data in mid-2008, but an even smaller proportion of these data is used in the surface analysis than over North America and Europe. No METAR data for relative humidity is used in many summer months.

Corresponding background and analysis departures from the statistics produced by ERA5's surface analysis scheme are presented in panel (c) of Figure 73. The largest changes in the background departures occur when the counts of manual observations change at the end of 1994 and in September 2009. This result appears be a consequence of changes in sampling rather than of fundamental changes in the quality of the background, as there are no significant changes in the analysis increment at these times. The analyses fit the observations very closely in the mean from 1982 to 2009, but the mean departure increases a little when observation counts rise in 2009. Background departures are a little higher for a few years after 2003 than for the preceding few years, but the difference is at most a percentage point or so. Analysis departures remain close to zero over these years.

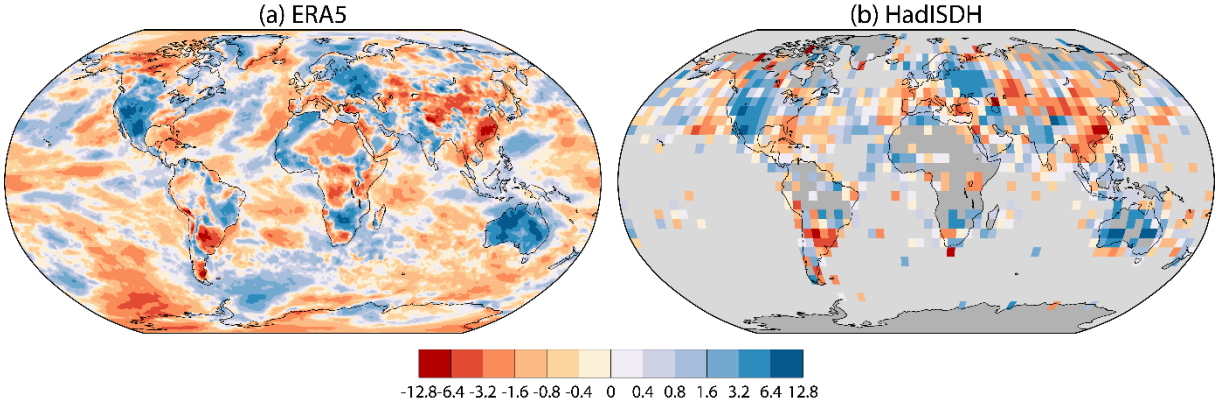

Figure 74 Average difference between 2004 and 2002 in two-metre relative humidity (%) for (a) ERA5 and (b) HadISDH.

Global maps of the mean differences in relative humidity between the years 2004 and 2002 for ERA5 and HadISDH are presented in Figure 74. Both datasets show a relatively large decrease in relative humidity from 2002 to 2004 over southern and eastern China, but less change, including some increase, over the rest of China. Agreement between the datasets is reasonably good elsewhere with regard to the larger changes. Several other regions show either increases or decreases of the same order of magnitude as seen over southern and eastern China.

In summary, the relative humidity observations analysed by ERA5 are reported to have come mainly from manual observation, and there is no indication that the behaviour of ERA5's surface analysis

scheme changes significantly in the early 2000s in direct response to a change in the bias of the observations. Instead, the decrease in relative humidity at that time is a property common to the background forecast and the analysed observations. Had the observed decrease been dominated by a change in measurement bias, it would be expected that the background forecast would have shown less of a change, although some influence via the soil moisture analysis could have been felt. Soil moisture does decline over the region in question, but precipitation also declines, relative to values from GPCC, as can be seen from the map presented as panel (b) of Figure 58. The decrease in relative humidity over southern and eastern China between 2002 and 2004 is seen not only in ERA5 but also in HadISDH, a dataset whose production employs homogenisation to reduce impacts of detected changes in measurement biases.

4.7.4 Problematic humidity over the Great Lakes

Some examples of areas with problematic surface air temperatures were presented in section 3.5. It is expected that such areas will also be problematic for surface humidity. This is illustrated here for the Great Lakes, whose temperatures in ERA5 were erroneously low in winter and high in summer, prior to 2014. The associated biases in surface air humidities result in climatologies for 1981-2010 that are unrepresentative of conditions over the Great Lakes from 2014 onwards. As a consequence, maps of recent anomalies in surface air humidity exhibit spurious features there.

Figure 75 provides an illustration. It presents anomalies relative to 1981-2010 of specific and relative humidity for February 2018 and July 2018. The issue is pronounced in summer. The maps for July show anomalously high relative humidity over almost all the major lakes, but low specific humidity, especially for Lake Superior. The issue is less obvious in winter, but the maps for February show a slightly high anomaly in specific humidity and low anomaly in relative humidity.

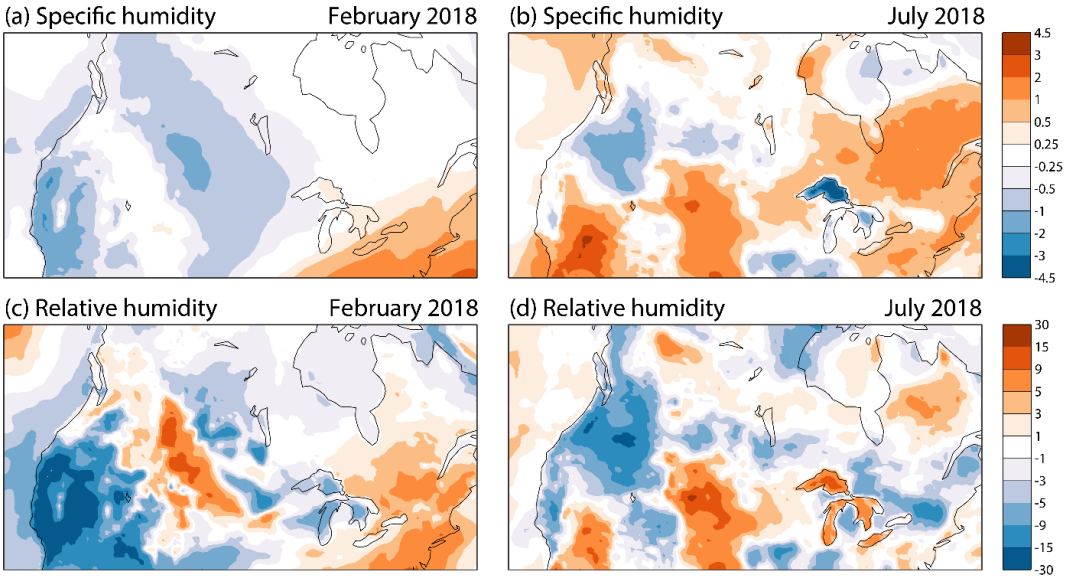

Figure 75 Anomalies relative to 1981-2010 of (a, b) ERA5 specific humidity (g/kg) and (c, d) ERA5 relative humidity (%), for (a, c) February 2018 and (b, d) July 2018.

5 Conclusions

ERA5 is a further step forward for comprehensive atmospheric reanalysis, improving in several ways on its predecessor ERA-Interim for the period 1979-2019, providing analyses back to 1950 and offering prompt updates for the current year within a few days of the time of observation. Its analyses of surface

air temperature and humidity, the focus of this report, form a data record covering more than 70 years that complements the records available from other reanalyses and from direct analysis of monthly climatological temperature records and of synoptic observations of humidity.

This report has provided examples of how comparing a new reanalysis such as ERA5 with other appropriate reanalyses, with alternative monthly climatological datasets for the variables of interest, with both the observations selected for analysis and the observations rejected by quality-control checks, and with the background fields provided by the data assimilation system, all have a valuable part to play in assessing the quality of the new product and pinpointing areas for future improvement. Time series of global or regional averages are an important part of this, and played a vital role in the routine monitoring of ERA5 production, enabling a number of potential issues to be nipped in the bud. It is also important, however, to assess geographical distributions in map form, as spatial averaging may either mask or give undue prominence to issues that are of local but not wider significance.

The trend and low frequency variability of global-mean temperature from ERA5 are largely consistent with values provided by other datasets from 1950 onwards. ERA5 is biased cold over land prior to 1967 because a cold bias of its background forecasts is less well constrained by analysed observations then than in later years. The effect on global-mean temperature, however, is smaller than the uncertainty in all datasets for the 1950s and 1960s that arises from differences in SST analysis, and comparable in magnitude with the difference that arises from using SST rather than the air temperature over sea, one of the characteristics that distinguishes the monthly climatological datasets from what is usually provided by the reanalyses.

Although ERA5 in general gives a reliable representation of surface air temperature (and humidity) there are regions and periods of time where this is not the case. Performance appears to be good throughout for Europe, but there is a little more uncertainty in earlier years, particularly in the mid-1960s when there are significant gaps in observational data coverage. The situation is far worse for Australia prior to the 1970s, and in this case the issue, which is shared with JRA-55, is more than simply one of missing or poorly used surface observations, as the heart of the problem lies in unusually large warm biases of the background forecasts. Poorer than usual agreement over Australia between the reanalyses and monthly climatological datasets also occurs occasionally later in the period, associated in part at least with biases that differ between dry and wet spells, and stem from different definitions of daily average temperature. This report has also discussed a number of issues elsewhere, of a more-local nature, related to periods of missing data, questionable representation of fractional sea-ice cover, inconsistent coastal SSTs and erroneous treatment of the temperatures of the Great Lakes.

Case studies and systematic comparisons of monthly temperature anomalies have been presented. Aside from the comparison with other reanalyses, ERA5 agrees best with GISTEMP and HadCRUT5 when compared with the monthly climatological datasets. Agreement is particularly close over the land masses of the extratropical northern hemisphere. Differences in the tropics and southern extratropics are more pronounced earlier in the period, when ERA5 suffers from an absence of synoptic data that is more acute than the absence of monthly climatological temperature data. On the other hand, ERA5 and other reanalyses have the advantage of supplying products for other variables and levels, in particular relating to atmospheric circulation, thereby supporting the understanding of monthly anomalies. They also provide a more physically sound calculation of temperature over ice-covered sea than the extrapolation of land values provided by a number of the monthly climatological datasets.

There are fewer surface air humidity products to compare, but the reanalyses and HadISDH give similar depictions of interannual variability and longer term changes from the early 1970s onwards, including the net long-term increase in specific humidity but decrease in relative humidity. Agreement is reasonable for the recently available marine values from HadISDH as well as for the longer-established

values over land. Interannual variability over sea and land are largely similar, with changes over sea preceding those over land by a month or so. These results confirm and extend to a longer period what was found earlier for the ERA-Interim and HadCRUH datasets. ERA5 is nevertheless moister over sea than HadISDH for recent years. It also appears to be biased dry as well as cold over land for the 1950s and 1960s. Long-term average values are drier over south-eastern Asia for ERA5 than for JRA-55 and HadISDH, but both reanalyses are drier than HadISDH over the Arabian Peninsula.

Much of the trend and low frequency variability in the reanalyses studied here is captured in the background fields of the data assimilation. Aside from their role in modifying land-surface conditions, the surface air analyses of temperature and humidity act largely to reduce biases in these variables. Issues arise where bias in the background values is large and observational coverage varies in time. It can be expected that biases will be driven down over time by the attention that is paid to the operational forecasting of these variables, though this may not include an emphasis on remote parts of the world, where biases can be particularly large.

Progress will also be made through better observation collection and consolidation of past observations. Effort was devoted in preparation of ERA5 to additional pre-1979 data that would benefit the 4D-Var data assimilation. Improvement of the coverage of the data directly used in the OI analysis of synoptic surface air temperature and humidity measurements was limited to acquiring temperature data for the 1950s held by NCEP, even though other sources were known to be available, notably for Australia, for which a number of issues have been documented in this report. Resources were instead devoted by C3S to supporting data rescue activities that will benefit reanalyses and other climatic applications in the longer term.

The most pressing need for improved datasets is for the period prior to 1967, more particularly for 1965-66 and before 1958. Effort also needs to be devoted to filling data gaps or remedying reduced data coverage for more recent decades. Several examples have been discussed in this report. It will also be necessary to address issues arising from the recent transition from traditional alphanumeric codes to the BUFR encoding of observations. This includes both dealing with data reported in both forms and returning to source for any data that have been not been archived by ECMWF during the transition.

Some applications require climate data for specific times of day, not simply for maximum, minimum or daily-average temperature. ERA5 is a step forward in this regard through its provision of hourly analyses. There is a need though for more-complete international exchange of hourly observations and inclusion of such data in the archives of past observations. Data from airport stations transmitted as METARs provide about half of the hourly surface air temperature observations analysed by ERA5 for recent years, but the coding of these data limits the temperature (dry bulb and dew point) to the nearest whole number of degrees. International transmission of such observations as SYNOPs, or in some other way with the same precision as SYNOPs, should be considered.

Changes to the surface analysis scheme should be implemented for future ECMWF analyses. The FGAT approach adopted by the Japan Meteorological Agency for JRA-55 should be implemented so that off-time observations are used better. This would in particular improve the use of Australian observations, the bulk of which are for non-standard synoptic hours. The surface analysis scheme should be reactivated over sea ice. Diagnosis of future reanalyses would be helped by collection of complete feedback on the scheme's hourly use of observations.

Adjustments may be made for the effects of bias in the background model on estimates of trends and low frequency variability. It has been illustrated how average analysis increments from relatively well-observed periods with little long-term change in bias may be used to adjust for biases when and where there are gaps in data coverage. There is a potential also to use the results of prior analysis of a well-

observed period to adjust the background values used in subsequent analyses. This could be done fully interactively, or through running the surface analysis scheme in a stand-alone mode using archived background fields from the 4D-Var assimilation. In the latter case any potential benefit to the 4D-Var from consequential adjustments of soil moisture and temperature would not be realised. Where there are monthly climatological station data but not synoptic data, consideration could also be given to using the monthly data either to adjust the monthly values from a reanalysis or more interactively in a second pass through either the surface analysis or the full data assimilation.

Acknowledgments

ERA5 relied on the work of many people, not only in its production team but also among those at ECMWF and from outside who contributed to the development and refinement of the ECMWF forecast model and data assimilation system over the years, those from outside who developed the ancillary analyses of sea-surface temperature and sea-ice concentration and other ancillary datasets used in the ERA5 production system, and all who contributed to the reprocessed and additional sets of observations that were used. ECMWF implements the Copernicus Climate Change Service on behalf of the European Union, and ERA5 was produced with funding from this Service. A wider set of acknowledgments is provided by Hersbach *et al*. (2020).

References

Batrak, Y., and M. Müller, 2019: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice. Nat. Commun. 10, 4170, doi: 10.1038/s41467-019-11975-3.

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

Bell, B., H. Hersbach, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, A. Simmons, C. Soci, J. Bidlot, L. Haimberger and J. Woollen, 2021: The ERA5 global reanalysis: 1950-2020. Paper in preparation.

Blunden, J. and D.S. Arndt, Eds., 2020: State of the Climate in 2019. Bull. Amer. Meteor. Soc., 101 (8), Si–S429, doi:10.1175/2020BAMSStateoftheClimate.1.

Cornes, R. C., G. van der Schrier, E.J.M. van den Besselaar and P.D. Jones, 2018: An ensemble version of the E-OBS temperature and precipitation data sets. Journal of Geophysical Research: Atmospheres, 123, 9391–9409, doi:10.1029/2017JD028200.

Cowtan K. and R.G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. Q.J.R. Meteorol. Soc., 140, 1935–1944, doi:10.1002/qj.2297.

Cowtan K., Z. Hausfather, E. Hawkins, P. Jacobs, M.E. Mann, S.K. Miller, B.A. Steinman, M.B. Stolpe and R.G. Way, 2015: Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures. Geophys. Res. Lett., 42, 6526–6534, doi:10.1002/2015GL064888.

Dee D.P., S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.-J. Morcrette, B.-

K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut and F. Vitart , 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828.

Donlon C.J., M. Martin, J. Stark, J. Roberts-Jones E. Fiedler and W. Wimmer, 2012: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. Remote Sens. Environ., 116: 140-158, doi: 10.1016/j.rse.2010.10.017.

Dutra, E., I. Sandu, G. Balsamo, A. Beljaars, H. Fréville, E. Vignon and E. Brun, 2015: Understanding the ECMWF winter surface temperature biases over Antarctica. ECMWF Tech Memo no. 762, 16pp, doi:10.21957/77ohp5te.

Freychet, N., S.F.B. Tett, Z. Yan and Z. Li, 2020: Underestimated change of wet-bulb temperatures over East and South China. Geophysical Research Letters, 47, e2019GL086140, doi: 10.1029/2019GL086140 .

Fréville H, E. Brun, G. Picard, N. Tatarinova, L. Arnaud, C. Lanconelli, C. Reijmer and M. van den Broeke, 2014: Using MODIS land surface temperatures and the Crocus snow model to understand the warm bias of ERA-Interim reanalyses at the surface in Antarctica. The Cryosphere, 8, 1361-1373, doi:10.5194/tc-8-1361-2014.

Gelaro, R., W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G. Kim, R. Koster, R. Lucchesi, D. Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, and B. Zhao, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate, 30, 5419–5454, doi:10.1175/JCLI-D-16-0758.1.

GCOS, 2015: Status of the global observing system for climate. GCOS publication no. 195, WMO, Geneva. 353pp. Available from http://gcos.wmo.int.

Hansen J., R. Ruedy, M. Sato and K. Lo, 2010: Global surface temperature change. Rev. Geophys., 48, RG4004, doi:10.1029/2010RG000345.

Harris, I., T.J. Osborn, P. Jones and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci Data 7, 109, doi:10.1038/s41597-020-0453-3.

Hersbach H., C. Peubey, A. Simmons, P. Berrisford, P. Poli and D. Dee, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble. Q.J.R. Meteorol. Soc., 141, 2350–2375, doi: 10.1002/qj.2528.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume and J.-N. Thépaut, 2020: The ERA5 global reanalysis. Q.J.R. Meteorol. Soc., 146, 1999–2049, doi:10.1002/qj.3803.

Hirahara S., M.A. Balmaseda and H. Hersbach, 2016: Sea Surface Temperature and Sea Ice Concentration for ERA5. ERA Report Series no. 26, 25pp, https://www.ecmwf.int/node/16555.

Huang, B., P.W. Thorne, V.F. Banzon, T. Boyer, G. Chepurin, J.H. Lawrimore, M.J. Menne, T.M. Smith, R.S. Vose and H-M. Zhang, 2017: Extended Reconstructed Sea Surface Temperature, Version 5

(ERSSTv5): Upgrades, Validations, and Intercomparisons. J. Climate, 30, 8179–8205, doi: 10.1175/JCLI-D-16-0836.1.

Ishii M, A. Shouji, S. Sugimotoand T. Matsumoto, 2005: Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the KOBE collection. Int. J. Climatol., 25, 865–879, doi: 10.1002/joc.1169.

Jones, P.D., M. New, D.E. Parker, S. Martin and I.G. Rigor,1999. Surface air temperature and its changes over the past 150 years. Rev. Geophys., 37, 173-199, doi:10.1029/1999RG900002.

Jones, P.D. and D.H. Lister, 2015: Antarctic near-surface air temperatures compared with ERA-Interim values since 1979. Int. J. Climatol., 35, 1354–1366, doi:10.1002/joc.4061.

Kennedy J.J., N.A. Rayner, R.O. Smith, M. Saunby and D.E. Parker, 2011: Reassessing biases and other uncertainties in sea-surface temperature observations since 1850 part 1: measurement and sampling errors. J. Geophys. Res., 116, D14103, doi:10.1029/2010JD015218.

Kennedy, J.J., N.A. Rayner, C.P. Atkinson and R.E. Killick, 2019: An ensemble data set of sea-surface temperature change from 1850: the Met Office Hadley Centre HadSST.4.0.0.0 data set. J. Geophys. Res., 124, 7719– 7763. https://doi.org/10.1029/2018JD029867.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, K. Takahashi, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. J. Meteorol. Soc. Japan, 93, 5-48, doi:10.2151/jmsj.2015-001.

Lenssen, N., G. Schmidt, J. Hansen, M. Menne, A. Persin, R. Ruedy, and D. Zyss, 2019: Improvements in the GISTEMP uncertainty model. J. Geophys. Res. Atmos., 124, 6307-6326, doi:10.1029/2018JD029522.

Li, Z., Z. Yan, Y. Zhu, N. Freychet and S. Tett, 2020: Homogenized daily relative humidity series in China during 1960–2017. Adv. Atmos. Sci. 37, 318–327, doi: 10.1007/s00376-020-9180-0.

Ma, Y. and P. Guttorp, (2013: Estimating daily mean temperature from synoptic climate observations. Int. J. Climatol., 33, 1264-1269, doi:10.1002/joc.3510.

Morice C.P., J.J. Kennedy, N.A. Rayner and P.D. Jones, 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.

Morice, C.P., J.J. Kennedy, N.A. Rayner, J.P. Winn, E. Hogan, R.E. Killick, R.J.H. Dunn, T.J. Osborn, P.D. Jones and I.R. Simpson,2020: An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. J. Geophys. Res., in press, doi:10.1029/2019JD032361.

Orsolini Y., M. Wegmann, E. Dutra, B. Liu, G. Balsamo, K. Yang, P. de Rosnay, C. Zhu, W. Wang, and R. Senan, 2019: Evaluation of snow depth and snow-cover over the Tibetan Plateau in global reanalyses using in-situ and satellite remote sensing observations. The Cryosphere, 13, 2221–2239, doi: 10.5194/tc-13-2221-2019.

Osborn, T.J., P.D. Jones, D.H. Lister, C.P. Morice, I.R. Simpson and I.C. Harris, 2020: Land surface air temperature variations across the globe updated to 2019: the CRUTEM5 dataset. J. Geophys. Res., in press, doi: 10.1029/2019JD032352.

Rennie, J.J, J. H. Lawrimore, B.E. Gleason, P.W. Thorne, C.P. Morice, M.J. Menne, C.N. Williams, W. Gambi de Almeida, J.R. Christ,y M. Flannery, M. Ishihara, K. Kamiguchi, A.M.G. Klein-Tank, A.

Mhanda, D.H. Lister, V. Razuvaev, M. Renom, M. Rusticucci, J. Tandy, S.J. Worley, V. Venema, W. Angel, M. Brunet, B. Dattore, H. Diamond, M.A. Lazzara, F. Le Blancq, J. Luterbacher, H. Mächel, J. Revadekar, R.S. Vose, X. Yin, 2014: The international surface temperature initiative global land surface databank: monthly temperature data release description and methods. Geosci. Data J., 1: 75-102, doi:10.1002/gdj3.8.

Rigor, I.G., R.L. Colony and S. Martin, 2000: Variations in Surface Air Temperature Observations in the Arctic, 1979–97. J. Climate, 13, 896–914, doi: 10.1175/1520-0442(2000)013<0896:VISATO>2.0.CO;2.

Rohde, R. and Z. Hausfather, 2020: The Berkeley Earth Land/Ocean Temperature Record, Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2019-259, accepted for publication.

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

Saha, S., S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H.-Y. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M.P. Mendez, H. van den Dool, Q. Zhang, W. Wang, M. Chen and E. Becker, 2014: The NCEP Climate Forecast System version 2, J. Climate, 27, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

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

Simmons A.J., K.M. Willett, P.D. Jones, P.W. Thorne and D.P. Dee, 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.

Simmons, A.J., P. Poli, D.P. Dee, P. Berrisford, H. Hersbach, S. Kobayashi S and C. Peubey, 2014: Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. Q.J.R. Meteorol. Soc., 140: 329-353, doi:10.1002/qj.2317.

Simmons A.J. and P. Poli, 2015: Arctic warming in ERA-Interim and other analyses. Q.J.R. Meteorol. Soc., 141, 1147-1162, doi:10.1002/qj.2422.

Simmons, A.J., P. Berrisford, D.P. Dee, H. Hersbach, S. Hirahara and J.-N. Thépaut, 2017: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. Quart. J. Roy. Meteorol. Soc., 143, 101-119, doi:10.1002/qj.2949.

Simmons, A.J., C. Soci, J. Nicolas, W. Bell, P. Berrisford, R. Dragani, J. Flemming, L. Haimberger, S. Healy, H. Hersbach, A. Horányi, A. Inness, J. Muñoz-Sabater, R. Radu and D. Schepers, 2020: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. ECMWF Tech Memo no. 859, 38pp. doi:10.21957/rcxqfmg0.

Smith, T.M., R.W. Reynolds, T.C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land–ocean surface temperatures analysis (1880–2006); Journal of Climate, 21, 2283–2296, doi:10.1175/2007JCLI2100.1.

Trewin, B., 2013: A daily homogenized temperature data set for Australia. International Journal of Climatology, 33, 1510–1529, doi:10.1002/joc.3530. See also the technical report on version 2, available at http://www.bom.gov.au/climate/data/acorn-sat/.

Uppala, S.M., P.W. Kållberg, A.J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K Gibson, J. Haseler, A. Hernandez, G.A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, R. Jenne, A.P. McNally, J.-F. Mahfouf, J.- J. Morcrette, N.A Rayner, R.W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen, 2005: The ERA-40 Reanalysis. Quart. J. Roy. Meteorol. Soc., 131, 2961-3012, doi:10.1256/qj.04.176.

Wang, C., R.M. Graham, K. Wang, S. Gerland and M.A. Granskog, 2019: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution, The Cryosphere, 13, 1661–1679, doi: 10.5194/tc-13-1661-2019.

Willett, K.M., R.J.H. Dunn, P.W. Thorne, S. Bell, M. de Podesta, D.E. Parker, P.D., Jones and C.N. Williams Jr, 2014): HadISDH land surface multi-variable humidity and temperature record for climate monitoring. Clim. Past, 10, 1983-2006, doi:10.5194/cp-10-1983-2014.

Willett, K.M., D.I. Berry, M.G. Bosilovich and A.J. Simmons, 2020a: [Global climate] Surface humidity [in “State of the Climate in 2019”]. Bull. Amer. Meteor. Soc., 101 (8), S42–S43, doi:10.1175/2020BAMSStateoftheClimate.1.

Willett, K. M., R.J.H. Dunn, J.J. Kennedy and D.I. Berry, 2020b: Development of the HadISDH marine humidity climate monitoring dataset. Earth Syst. Sci. Data, 12, 2853-2880, doi: 10.5194/essd-12-2853-2020.

Woodruff, S. D., S.J. Worley, S.J. Lubker, Z. Ji, J.E. Freeman, D.I. Berry, P. Brohan, E.C. Kent, R.W. Reynolds, S.R. Smith and C. Wilkinson, 2011: ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol., 31, 951–967, doi: 10.1002/joc.2103.

Zhang, H.-M., J.H. Lawrimore, B. Huang, M.J. Menne, X. Yin, A. Sánchez-Lugo, B.E. Gleason, R. Vose, D. Arndt, J.J. Rennie and C.N. Williams, 2019: Updated temperature data give a sharper view of climate trends, Eos, 100, doi:10.1029/2019EO128229.