Consistency of Temperature and Precipitation Extremes across Various Global Gridded In Situ and Reanalysis Datasets*

MARKUS G. DONAT

Climate Change Research Centre, and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia

JANA SILLMANN

University of Victoria, Victoria, British Columbia, Canada

SIMON WILD

School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom

LISA V. ALEXANDER AND TANYA LIPPMANN

Climate Change Research Centre, and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia

FRANCIS W. ZWIERS

Pacific Climate Impacts Consortium, Victoria, British Columbia, Canada

(Manuscript received 10 July 2013, in final form 13 December 2013)

ABSTRACT

Changes in climate extremes are often monitored using global gridded datasets of climate extremes based on in situ observations or reanalysis data. This study assesses the consistency of temperature and precipitation extremes between these datasets. Both the temporal evolution and spatial patterns of annual extremes of daily values are compared across multiple global gridded datasets of in situ observations and reanalyses to make inferences on the robustness of the obtained results.

While normalized time series generally compare well, the actual values of annual extremes of daily data differ systematically across the different datasets. This is partly related to different computational approaches when calculating the gridded fields of climate extremes. There is strong agreement between extreme temperatures in the different in situ-based datasets. Larger differences are found for temperature extremes from the reanalyses, particularly during the presatellite era, indicating that reanalyses are most consistent with purely observational-based analyses of changes in climate extremes for the three most recent decades. In terms of both temporal and spatial correlations, the ECMWF reanalyses tend to show greater agreement with the gridded in situ-based datasets than the NCEP reanalyses and Japanese 25-year Reanalysis Project (JRA-25). Extreme precipitation is characterized by higher temporal and spatial variability than extreme temperatures, and there is less agreement between different datasets than for temperature. However, reasonable agreement between the gridded observational precipitation datasets remains. Extreme precipitation patterns and time series from reanalyses show lower agreement but generally still correlate significantly.

*Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-13-00405.s1>.

Corresponding author address: Markus G. Donat, Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Level 4 Mathews Building, Sydney NSW 2052, Australia.
E-mail: m.donat@unsw.edu.au

1. Introduction

Climate extremes often have severe impacts on societies and ecosystems. Different types of extremes, such as heat waves and droughts, are expected to increase in frequency and intensity along with anthropogenic climate change (Seneviratne et al. 2012). To monitor risk and changes in temperature and precipitation extremes, several global observational datasets of climate extremes indices based on daily records of temperature and precipitation have recently been developed (Donat et al. 2013a,b). Because of frequently existing restrictions on the exchange of the actual daily temperature and precipitation data between countries, the joint World Meteorological Organization (WMO) Commission for Climatology (CCL)/Climate Variability and Predictability (CLIVAR)/Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) Expert Team on Climate Change Detection and Indices (ETCCDI; see e.g., Zhang et al. 2011) have recommended a set of indices derived from daily data that represent some aspects of climate extremes and bypass these data exchange issues. Many countries who are not willing to share their daily observational data agree, however, to share the calculated indices. Hence, these indices often remain the only source of information about extremes in some regions. Gridded datasets have then been developed from indices calculated from station data. However, these datasets based on daily in situ observations generally lack spatial coverage (e.g., Caesar et al. 2006; Alexander et al. 2006; Donat et al. 2013a), as suitable long-term records are sparse for some regions, particularly for South America and Africa.

Atmospheric reanalyses are produced by assimilating certain types of observational data into numerical models, whereby surface temperature and precipitation are usually not directly used for assimilation. Generally based on operational forecast models, they provide complete spatial and temporal coverage of the globe with physically consistent data. This makes reanalysis data a popular reference for climate model evaluation, for example (e.g., Frankignoul et al. 2004; Bengtsson and Hodges 2011; Svensson and Karlsson 2011; Sillmann et al. 2013). Reanalyses are also often used in conjunction with in situ-based products for real-time monitoring of the climate system: for example, the European Reanalysis and Observations for Monitoring project (EURO4M; <http://www.euro4m.eu>). Impact models in various fields can be trained and/or validated through the use of reanalyses to estimate damages, monetary losses, and casualties through extremes in climate phenomena such as heat waves (Orlowsky and Seneviratne 2011) or wind storms (Donat et al. 2011).

As pointed out by Dee et al. (2011b), reanalyses represent the most homogeneous and accurate observational-based datasets for the last 20 yr. However, because of inhomogeneities in the observational input data available before the last two decades, imperfections in data assimilation schemes, and model uncertainty, reanalysis data may not be suitable for long-term climate assessments (Bengtsson et al. 2004; Thorne and Vose 2010). Indeed, reanalyses producers themselves highlight some of the inhomogeneities that are detected in 1958 (introduction of radiosondes) and 1979 (introduction of satellite data) in their products (e.g., Kistler et al. 2001).

A number of studies have compared average conditions between reanalyses and observations. Despite indisputable uncertainty between the data products (e.g., Thorne et al. 2005), comparisons reveal broad agreement in long-term trends of temperature time series (e.g., Simmons et al. 2004, 2010; Santer et al. 2008; Vose et al. 2012). However, climate extremes have received less attention. For 20-yr return values of extreme temperature and precipitation, Kharin et al. (2005, 2013) pointed to partly substantial differences in actual values between extremes derived from models and observations. For Europe, Cornes and Jones (2013) recently demonstrated generally good agreement in extreme temperature trends between gridded observations and the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) over the past 30 yr.

Precipitation trends and variability show a far more heterogeneous picture. Differences between ground-based gauge data and precipitation estimates from satellite measurements and reanalyses are substantial (Simmons et al. 2010; Trenberth 2011). Simmons et al. (2010) show pronounced biases in absolute precipitation values with the sign dependent on the region considered. A significant trend in precipitation in either observations or reanalysis data could not be identified in that study.

Investigating extreme temperature and precipitation, Sillmann et al. (2013) used gridded datasets of in situ observations and reanalysis data to evaluate the representation of these extremes in simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5). They highlight the large spread in absolute values of temperature and precipitation extremes between the different reanalysis products, comparable to the spread between different climate models. It is, however, still an open question as to what extent reanalyses are suitable to investigate changes in climate extremes during recent decades. In particular, given the limited coverage of purely in situ observations-based datasets, it is necessary to determine whether reanalyses may be suitable to fill in these gaps. Also the robustness of results from

TABLE 1. Overview of gridded observational and reanalysis datasets used in this study.

Dataset	Description/approach	Temporal coverage	Spatial resolution	Reference
Interpolated in situ observations				
HadEX2	Gridded fields of indices calculated at station locations	1901–2010	$3.75^{\circ} \times 2.5^{\circ}$	Donat et al. (2013a)
GHCNDEX	Gridded fields of indices calculated at station locations	1951–2012	$2.5^{\circ} \times 2.5^{\circ}$	Donat et al. (2013b)
HadGHCND	Indices calculated from daily gridded fields	1951–2010	$3.75^{\circ} \times 2.5^{\circ}$	Caesar et al. (2006)
Reanalysis data				
NCEP-1	Indices calculated from daily gridded fields	1948–2012	T62 spectral truncation (data provided on $2.5^{\circ} \times 2.5^{\circ}$ grid)	Kalnay et al. (1996)
ERA-40	Indices calculated from daily gridded fields	1958–2001	T159 spectral truncation (data provided on $2.5^{\circ} \times 2.5^{\circ}$ grid)	Uppala et al. (2005)
NCEP-2	Indices calculated from daily gridded fields	1979–2012	T62 spectral truncation (data provided on $1.875^{\circ} \times 1.875^{\circ}$ grid)	Kanamitsu et al. (2002)
ERA-Interim	Indices calculated from daily gridded fields	1979–2012	T255 spectral truncation (data provided on $1.5^{\circ} \times 1.5^{\circ}$ grid)	Dee et al. (2011a)
JRA-25	Indices calculated from daily gridded fields	1979–2012	T106 spectral truncation (data provided on $2.5^{\circ} \times 2.5^{\circ}$ grid)	Onogi et al. (2007)

the different observational datasets remains to be verified.

This study investigates the robustness of changes in extreme temperature and precipitation based on different datasets of gridded in situ observations and several reanalyses. We also evaluate the effect of different computational approaches in which the order of operation of index calculation is varied. On one hand, climate indices are first calculated at station locations before gridding the indices. This is done in most of the interpolated datasets of in situ observations to improve coverage for regions where daily data are not readily available but where national meteorological and hydrological services or researchers are able to share the calculated indices. On the other hand, climate extremes indices are often computed from daily gridded fields, as these are provided by reanalyses and climate model output. We consider whether there are systematic differences between the datasets calculated in the two ways, and we examine how well the different datasets agree in the temporal and spatial representation of extreme temperature and precipitation values. This evaluation is essential for potential applications incorporating changes in these kinds of climate extremes from reanalysis data.

This paper is structured as follows: We describe the data used and methods applied in [section 2](#). The comparison of the climate extremes from the different datasets is presented in [section 3](#), followed by a discussion of the results and a summary of the most important conclusions in [section 4](#).

2. Data and methods

We investigate and compare extreme temperature and precipitation values from three global gridded datasets of in situ observations (interpolated observations hereafter) and five reanalyses ([Table 1](#)). We note that more reanalysis products are available. However, for the purpose of this study, a general comparison between reanalyses and in situ–based extremes, the chosen reanalyses represent products that are commonly used in the scientific literature. Furthermore, in the case of National Centers for Environmental Prediction (NCEP) and ECMWF reanalyses, we can compare different reanalysis generations from the same modeling center.

Two of the interpolated observational datasets [Hadley Centre Global Climate Extremes Index 2 (HadEX2) and National Climatic Data Center (NCDC) Global Historical Climatology Network (GHCN)-Daily climate extremes (GHCNDEX)] are produced by gridding climate extremes indices calculated at the station locations onto a global grid. Both are based on largely independent sets of input stations but use identical gridding methods. While only limited quality control and correction is applied to the data in the GHCN-Daily archive used for GHCNDEX ([Donat et al. 2013b](#)), HadEX2 is based on input stations from several sources (including ETCCDI workshops and local contacts), with more rigorous quality control applied ([Donat et al. 2013a](#)). In contrast, climate extremes indices of all other datasets are calculated from daily gridded

fields. Hadley Centre GHCN-Daily (HadGHCND) provides daily fields gridded from station observations for temperature only. As this dataset is based largely on the same station data as GHCNDEX, the comparison of these two datasets allows for the estimation of some of the structural uncertainty related to the two different approaches.

In this study, all climate extremes indices are calculated from daily maximum and daily minimum temperatures and daily precipitation amounts over land. In particular, we use the annual maxima of daily maximum temperatures (TXx) and daily minimum temperatures (TNx) and the annual minima of daily maximum (TXn) and daily minimum temperatures (TNn). As measures of extreme precipitation, we consider annual maximum daily precipitation (Rx1day) and annual maximum consecutive 5-day precipitation (Rx5day). Note that daily temperature extremes (TN and TX) are not provided for the 40-yr ECMWF Re-Analysis (ERA-40) and ERA-Interim; hence, for these datasets they were approximated by using the daily minimum and maximum of instantaneous 6-hourly near-surface temperature values.

When gridding the observed precipitation extremes, the decorrelation length scale is generally higher for monthly compared to annual extremes indices, which leads to a more complete spatial coverage in the monthly gridded index fields than in the annual grids (e.g., Donat et al. 2013a). Therefore, in order to have a coverage for the intercomparison of precipitation extremes that is as complete as possible, we calculate the annual maximum daily and 5-day precipitation amounts as maximum of the monthly Rx1day and Rx5day fields, respectively. For temperature extremes, this seasonal variation in spatial coverage is less pronounced, as annual temperature maxima generally occur in summer and temperature minima in winter.

Most of the datasets have different temporal coverage and spatial resolution (Table 1), and spatial coverage of the interpolated observations datasets is incomplete. A fair comparison must be restricted to time periods and regions included in all data. Therefore, we only consider the common period 1979–2010 (note that ERA-40 annual extremes are only available until 2001 and are therefore not included in the comparison of trends) for our comparisons and restrict the comparisons to identical regions. To this end, we mask the fields to include only grid boxes common to all datasets and which additionally fulfill a strict completeness criterion: we only use grid boxes that have valid data in all interpolated observations for at least 30 yr out of the 32-yr period 1979–2010. For the extreme precipitation indices, there is a considerable decline in spatial observational coverage in 2009 and 2010. In relation to the stronger spatial

heterogeneity of precipitation compared to the temperature indices, this decline may introduce artificial trends when considering estimates of global and continental average time series. Therefore, we restrict the dataset comparison for the precipitation indices to the 30-yr period 1979–2008 and use only grid boxes with at least 28 yr of data during this period (the mask of common grid boxes is shown by the green area in the bottom-right panels of Figs. 3, 4, and 8). For masking, all gridded datasets are interpolated to the coarsest grid used here (i.e., $3.75^\circ \times 2.5^\circ$: HadEX2 and HadGHCND), using a first-order area-conservative remapping technique (Jones 1999) as implemented in the Climate Data Operators software (<https://code.zmaw.de/projects/cdo>). Besides global averages, we also compare area-averaged values for four continental subregions: northern Africa and Europe combined; North America; Asia; and Australia, which is the only region in the Southern Hemisphere with sufficient observational data for comparison. All area averages are calculated by area weighting the individual gridbox contributions. Note that, because of limited observational coverage, our global time series are representative only of the Northern Hemisphere extratropics and Australia.

Linear trends in this comparison study are calculated using Sen's trend estimator (Sen 1968), and trend significance is estimated using the Mann–Kendall test (Kendall 1975). We use these methods to allow comparability with earlier studies (e.g., Alexander et al. 2006; Donat et al. 2013a) and because they are non-parametric, and some of the annual extremes do not follow a Gaussian distribution, particularly at specific locations. For the same reason, the temporal correlation coefficients are calculated using Spearman's rank correlation (Spearman 1904). No significant autocorrelations were found in the observational datasets of annual extremes.

We calculate the pattern correlation for each year, using the fields from HadEX2 as reference, as this is assumed to be the interpolated observations dataset of highest quality because of most intense data quality and homogeneity controls. Pattern correlations are calculated using Spearman's rank correlation and are based on anomalies from the 1979–2010 average at each grid box rather than the actual values. This is to avoid unreasonably high correlation values because of obvious global distribution of temperature (i.e., warm in low latitudes and cooler in high latitudes) and precipitation in all datasets. Hence, we compare agreement in local anomalies rather than agreement in the global distribution.

We restrict this comparison study to indices representative of annual maximum or minimum values of

FIG. 1. Time series of area-averaged values of TXx in the different datasets over global land and four continental subregions: (left) actual values and (right) anomalies from the 1981–2000 average ($^{\circ}\text{C}$).

temperature and precipitation, as they can be calculated consistently across all datasets. Some additional extremes indices are defined relative to climatological percentiles during a specific base period. However, this leads to unavoidable inconsistencies between the interpolated observations and some reanalysis datasets, as the interpolated data commonly use 1961–90 as the base period for percentile calculations, which cannot be covered by the latest reanalyses only starting in 1979. To allow for comparison over the longest possible period, we also do not include HadEX (Alexander et al. 2006), the first global gridded dataset of climate extremes, which covers the period 1951–2003. However, a previous comparison has shown that HadEX generally compares well to HadEX2 for the period when both provide data (Donat et al. 2013a).

3. Results

a. Extreme temperatures

1) TEMPORAL EVOLUTION AND SPATIAL TREND PATTERNS

The global and regional average values of extreme temperatures show a large spread between the different datasets (Figs. 1, 2 and Figs. S1, S2 in the supplementary material). For TXx, for example, the difference between the warmest and the coldest dataset is approximately 5°C in most subregions and globally and as large as 15°C over Australia. The highest TXx values are found in the HadEX2 and GHCNDEX interpolated observations and the NCEP–National Center for Atmospheric Research (NCAR) reanalysis (NCEP-1), whereas lowest

TXx values are generally found in the Japanese 25-year Reanalysis Project (JRA-25; Fig. 1). Differences in TNn are even larger (15°C over most regions) and, in contrast to the warm extremes, somewhat smaller over Australia. Temperatures on the coldest nights of the year are warmest in JRA-25 in most regions and generally coldest in NCEP-1 and the NCEP–U.S. Department of Energy (DOE) Atmospheric Model Intercomparison Project 2 (AMIP-II) reanalysis (NCEP-2; Fig. 2). Besides showing the smallest intraannual temperature range (i.e., difference between highest maximum and lowest minimum temperature annually), JRA-25 also stands out as the dataset with the lowest diurnal temperature range in most areas (not shown).

The comparison of extreme temperature values between GHCNDEX and HadGHCND is of particular interest, as both datasets are based largely on the

same input stations but follow two different computational approaches. The HadGHCND annual extremes are based on gridded daily maximum and minimum temperatures, whereas the GHCNDEX annual extremes are first calculated for each station and then gridded. Both datasets use the same gridding method. Results show that the annual maxima of daily minimum and maximum temperatures are generally higher in GHCNDEX compared to HadGHCND (Figs. 1, S1). In contrast, the annual minimum values are generally lower in GHCNDEX than in HadGHCND (Figs. 2, S2). This means “more extreme” temperature values (i.e., higher maxima and lower minima) are found when gridding the in situ calculated extremes indices and demonstrates how local extreme values are smoothed when averaging them to calculate daily grid values.

This comparison of the two computational approaches is of particular interest when comparing climate model or reanalysis data to interpolated observations of climate extremes (e.g., Min et al. 2011; Sillmann et al. 2013; Kharin et al. 2013). The climate model grids also represent spatial averages of daily values (like HadGHCND) rather than spatial averages of local annual extremes (like GHCNDEX). Therefore, systematic biases are to be expected, with temperature ranges between cold and warm extremes being smaller in climate model and reanalysis data or daily datasets of gridded in situ observations, compared to grids of observed extremes. This effect accounts for differences of approximately 2°–3°C for annual maximum and minimum temperature indices in most regions. Note that, while the order of operation creates some systematic discrepancies regarding the actual values of extremes, we would not expect those differences to substantially affect trends. The extreme temperatures from HadEX2 show similar values to those from GHCNDEX for most regions. For these two datasets the same method is used for gridding the extremes indices, which are first calculated at the stations locations.

Normalizing the temperature time series by removing their average over the 1981–2000 period brings them to a comparable data range and allows easier comparison in terms of temporal changes (right panels in Fig. 1, 2, S1, S2). Globally and for most subregions, all datasets (apart from NCEP-1 for warm extremes, see discussion below) show robust warming trends during the past three to four decades. Global average trends in the different datasets (without NCEP-1) during 1979–2010 range between 0.3° and 0.4°C decade^{−1} for TXx, for example, and global average TNn trends for this period are about 0.6°C decade^{−1} in most datasets, with only NCEP-2 showing a stronger TNn warming (1.0°C decade^{−1}). While the three interpolated observational datasets show very good agreement (almost identical temperature anomalies) throughout the entire period 1951–2010, some of the reanalyses show larger differences, particularly during earlier years. This is most obvious for NCEP-1 and ERA-40, which have a tendency to show higher than observed temperatures during the 1950s, 1960s, and 1970s, before satellite data are used for assimilation, and can be seen most clearly in TNn, TXn, and TNx (Figs. 2, S1, S2). However, also during the satellite era, some reanalyses show large deviations from observed extreme temperature anomalies, particularly over Australia.

The results also point to likely erroneous maximum temperatures in NCEP-1, with annual maximum temperatures showing a jump toward higher values in the early 1980s and jumping back to cooler conditions in the

mid-1990s. This effect is also apparent in other measures based on daily maximum temperatures and was previously discussed by Kharin et al. (2005) and Sillmann et al. (2013). The unreasonable high maximum temperatures are more pronounced in tropical regions and may be related to problems in the formulation of boundary layer processes as mentioned in NCEP–NCAR (2011). It is, however, not clear why this effect occurs most strongly only in the 1980s and 1990s. A thorough investigation of the input data used for assimilation may help to understand this inhomogeneity.

The spatial patterns of trends generally agree well between the different interpolated observations, and most of the reanalyses also reproduce the key features of the observations. Annual maximum temperatures, for example, have significantly increased over large areas in Europe and East Asia during the 1979–2010 period (Fig. 3). There is less consensus between the reanalyses over areas not sufficiently covered by the observational datasets, in particular Africa and South America, where the reanalyses show partly comparable signs of change but at different magnitudes. This includes a general warming over large parts of Africa; however, the extent of cooling over the southern tip of Africa varies strongly between the different reanalyses. Over South America, all reanalyses show increasing TXx values over the northern part of Brazil; however, there is an obvious discrepancy regarding cooling trends along the South American west coast shown by some reanalyses. NCEP-1 shows strong reductions in TXx over large areas, most significantly in low latitudes. This reflects the erroneous jumps in TXx values described above. This effect occurs mainly during the first half of the trend period shown and explains the decreasing trends, largely in contradiction to all other datasets.

Annual minimum temperatures show strong increases during 1979–2010 over North America and northern parts of Europe and Asia. This is consistently shown by all interpolated observations and most reanalysis datasets (Fig. 4). There is also an area with cooling trends in lower latitudes over parts of eastern and central Asia, which differs in extent and magnitude among the datasets. Again, NCEP-1 shows converse trends to all other datasets by indicating TNn decreases over Europe. Similarly, we find generally robust trend patterns for TXn and TNx among the three interpolated observations datasets, and the most significant features also apparent in most reanalyses (Figs. S4, S5 in the supplementary material). Particularly over Africa and South America, the regions with insufficient coverage in the interpolated observations, there are also larger variations among the different reanalyses.

TXx decadal trend 1979–2010

FIG. 3. Maps of linear decadal trends in TXx during 1979–2010 ($^{\circ}\text{C decade}^{-1}$). Trends are calculated only for land grid boxes that have at least 10 yr of data, and hatching indicates significant trends ($p \leq 0.05$; Mann–Kendall test). (bottom right) The mask shows the grid boxes that provide almost complete time series (30 out of 32 yr) in all gridded datasets of in situ observations and is used for comparison of the datasets. Continental subregions used for further analysis are also shown.

2) TEMPORAL CORRELATION ANALYSIS

We investigate the agreement among the global and subregion time series (e.g., Figs. 1, 2) from the different datasets in more detail by calculating the temporal correlations between all pairs of datasets. The resulting correlation coefficients for all combinations are shown in the matrices in Fig. 5. The correlation matrices confirm the generally good agreement between the three interpolated observational datasets, with correlation coefficients being above 0.9 for most extreme temperature indices. Also the reanalyses show mostly good agreement with both interpolated observations and other reanalyses, with correlation coefficients that are above 0.8 in most cases. Somewhat lower correlations are found for the global time series of TNn, and regional analysis (Fig. 5, right) shows that for this index the agreement between most datasets seems to be

particularly low over Australia. For NCEP-1, particularly low correlation values of less than 0.5 are obvious in comparison to all other datasets for TXx, reflecting the NCEP-1 problems as discussed above. In general, correlations tend to be higher when comparison is restricted to the common period of all datasets, covering the most recent three decades.

Consideration of the agreement between the regional time series indicates that on average the correlations are highest for North America and Europe and lowest for Australia. Attempting to identify “the best” reanalysis in terms of highest correlations with the interpolated observations, we find that on average over all subregions ERA-Interim and ERA-40 seem to show the best agreement with observed annual extreme temperatures. Since ERA-40 is not available after 2001, our results suggest that ERA-Interim seems to be the reanalysis that most closely reproduces variations in observed temperature extremes

TNn decadal trend 1979–2010

FIG. 4. As in Fig. 3, but for TNn ($^{\circ}\text{C decade}^{-1}$).

during the past three decades. It is worth noting that ERA-Interim indirectly assimilates near-surface air temperature and humidity to initialize the soil moisture fields, which may account for the better agreement.

3) SPATIAL CORRELATION ANALYSIS

We investigate spatial correlations of local annual anomalies in comparison to HadEX2 to evaluate the robustness of spatial patterns between the different datasets. There is generally good agreement between annual extreme temperatures in the interpolated observations datasets (Figs. 6a–d), with spatial correlations above 0.8 in most years for GHCNDEX (0.75 for HadGHCND). Correlation values are generally highest for GHCNDEX, but for most indices there is also a slight decline in correlation for the most recent years after 2005.

The spatial correlations for temperature extremes calculated from the reanalysis datasets in comparison to HadEX2 are between 0.5 and 0.8 for most years. On average, best agreement with the interpolated observations data is found for ERA-40 and ERA-Interim, whereas

NCEP-1, JRA-25, and NCEP-2 generally show the lower correlations. Correlations are particularly low (down to 0.4) for TXx calculated from NCEP-1 during the 1980s and early 1990s. This is related with the erroneous maximum temperature values in this dataset, which are discussed above.

For TNn and less so for TXn, the agreement with HadEX2 declines in 2009 and 2010 for most of the datasets (Figs. 6b,d), as reflected by the reduced correlations. All interpolated observations show comparable patterns over North America and Europe in 2009 (Fig. S6 in the supplementary material), with cold anomalies over extended parts of North America and western and central Europe, strong warm anomalies over Scandinavia (which extend farther into eastern Europe in HadEX2), and warm anomalies over northeast Asia. However, while HadEX2 shows anomalously warm minimum temperatures over central Asia, GHCNDEX shows particularly strong cold anomalies over this region. These differences are most likely related to changes in the station network used for interpolation. As some of

FIG. 5. Correlations of the area-averaged time series for the temperature indices from the different interpolated observational and reanalysis datasets. Above the diagonal of each matrix shows the correlations for all years available in each pair of correlated datasets, and below the diagonal of the matrix shows correlations for the common years 1979–2010 in all datasets (* except ERA-40, ending in 2001). Correlations are shown for (left) global averages and (right) four continental subregions.

FIG. 6. Pattern correlation of annual fields in the different datasets with HadEX2. Only the grid boxes shown in the masks in Figs. 3, 4 and 8 (and Figs. S4, S5, and S7, respectively) are used for calculating the correlations.

the station time series used to produce HadEX2 end by 2009, the number of stations providing data for the most recent years is reduced. Hence, more remote stations receive increased weight when calculating the local gridbox values (see Alexander et al. 2006; Donat et al. 2013a), which leads to biases in the actual temperature values. The reanalyses show a somewhat mixed pattern in this region, but also indicate a warm anomaly in central Asia (in agreement with HadEX2). However, the reanalyses in general also have more spatially heterogeneous anomaly field patterns (see, e.g., NCEP-2 in Fig. S6) compared to the smoother fields of the interpolated observations.

b. Extreme precipitation

1) TEMPORAL EVOLUTION AND SPATIAL TREND PATTERNS

For extreme precipitation amounts, there is a substantial spread between the different datasets (Figs. 7

and S3 in the supplementary material), with the highest precipitation intensities in most regions in HadEX2, GHCNDEX, and NCEP-2 and the lowest actual values in ERA-40. Estimates of the global land area mean intensity of extreme precipitation differ by almost a factor of 2 between the datasets. Note that no global interpolated observational dataset providing daily gridded precipitation is available to compare the different computational approaches. However, the finding that GHCNDEX and HadEX2 (both gridding the extremes first calculated at the station locations) show mostly higher extreme precipitation than the reanalyses (for which extremes are calculated from daily gridbox values) is likely influenced by the smoothing effects related to daily grids, in addition to general deficiencies in the simulation of precipitation amounts (e.g., Kharin et al. 2013). Note also that precipitation is only a type “C” variable in reanalyses (Kalnay et al. 1996), meaning that precipitation is dominantly influenced by the

FIG. 7. As in Fig. 1, but for Rx5day (mm).

underlying atmospheric model without being constrained by actual precipitation observations.

The normalized area-averaged time series also show a wide range of behaviors for the different datasets. HadEX2 and GHCNDEX generally show no significant changes in the intensity of extreme precipitation, apart from a slight increase in the 1950s and 1960s, which is more obvious in HadEX2. This increase is mainly found in Asia and North America but is also visible in the global average. These reanalysis data show strong changes in the presatellite era: for example, ERA-40 shows a strong increase in extreme precipitation over North America and Australia. During the common data period 1979–2008 most reanalyses show no significant trends for the areas covered by observations, consistent with the observational datasets. An exception is NCEP-2, which shows changes toward

more intense precipitation extremes in the two decades since 1990, globally and for all Northern Hemispheric subregions compared here.

Regarding the spatial trend patterns, the interpolated observations generally do not show locally significant changes in grid boxes with sufficient data coverage, but there are some consistent tendencies toward less extreme precipitation (e.g., over East Asia and southern Africa) and toward more extreme precipitation (e.g., over the tropical north of Australia and northern Europe; Figs. 8 and S7 in the supplementary material). These tendencies are only partly reproduced by the reanalyses. In particular, the extended area with an observed tendency toward drier conditions over South Africa is not found in any of the reanalyses. In contrast, all reanalyses indicate a tendency toward wetter precipitation extremes in this area. The reanalyses show

Rx5day decadal trend 1979–2008

FIG. 8. As in Fig. 3, but for Rx5day during 1979–2008 (mm decade⁻¹). (bottom right) The mask shows the grid boxes that provide almost complete time series (28 out of 30 yr) in all gridded datasets of in situ observations and is used for comparison of the datasets. Continental subregions used for further analysis are also shown.

strongest and locally significant trends in extreme precipitation amounts over tropical regions. While they consistently (apart from NCEP-1) show an extended area with trends toward less extreme precipitation over tropical Africa and mostly increasing trends over Indonesia, Papua New Guinea, and northern Australia, there are inconsistent patterns over tropical South America. ERA-Interim, for example, shows significant increases in extreme precipitation amounts in this area (except for the northeastern part of Brazil), whereas NCEP-2 shows predominantly negative trends here. Also, the other reanalyses do not show consistent trend patterns for South America. Further in-depth analysis would be required to explain the differences between the different reanalyses and future research is warranted.

2) TEMPORAL CORRELATION ANALYSIS

Correlations between the area-averaged precipitation time series are lower than corresponding temporal

correlations for temperature time series but are generally still significant (Fig. 9). For both precipitation indices, the correlation between the global average time series from the two interpolated observational datasets is about 0.7 but slightly higher for the maximum 5-day accumulation. The reanalyses show some agreement with the observed precipitation extremes, with correlations between global average time series in the range of 0.4–0.7. Regional analysis reveals that for most reanalyses the agreement with interpolated observations is highest for Australia and North America, whereas correlations are lower on average for Europe and Asia. The correlation matrices also show particularly good agreement between reanalyses produced by the same modeling center: that is, between NCEP-1 and NCEP-2 and between ERA-40 and ERA-Interim. This is most strongly found for Rx1day and likely reflects common model components or assimilation techniques used. In terms of identifying the reanalysis most closely

FIG. 9. As in Fig. 5, but for the extreme precipitation indices, Rx1day and Rx5day.

resembling the interpolated observations, on average over all subregions ERA-Interim and ERA-40 again show slightly higher correlations with HadEX2 and GHCNDEX than the other reanalyses datasets.

3) SPATIAL CORRELATION ANALYSIS

As precipitation fields are generally more heterogeneous than temperature fields, the spatial pattern correlations of extreme annual precipitation amounts in the different datasets are lower than for extreme temperatures. This also reflects the largely inconsistent trend patterns discussed above. However, there is still reasonable agreement between the two interpolated observations datasets that provide precipitation extremes, HadEX2 and GHCNDEX, with spatial correlations around 0.5 (0.6) for Rx1day (Rx5day) in most years (Figs. 6e,f). The reanalyses show generally lower but mostly still significant ($p \leq 0.05$) agreement with HadEX2, with spatial correlations in a range between 0.2 and 0.4

in most years. The agreement is better for Rx5day than for Rx1day, with correlations in the range from 0.3 to 0.5. This is because Rx5day reflects more large-scale precipitation extremes rather than heavy (but short) convective events, which are more likely to be represented in Rx1day. As with extreme temperatures and the temporal correlations, ERA-Interim and ERA-40 again have somewhat higher spatial correlations with the interpolated observations than the other reanalyses.

4. Summary, discussion, and conclusions

We compare annual temperature and precipitation extremes across multiple interpolated observations and reanalysis datasets. In total, we use three gridded datasets of observational temperature extremes, two sets of gridded observational precipitation indices, and five different reanalyses from three different modeling centers. We analyze differences in actual values and compare

agreement in both temporal evolution and spatial pattern. The comparison is performed by considering temporal correlations of global and continental-scale area-averaged time series, trend maps, and annual pattern correlations.

We find substantial differences in the actual values of the annual extremes, which are partly related to different computational approaches for calculating the gridded extreme indices fields. A comparison of GHCNDEX and extremes in HadGHCND, both largely based on identical input stations, shows that the values tend to be more extreme (e.g., higher maxima and lower minima) if the extremes indices are first calculated for the station time series before calculating the grids. This shows how daily local extremes may be smoothed when calculating annual extremes from daily gridded fields (as, e.g., from climate model or reanalysis data). However, in addition there are also substantial differences between extreme values from the different reanalysis datasets, which all calculate the annual extremes based on daily gridded fields. Normalizing the different datasets relative to their specific climatology shows that temporal variability in the global and regional time series mostly compares well.

Investigating temporal changes in extremes, we find that there is high coherence between the interpolated observational datasets since the middle of the twentieth century, particularly for temperature extremes. Reanalyses, however, show larger inconsistencies compared to in situ-based observational data regarding temporal changes in climate extremes prior to the three most recent decades when satellite data are available for assimilation. Although all datasets are likely to be affected by inhomogeneities and data errors, we assume the interpolated observations to be more realistic in their long-term changes than the reanalyses. This is justified because, in producing the interpolated in situ-based extremes datasets, input station data are carefully checked for homogeneity (HadEX2), and only long-term stations are used (GHCNDEX) to minimize inhomogeneities from variability in station density. It is important to note that even long-term high-quality station time series might be affected by inhomogeneities related to instrumentation, station surroundings, or small changes in location, which are not always documented. However, those specific inhomogeneities are largely cancelled out when calculating the gridbox averages, particularly in data-rich areas. In general, consistency between the different datasets is higher for temperature extremes than for precipitation extremes, as may be expected since temperature is more homogeneous spatially than precipitation. In addition, while temperature observations are assimilated when calculating the

reanalysis fields (although near-surface temperatures are generally not directly assimilated and are thus strongly influenced by the specific atmospheric model), generally no precipitation observations are used for assimilation, and precipitation is calculated as a model output only (e.g., Kalnay et al. 1996).

Extreme temperatures from all datasets agree well in showing warming trends over most regions for the common data period 1979–2010. Only NCEP-1 shows largely inconsistent behavior in comparison to the other datasets, particularly for hot temperature extremes.

Also the spatial patterns of trends largely agree among the datasets during the comparison period in regions where in situ observations are available. However, the trend maps display a larger variety of changes in reanalyses data over regions not sufficiently covered by observations, such as Africa and South America. This reflects the weaker observational constraints on reanalyses in data-sparse regions. Therefore, it is difficult to estimate which of the reanalyses has a more realistic behavior than others in these regions. In the regions covered by observations, ERA-Interim on average shows highest temporal and spatial correlations with the interpolated observations.

Precipitation extremes are compared for the common period 1979–2008 when area averages from the interpolated observations are sufficiently robust. Changes in extreme precipitation amounts during this period are generally not significant. Note that some areas with significant trends were identified over the longer period 1951–2010 (Donat et al. 2013a). The reanalysis data display their strongest trends in tropical regions, where the interpolated observations generally do not provide sufficient coverage. These include largely consistent trends toward less extreme precipitation over tropical Africa and toward higher extreme precipitation amounts over Indonesia and northern Australia. However, trend patterns from the different reanalyses over South America are largely inconsistent.

In conclusion, we find a high level of consistency between the different interpolated observations datasets of temperature and precipitation extremes over the past 60 yr. Most reanalyses reproduce observed changes and spatial patterns reasonably well for the period after 1979 when satellite data are available for assimilation. Over areas with poor coverage of in situ observations, such as Africa and South America, the different reanalyses show differences in patterns of changes, such that regional results become strongly dependent on the specific dataset used.

Thus, our results indicate that current reanalyses confirm the observational trends in extremes in data-rich areas over the most recent three decades. However,

they seem to be of more limited use for the analysis of past trends in climate extremes over areas with sparse observational coverage. The results highlight the need for high-quality observational datasets to monitor changes in climate extremes, particularly the need to fill the gaps in the interpolated observations datasets by collecting high-quality climate data from the data-sparse regions in large parts of Africa and South America. Activities to obtain data, perform quality control, and analyze past changes in climate extremes are continuously supported by the joint WMO CCL/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (e.g., [Peterson and Manton 2008](#)).

Acknowledgments. This study was supported by Australian Research Council grants LP100200690 and CE110001028. J. Sillmann is funded by the German Research Foundation (DFG Grant Si 1659/1-1) and the AERO-CLO-WV project (184714/S30) funded by the Norwegian Research Council.

REFERENCES

- Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, doi:[10.1029/2005JD006290](#).
- Bengtsson, L., and K. I. Hodges, 2011: On the evaluation of temperature trends in the tropical troposphere. *Climate Dyn.*, **36** (3–4), 419–430, doi:[10.1007/s00382-009-0680-y](#).
- , S. Hagemann, and K. I. Hodges, 2004: Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11111, doi:[10.1029/2004JD004536](#).
- Caesar, J., L. Alexander, and R. Vose, 2006: Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *J. Geophys. Res.*, **111**, D05101, doi:[10.1029/2005JD006280](#).
- Cornes, R. C., and P. D. Jones, 2013: How well does the ERA-Interim reanalysis replicate trends in extremes of surface temperature across Europe? *J. Geophys. Res.*, **118**, 10262–10276, doi:[10.1002/jgrd.50799](#).
- Dee, D. P., and Coauthors, 2011a: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:[10.1002/qj.828](#).
- , E. Källén, A. J. Simmons, and L. Haimberger, 2011b: Comments on “Reanalyses suitable for characterizing long-term trends.” *Bull. Amer. Meteor. Soc.*, **92**, 65–70, doi:[10.1175/2010BAMS3070.1](#).
- Donat, M. G., G. C. Leckebusch, S. Wild, and U. Ulbrich, 2011: Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations. *Nat. Hazards Earth Syst. Sci.*, **11**, 1351–1370, doi:[10.5194/nhess-11-1351-2011](#).
- , and Coauthors, 2013a: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, doi:[10.1002/jgrd.50150](#).
- , L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013b: Global land-based datasets for monitoring climatic extremes. *Bull. Amer. Meteor. Soc.*, **94**, 997–1006, doi:[10.1175/BAMS-D-12-00109.1](#).
- Frankignoul, C., E. Kestenare, M. Botzet, A. Carril, H. Drange, A. Pardaens, L. Terray, and R. Sutton, 2004: An intercomparison between the surface heat flux feedback in five coupled models, COADS and the NCEP reanalysis. *Climate Dyn.*, **22**, 373–388, doi:[10.1007/s00382-003-0388-3](#).
- Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, doi:[10.1175/1520-0493\(1999\)127<2204:FASOCR>2.0.CO;2](#).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](#).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, doi:[10.1175/BAMS-83-11-1631](#).
- Kendall, M. G., 1975: *Rank Correlation Methods*. Charles Griffin, 272 pp.
- Kharin, V. V., F. W. Zwiers, and X. Zhang, 2005: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate*, **18**, 5201–5223, doi:[10.1175/JCLI3597.1](#).
- , —, —, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, doi:[10.1007/s10584-013-0705-8](#).
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267, doi:[10.1175/1520-0477\(2001\)082<0247:TNNYRM>2.3.CO;2](#).
- Min, S. K., X. Zhang, F. W. Zwiers, and G. C. Hegerl, 2011: Human contribution to more-intense precipitation extremes. *Nature*, **470**, 378–381, doi:[10.1038/nature09763](#).
- NCEP–NCAR, cited 2011: NCEP/NCAR reanalysis problems list. [Available online at <http://www.esrl.noaa.gov/psd/data/reanalysis/problems.shtml>.]
- Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432, doi:[10.2151/jmsj.85.369](#).
- Orlowsky, B., and S. I. Seneviratne, 2011: Investigating spatial climate relations using CARTs: An application to persistent hot days in a multimodel ensemble. *J. Geophys. Res.*, **116**, D14106, doi:[10.1029/2010JD015188](#).
- Peterson, T. C., and M. J. Manton, 2008: Monitoring changes in climate extremes: A tale of international collaboration. *Bull. Amer. Meteor. Soc.*, **89**, 1266–1271, doi:[10.1175/2008BAMS2501.1](#).
- Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722, doi:[10.1002/joc.1756](#).
- Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall’s Tau. *J. Amer. Stat. Assoc.*, **63**, 1379–1389, doi:[10.1080/01621459.1968.10480934](#).
- Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230.
- Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res.*, **118**, 1716–1733, doi:[10.1002/jgrd.50203](#).

- Simmons, A. J., and Coauthors, 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](https://doi.org/10.1029/2004JD005306).
- , 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 data sets. *J. Geophys. Res.*, **115**, D01110, doi:[10.1029/2009JD012442](https://doi.org/10.1029/2009JD012442).
- Spearman, C., 1904: The proof and measurement of association between two things. *Amer. J. Psychol.*, **15**, 72–101, doi:[10.2307/1412159](https://doi.org/10.2307/1412159).
- Svensson, G., and J. Karlsson, 2011: On the Arctic wintertime climate in global climate models. *J. Climate*, **24**, 5757–5771, doi:[10.1175/2011JCLI4012.1](https://doi.org/10.1175/2011JCLI4012.1).
- Thorne, P. W., and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends. Are they really achievable? *Bull. Amer. Meteor. Soc.*, **91**, 353–361, doi:[10.1175/2009BAMS2858.1](https://doi.org/10.1175/2009BAMS2858.1).
- , D. Parker, J. Christy, and C. Mears, 2005: Uncertainties in climate trends—Lessons from upper-air temperature records. *Bull. Amer. Meteor. Soc.*, **86**, 1437–1442, doi:[10.1175/BAMS-86-10-1437](https://doi.org/10.1175/BAMS-86-10-1437).
- Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Res.*, **47**, 123–138, doi:[10.3354/cr00953](https://doi.org/10.3354/cr00953).
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:[10.1256/qj.04.176](https://doi.org/10.1256/qj.04.176).
- Vose, R. S., S. Applequist, M. J. Menne, C. N. Williams Jr., and P. Thorne, 2012: An intercomparison of temperature trends in the U.S. Historical Climatology Network and recent atmospheric reanalyses. *Geophys. Res. Lett.*, **39**, L10703, doi:[10.1029/2012GL051387](https://doi.org/10.1029/2012GL051387).
- Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. Klein Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, doi:[10.1002/wcc.147](https://doi.org/10.1002/wcc.147).