JGR Atmospheres

RESEARCH ARTICLE
10.1029/2019JD032263

Key Points:
- We present an updated data set of gridded temperature and precipitation extremes
- Spatio-temporal coverage over 1901–2018 for 29 extremes indices was improved over previous version
- We find increased (decreased) intensity and frequency of warm (cool) extremes and increased intensity of heavy precipitation events

Supporting Information:
- Supporting Information S1

Correspondence to:
R. J. H. Dunn,
robert.dunn@metoffice.gov.uk

Citation:
Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., et al. (2020). Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD032263. https://doi.org/10.1029/2019JD032263

Received 16 DEC 2019
Accepted 26 MAY 2020
Accepted article online 2 JUL 2020

Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3

Robert J. H. Dunn[1], Lisa V. Alexander[2,3], Markus G. Donat[4], Xuebin Zhang[5], Margot Bador[2,3], Nicholas Herold[6], Tanya Lippmann[2,3,7], Rob Allan[1], Enric Aguilar[8], Abdoul Aziz Barry[8], Manola Brunet[8,9], John Caesar[1], Guillaume Chagnaud[10], Vincent Cheng[5], Thelma Cinco[11], Imke Durre[12], Rosaline de Guzman[11], Tin Mar Htay[13], Wan Maisarah Wan Ibadullah[14], Muhammad Khairul Izzat Bin Ibrahim[15], Mahbobeh Khoshkam[16], Andries Kruger[17,18], Hisayuki Kubota[19], Tan Wee Leng[20,21], Gerald Lim[20], Lim Li-Sha[20], Jose Marengo[22], Sifiso Mbatha[17], Simon McGree[23], Matthew Menne[12], Maria de los Milagros Skansi[24], Sandile Ngwenya[17], Francis Nkrumah[25], Chalump Oonariya[26], Jose Daniel Pabon-Caicedo[27], Gérémy Panthou[10], Cham Pham[28], Fatemeh Rahimzadeh[29], Andrea Ramos[30], Ernesto Salgado[31], Jim Salinger[32], Youssouph Sané[33], Ardhasena Sopaheluwakan[34], Arvind Srivastava[35], Ying Sun[36], Bertrand Timbal[20,23], Nichanun Trachow[26], Blair Trewin[23], Gerard van der Schrier[37], Jorge Vazquez-Aguirre[38], Ricardo Vasquez[39], Claudia Villarroel[39], Lucie Vincent[5], Theo Vischel[10], Russ Vose[12], and Mohd Noor'Arifin Bin Hj Yussof[15]

[1]Met Office Hadley Centre, Exeter, UK, [2]Climate Change Research Centre, UNSW Sydney, Sydney, New South Wales, Australia, [3]ARC Centre of Excellence for Climate Extremes, UNSW Sydney, Sydney, New South Wales, Australia, [4]Barcelona Supercomputing Center, Barcelona, Spain, [5]Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada, [6]Department of Planning, Industry and Environment, Sydney, New South Wales, Australia, [7]Department of Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, [8]Centre for Climate Change, Department of Geography, University Rovira i Virgili, Tarragona, Spain, [9]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK, [10]Univ. Grenoble Alpes, IRD, CNRS, Grenoble INP, IGE, Grenoble, France, [11]PAGASA-DOST, Manila, Philippines, [12]NOAA/NCEI, Asheville, NC, USA, [13]Department of Meteorology and Hydrology, Ministry of Transport and Communications, Naypyitaw, Myanmar, [14]Malaysian Meteorological Department, Petaling Jaya, Malaysia, [15]Brunei Darussalam Meteorological Department, Bandar Seri Begawan, Brunei Darussalam, [16]I.R. of Iran Meteorological Organization, Tehran, Iran, [17]Climate Service, South African Weather Service, Pretoria, South Africa, [18]Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa, [19]Department of Earth and Planetary Sciences, Hokkaido University, Sapporo, Japan, [20]Centre for Climate Research Singapore, Meteorological Service Singapore, Singapore, [21]University of Reading, Reading, UK, [22]National Center for Monitoring and Early Warning of Natural Disasters CEMADEN, Sao Paulo, Brazil, [23]Bureau of Meteorology, Melbourne, Victoria, Australia, [24]Servicio Meteorológico Nacional, Buenos Aires, Argentina, [25]Department of Physics, University of Cape Coast, Cape Coast, Ghana, [26]Climate Centre, Thai Meteorological Department, Bangkok, Thailand, [27]Department of Geography, National University of Colombia, Bogota, Colombia, [28]National Center for Hydro-Meteorological Forecasting of Vietnam, Vietnam Meteorology Hydrology Administration, Hanoi, Vietnam, [29]formerly Atmospheric Science and Meteorological Research Center, Tehran, Iran, [30]Instituto Nacional de Meteorologia (INMET), Corumbá, Brazil, [31]Servicio Meteorológico Nacional, Tegucigalpa, Honduras, [32]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Florence, Italy, [33]Agence Nationale de l'Aviation Civile et de la Météorologie (ANACIM), Dakar, Senegal, [34]Center for Research and Development, Agency for Meteorology Climatology and Geophysics (BMKG), Jakarta, Indonesia, [35]Climate Research and Services, IMD Pune, Pune, India, [36]National Climate Center, China Meteorological Administration, Beijing, China, [37]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands, [38]Licenciatura en Ciencias Atmosféricas, Universidad Veracruzana, Xalapa, Mexico, [39]Dirección Meteorológica de Chile (Chilean National Weather Service), Santiago, Chile

Abstract We present the second update to a data set of gridded land-based temperature and precipitation extremes indices: HadEX3. This consists of 17 temperature and 12 precipitation indices derived from daily, in situ observations and recommended by the World Meteorological Organization (WMO) Expert Team on Climate Change Detection and Indices (ETCCDI). These indices have been calculated at around 7,000 locations for temperature and 17,000 for precipitation. The annual (and monthly) indices have been interpolated on a 1.875°×1.25° longitude-latitude grid, covering 1901–2018. We show changes in these

indices by examining "global"-average time series in comparison with previous observational data sets and also estimating the uncertainty resulting from the nonuniform distribution of meteorological stations. Both the short and long time scale behavior of HadEX3 agrees well with existing products. Changes in the temperature indices are widespread and consistent with global-scale warming. The extremes related to daily minimum temperatures are changing faster than the maximum. Spatial changes in the linear trends of precipitation indices over 1950–2018 are less spatially coherent than those for temperature indices. Globally, there are more heavy precipitation events that are also more intense and contribute a greater fraction to the total. Some of the indices use a reference period for calculating exceedance thresholds. We present a comparison between using 1961–1990 and 1981–2010. The differences between the time series of the temperature indices observed over longer time scales are shown to be the result of the interaction of the reference period with a warming climate. The gridded netCDF files and, where possible, underlying station indices are available from www.metoffice.gov.uk/hadobs/hadex3 and www.climdex.org.

1. Introduction

Under a changing climate, climate extremes are projected to pose some of the most critical challenges on human physiological thresholds, society and infrastructure, and the natural environment. To understand how these extremes are currently changing, up-to-date observational products are required. Several international efforts exist to discover, collate, and process climate data and make them available to users. For almost two decades, the former WMO/WCRP/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) was one of the leading groups to develop a standard set of indices representing temperature and precipitation extremes for the terrestrial environment, representing 29% of the surface of the globe (Frich et al., 2002; Karl et al., 1999; Peterson, 2005). The development of these standard climate indices based on daily data has allowed derived products to be more easily and widely shared than the underlying measurements themselves (Alexander et al., 2019; Thorne et al., 2017; Zhang et al., 2011), because in many cases, these raw daily observations are not openly available. The ETCCDI facilitated several regional workshops, where representatives of National Meteorological and Hydrological Services (NHMSs) could attend and calculate the climate extremes indices from their daily data without needing to release those data publicly (Aguilar et al., 2009; Caesar et al., 2011; Donat et al., 2014; Peterson & Manton, 2008; Meteorological Service Singapore, 2019; Vincent et al., 2011). This approach has ensured the availability of climate indices from regions where daily data are not readily available free of charge to the scientific community. The development of standard software has further simplified this process, allowing contributors to calculate the indices without needing to travel to a workshop. The ETCCDI approach has now been adopted by other groups such as the WMO Expert Team for Sector-specific Climate Indices (ET-SCI) who have also run workshops and developed sector-focused (e.g., health and agriculture) indices and software (Alexander & Herold, 2015; Herold & Alexander, 2016).

The first global data set to contain all 27 indices defined by the ETCCDI was HadEX (Alexander et al., 2006), which covered the period 1951–2003. As a result of the way that part of the data were collected and processed for inclusion in HadEX (through the aforementioned regional workshops), this data set was released as a static product covering a fixed time period. A follow-up effort a few years later resulted in the updated HadEX2 (Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013), which extended the length of the data set, from 1901 to 2010. The number of land-surface stations contributing to HadEX2 was larger than that in HadEX, resulting in a larger proportion of the land-surface being covered by HadEX2. The underlying station indices were also made available where their license made this possible. Both these products have underpinned past Intergovernmental Panel on Climate Change (IPCC) assessments on global changes in climate extremes. They are available on an annual, 3.75°×2.5° longitude-latitude grid (chosen to match that of the HadCM3 model when developing HadEX), with some indices available on a monthly time scale as well (through www.climdex.org). Both HadEX and HadEX2 included three additional indices over the 27 from the ETCCDI (ETR, R95pTOT, and R99pTOT, see Table 1). One index (Rnnmm, using user-defined thresholds) is difficult to combine between data sources and so was not included in these two data sets.

As HadEX2 was again a static data set and is not easy to keep updated, another data set, GHCNDEX, was developed by Donat, Alexander, Yang, Durre, Vose, & Caesar (2013), which uses the same indices and a

Table 1
Details of the ETCCDI Indices and the Numbers of Stations Available for Each That Were Taken Forward Into the Gridded Product

Index	Name	Description	Units	Stations (1961–1990 ref. period)	Stations (1981–2010 ref. period)
TXx	Hottest day	Monthly and annual highest value of daily max temperature	°C	7,191	
TNx	Warmest night	Monthly and annual highest value of daily min temperature	°C	7,209	
TXn	Coldest day	Monthly and annual lowest value of daily max temperature	°C	7,061	
TNn	Coldest night	Monthly and annual lowest value of daily min temperature	°C	7,069	
TN10p	Cool nights	Percentage of time when daily min temperature <10th percentile	%	4,869	4,615
TX10p	Cool days	Percentage of time when daily max temperature <10th percentile	%	4,891	4,624
TN90p	Warm nights	Percentage of time when daily min temperature >90th percentile	%	4,867	4,615
TX90p	Warm days	Percentage of time when daily max temperature >90th percentile	%	4,891	4,623
DTR	Diurnal temperature range	Annual mean difference between daily max and min temperature	°C	7,056	
GSL	Growing season length	Annual (1 January to 31 December in Northern Hemisphere, 1 July to 30 June in Southern Hemisphere) count between first span of at least 6 days with TG > 5°C and first span after 1 July (1 January in SH) of 6 days with TG < 5°C (where TG is daily mean temperature)	days	5,503	
ID	Ice days	Annual count when daily maximum temperature <0°C	days	6,996	
FD	Frost days	Annual count when daily minimum temperature <0°C	days	7,019	
SU	Summer days	Annual count when daily max temperature >25°C	days	7,180	
TR	Tropical nights	Annual count when daily min temperature >20°C	days	7,203	
WSDI	Warm spell duration index	Annual count when at least six consecutive days of max temperature >90th percentile	days	5,992	6,358
CSDI	Cold spell duration index	Annual count when at least six consecutive days of min temperature <10th percentile	days	5,813	6,328

very similar methodology but can be continuously updated (either annually or monthly). GHCNDEX uses the Global Historical Climatology Network Daily (GHCND) data set (Menne et al., 2012) as input rather than the mix of national and regional collections in HadEX2. The ability to update GHCNDEX has allowed its inclusion in climate monitoring reports (e.g., King et al., 2019; Perkins-Kirkpatrick et al., 2018), although the coverage of GHCNDEX is not as complete as HadEX2. However, as GHCN-Daily station data are freely available, the GHCNDEX data set is fully traceable, whereas for HadEX and HadEX2, some of the station time series that have been provided could not (and unfortunately still cannot) be made available free of charge.

By comparing HadEX2 and GHCNDEX, it became possible to see the impact of the input station network and coverage on the final annual (and monthly) fields. This has been further studied using indices calculated from reanalysis products (Donat et al., 2014) and historical simulations from the CMIP5 archive (Sillmann et al., 2013). Other regional investigations concerning the impact of different gridding methods on the final gridded fields were carried out by, for example, Avila et al. (2015) and Contractor et al. (2015). These showed that although overall, large-scale relative patterns were consistent, regional differences were apparent. They highlighted that the order of operation (first calculating the indices from the station time series, then interpolating the indices to a grid versus first gridding the daily station time series, and then calculating the indices on the gridded fields) resulted in different values of the gridded extremes depending on the gridding method used. The former has been done in HadEX3 and its predecessors but the latter in HadGHCND; Caesar et al. (2006) created gridded fields of maximum and minimum temperatures, which have also been used to calculate grids of temperature extreme indices in comparison to gridding station extremes (Donat, Alexander, Yang, Durre, Vose, & Caesar, 2013), allowing this difference to be investigated. However, for long-term trends and interannual variability, these effects appear to be much smaller. A wider assessment of the effects of both structural (i.e., use of different methods) and parametric (i.e., from within method choices) uncertainties concluded that the largest effects arose from the choice of the gridding method and the density of the input station network (Dunn et al., 2014). In combination, these studies indicate that the HadEX family of data sets do have uncertainties within them. But, for indices showing large changes over time, and especially in regions with a dense station network, these uncertainties are small, as the spread between the different products is small

Table 1
(Continued)

Index	Name	Description	Units	Stations (1961–1990 ref. period)	Stations (1981–2010 ref. period)
**ETR*	Extreme temperature range	TXx - TNn	°C	7,068	
Rx1day	Max 1 day precipitation amount	Monthly and annual maximum 1-day precipitation	mm	17,524	
Rx5day	Max 5 day precipitation amount	Monthly and annual maximum consecutive 5-day precipitation	mm	17,529	
SDII	Simple daily intensity index	The ratio of annual total precipitation to the number of wet days (≥1 mm)	mm/day	17,763	
R10mm	Number of heavy precipitation days	Annual count when precipitation ≥10 mm	days	17,705	
R20mm	Number of very heavy precipitation days	Annual count when precipitation ≥20 mm	days	17,705	
CDD	Consecutive dry days	Highest number of consecutive days when precipitation < 1 mm	days	17,776	
CWD	Consecutive wet days	Highest number of consecutive days when precipitation ≥1 mm	days	17,786	
R95p	Very wet days	Annual total precipitation from days >95th percentile	mm	14,870	7,912
R99p	Extremely wet days	Annual total precipitation from days >99th percentile	mm	14,856	7,880
PRCPTOT	Annual total wet day precipitation	Annual total precipitation from days ≥1 mm	mm	17,778	
*R95pTOT	Contribution from very wet days	100 * R95p / PRCPTOT	%	14,801	7,635
*R99pTOT	Contribution from extremely wet days	100 * R99p / PRCPTOT	%	14,798	7,621

Note. The indices labeled in italics are available on a monthly as well as annual basis. The three indices marked with an asterisk are additional indices which were included in HadEX2 and HadEX.

compared to the longer-term changes. Conversely, indices showing small long-term changes or regions containing few stations showed a greater spread (relative to the long-term change signal) between the different versions of the data sets.

In this manuscript, we outline the second update to the HadEX family, presenting HadEX3. The latest round of data gathering is presented in section 2, with a recap of the methods in section 3. We highlight some of the global and regional behaviors of the indices over the recent past in section 4 and compare to other gridded data sets of extremes indices. Our discussions and summary are presented in sections 5 and 6.

2. Data

Regional data collections and contributions from individuals and organizations, as both indices and raw daily temperature and precipitation observations, have been combined to form HadEX3. Below we outline the characteristics of these collections. Major collections are as follows, with the data span included in HadEX3 given in parentheses at the end of each entry:

- The European Climate Assessment and Dataset (ECAD, https://www.ecad.eu//indicesextremes/index.php, Klein Tank et al., 2002; Klok & Klein Tank, 2009) contains indices from around 5,000 temperature stations and almost 15,000 precipitation stations across Europe, western Asia, and northern Africa. The ECAD use slightly different definitions for some of their indices (TX90p, TX10p, TN90p, TN10p, R95p, and R99p). We adjust these to match the definitions in Table 1as part of the processing of HadEX3. These indices are based on the blended data set (see http://www.ecad.eu/FAQ/index.php\#3), which uses neighboring stations to infill data gaps (1901–2018).
- The Southeast Asia Climate Assessment and Dataset (SACAD, https://sacad.database.bmkg.go.id/indicesextremes/index.php, Klein Tank et al., 2006) contains indices from around 400 temperature stations and

around 4,000 precipitation stations from countries in Southeast Asia. We do not include SACAD stations in Australia as separate data sets were used for that region in combination with GHCNDEX. (1952–2011).
- The Latin American Climate Assessment and Dataset (LACAD, https://lacad.ciifen.org/indicesextremes/index.php Van Den Besselaar et al., 2015) contains indices from 200 temperature stations and 1,200 precipitation stations from countries in western South America. (1932–2014).

These three data sets contain precalculated indices referenced to the 1961–1990 period. As we also calculate some indices using a 1981–2010 reference period (see sections 2.2 and 4.3), we also use the daily station measurements from these collections (named EOBS, SAOBS, and LAOBS) for those calculations.

- The GHCND (Menne et al., 2012) is a comprehensive data set of over 100,000 stations across the globe (https://www.ncdc.noaa.gov/ghcnd-data-access). We use the same homogeneous subset from Peterson et al. (2008) as was used in HadEX2, which are those National Weather Service Cooperative and first-order stations with reasonably long records over the United States. These were determined to be free from significant discontinuities after 1950. After selecting for data record (section 3.1), there are 1,320 GHCND stations available for calculating indices (1901–2018).
- A subset of those stations from GHCNDEX (Donat, Alexander, Yang, Durre, Vose, & Caesar, 2013) where there are at least 40 years of data after 1950 and over Australia only were included (https://www.climdex.org). These supplement the network of supplied precipitation stations in that region and in effect contain indices from all precipitation stations for Australia (1901–2018).
- The publicly available observation data from the DECADE project (Andrade et al., 2018) for stations in the central Andes (https://www.geography.unibe.ch/research/climatology_group/research_projects/decade/index_eng.html). These data have had their homogeneity assessed using the ACMANT code (Domonkos, 2011), and we masked time periods where the gross inhomogeneity frequency of a station is moderate or high (1919–2015).

The remaining regional and national collections have been submitted by authors through either their personal research, via the National Meteorological Service in that country, or from a number of workshops held over the last decades.

- Six hundred stations over the former USSR as an update to the stations described in Razuvaev et al. (1993) (http://meteo.ru/english/climate/d_temp.php) (1901–2018).
- Indices from 615 precipitation stations over Western Africa from Panthou et al. (2012) were included. These have been subject to a visual and automatic check of data quality to identify spurious and erroneous data (Panthou et al., 2018) (1901–2016).
- Indices from around 260 temperature stations and 280 precipitation stations from countries in South America for 1950–2009 in the South America assessment and data set (de los Milagros Skansi et al., 2013). Time series were quality controlled used RClimDex and extra QC checks. The temperature time series were subjected to homogenization by using the RSNHT software, whereas RHtestV3 (Wang & Feng, 2013) was used to test homogeneity on the precipitation series (1950–2017).
- A further set of around 350 temperature and precipitation stations from 1961 from the Regional Climate Center for the South of South America was also included. These data were quality controlled using a suite of quality control (Podestá et al., 2013; Veiga et al., 2014) (1950–2017).
- Twenty-two temperature and precipitation stations over Spain, which have been quality controlled using the RClimDex-extra QC (http://www.c3.urv.cat/softdata.php) and homogenized at the daily scale following procedures described in Brunet et al. (2006) and Sigró et al. (2015) (1901–2017).
- One hundred twelve homogenized temperature stations from the Australian Climate Observations Reference Network Surface Air Temperature (ACORN-SAT) data set, version 2 ((Trewin et al., 2020), https://www.bom.gov.au/climate/data/acorn-sat/) and precipitation data from those stations (1901–2017).
- Two hundred fifty-six stations from the ACRE project (Allan et al., 2011). Many have records which end before 1901, and so indices were only calculated for 46 stations. As all end early in the 20th century, none are carried through into the final gridded product (1782–1950).
- Indices from 319 stations over China, quality controlled and homogenized as detailed in Xu et al. (2013) 1951–2018).
- Three hundred thirty-eight homogenized daily temperatures up to 2017 (Vincent et al., 2012) and 464 high-quality daily total precipitation stations (adjusted for known measurements issues Mekis &

Vincent, 2011) for Canada. Homogeneity issues related to station relocation, changes in observing time, and joining of observations from colocated sites were addressed using the observations of neighboring stations. Daily total precipitation was the sum of the daily rainfall and snowfall water equivalent amounts (1901–2017).

- Eighteen stations over Chile with quality control carried out by a visual inspection and automated tests. The homogenization of the series was carried out by means of the SNHT homogeneity test with the Climatol package, which corrects and fills the series (Guijarro, 2016) (1961–2018).
- Twenty-two stations from Colombia which have been used for a national assessment published by Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) (2018) (1980–2017).
- Indices from 26 temperature stations and 84 precipitation stations over South Africa. These are an update to the data sets used in Kruger and Nxumalo (2017a) for surface temperature and Kruger and Nxumalo (2017b) for precipitation (1921–2018).
- Around 5,300 stations in Mexico were received from full national climatic database of the Mexican National Meteorological Service (SMN) of the National Water Commission (CONAGUA). Basic quality control is routinely applied to the data by SMN using CLICOM and other basic validations. No homogeneity procedures were applied to the data (1901–2018).
- Six hundred fourteen stations in Brazil, from the archives of the National Institute of Meteorology (INMET) (1933–2018).
- Indices from 61 stations over North Africa and Arabian Peninsula from a workshop in Morocco in 2012 (Donat et al., 2014) (1940–2011).
- Indices from 29 temperature stations and 52 precipitation stations from 20 western Pacific Islands countries and territories (McGree et al., 2019) (1951–2015).
- Indices from 154 precipitation and 123 temperature stations from past workshops held in Mauritius, Vietnam, and Congo that were also used in HadEX2 (Aguilar et al., 2009; Caesar et al., 2011; Vincent et al., 2011) (1922–2008).
- Indices from 169 stations in West Africa arising from a workshop held in the Gambia in 2011 (Barry et al., 2018) (1960–2010).
- Indices from 112 stations over Southeast Asia, calculated during a workshop in Singapore in 2019 (Brunei: 1, Singapore: 3, Philippines: 4, Thailand: 7, Vietnam: 14, Malaysia: 15, Myanmar: 17, Indonesia: 51 Meteorological Service Singapore, 2019) (1951–2018).
- Ninety-six stations over India (1901–2016).
- Fifty-eight long-record stations over Iran with updated locations (Rahimzadeh & Nassaji Zavareh, 2014) (1951–2018).
- Twenty-four stations from New Zealand and surrounding islands (1901–2017).
- Six stations over Honduras (1951–2005).
- Ten stations over Japan (1901–2018).

We note that the level of quality control and homogeneity adjustment or assessment applied to each station varies between the sources. We do no quality checks on the observed data but do apply further tests on the station-level indices (section 3.2). Data sources which have been homogenized are Australia (temperature only), Canada, China, DECADE, South America assessment and data set, and Spain. We note that the homogenization techniques applied may not be well suited for extremes. Furthermore, some changes, for example, from manual to automatic stations, may have occurred across a wide region contemporaneously, and so (likely small) systematic biases may still be present. The maximal number of stations for each index is shown in Table 1. For sources where indices have been provided, not all are available for both reference periods (sections 2.2 and 4.3), which may be the result of, for example, the station record length or effort available to perform the calculations on archived data.

2.1. Indices

The ETCCDI indices have a long heritage in their use for monitoring, analyzing, and projecting moderate extreme events and absolute values (Alexander et al., 2019; Zhang et al., 2011). They are calculated from daily maximum and minimum temperature as well as daily precipitation amount. In many cases, these indices have been calculated by data contributors using the RClimdex, climdex.pcic, or FClimdex software (Zhang & Feng, 2004). When data have been submitted as daily observations, we have used the Climpact2

v1.2.7 software (Herold & Alexander, 2016) to calculate the indices. Climpact2 is the most recent instance of code used at regional workshops to calculate the ETCCDI and ETSCI indices and builds on the earlier codes (e.g., climdex.pcic). It is available from https://github.com/ARCCSS-extremes/climpact2 and https://climpact-sci.org/to enable others to calculate the indices.

Some of the indices, being summaries of the underlying daily observations, require careful interpretation, especially when gridded or expressed as trends. In the case of indices using a fixed threshold (SU, TR, FD, ID, R10mm, and R20mm), some parts of the world have climates which either very rarely or almost continually trigger these indices, and so conclusions about trends need to be combined with understanding about the local climate. The temperature indices providing annual (or monthly) actual temperature extremes (TXx, TXn, TNx, and TNn) can be affected by a changing station network during the gridding. As elevation influences the temperatures recorded at each station, if a mountaintop station drops out of the network at the end of the series, then the weighted average value recorded will increase as a result and may appear as a step-change in the time series. We attempt to mitigate this effect with the station selection criteria (see section 3).

The percentile indices (TX90p, TX10p, TN90p, TN10p, R95pTOT, and R99pTOT), although being easily linked to common descriptions ("warm days," etc.), also need careful interpretation. By design, these indices have an average value of 10% (temperature) or 5% or 1% (precipitation) over the reference period (using daily rather than seasonal values for the calculation Zhang et al., 2005). Given the globally warming climate (Stocker et al., 2014), then the number of cool days and nights (TX10p and TN10p) will decrease on average over time (e.g., Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013). But the minimum number of these is bounded by zero. Conversely, the number of warm days and nights (TX90p and TN90p) have much greater freedom to increase, although they are bounded by the number of days in a year or month. Also, in regions with relatively low interannual variability, small changes in these indices can be more prominent than in regions with higher year-to-year variation. Finally, the use of a reference period can introduce artificial inhomogeneities in the time series of these indices (Zhang et al., 2005).

More generally, other data measurement changes, for example, changes in reporting resolution or in the observation/accumulation times, can also affect the homogeneity of the index time series (e.g., Rhines et al., 2015).

2.2. Notes on the Reference Period

A number of the ETCCDI indices (TX90p, TX10p, TN90p, TN10p, CSDI, WSDI, R99p, R95p, R95pTOT, and R99pTOT) use a reference period for the calculation of the thresholds used (e.g., the 10th percentile for TX10p). These indices summarize exceedances over/under these thresholds. If the global climate had no long-term anthropogenic forcing, then changing the reference period would have almost no effect on the indices and their long-term trends. Only small differences would arise from changes in the station network resulting from the ability to include stations with records which only overlap one or another reference period. As the climate is not static (Stocker et al., 2014), the difference of behavior of these indices under different reference periods is complex, being more than just a shift in the mean. A small change in the threshold can result in a much larger change in the exceedance (cumulative probability). Furthermore, as noted in the previous section, the bounded nature of some indices also affects the values when comparing between reference periods.

HadEX and HadEX2 (and GHCNDEX) used the reference period 1961–1990, which coincided with the peak of available data and stations at the time, as well as being a well-established base period and used extensively across climate science and in IPCC reports. However, a number of data contributors to HadEX3 have supplied indices referenced to the more recent period, 1981–2010, as a better match to the station records. The advantages of this more recent period are two fold: (i) that the peak of data availability has now moved to later in the time series, and so for individual (especially national) sources, more stations may be available with coverage over this period, and (ii) it also aligns with the current WMO standard climatological period. The disadvantages are that this is a change from the procedure used in HadEX and HadEX2, and some sources only have these indices calculated for a single reference period. Several large collections and regions are not available with the updated reference period (e.g., ECAD and GHCNDEX), and these dominate the change in global station counts. We have attempted to balance the spatial coverage of the gridded data set between the two

reference periods, but the underlying station networks are different in, for example, Western Africa, South America, Southeast Asia, and Australia. Furthermore, without access to underlying observations, we are not able to amend the base period of precalculated indices, and so each choice of the reference period restricts the use in HadEX3 of stations with precalculated indices (that rely on base period calculations).

Our approach for HadEX3 has been to produce two versions of the percentile-based indices: one using the 1961–1990 reference period and another using 1981–2010. Although this makes the data set a little more complex for users as it will require a decision on which version to use, it also offers users greater freedom by allowing them to choose the version that is most applicable to their needs. By analyzing both versions herein, we are able to assess the differences between them. We are also able to include as much of the submitted data as possible, which improves the spatio-temporal coverage for some regions for the indices which rely on a reference period (e.g., western and southern Africa and Southeast Asia). Submissions of daily data rather than indices can be used for both versions, with only small changes to these station networks.

3. Methods

Having obtained data from almost 37,000 stations, we select those which have sufficiently long records. The quality of these is then assessed, giving the final set of stations available to combine into the gridded product. We use the same gridding method as in HadEX and HadEX2, which combines stations within a search radius of the grid box center.

3.1. Station Selection

As in HadEX2, HadEX3 comprises two parts. The primary product is gridded, using spatial interpolation/gridding to combine the station-based indices from the different data sources, and is the main focus here. The secondary products are the underlying station-based indices, where these can be made available. Data from almost 37,000 stations have been collected as part of HadEX3 (see Figure A1), but many of these will have short records, data gaps, inconsistencies in the indices, and other issues and hence do not contribute to the gridded product.

If there are data present in the index time series for a station, then we require the equivalent of at least 20 years of data for it to be included in the gridded product. Stations which only have data prior to the 20th century are not included in this gridded data set. The gridded data set is intended for long-term monitoring, and so a changing station network, especially toward the more recent years, can result in inhomogeneities in the final grids as stations drop out as their record ends. Therefore, to reduce such inhomogeneities (which can result in spurious trends), we only select stations which end in 2009 or later for the gridded product. We note that this reduces the number of stations included.

3.2. Quality Control

3.2.1. Climpact2

Climpact2 produces diagnostic information and performs a variety of tests to help the user conduct quality control (QC) on their daily temperature and precipitation data. These tests include identifying

- statistical outliers (based on multiples of the inter-quartile range),
- potential biases in the rounding of decimal values,
- multiple consecutive days with identical temperatures,
- duplicated dates,
- unlikely large changes in temperatures between days (when minimum or maximum temperatures change by 20°C or more),
- days where minimum temperature exceeds maximum temperature, and
- uncommonly large precipitation and temperature values (exceeding 200 mm and 50°C, respectively).

These tests are designed to focus the user's attention on common issues in station data without the need for manual inspection of each value.

A number of automated quality control tests are run on the index time series of the indices after the initial selection. These aim to identify logical issues with the indices, as well as errors in the station locations. Many of these checks identify systematic issues in the index time series rather than single values being at fault, and hence many of these quality control tests are not triggered by the data. Where an

issue was identified by these index-based tests, the whole station is withheld from further processing for that index. The final number of stations available for the annual indices after QC and selection are shown in Table 1.

3.2.2. File Checks (All)

Although a redundant check, coming as it does after the station selection step, this test confirms that the file for a specific station, time scale, and index is present and contains some data.

3.2.3. World Record Check (TXx, TNn, Rx1day, and PRCPTOT)

These four indices are compared to the WMO record values for maximum and minimum temperature and daily and annual rainfall accumulations (https://wmo.asu.edu/#continental). We only assess the world, rather than continental records, aiming to catch factor-of-10 errors, issues with unit conversions, or unusual missing data indicators.

3.2.4. Reference Period Coverage (TX10p, TX90p, TN10p, TN90p, R95p, R99p, R95pTOT, and R99pTOT)

These eight indices use a specified reference period (e.g., 1961–1990) to calculate percentile values used to derive the final quantities. These percentiles should only be calculable if there are sufficient years within the 30-year reference period. For HadEX2, this was 70% (21 years), and so a flag is set if there are fewer than 21 years in which this index is reported during the reference period.

3.2.5. Annual in Monthly (TXx, TXn, TNx, TNn, Rx1day, and Rx5day)

As these six indices measure the absolute maximum (or minimum) values, this test cross-checks that the annual value for each index can be found as one of the 12 separate monthly values.

3.2.6. Negative Values (all but TXx, TXn, TNx, and TNn)

This test checks for negative values in time series of these indices. All of these indices have a lower bound of zero, and so negative values should not be present.

3.2.7. Consistency Check (TXx [TXn, TNx], TNx [TNn], TXn [TNn], R10mm [R20mm], R95p [R99p], PRCPTOT [R95p], and Rx5day [Rx1day])

This test works on pairs of indices. For each of the pairs (primary and secondary given in the list above), the former is expected to have a larger value than the latter; for example, the highest T_{max} in a month should be larger than the lowest T_{max} in a month. Stations where this is not the case are flagged for both indices as we cannot tell which one contains the erroneous values.

3.2.8. Metadata Checks (All)

This checks for exact matches in the latitude and longitude and so will identify stations which are apparent duplicated stations, either arising because the station appears in two or more input data sources or because there is a problem in one of the sets of metadata. In dense station networks, limitations on the accuracy of the positional metadata could result in two distinct stations being identified as duplicates by this test.

After candidate pairs are identified, these are then resolved, to retain only one in the data set. We deprioritize sources which are collections (ECAD, SACAD, LACAD, EOBS, SAOBS, LAOBS, GHCNDEX, and GHCND) and favor sources from individual countries or other regional programmes as they are likely to have had more detailed quality control applied. If this step does not result in a single station being selected (e.g., both are from collections or both are not), we then keep the source that has a longer record, or, if both have the same record length, then we take the one which ends more recently. If none of these result in just a single source being kept, we just retain the first in the list.

3.2.9. Correlation Checks (all except FD, ID, SU, TR, R10mm, and R20mm)

During early testing of the code, we found some instances where widely separated station pairs had a high correlation (>0.99). This could indicate a metadata problem as well as issues in the index data. We calculate the correlation between all station pairs, excluding any station where the metadata checks have set a flag. Then if there are high correlation values at separations of greater than 250 km, a flag is set.

The above six indices specified are excluded from this process as, for some parts of the world, these threshold values are almost-always or almost-never reached. A sequence of almost all zero or 365 days will correlate highly at large separations and result in many erroneous flags being set.

3.3. Gridding Method

The gridding method used for HadEX and HadEX2 was the angular distance weighting (ADW) scheme (Shepard, 1968). It has also been used in a number of other data sets (e.g., Caesar et al., 2006; Contractor

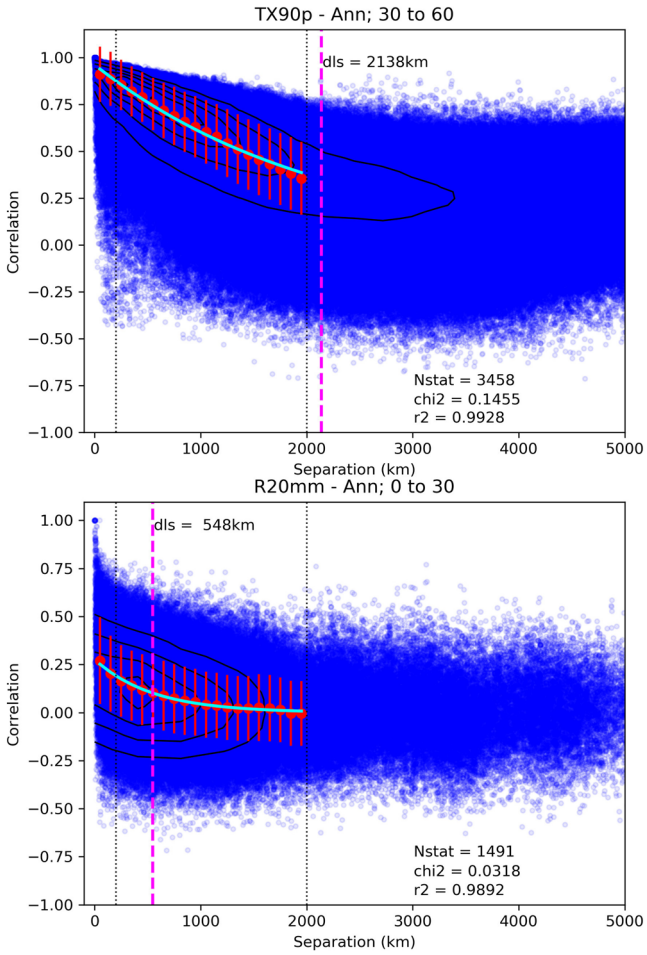

Figure 1. Estimation of the decorrelation length scale (DLS), for TOP TX90p between 30N and 60N and BOTTOM R20mm between 0 and 30N for annual values. Blue points show the correlation and separations for each station pair within the latitude band. Black contours show the regions where most points are concentrated. Red points and error bars are the averaged values which are fitted with the cyan line. The DLS is marked by the vertical dashed magenta line. Vertical dotted lines mark the upper and lower allowed values of the DLS. The statistics of the fit (R^2 and χ^2) as well as the number of stations in the latitude band are stated on the bottom right.

et al., 2015). This method estimates values at the central point of each grid box, blending stations within a search radius from that central point. This search radius is the decorrelation length scale (DLS), calculated from the correlation properties of station pairs. Although the value that ADW produces is likely to be representative of the grid box, it is not the average over the grid box area; especially in region/index combinations where elevation is important and varies, the representativity is likely to be less than in other regions. When comparing HadEX3 to reanalysis or model products, care needs to be taken as these other products provide grid-box averages and so there is an order-of-operation difference between them and HadEX3. For further discussion on the effect of this, see section 5.1 and, for example, Avila et al. (2015). One advantage of gridding the ETCCDI indices (rather than gridding the underlying observations and then calculating the indices) is that the indices tend to correlate over larger distances. However, a changing station network can still have adverse impacts for some indices which are closer to the underlying observations than others (see section 2.1).

3.3.1. Decorrelation Length Scale

The ADW scheme (Shepard, 1968) uses a search radius to identify stations close enough to the grid box center to contribute to the weighted average. As in HadEX and HadEX2, we use the DLS (also known as the correlation decay distance). This is a measure of how the correlation between stations reduces as their spatial separation increases and ensures the search radius is set at an appropriate value for the index (compared to using a fixed value for all indices).

The correlation of all station pairs with sufficient overlap (at least 30 years of data after 1951) is calculated, then placed into 100-km bins, and the median is taken. We fit an exponential decay curve of the form

$$Y = A.\exp(-x/b) + c \quad (1)$$

to these binned values if there are at least 30 stations with enough overlap. The form of the function fitted fixes c=0. The DLS is the point at which this curve has reached a value of A/e (the e-folding distance). The minimum DLS globally is set to 200 km and the maximum to 2,000 km (Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013). The standard deviations of the bin values are used as the uncertainties on the points when passed to the fitting procedure. This allows both the R^2 and chi-squared values to be calculated. If the fit is insufficiently good as determined by the chi-square value at 5%, then the calculated DLS is rejected and the minimum value of 200 km used.

The DLS is calculated for each index separately (annual and monthly where appropriate). As in HadEX and HadEX2, the DLS is also estimated in zonal bands, four 30° bands between 90°N and 30°S and then one 60° band for 30°S and 90°S. The larger band for the southern part of the Southern Hemisphere reflects the smaller fraction of land present in this region, as well as the much lower station density. We show in Figure 1 examples of the correlation-separation plots with the fitted decay curve and calculated DLS value. To avoid discontinuities at the zonal band boundaries, the DLS values are linearly interpolated onto the latitudes of the grid box centers.

In HadEX2, as well as the average correlations for each 100-km bin, a point was added at zero distance with perfect correlation (x=0 km, r=1), the theoretically expected value for two perfect instruments at the same location. However, in practice, this would rarely be the case, as demonstrated by a "nugget effect" (e.g., Cressie, 1992), which is clearly visible in Figure 1, as well as other studies in, for example, global air temperature (Rohde et al., 2013) and marine surface data (Kent et al., 1999). Therefore, and especially with indices

which use a fixed threshold, resulting correlations would be less than perfect at short separations. Hence, for HadEX3, we relax this criterion as although accurate characterization of the correlation behaviors at short distances is important for some applications, when creating HadEX3, we are interested in the long distance (more than a few km) behavior to calculate the DLS (minimum value of 200 km).

3.3.2. Angular Distance Weighting

This method (Shepard, 1968) has been shown to be an appropriate method for gridding irregularly spaced data (New et al., 2000). We increase the spatial resolution of HadEX3 over its predecessor by a factor of 4, using a 1.875°×1.25° longitude-latitude grid. All stations which lie within the DLS of the grid box center are selected, and then weights are assigned to each station depending on both the distance and the angular distribution of these selected stations. This method has been described in many other contexts (e.g., Caesar et al., 2006; Hofstra & New, 2009; New et al., 2000), but for reference, we summarize it in Appendix A1.

At least three stations with valid data need to be present within the DLS for a value to be set for a grid box at that time period, with a missing value placed if there are fewer. This is to ensure that stations with inhomogeneities or other data issues not already identified or removed cannot overly influence the final fields in regions of sparse station density. Dunn et al. (2014) relaxed this requirement down to a single station, which has the greatest effect in regions with a low station density (e.g., Africa), the early part of the record when few stations are available globally, or indices with a small DLS (e.g., R99p). Indices with strong long-term changes (e.g., TX10p) were less affected.

Alternative gridding methods for the ETCCDI extremes indices have been presented in, for example, Dunn et al. (2014), Yin et al. (2015), Avila et al. (2015), and Contractor et al. (2015). These range from a simple approach by averaging all stations within the grid box boundaries, which has lower coverage as it does not interpolate, through to ordinary kriging, which can be adapted to include other predictors. We leave an assessment between different gridding methods for a future study.

3.4. Coverage Uncertainty

As the spatial coverage of HadEX3 is not globally complete (for the land surface), we have included an assessment of the uncertainty in the global time series arising from this (see section 4). We follow the methodology of Brohan et al. (2006) when calculating the spatial coverage uncertainty, using the new ERA5 reanalysis (Hersbach et al., 2019). The predecessor of ERA5, ERA-Interim (Dee et al., 2011), showed some of the highest spatial and temporal correlations with observational data sets (Donat et al., 2014), with the temperature indices having greater correlations (r∼0.9) than the precipitation indices (r∼0.6). Although ERA5 has not undergone such an assessment, the high correlations between reanalysis products from the same modeling center (Donat et al., 2014) suggest that ERA5 may show reasonable spatio-temporal correlations with HadEX3.

We calculate the ETCCDI indices from ERA5 by aggregating the hourly temperature and precipitation values to daily values, obtaining the maximum/minimum and sum, respectively, for each 00-23Z period and passing these through the Climpact2 software. Then, for each year (and month if appropriate) in HadEX3, the ERA5 fields (for the matching calendar month if appropriate) are selected. These ERA5 fields are regridded to match the resolution of HadEX3 and then are masked by the spatial coverage of the HadEX3 field, and the global average (cosine-weighted) is calculated. The residuals between the masked and unmasked averages are taken, and the standard deviation of this distribution of residuals is used as the spatial coverage uncertainty in the global average time series in section 4. As ERA5 has not been assessed for use with the ETCCDI indices (unlike other products in, e.g., Donat et al., 2014), we do not use this in our comparison of the time series of different products but only for estimating the uncertainty arising from incomplete spatial coverage.

4. Results

We present the inferences from the version of HadEX3 using a 1961–1990 reference period, on a 1.875°×1.25° longitude-latitude grid (see section 4.3 for comparisons to the 1981–2010 reference period for the indices affected).

Time series of the ”global-land” average values for HadEX3 and the other data sets (HadEX, HadEX2, and GHCNDEX) are derived using cosine-weighting for the grid box center latitudes. Only grid boxes with at

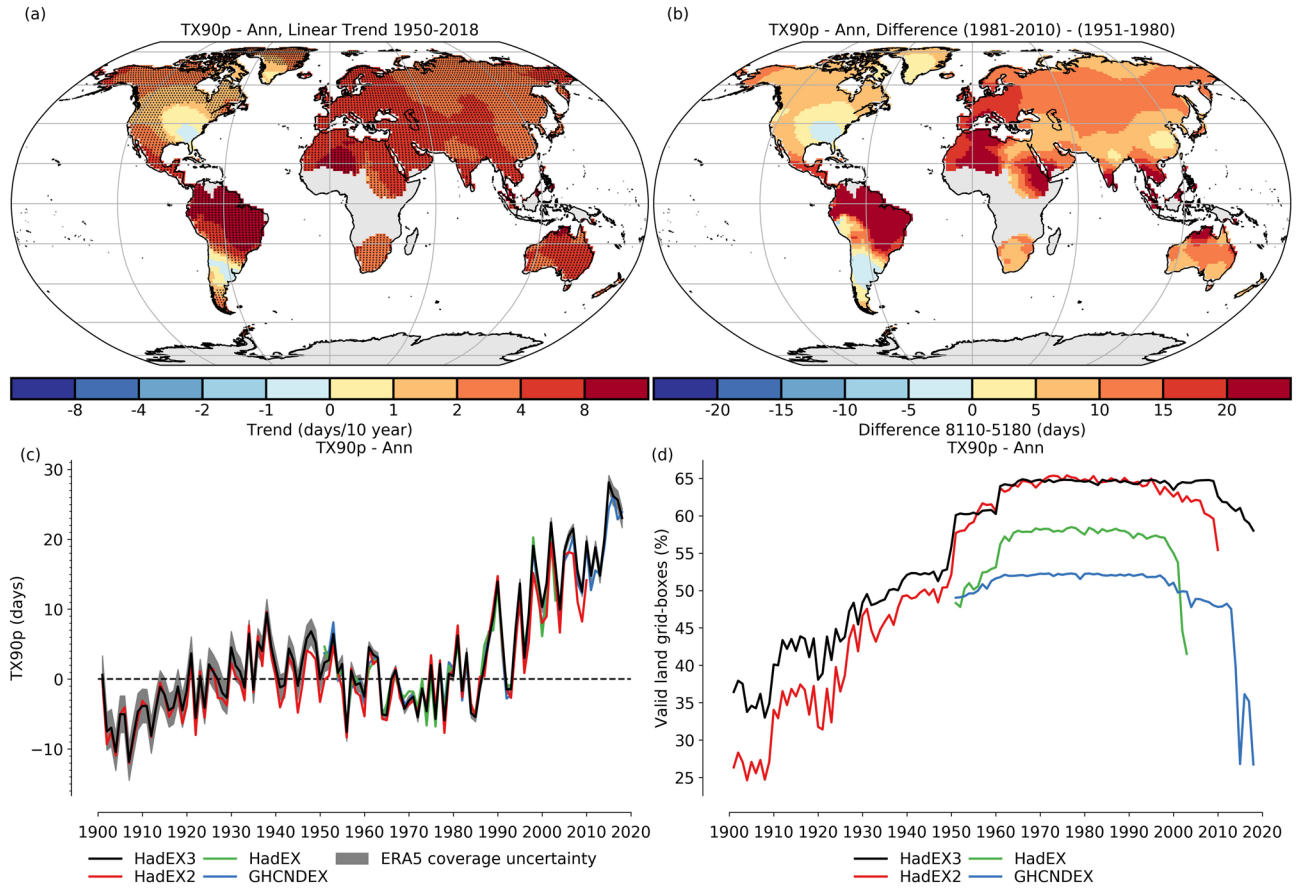

Figure 2. (a) Linear trends in annual series of TX90p over 1950–2018 (days/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Trends only calculated for grid boxes with sufficient data (at least 66% of years having data and the last year of the series being 2009 or later). Significant trends are indicated with stippling. The global time series is calculated from grid boxes with at least 90% temporal completeness (106 years over 1901–2018). Gray shading in panel (c) is the coverage uncertainty calculated using the ERA5 reanalysis. All panels use a reference period of 1961–1990, with maps presented on a 1.875° x 1.25° longitude-latitude grid. We have converted the units for this index from % into days for clarity.

least 90% temporal coverage (106 years) are used. Hence, the fraction of the global land surface represented in these time series varies between the indices (even within the different "families"). Also, as the long-term behavior of some indices varies regionally, these time series only give a flavor of the changes over time. We also show the time series of the percentage of the land grid boxes which contain data for the same four data sets. The time series for each data set shown in the following figures have been normalized using a common period of 1961–1990.

Although we do not expect the changes over time of these indices to be linear, we show maps of linear trends to provide a general picture of changes that have occurred. We use the median of pairwise slopes method (Lanzante, 1996; Sen, 1968; Thiel, 1950) to calculate the trends. To give an indication of the confidence in the trends, grid boxes where the 95% confidence range in the trends excludes zero are shown stippled. Given the sparser spatial coverage before the 1950s, we show linear trends over 1951–2018. For a trend to be calculated in a grid box, values have to be available for at least 66% of the years (45 years), and the final valid value of the grid box has to occur in 2009 or later.

Along with trend maps over 1951–2018 and global average time series for the indices and the coverage, we also show difference maps between two 30-year periods (1951–1980 and 1981–2010). At least 15 years have to be present in each 30-year period to enable a difference to be shown. As these maps do not include data from the last 8 years, they are not as affected as the trend maps by the data drop-off that has started in the mid-2000s and so can exhibit in greater spatial coverage.

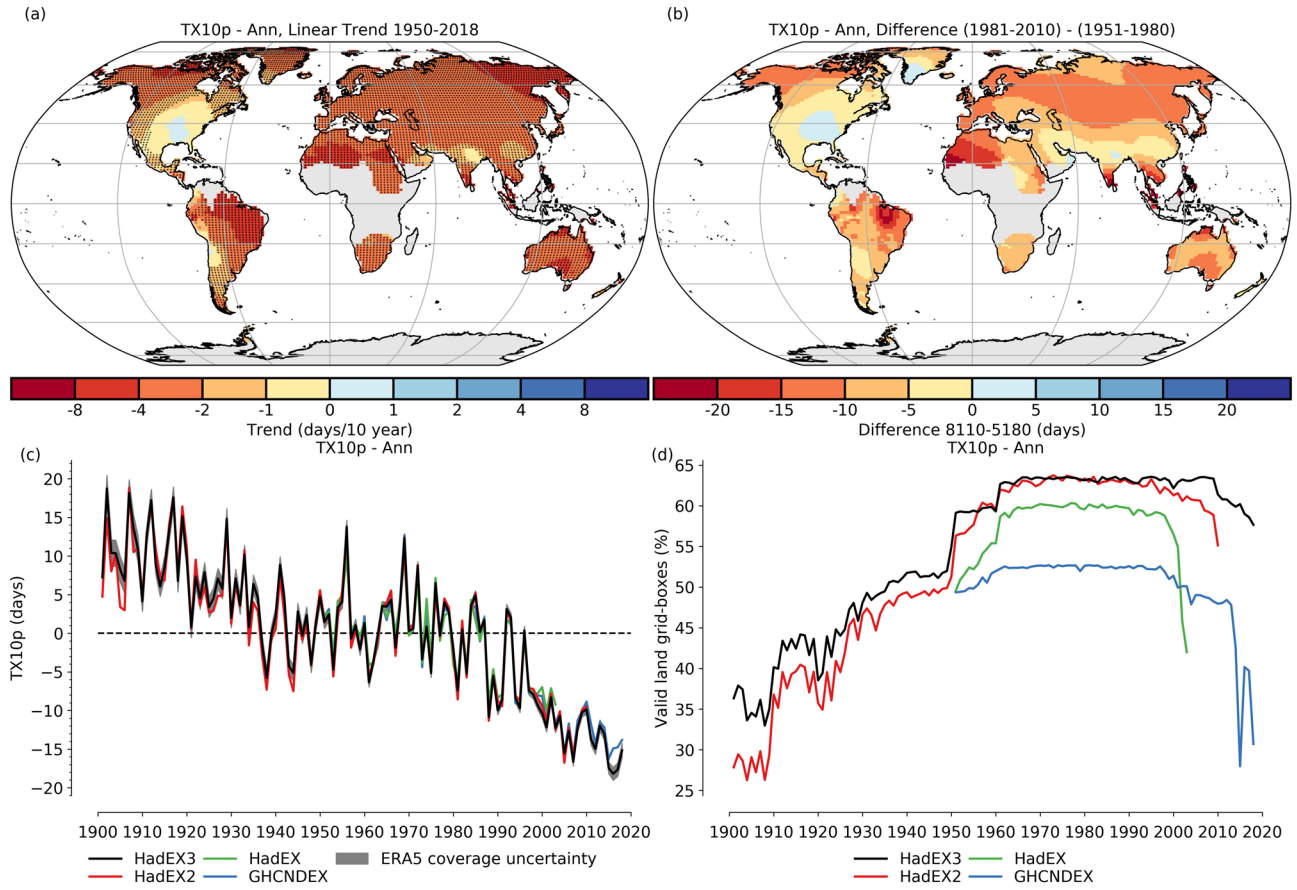

Figure 3. (a) Linear trends in annual series of TN10p over 1950–2018 (days/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Details of trend and time series calculation are described in Figure 2. We have converted the units for this index from % into days for clarity.

The time series, maps, and station location plots for each index can be found in the supporting information.

4.1. Trends in Annual Temperature Indices

As expected from the wide variety of previous studies (e.g., Alexander et al., 2006; Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013; Donat, Alexander, Yang, Durre, Vose, & Caesar, 2013) and the ongoing warming of the land surface (e.g., Stocker et al., 2014), all temperature indices show global trends and changes consistent with increasing surface temperatures. The increased spatial coverage of HadEX3 has not greatly changed either the long-term behavior or the short-term variability of these indices for the overlapping period with HadEX2 or GHCNDEX.

Starting with the percentile-based indices (TX90p, TX10p, TN90p, and TN10p, Figures 2,3, and S1–S4), these have relatively complete spatial coverage except, for example, for some parts of Africa. Widespread changes are consistent with the warming global temperatures, and only a few regions show slight cooling. As in HadEX2, changes in indices from the daily minimum temperatures (TN90p and TN10p) are on average stronger than those from daily maximum temperatures (TX90p and TX10p). In tropical regions of South America, Northern Africa, and through Asia, the numbers of warm nights have increased by over 8 days per decade, leading to a doubling since the late 1970s (the average annual frequency during the 1961–1990 base period is 36.5 days) to between 70 and 80 days per year during the recent decade. There have also been strong decreases in the numbers of cool nights over the same period, down to around only 15 days per year. Maps of the differences between two 30-year periods (1981–2010 and 1951–1980) show a little more structure for these two indices. For example, the greatest difference for TN10p are in the Amazon, parts of Africa just north of the Equator, and northern China.

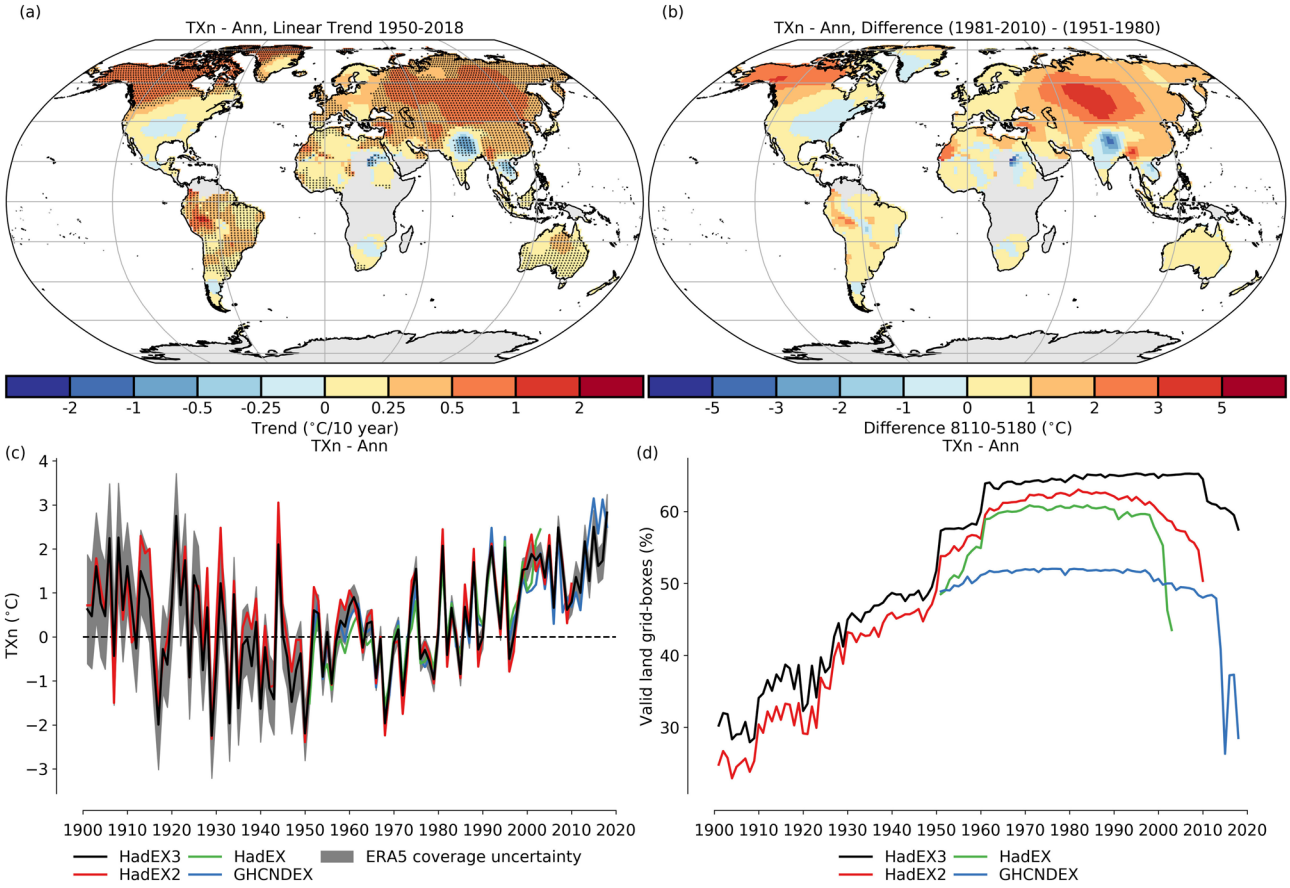

Figure 4. (a) Linear trends in annual series of TXn over 1950–2018 (°C/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Details of trends and time series calculation are described in Figure 2.

For the indices based on daily maximum temperature, the increase in the number of warm days (TX90p, around 30 days since the late 1970s and around 40 days since 1901) is larger than the decrease in the number of cool days (TX10p, around 15 days since the late 1970s and also around 40 days since 1901), and this is reflected in the maps of trends. Furthermore, in a warming climate with a fixed reference period, the number of cool days is zero-bounded, whereas the number of warm days is less affected by the upper limit (365 days; see section 2.1). The ”warming hole” over the southern-central United States is seen in both indices but more so in TX10p. This area of reduced warming, or even cooling, may be the result of internal variability (Meehl et al., 2012), an effect of the changing land use of this area as plantations have been repurposed (Portmann et al., 2009), or changes to the local hydrological cycle (Pan et al., 2004). A region with similar behavior in TX90p is seen in the southern parts of South America. Both these indices are spatially more heterogeneous than those based on the minimum temperature, but the areas which are showing the strongest trends are again the Amazon and large swathes of Africa, along with Southeast Asia.

Along with the long-term behavior of the time series for these four indices, the plots in Figures 2,3, and S1–S4 also show the year-to-year variation in comparison to other data sets. HadEX3 shows very good agreement with other observational data sets (HadEX, HadEX2, and GHCNDEX) over the common periods, for both the long- and short-term behavior. We do not match the coverage of these four data sets, so some of the differences are the result of the different spatio-temporal coverage of each one. For all four indices, the strongest period of change started in the late 1970s, with smaller amplitude changes in either direction before then. Although the coverage uncertainty (gray shading) increases toward the beginning of the record, it remains small compared to the magnitude of both the short and long time scale variations.

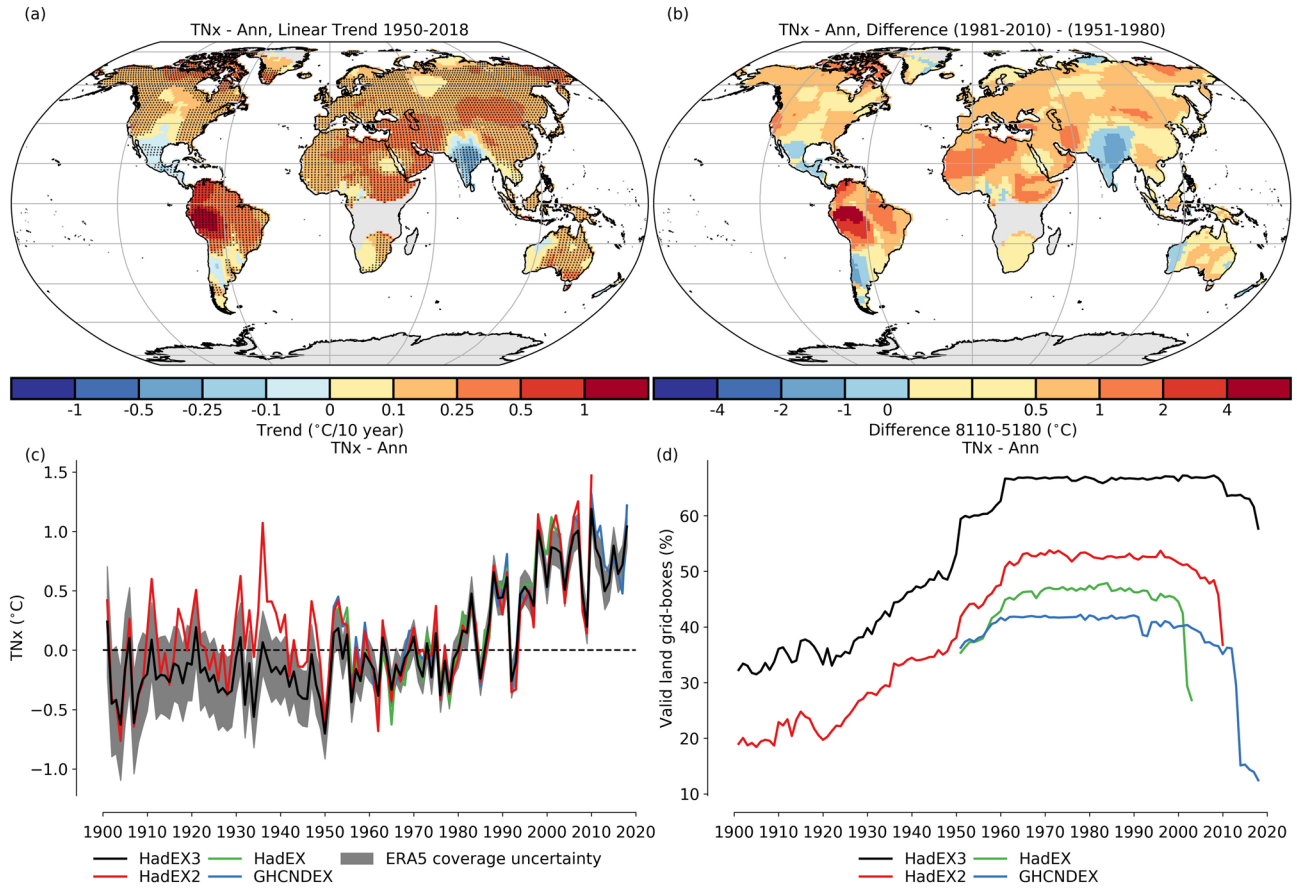

Figure 5. (a) Linear trends in annual series of TNx over 1950–2018 (°C/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Details of trend and time series calculation are described in Figure 2.

The time series of the fraction of land grid-boxes with data show how the best coverage is usually between around 1960 to the mid-2000s, with a decline thereafter. This does unfortunately limit what can be inferred on a global level for the early and most recent periods in this data set. However, by including the coverage uncertainty from ERA5 (gray shading), we can place some level of confidence on global changes in these four indices over the last four to five decades.

The indices based on the absolute warmest and coldest temperatures (TXx, TNx, TXn, and TNn, Figures 4,5, and S5–S8) are spatially more heterogeneous than the percentile indices discussed above. Again, the strongest changes over the last 50 years or so have been in the minimum temperatures, with almost a 4°C change globally since the middle of the last century for the temperature of the coldest night (TNn, Figure S8). The largest increases are seen in the northern high latitudes with changes of up to 1°C per decade. The coldest day (TXn) shows similar changes but of smaller magnitude. Changes in the warmest day and night (TXx and TNx) are smaller, being globally around 1°C warmer than the 1960s to 1970s. They also have a more heterogeneous distribution than the other two indices. The effect of the coverage uncertainty in the early years of the 20th century on these global series is larger, a result of both the more heterogeneous nature of these indices as well as the lower spatial coverage at this time compared to the previous four indices.

What is clear from all the temperature indices is that all have continued to change consistent with recent past behavior (seen in HadEX2) and the ongoing warming of global surface temperatures. All these maps and time series are very similar to those presented for HadEX2 (Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013). Larger changes in minimum temperatures than in maximum temperatures have been noted by other studies (e.g., Donat & Alexander, 2012; McKinnon et al., 2016). As the maximum and

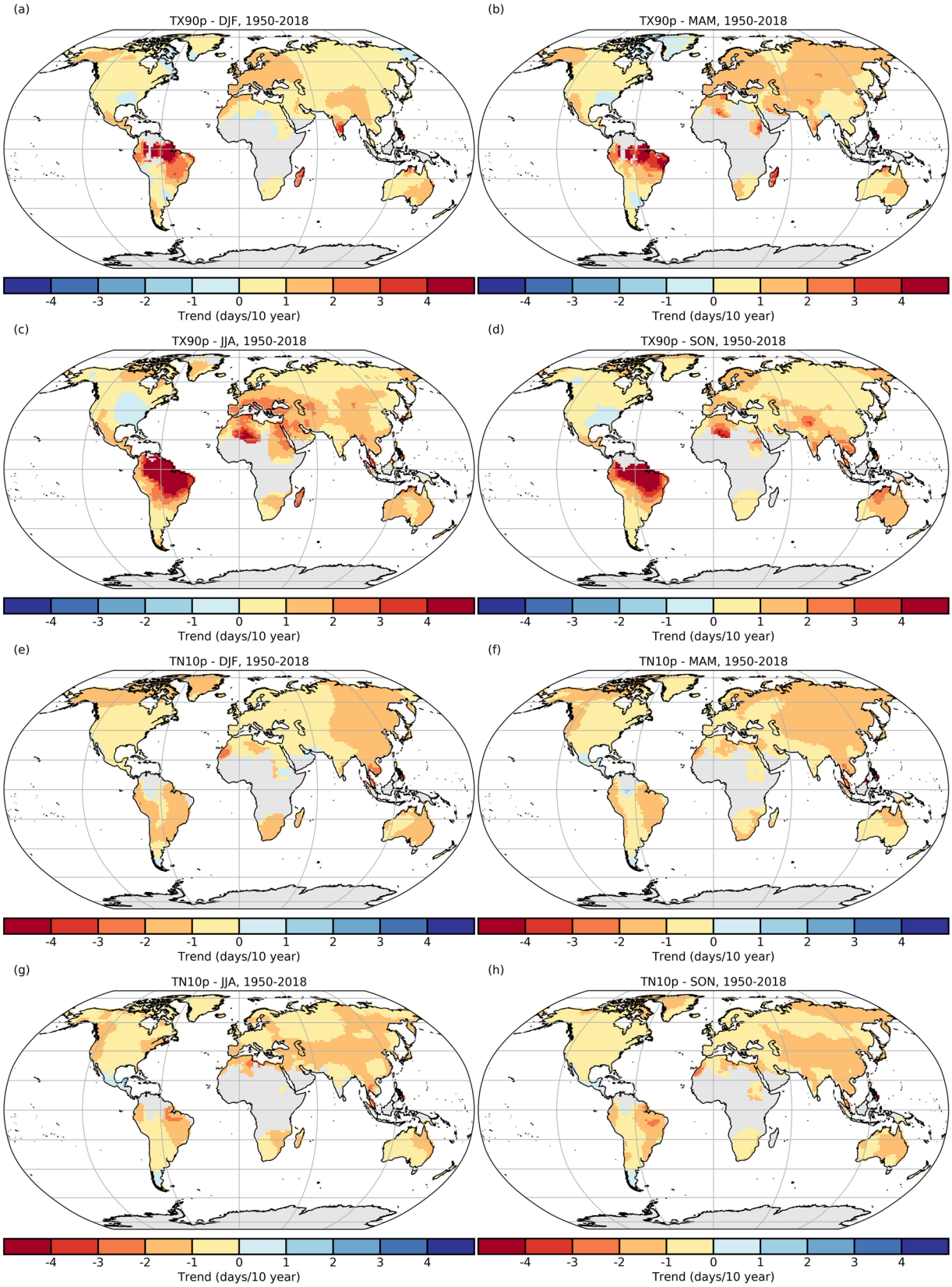

Figure 6. Linear trends in seasonal series of TX90p (a–d) and TN10p (e–h) (days per decade) over 1950–2018, using a reference period of 1961–1990. We have converted the units for this index from % into days for clarity.

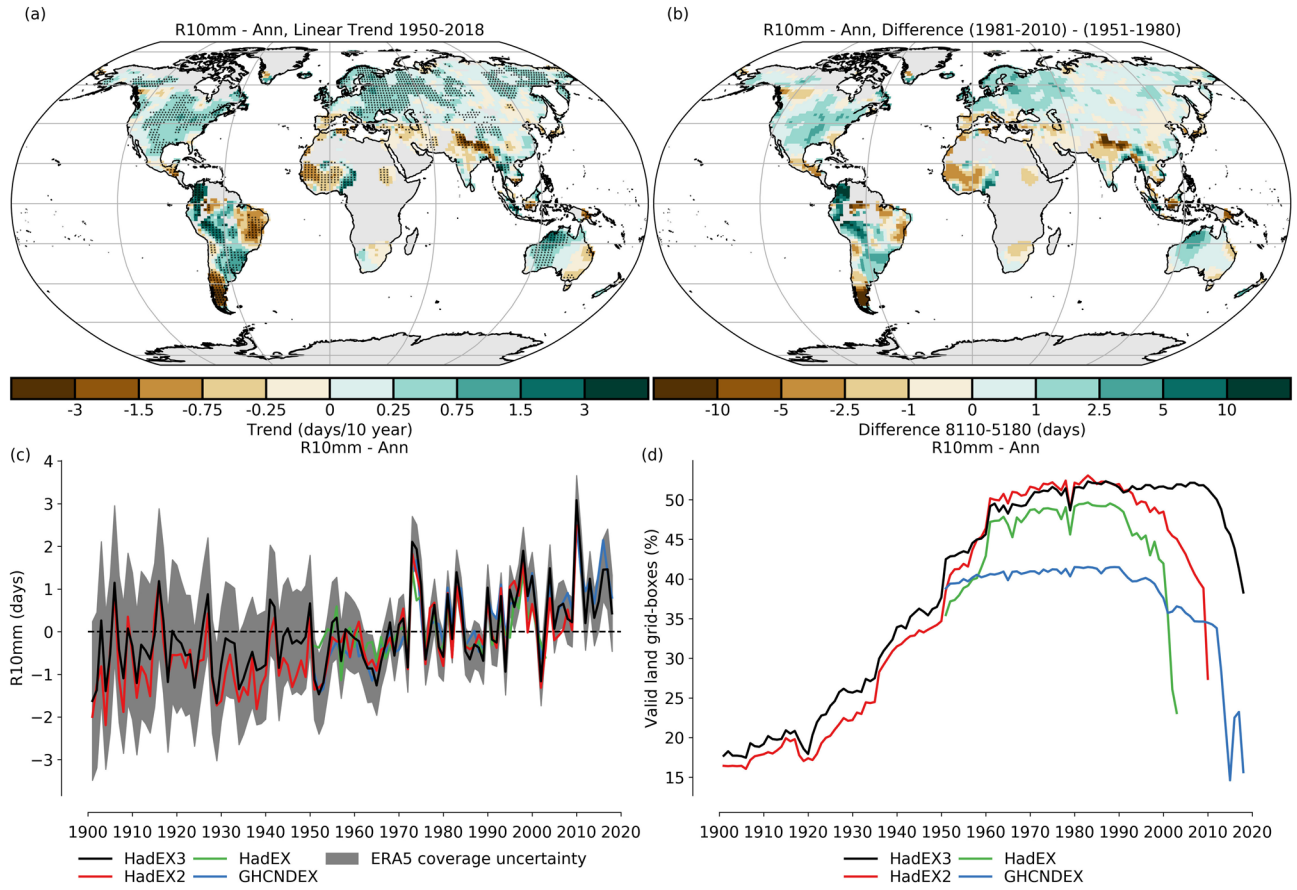

Figure 7. (a) Linear trends in annual series of R10mm over 1950–2018 (days/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Details of trend and time series calculation are described in Figure 2.

minimum temperatures captured in the indices discussed so far are likely to have come from different events, it is not possible to deduce any changes in the diurnal temperature range (DTR) from these indices alone. The DTR (Figure S9) index itself shows there has been a reduction since the middle of the last century, with most of the change occurring before 1980, also supported by other studies (Thorne, Menne, et al., 2016; Thorne, Donat, et al., 2016). The extreme temperature range (ETR, Figure S17), however, does have a strong link to the temperature indices and does show a clear decrease over the last 80 years.

4.1.1. Seasonal Temperature Indices

Several of the temperature indices are calculated on a monthly as well as an annual basis. Here we show maps of seasonal changes, where at least two of the three months need to be present for a seasonal value to be calculated (seasonal sum, maximum or minimum). For TX90p (Figure 6), the strongest changes are in the tropics of South America during JJA. In fact, this region has the largest area of strong increases in TX90p in all four seasons, as a consequence of the low seasonal and interannual temperature variability in tropical regions possibly combined with land-use change effects (Cohn et al., 2019). The strongest increases in northern Asia are seen in MAM, with southern and southeastern Asia showing their largest increases during the other three seasons of the year. In Europe, MAM and JJA show the strongest increases, with the focus of the JJA increases being around the Mediterranean and Black Seas, but for MAM it is western/Atlantic Europe that sees the largest increases. Over North America, increases are lower than in most other places with data, and the warming hole is prominent in JJA and to a smaller extent during MAM.

In contrast, TN10p (Figure 6) shows much less variation across the regions of the globe where trends have been calculated. South America has some of the strongest changes in all seasons, of a decrease of up to 3 days

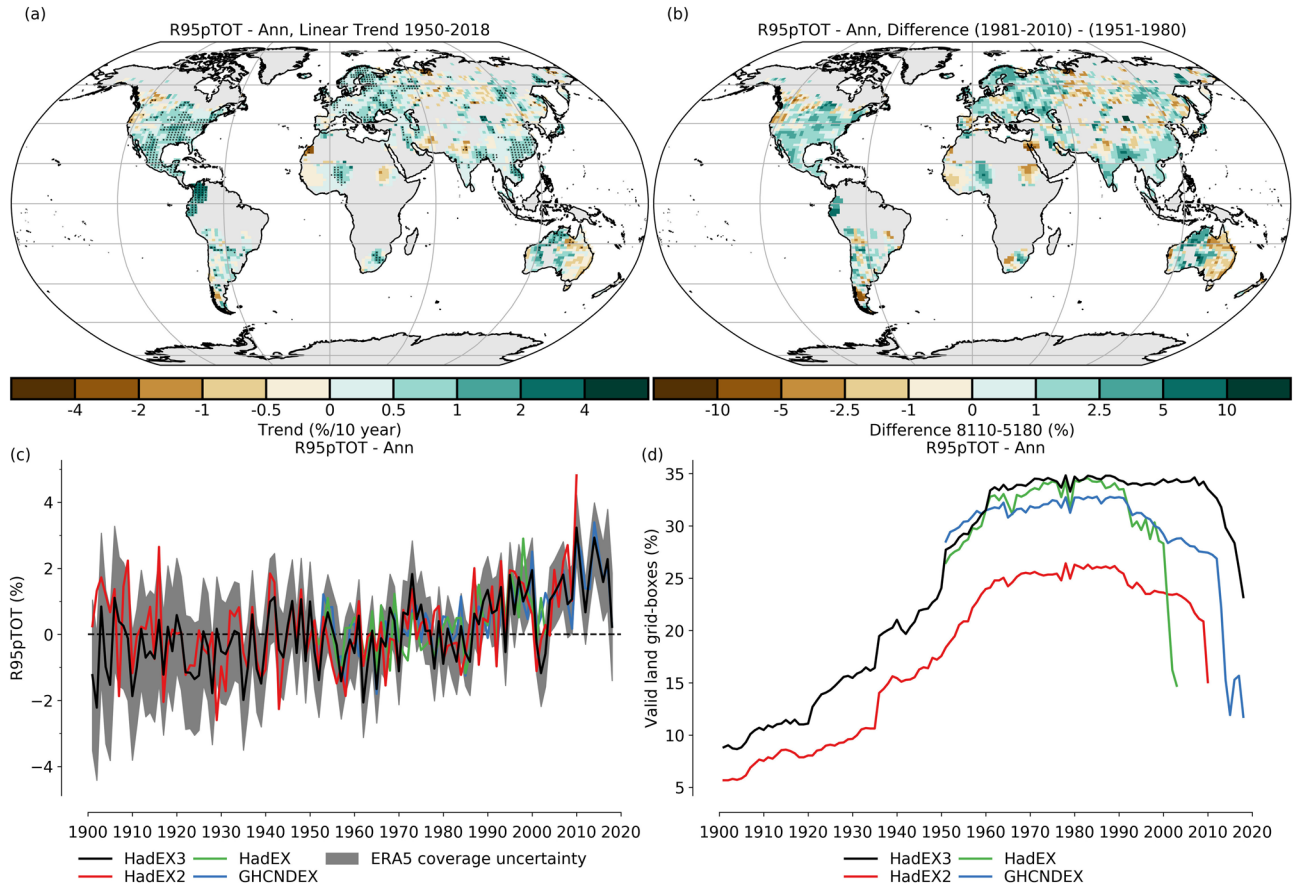

Figure 8. (a) Linear trends in annual series of R95pTOT over 1950–2018 (%/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Details of trend and time series calculation are described in Figure 2.

per decade. In other parts of the world, the decreases are lower, of 1 or 2 days per decade, and with no clear seasonal variation. The remaining indices are shown in Figures S30–S43.

4.2. Trends in Precipitation Indices

As in HadEX2, the lower correlation length scales of precipitation accumulations and events results in fewer stations falling within the search radius for the ADW gridding and reduced spatial coverage compared to temperature indices, despite having more stations contributing. Furthermore, when showing the maps of trends in Figures 7–9, we require 66% completeness, and so intermittent grid-boxes are excluded from this overview analysis. Generally, these indices show spatially more heterogeneous patterns than the temperature indices, also a result of a smaller DLS. The relative magnitude of the coverage uncertainty on the global time series is also consequently larger, in many cases being comparable to both the long- and short-term variability.

Heavy precipitation days (R10mm, Figure 7) show very little change globally over the entire period of the data set, though there are indications in the last decades of a gentle increase. However, regionally, there are contiguous areas of positive and negative change. The strongest of these are in South America, where over the Andes and the eastern Amazon there are reductions, but increases in a swath from northern Argentina up to the Caribbean coast. A uniform and moderate increase in the number of days heavy precipitation is seen over North America, along with the Eurasian high latitudes. Larger increases are seen in Southeast Asia and through into central Australia. However, there are decreases around the Mediterranean and through into the Middle East.

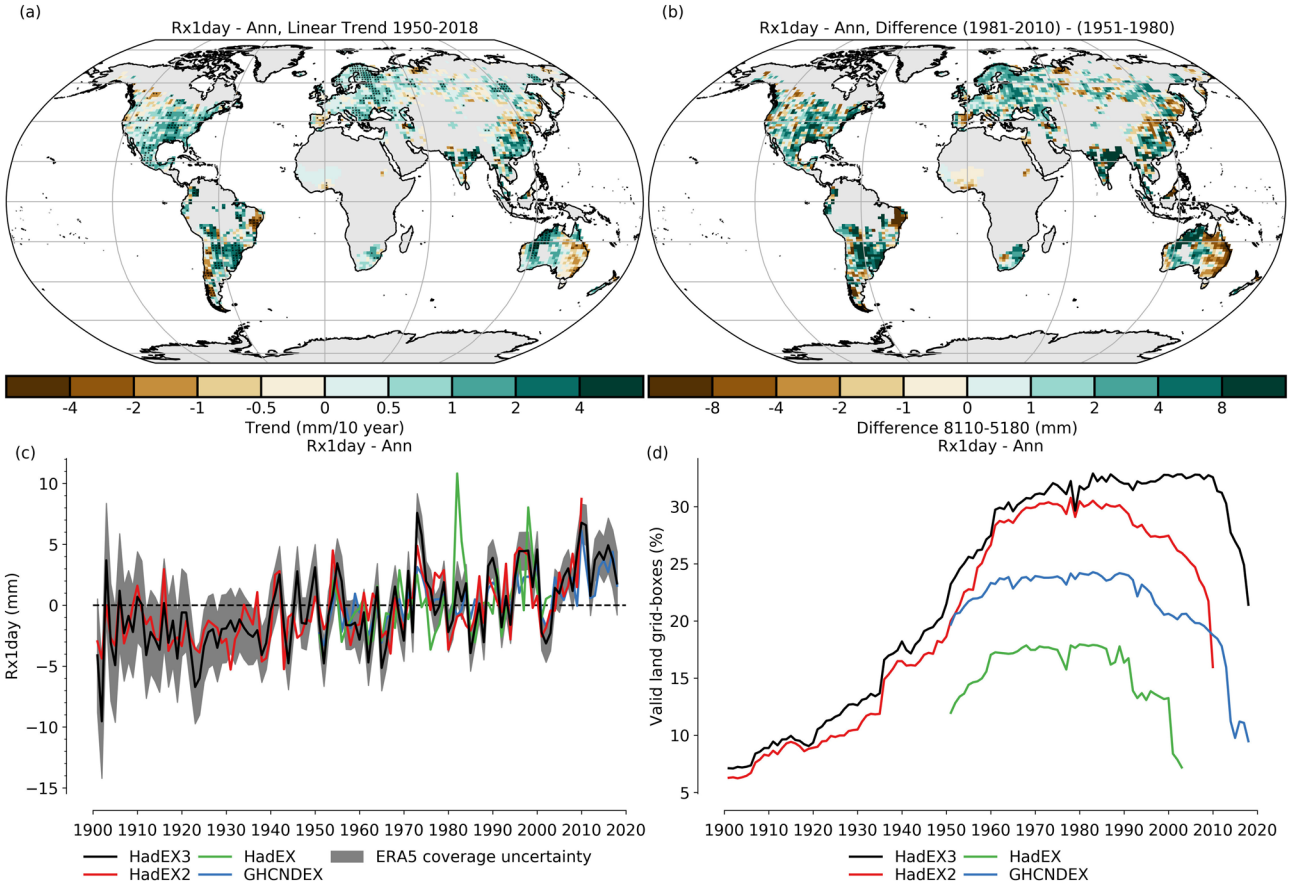

Figure 9. (a) Linear trends in annual series of Rx1day over 1950–2018 (mm/decade); (b) difference between two 30-year periods (1981–2010 and 1951–1980); (c) time series of HadEX3 compared with HadEX, HadEX2, and GHCNDEX; (d) time series of land fraction containing grid boxes valid data. Details of trend and time series calculation are described in Figure 2.

The contribution from very wet days (days exceeding the 95th percentile of daily precipitation, R95pTOT, Figure 8) is increasing globally, with now an extra 1–2% of precipitation falling during very wet days. Where we have data, the signal is for a relatively uniform increase, with only a few grid boxes showing decreases. The global time series shows an obvious increase from the 1970s. This agrees with the theoretically expected behavior in a warming climate (e.g., Allen & Ingram, 2002; Groisman et al., 2005; Trenberth et al., 2003) and with the findings of Donat, Alexander, Yang, Durre, Vose, Dunn, et al. (2013).

As well as cumulative effects, short-term downpours can have strong impacts, and maximum 1-day precipitation, Rx1day (Figure 9), shows strong increases of around 2 mm per decade in the eastern half of North America, as well as the eastern parts of southern South America, parts of India, and China. Smaller increases are seen over Europe. This is reflected in the global time series with around 2–3 mm more falling in recent years than in 1961–1990. This supports the more detailed study of changes in Rx1day over 1900–2009 by Westra et al. (2013), who found significant increases on a global scale, and has been shown to be the result of anthropogenic changes (Min et al., 2011).

4.3. Comparison of Reference Periods

As noted in section 2.2, two versions of HadEX3 have been created with different references periods enabling the use of a wider range of submitted data and also to align with the current WMO climatology period (WMO, 2007). In the supporting information, we repeat the panels of time series and maps for these indices (TX90p, TX10p, TN90p, TN10p, CSDI, WSDI, R99p, R95p, R95pTOT, and R99pTOT) for this second reference period (Figures S51–S60). Comparing between the two reference periods shows the difference in the station networks available, with 1961–1990 having greater coverage over Australia, India, and northern South America and 1981–2010 over Tropical South America, Africa, and Southeast Asia.

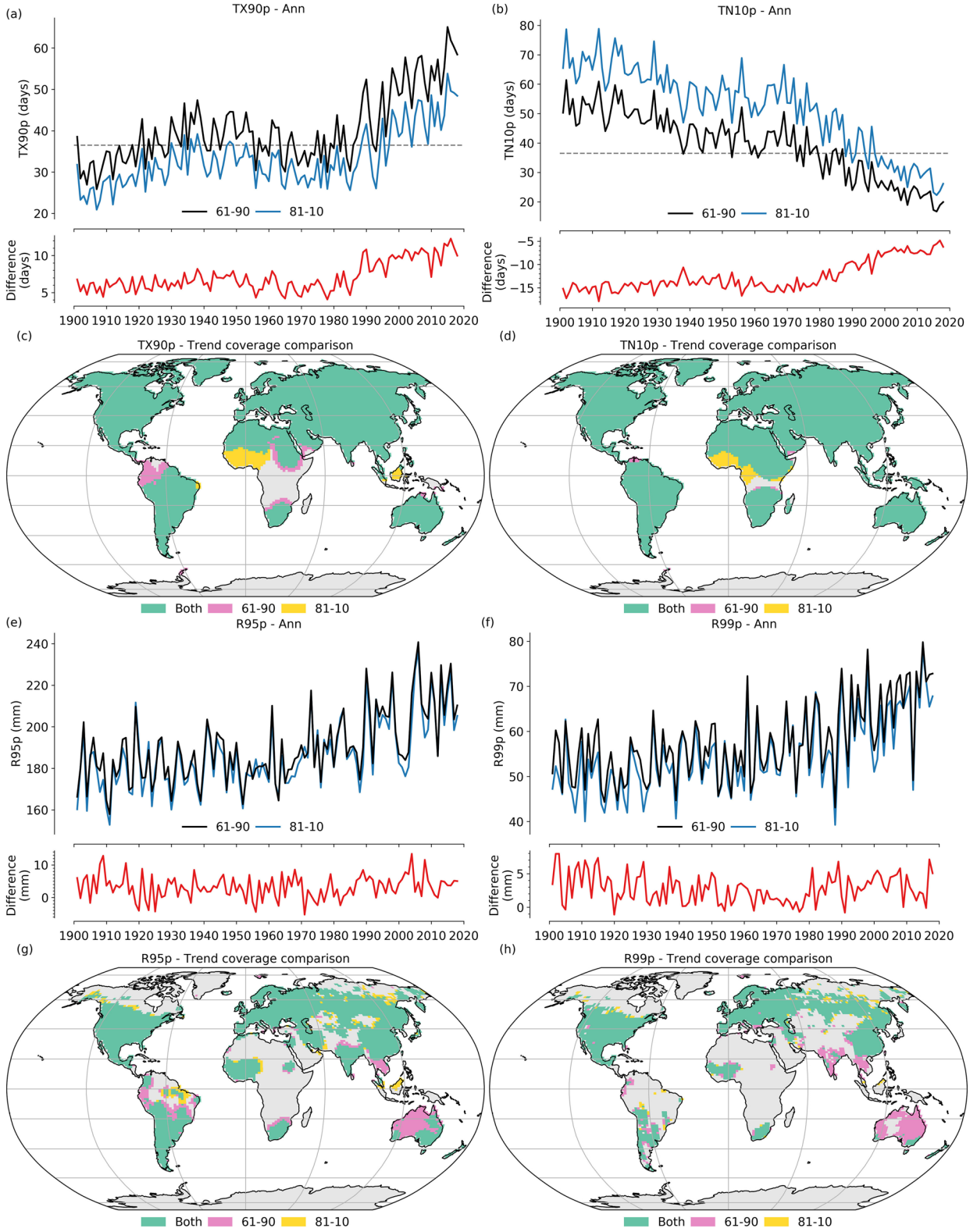

Figure 10. Upper time series panels: globally averaged time series for (a) TX90p, (b) TN10p, (e) R95p, and (f) R99p referenced to black (1961–1990) and blue (1981–2010) with the difference between the two reference periods (1961–1990 minus 1981–2010) in red. The time series are calculated from only those grid boxes which have data in both versions (roughly corresponding to the green shading in the maps) (c, d, g, h). We show the expected value for TX90p and TN10p over their respective reference period by the horizontal dashed line. Yellow (1981–2010) and pink (1961–1990) regions on the maps show where only one has coverage.

In Figure 10, we show comparisons between the two reference periods for four of these 10 indices using both the "globally" averaged time series and maps of parts of the land surface where trends could be calculated for each period independently and for both. Note the requirement that for a grid box trend to be calculated, 66% of the years need to be present and the last year must be 2009 or later (section 4). We have not normalized these time series to show the effect of a change in reference period more clearly, but we have used a common mask for both versions so that changes in the coverage do not affect the time series. A time series plot of the difference is shown for each index.

For the four temperature percentile indices (TX90p and TN10p, Figure 10; for TX10p and TN90p, see Figure S62), the maps show that the largest changes in coverage are in equatorial Africa and northern South America.

The thresholds used, for example, for TX90p, are determined from the percentiles of the distribution calculated over the specified reference period. These thresholds will be higher in the later 1981–2010 period than over 1961–1990, given the warming seen globally between these time periods. However, the daily maximum temperatures being assessed by the use of these indices are the same, so, as the thresholds are on average warmer, fewer days exceed them, resulting in fewer warm days being measured, and hence lower values for TX90p when using 1981–2010.

As the climate is not changing monotonically, the differences between the indices depend on the value of the indices themselves. If within a single year, the maximum temperatures form a perfect Gaussian distribution, then with two thresholds (one from 1961–1990 and one from 1981–2010), the number of exceedances of the higher threshold is smaller than that of the lower one. If the distribution now moves to warmer temperatures, although the exceedances over both thresholds increase, there are relatively more days which exceed the lower threshold. Hence, as the climate warms, TX90p for 1961–1990 is rising faster than the same index measured relative to 1981–2010 (seen in Figure 10). For the cool end of the distribution, the TN10p index is decreasing more slowly for 1961–1990 relative to 1981–2010 (shown by the difference curve becoming less negative).

The precipitation indices (Figures 10 and S64) show very little difference between the two versions of this data set, both those showing long-term changes and those with little change over the record. However, in all cases, the year-on-year, short time scale variability is well reproduced. For the precipitation indices (R95p, R99p, R95pTOT, and R99pTOT), the interpretation is more complex. Starting with R95p and R99p, these sum the precipitation in the wettest 5% and 1% of days, respectively, determined from daily rather than on total accumulation. Hence, there is no expectation that R95p or R99p represent 5% or 1% of the total accumulation, and hence R95pTOT and R99pTOT do not have simple expected values.

For these indices, despite there being indications of a long-term trend, the difference between the two reference periods is stable over the period of record. There are year-to-year differences, which are likely to arise from variations in the underlying stations (rather than the gridded coverage). The precipitation from wet days (R95p and R99p) has increased over time, so between the two reference periods, the threshold set by the wettest N-percent of days has increased. Hence, fewer days exceed the higher, later threshold. But the change is small, and so the effect of the change in shape of the distribution of daily accumulations does not stand out in the difference time series. As R95pTOT and R99pTOT are just the R95p and R99p indices normalized by the total precipitation (PRCPTOT), then these also only show a constant offset.

5. Discussion

As its predecessors, HadEX3 supports the results from previous global assessments of land areas (Alexander et al., 2006; Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013) of the change in the ETCCDI indices over the 20th and 21st centuries. The changes in the temperature extremes are consistent with the warming world, and changes in indices based on the daily minimum temperatures are occurring faster than those based on daily maximum temperatures. Hence, there is not only a shift in the temperature distributions (a change in the mean), but also a change in higher order moments (e.g., variance and skew; see, e.g., Dunn et al., 2019). There are also indications on a global scale of changes to extreme precipitation over land areas. As has been shown in previous observational studies (Donat, Alexander, Yang, Durre, Vose, Dunn, et al., 2013; Westra et al., 2013), increases in the annual daily maximum intensity (Rx1day) are increasing,

with these increases and those in the fraction falling in heavy events (R95pTOT) being linked to anthropogenic climate change (Min et al., 2011). Many of the large-scale temperature changes are easily linked to the ongoing changes in anthropogenic greenhouse gases (e.g., Stocker et al., 2014), but other mechanisms will also have played a role, especially at regional scales, for example, land-use change (e.g., Pan et al., 2004; Portmann et al., 2009) and changes in large-scale circulation patterns (e.g., Barrucand et al., 2008; Kenyon & Hegerl, 2008; Ning & Bradley, 2015; Otero et al., 2018; Scaife et al., 2008; Sillmann & Croci-Maspoli, 2009).

We have already noted how the definition and construction of these indices can affect how the gridded product behaves with a changing station network, and especially so when comparing products derived from the two reference periods outlined above. However, despite these extra considerations, the agreement between the "global" HadEX3 and other products is very good, and the uncertainties arising from the incomplete coverage frequently encompass all of the products. This increases our confidence in the changes seen over time.

The gridded HadEX3 product presented and analyzed here is one of a wide number of possible versions of a data set of gridded climate extremes indices. A specific set of values, methods, and algorithms has been chosen for good reasons, but other combinations could be equally valid. These parametric and structural uncertainties, as assessed for HadEX2 in Dunn et al. (2014), are yet to be fully explored for HadEX3. We expect that changes to the methods used in the creation of HadEX3 (parametric) would have a smaller effect than the use of different methods entirely (structural). Although the underlying station data would remain the same, the spatial representation resulting from the gridding method used herein (ADW) compared to others or by inverting the order of operation is very likely to be different, especially at local and regional scales. Investigations into the effects of different gridding and interpolation methods show that the method used for HadEX3 has a high level of smoothing, resulting in loss of detail, especially in topographically varying areas, and the order of operation changes the spatial patterns in the final grids (Avila et al., 2015; Contractor et al., 2015; Yin et al., 2015). For indices which show strong, widespread, and coherent changes over time, it is unlikely that use of different methods will change the conclusions drawn (Dunn et al., 2014). We have presented two possible versions of HadEX3, one for each of the two reference periods, but other options have been explored as part of this process (e.g., grid box size and gridding method). To ensure that this manuscript is focussed on the update to and changes since the release of HadEX2, we have not presented these alternative versions here but leave them for a future study.

5.1. Uses of HadEX3

As noted in earlier sections, the gridded HadEX3 product contains spatially smoothed ETCCDI indices estimated at the grid box centers. The underlying station network also changes over the period of record, with many fewer stations in the early period, but also a drop off in the most recent years. Hence, "global" or regionally averaged time series should be accompanied by plots showing how the coverage has changed over that time. Furthermore, even in regions which have a dense network and comparatively consistent coverage, changes in the underlying network can result in inhomogeneities in the grid box values, and some indices are more susceptible to this than others (see example of a mountain top station for TXx in section 2.1). We have endeavored to mitigate this effect with the station selection but will not have been able to remove it entirely.

HadEX3 has been designed to estimate long-term changes in observed extremes. The order of operations in going from daily station observations to the gridded product (calculate indices then grid) results in this data set comprising estimates at grid box centers. The output from climate models or reanalyses are most typically representative of the average over a grid box. Although this may not have a noticeable affect for some temperature indices, for maxima and minima (e.g., TXx) and especially for precipitation indices, these values may differ (e.g., Chen & Knutson, 2008). Therefore, comparisons with products where gridded values are the average value of the indices for that grid box (rather than being representative of the center and some region surrounding it) should be undertaken with care. We note that creating a version of HadEX3 using the alternative order of operations (grid observations then calculating indices) is likely to result in oversmoothing the extremes (Haylock et al., 2008; Hofstra et al., 2010), but this would allow for more direct comparisons to climate models and reanalyses.

6. Summary

We present the creation of the next version of a data set of extreme temperature and precipitation indices, HadEX3. Station data have been gathered from a number of open global and regional sources as well as from individuals from a large number of nations and territories. The ETCCDI extremes indices have been calculated from daily observations resulting in almost 37,000 stations of varying record length and completeness. For a gridded product, the stations with the longest records ending in 2009 or later have been selected and combined using an ADWinterpolation algorithm. The final gridded data set is available on a 1.875°×1.25° longitude-latitude grid from 1901–2018 inclusive, and some indices are available on a monthly as well as annual resolution.

Analysis of the long-term changes of both temperature and precipitation indices agree very well with previous (HadEX2) and current monitoring (GHCNDEX) products. Changes in the temperature indices are consistent with the ongoing warming of the globe, with greater changes in the indices derived from minimum temperature than those from the maximum temperatures. Indices summarizing the intensity of heavy precipitation also show a slight increase over recent decades. Spatially, the story is more complex, but temperature indices are overall more homogeneous than precipitation indices, and those using percentiles/anomalies more so than those based on actuals values.

Spatially, precipitation is more heterogeneous than temperature and often has lower coverage. However, on a global level, a number of the indices show increases (e.g., R10mm, Rx1day, and R95pTOT), indicating that there are more heavy precipitation events that are also more intense and contribute a greater fraction to the total.

As some indices use a reference period to derive threshold values, we have created versions using the original (HadEX2 and HadEX) period of 1961–1990 (which corresponds to the WMO period for long-term climate change assessments) and one using the current WMO reference period of 1981–2010, with fractionally different data submitted for each. Although the global results are qualitatively very similar, there are some intricate differences for the temperature percentile indices which will require careful thought by users.

Despite our best efforts, there are still parts of the world where data are sparse or the temporal coverage is inadequate for a data set designed for long-term monitoring. This is partly a result of the approach to use regional workshops for data collection, which, although allowing access to the underlying data which cannot be shared openly, can result in station index time series for these regions ending much earlier than the data set and so limits their usefulness. Efforts are underway to augment current global collections of data to improve the data available for all users.

Appendix A

The ADW calculations are as follows. The distance term is equal to

$$w_i = (e^{-x/DLS})^m, \tag{A1}$$

where w_i is the weight for station i, and x is the distance between station i and the center of the grid box. The constant, m, is set to 4 as in the other data sets mentioned above. In the study by Dunn et al. (2014), changing this decay parameter had only a minor impact on the final gridded fields, which had also been mentioned in earlier studies, so we keep it consistent with the previous data sets.

The angular term scales this distance term, using the bearing of the station i relative to north from the central point (θ_i), compared to the bearing of all other stations, θ_k, which have been selected to contribute to this grid box.

$$W_i = w_i \left\{ 1 + \frac{\sum_k w_k [1 - \cos(\theta_k - \theta_i)]}{\sum_k w_k} \right\}, \; i \neq k. \tag{A2}$$

This method ensures that stations closest to the center of the grid-box and those which are relatively isolated have greater weight than stations further from the center or in a cluster. As the DLS can extend up to 2,000

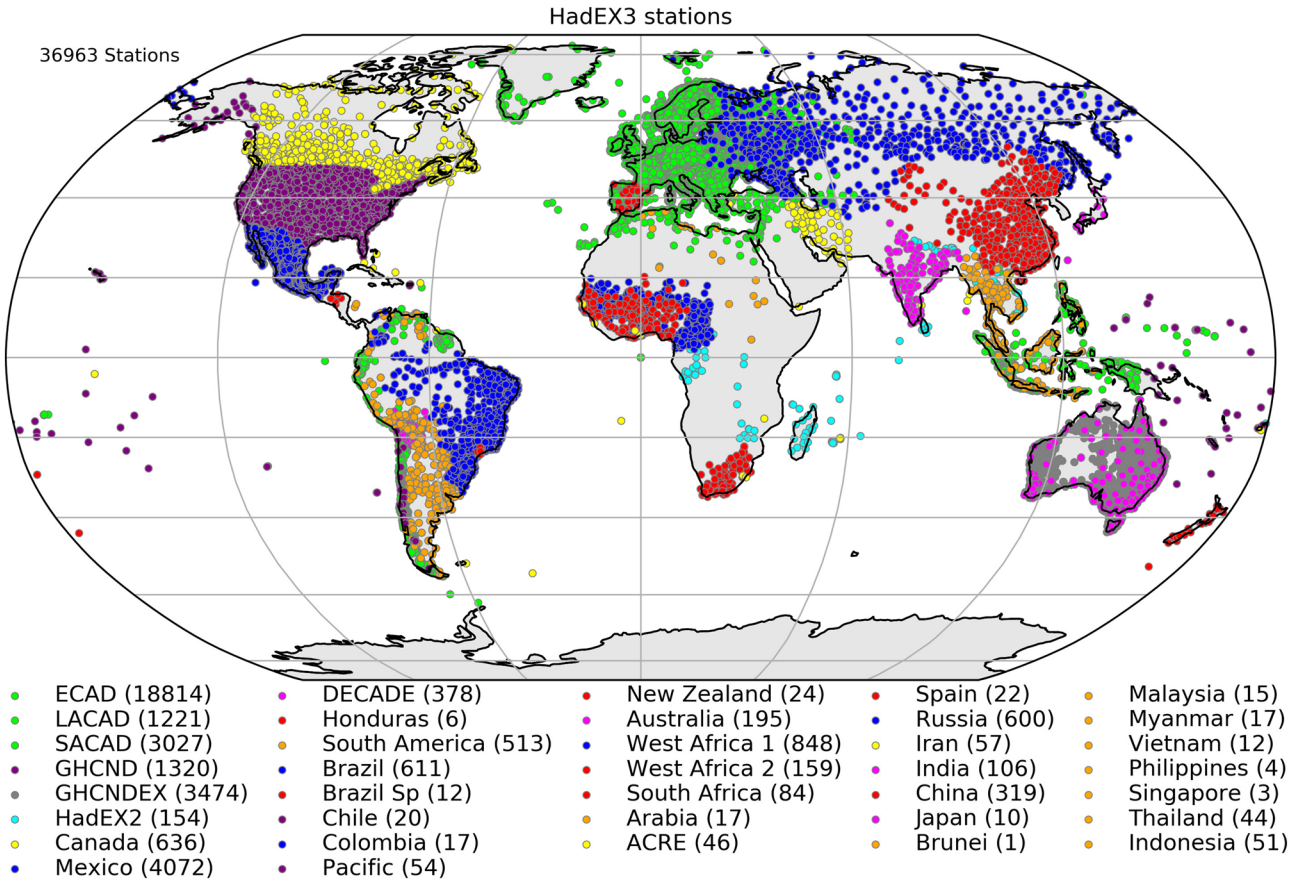

Figure A1. All the stations submitted to HadEX3. GHCND and GHCNEX are subject to the restrictions as outlined in the text. ECAD/SACAD/LACAD are the complete data sets. No QC flags have been applied to these stations.

km (especially for some temperature indices), this means that many stations fall within the selection radius. The result is that this method interpolates into regions which contain no stations. Conversely, for indices which have a small DLS (some concerning precipitation), then the 200-km lower limit can be smaller than the extent of the grid box (1.875°×1.25° longitude-latitude) and the lowest latitudes.

We show in Figure A1 all stations which were submitted to HadEX3 (though not all were selected for use in the final gridded fields), and those where we are allowed to do so will be made available to users.

Data Availability Statement

The gridded data set is available at www.metoffice.gov.uk/hadobs/hadex3/ and at www.climdex.org. In addition, a version will be lodged on the CEDA archive (https://archive.ceda.ac.uk/).

The underlying station indices will be made available on www.climdex.org where we are allowed to do so. For some collections, we are not allowed to make the underlying station data public under terms of their license, and we encourage readers to contact the appropriate co-author.

All software has been written in Python 3 (Python Software Foundation, Python Software Foundation), with the exception of Climpact2 which is in R (Ihaka & Gentleman, 1996; R Core Team, 2013). The dependencies and interplay between them have been controlled using a Rose (Shin et al., 2019)/Cylc suite (Oliver et al., 2018). The Python scripts are available at a Github repository available through https://doi.org/10.5281/zenodo.3886492, but we are not able to offer any technical support for users. The ERA5 data (Hersbach et al., 2019) are available at the Copernicus Climate Change Service (C3S) Climate Data Store (CDS) at https://cds.climate.copernicus.eu.

Acknowledgments

Robert Dunn was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra (GA01101) and thanks Nick Rayner and Lizzie Good for helpful comments on the manuscript. Lisa Alexander is supported by the Australian Research Council (ARC) Grants DP160103439 and CE170100023. Markus Donat acknowledges funding by the Spanish Ministry for the Economy, Industry and Competitiveness Ramón y Cajal 2017 Grant Reference RYC-2017-22964. Mohd Noor'Arifin Bin Hj Yussof and Muhammad Khairul Izzat Bin Ibrahim thank the Brunei Darussalam Meteorological Department (BDMD). Ying Sun was supported by China funding agencies 2018YFA0605604 and 2018YFC1507702. Fatemeh Rahimzadeh and Mahbobeh Khoshkam thank I.R. of Iranian Meteorological Organization (IRIMO) and the Atmospheric Science and Meteorological Organization Research Center (ASMERC) for Data and also sharing their experiences, especially Abbas Rangbar. Jose Marengo was supported by the National Institute of Science and Technology for Climate Change Phase 2 under CNPq Grant 465501/2014-1, FAPESP Grants 2014/50848-9 and 2015/03804-9, and the National Coordination for High Level Education and Training (CAPES) Grant 88887.136402-00INCT. The team that worked on the data in West Africa received funding from the UK's National Environment Research Council (NERC)/Department for International Development DFID) Future Climate For Africa programme, under the AMMA-2050 project (Grants NE/M020428/1 and NE/M019969/1). Data from Southeast Asia (excl. Indonesia) was supported by work on using ClimPACT2 during the Second Workshop on ASEAN Regional Climate Data, Analysis and Projections (ARCDAP-2), 25–29 March 2019, Singapore, jointly funded by Meteorological Service Singapore and WMO through the Canada-Climate Risk and Early Warning Systems (CREWS) initiative. This research was supported by Thai Meteorological Department (TMD) and Thailand Science Research and Innovation (TSRI) under Grant RDG6030003. Daily data for Mexico were provided by the Servicio Meteorológico Nacional (SMN) of Comisión Nacional del Agua (CONAGUA). We acknowledge the data providers in the ECA&D project (https://www.ecad.eu), the SACA&D project (https://saca-bmkg.knmi.nl), and the LACA&D project (https://ciifen.knmi.nl). We thank the three anonymous reviewers for their detailed

Comprehensive reports from the Second Workshop on ASEAN Regional Climate Data, Analysis and Projections (ARCDAP-2) held in Singapore 25-29 March 2019, are available at CCRS: https://ccrs.weather.gov.sg/arcdap-2_workshop/and WMO: https://public.wmo.int/en/events/workshops/second-workshop-asean-regional-climate-data- analysis-and-projections-arcdap%E2%80%932

References

Aguilar, E., Aziz Barry, A., Brunet, M., Ekang, L., Fernandes, A., Massoukina, M., et al. (2009). Changes in temperature and precipitation extremes in western Central Africa, Guinea Conakry, and Zimbabwe, 1955–2006. *Journal of Geophysical Research*, *114*, D02115. https://doi.org/10.1029/2008JD011010

Alexander, L. V., Fowler, H. J., Bador, M., Behrangi, A., Donat, M. G., Dunn, R., et al. (2019). On the use of indices to study extreme precipitation on sub-daily and daily timescales. *Environmental Research Letters*, *14*(12), 125008. https://doi.org/10.1088/1748-9326/ab51b6

Alexander, L. V., & Herold, N. (2015). Climpactv2 indices and software. A document prepared on behalf of the Commission for Climatology (CCL) Expert Team on Sector-specific Climate Indices (ET-SCI). Retrieved from https://epic.awi.de/id/eprint/49274/1/ClimPACTv2_manual.pdf

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, *111*, D05109. https://doi.org/10.1029/2005JD006290

Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., & Brönnimann, S. (2011). The international Atmospheric Circulation Reconstructions over the Earth (acre) initiative. *Bulletin of the American Meteorological Society*, *92*(11), 1421–1425.

Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, *419*(6903), 228.

Andrade, M. F., Moreno, I., Calle, J. M., Ticona, L., Blacutt, L., Lavado-Casimiro, W., et al. (2018). Atlas-clima y eventos extremos del altiplano central perú-boliviano. https://doi.org/10.4480/GB2018.N01

Avila, F. B., Dong, S., Menang, K. P., Rajczak, J., Renom, M., Donat, M. G., & Alexander, L. V. (2015). Systematic investigation of gridding-related scaling effects on annual statistics of daily temperature and precipitation maxima: A case study for south-east Australia. *Weather and Climate Extremes*, *9*, 6–16.

Barrucand, M., Rusticucci, M., & Vargas, W. (2008). Temperature extremes in the south of South America in relation to Atlantic Ocean surface temperature and Southern Hemisphere circulation. *Journal of Geophysical Research*, *113*, D20111. https://doi.org/10.1029/2007JD009026

Barry, A. A., Caesar, J., Klein Tank, A. M. G., Aguilar, E., McSweeney, C., Cyrille, A. M., et al. (2018). West Africa climate extremes and climate change indices. *International Journal of Climatology*, *38*, e921–e938.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *Journal of Geophysical Research*, *111*, D12106. https://doi.org/10.1029/2005JD006548

Brunet, M., SaladiÉ, O., Jones, P., SigrÓ, J., Aguilar, E., Moberg, A., et al. (2006). The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850–2003). *International Journal of Climatology*, *26*(13), 1777–1802.

Caesar, J., Alexander, L. V., Trewin, B., Tse-ring, K., Sorany, L., Vuniyayawa, V., et al. (2011). Changes in temperature and precipitation extremes over the Indo-Pacific region from 1971 to 2005. *International Journal of Climatology*, *31*(6), 791–801.

Caesar, J., Alexander, L., & Vose, R. (2006). Large-scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research*, *111*, D05101. https://doi.org/10.1029/2005JD006280

Chen, C.-T., & Knutson, T. (2008). On the verification and comparison of extreme rainfall indices from climate models. *Journal of Climate*, *21*(7), 1605–1621.

Cohn, A. S., Bhattarai, N., Campolo, J., Crompton, O., Dralle, D., Duncan, J., & Thompson, S. (2019). Forest loss in Brazil increases maximum temperatures within 50 km. *Environmental Research Letters*, *14*(8), 84047.

Contractor, S., Alexander, L. V., Donat, M. G., & Herold, N. (2015). How well do gridded datasets of observed daily precipitation compare over Australia? *Advances in Meteorology*, *2015*, 325,718.

Cressie, N. (1992). Statistics for spatial data. *Terra Nova*, *4*(5), 613–617.

de los Milagros Skansi, M., Brunet, M., Sigró, J., Aguilar, E., Groening, J. A. A., Bentancur, O. J., et al. (2013). Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global and Planetary Change*, *100*, 295–307.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Domonkos, P. (2011). Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). *International Journal of Geosciences*, *2*(03), 293.

Donat, M. G., & Alexander, L. V. (2012). The shifting probability distribution of global daytime and night-time temperatures. *Geophysical Research Letters*, *39*, L14707. https://doi.org/10.1029/2012GL052459

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., & Caesar, J. (2013). Global land-based datasets for monitoring climatic extremes. *Bulletin of the American Meteorological Society*, *94*(7), 997–1006.

Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., Dunn, R. J. H., et al. (2013). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres*, *118*, 2098–2118. https://doi.org/10.1002/jgrd.50150

Donat, M. G., Peterson, T. C., Brunet, M., King, A. D., Almazroui, M., Kolli, R. K., et al. (2014). Changes in extreme temperature and precipitation in the arab region: Long-term trends and variability related to ENSO and NAO. *International Journal of Climatology*, *34*(3), 581–592.

Donat, M. G., Sillmann, J., Wild, S., Alexander, L. V., Lippmann, T., & Zwiers, F. W. (2014). Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, *27*(13), 5019–5035.

Dunn, R. J. H., Donat, M. G., & Alexander, L. V. (2014). Investigating uncertainties in global gridded datasets of climate extremes. *Climate of the Past*, *10*(6), 2171–2199.

comments which improved the manuscript.

Dunn, R. J. H., Willett, K. M., & Parker, D. E. (2019). Changes in statistical distributions of sub-daily surface temperatures and wind speed. *Earth System Dynamics*, *10*(4), 765–788.

Frich, P., Alexander, L. V., Della-Marta, P. M., Gleason, B., Haylock, M., Tank, A. M. G. K., & Peterson, T. (2002). Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, *19*(3), 193–212.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., & Razuvaev, V. N. (2005). Trends in intense precipitation in the climate record. *Journal of Climate*, *18*(9), 1326–1350.

Guijarro, J. A. (2016). Automatización de la homogeneización de series climáticas: Nuevas funciones del paquete climatol 3.0. Retrieved from http://rua.ua.es/dspace/bitstream/10045/57982/1/2016_X-Congreso-AEC-Alicante_14.pdf

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, *113*, D20119. https://doi.org/10.1029/2008JD010201

Herold, N., & Alexander, L. (2016). Climpact 2. https://github.com/ARCCSS-extremes/climpact2

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Sabater, J., Nicolas, J., et al. (2019). Global reanalysis: Goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, *159*, 17–24.

Hofstra, N., & New, M. (2009). Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *29*(12), 1872–1880.

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, *35*(5), 841–858.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) (2018). Variabilidad climática y cambio climático en colombia.

Karl, T. R., Nicholls, N., & Ghazi, A. (1999). CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes workshop summary, *Weather and climate extremes* (pp. 3–7): Springer.

Kent, E. C., Challenor, P. G., & Taylor, P. K. (1999). A statistical determination of the random observational errors present in voluntary observing ships meteorological reports. *Journal of Atmospheric and Oceanic Technology*, *16*(7), 905–914.

Kenyon, J., & Hegerl, G. C. (2008). Influence of modes of climate variability on global temperature extremes. *Journal of Climate*, *21*(15), 3872–3889.

King, A. D., Donat, M. G., & Dunn, R. J. H. (2019). Land surface temperature extremes [in "state of the climate in 2018"]. *Bulletin of the American Meteorological Society*, *100*(9), S14–S16.

Klein Tank, A. M. G., Peterson, T. C., Quadir, D. A., Dorji, S., Zou, X., Tang, H., et al. (2006). Changes in daily temperature and precipitation extremes in central and south Asia. *Journal of Geophysical Research*, *111*, D16105. https://doi.org/10.1029/2005JD006316

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *22*(12), 1441–1453.

Klok, E. J., & Klein Tank, A. M. G. (2009). Updated and extended European dataset of daily climate observations. *International Journal of Climatology*, *29*(8), 1182–1191.

Kruger, A. C., & Nxumalo, M. (2017a). Surface temperature trends from homogenized time series in South Africa: 1931–2015. *International Journal of Climatology*, *37*(5), 2364–2377.

Kruger, A. C., & Nxumalo, M. P. (2017b). Historical rainfall trends in South Africa: 1921–2015. *Water SA*, *43*(2), 285–297.

Lanzante, J. R. (1996). Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *16*(11), 1197–1226.

McGree, S., Herold, N., Alexander, L., Schreider, S., Kuleshov, Y., Ene, E., et al. (2019). Recent changes in mean and extreme temperature and precipitation in the western Pacific islands. *Journal of Climate*, *32*(16), 4919–4941.

McKinnon, K. A., Rhines, A., Tingley, M. P., & Huybers, P. (2016). The changing shape of Northern Hemisphere summer temperature distributions. *Journal of Geophysical Research: Atmospheres*, *121*, 8849–8868. https://doi.org/10.1002/2016JD025292

Meehl, G. A., Arblaster, J. M., & Branstator, G. (2012). Mechanisms contributing to the warming hole and the consequent us east-west differential of heat extremes. *Journal of Climate*, *25*(18), 6394–6408.

Mekis, E., & Vincent, L. A. (2011). An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean*, *49*(2), 163–177.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An overview of the Global Historical Climatology Network-Daily database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910.

Meteorological Service Singapore (2019). Second workshop on ASEAN Regional Climate Data, Analysis and Projections (ARCDAP-2): Workshop report. Retrieved from http://ccrs.weather.gov.sg/arcdap-2workshop/

Min, S.-K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378.

New, M., Hulme, M., & Jones, P. (2000). Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *Journal of Climate*, *13*(13), 2217–2238.

Ning, L., & Bradley, R. S. (2015). Winter climate extremes over the northeastern United States and southeastern Canada and teleconnections with large-scale modes of climate variability. *Journal of Climate*, *28*(6), 2475–2493.

Oliver, H. J., Shin, M., & Sanders, O. (2018). Cylc: A workflow engine for cycling systems. *Journal of Open Source Software*, *3*(27), 737.

Otero, N., Sillmann, J., & Butler, T. (2018). Assessment of an extended version of the Jenkinson-Collison classification on CMIP5 models over Europe. *Climate Dynamics*, *50*(5-6), 1559–1579.

Pan, Z., Arritt, R. W., Takle, E. S., Gutowski Jr, W. J., Anderson, C. J., & Segal, M. (2004). Altered hydrologic feedback in a warming climate introduces a "warming hole". *Geophysical Research Letters*, *31*, L17109. https://doi.org/10.1029/2004GL020528

Panthou, G., Lebel, T., Vischel, T., Quantin, G., Sane, Y., Ba, A., et al. (2018). Rainfall intensification in tropical semi-arid regions: The Sahelian case. *Environmental Research Letters*, *13*(6), 064013.

Panthou, G., Vischel, T., Lebel, T., Blanchet, J., Quantin, G., & Ali, A. (2012). Extreme rainfall in west africa: A regional modeling. *Water Resources Research*, *48*, W08501. https://doi.org/10.1029/2012WR012052

Perkins-Kirkpatrick, S. E., Donat, M. G., & Dunn, R. J. H. (2018). Land surface temperature extremes [in "state of the climate in 2017"]. *Bulletin of the American Meteorological Society*, *99*(8), S15–S16.

Peterson, T. C. (2005). Climate change indices. *World Meteorological Organization Bulletin*, *54*(2), 83–86.
Peterson, T. C., & Manton, M. J. (2008). Monitoring changes in climate extremes: A tale of international collaboration. *Bulletin of the American Meteorological Society*, *89*(9), 1266–1271.
Peterson, T. C., Zhang, X., Brunet-India, M., & Vázquez-Aguirre, J. L. (2008). Changes in North American extremes derived from daily weather data. *Journal of Geophysical Research*, *113*, D07113. https://doi.org/10.1029/2007JD009453
Podestá, G., Skansi, M., Herrera, N., Veiga, H., & Rovere, S. (2013). Reporte técnico crc-sas-2013-001: Diseño del proceso de control de calidad de datos climáticos diarios en el centro regional del clima para el sur de américa del sur. Retrieved from http://www.crc-sas.org/es/pdf/reporte_tecnico_CRC-SAS-2013-001.pdf
Portmann, R. W., Solomon, S., & Hegerl, G. C. (2009). Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States. *Proceedings of the National Academy of Sciences*, *106*(18), 7324–7329.
Python Software Foundation. Python language reference, version 3.6.8. https://www.python.org
R Core Team (2013). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Rahimzadeh, F., & Nassaji Zavareh, M. (2014). Effects of adjustment for non-climatic discontinuities on determination of temperature trends and variability over Iran. *International Journal of Climatology*, *34*(6), 2079–2096. https://doi.org/10.1002/joc.3823
Razuvaev, V. N., Apasova, E. G. E., Martuganov, R. A., Vose, R. S., & Steurer, P. M. (1993). Daily temperature and precipitation data for 223 USSR stations: Oak Ridge National Lab., TN (United States).
Rhines, A., Tingley, M. P., McKinnon, K. A., & Huybers, P. (2015). Decoding the precision of historical temperature observations. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 2923–2933.
Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., et al. (2013). Berkeley Earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview*, *1*(2), 1–13.
Scaife, A. A., Folland, C. K., Alexander, L. V., Moberg, A., & Knight, J. R. (2008). European climate extremes and the North Atlantic Oscillation. *Journal of Climate*, *21*(1), 72–83.
Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association*, *63*(324), 1379–1389.
Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd acm national conference*, ACM, pp. 517–524.
Shin, M., Fitzpatrick, B., Clark, A., Sanders, O., Bartholomew, S. L., Whitehouse, S., et al. (2019). Rose software. https://doi.org/10.5281/zenodo.3458111
Sigró, J., Brunet, M., Domokos, P., Aguilar, E., Gilabert, A., Lister, D., et al. (2015). Long-term temperature change over mainland Spain: An update to 2014 and reassessment of the Spanish Daily Adjusted Temperature Series (SDATS). In *Proceeding of the climate-es workshop*.
Sillmann, J., & Croci-Maspoli, M. (2009). Present and future atmospheric blocking and its impact on European mean and extreme climate. *Geophysical Research Letters*, *36*, L10702. https://doi.org/10.1029/2009GL038259
Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, *118*, 1716–1733. https://doi.org/10.1002/jgrd.50203
Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., et al. (2014). *Climate change 2013: The physical science basis: Working group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*: Cambridge University Press.
Thiel, H. (1950). A rank-invariant method of linear and polynomial regression analysis, Parts 1–3. In *Proceedings of koninalijke nederlandse akademie van weinenschatpen a*, *53*, pp. 386–392, 521–525, 1397–1412.
Thorne, P. W., Allan, R. J., Ashcroft, L., Brohan, P., Dunn, R. J. H., Menne, M. J., et al. (2017). Toward an integrated set of surface meteorological observations for climate science and applications. *Bulletin of the American Meteorological Society*, *98*(12), 2689–2702.
Thorne, P. W., Donat, M. G., Dunn, R. J. H., Williams, C. N., Alexander, L. V., Caesar, J., et al. (2016). Reassessing changes in diurnal temperature range: Intercomparison and evaluation of existing global data set estimates. *Journal of Geophysical Research: Atmospheres*, *121*, 5138–5158. https://doi.org/10.1002/2015JD024584
Thorne, P. W., Menne, M. J., Williams, C. N., Rennie, J. J., Lawrimore, J. H., Vose, R. S., et al. (2016). Reassessing changes in diurnal temperature range: A new data set and characterization of data biases. *Journal of Geophysical Research: Atmospheres*, *121*, 5115–5137. https://doi.org/10.1002/2015JD024583
Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, *84*(9), 1205–1218.
Trewin, B., Braganza, K., Fawcett, R., Grainger, S., Jovanovic, B., Jones, D., et al. (2020). An updated long-term homogenized daily temperature data set for Australia. *Geoscience Data Journal*. https://doi.org/10.1002/gdj3.95
Van Den Besselaar, E. J. M., Klein Tank, A. M. G., Van Der Schrier, G., Abass, M. S., Baddour, O., Van Engelen, A. F. V., et al. (2015). International climate assessment & dataset: Climate services across borders. *Bulletin of the American Meteorological Society*, *96*(1), 16–21.
Veiga, H., Herrera, N., delos MilagrosSkansi, M., & Podestá, G. (2014). Reporte técnico crc-sas-2014-001: Descripción de controles de calidad de datos climáticos diarios implementados por el centro regional del clima para el sur de américa del sur. *Centro Regional de Clima para el sur de Amércia del Srur, Buenos Aires*, *2*, 2014. http://www.crc-sas.org/es/pdf/reporte_tecnico_CRC-SAS-2014-001.pdf
Vincent, L. A., Aguilar, E., Saindou, M., Hassane, A. F., Jumaux, G., Roy, D., et al. (2011). Observed trends in indices of daily and extreme temperature and precipitation for the countries of the western Indian Ocean, 1961–2008. *Journal of Geophysical Research*, *116*, D10108. https://doi.org/10.1029/2010JD015303
Vincent, L. A., Wang, X. L., Milewska, E. J., Wan, H., Yang, F., & Swail, V. (2012). A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research*, *117*, D18110. https://doi.org/10.1029/2012JD017859
WMO (2007). The role of climatological normals in a changing climate, wmo/td-no. 1377. Retrieved from https://library.wmo.int/index.php?lvl=notice_display&id=16659
Wang, Y., & Feng, Y. (2013). Rhtestsv4 software. https://etccdi.pacificclimate.org/software.shtml
Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, *26*(11), 3904–3918.
Xu, W., Li, Q., Wang, X. L., Yang, S., Cao, L., & Feng, Y. (2013). Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *Journal of Geophysical Research: Atmospheres*, *118*, 9708–9720. https://doi.org/10.1002/jgrd.50791
Yin, H., Donat, M. G., Alexander, L. V., & Sun, Y. (2015). Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. *International journal of climatology*, *35*(10), 2809–2827.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 851–870.
Zhang, X., & Feng, Y. (2004). Rclimdex and fclimdex software. Retrieved from http://etccdi.pacificclimate.org/software.shtml
Zhang, X., Hegerl, G., Zwiers, F. W., & Kenyon, J. (2005). Avoiding inhomogeneity in percentile-based indices of temperature extremes. *Journal of Climate*, *18*(11), 1641–1651.