HadEX.3.0.4: Product User Guide

Robert Dunn

robert.dunn@metoffice.gov.uk

October 15, 2024

Citation:

Dunn, R.J.H, Alexander, L.V., Donat, M.G. 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, https://doi.org/10.1029/2019JD032263

Dunn, R.J.H, Herold, N., Alexander, L.V., Donat, M.G. et al, (2024), Observed Global Changes in Sector-Relevant Climate Extremes Indices - An Extension to HadEX3, Earth and Space Science, https://doi.org/10.1029/2023EA003279

EMERGENCY TWO PAGE QUICK START GUIDE FOR HadEX.3.0.4

HadEX3 (HadEX.3.0.4) is a dataset of temperature and precipitation extremes indices developed by the Met Office Hadley Centre and the ARC Centre of Excellence for Climate Extremes at the University of New South Wales (Sydney). Both ETCCDI and ET-SCI indices are available as annual, and for some, monthly quasi-global land surface fields. The dataset covers 1901-2018 on a $1.25° \times 1.875°$ latitude-longitude grid.

Some indices use a reference period to determine threshold values (e.g. a percentile of the distribution), and to enable easy comparison with earlier datasets, this is 1961-90. However, a version is available with a 1981-2010 reference period for these indices, with a comparable spatio-temporal coverage.

This dataset is static and we do not envisage any updates. It has been created from a wide variety of data sources, from public facing repositories of daily observations through to data available from individual researchers.

What are the climate extremes indices in HadEX.3.0.4?

The WMO Expert Team on Climate Change Detection and Indices (ETCCDI) defined a set of moderate climate extremes indices based on daily precipitation, maximum and minimum temperatures. A full list of the 29 indices from the initial HadEX3 release is available in Table 2 (Dunn, R.J.H, Alexander, L.V., Donat, M.G. 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, https://doi.org/10.1029/2019JD032263).

Subsequently, the WMO Expert Team on Sector-specific Climate Indices (ET-SCI) defined a further set of sector relevant indices, with a focus on agriculture, health, and water-resources. These indices were included in the extension to HadEX3, and are shown in Table ?? (Dunn, R.J.H, Herold, N., Alexander, L.V., Donat, M.G. et al, (2024), Observed Global Changes in Sector-Relevant Climate Extremes Indices - An Extension to HadEX3, Earth and Space Science, https://doi.org/10.1029/2023EA003279).

What products are available?

The primary product are gridded indices available at: http://www.metoffice.gov.uk/hadobs/hadex3/ and http://www.climdex.org/

Further diagnostic grids of station counts are available at: http://www.metoffice.gov.uk/hadobs/hadex3/

Finally, station time series files for the subset of stations that we are able to share are available at: http://www.climdex.org/

How do I obtain the data?

Data are available from http://www.metoffice.gov.uk/hadobs/hadex3/ and http://www.climdex.org/. There are also versions available at the CEDA Archive https://catalogue.ceda.ac.uk/

How do I read the HadEX.3.0.4 data?

Grids

The gridded data are stored in NetCDF format files. NetCDF files are a platform-independent, self-describing binary format and there are a number of common tools (Section 2.1) that can be used to access the data. Some basic python code is provided in Section 4 to show worked examples of reading the data and performing some simple calculations and processing.

Through http://www.metoffice.gov.uk/hadobs/hadex3/ there are files available in two formats. One matches that from HadEX2 and HadEX, the other matches those available at http://www.climdex.org/.

Station series

The station series are available as `csv` (comma separated value) which can be read by a wide variety of tools. These are available at http://www.climdex.org/.

Diagnostic Plots

A number of automatically generated plots of trends, differences and coverage are available at http://www.metoffice.gov.uk/hadobs/hadex3/.

What tools are available for these products?

Some basic Python code is provided in Section 4 to show worked examples of reading in the data and performing some simple calculations and processing.

HadEX3 Changelog

Version number changes are outlined below (most recent at the top). The release of the ET-SCI extension did not affect the ETCCDI indices, and so the version numbers were aligned at this point.

Version	Date	Description
3.0.4	2024-04-04	Extension of ET-SCI indices released.
3.0.4	2022-01-07	Replaced GSL files.
3.0.3	2020-12-16	Corrected Rx1day and Rx5day from West African (indices) data. Included some GHCND flags when reformatting USA station files (all indices).
3.0.2	2020-11-01	Corrected land-sea mask (all indices and reference periods).
3.0.1	2020-09-28	Replaced DTR and TN90p (61-90) files. Erroneous stations in SE Asia withheld.
3.0.0	2020-07-01	Initial release

How to cite the data set

Please cite the following references when using HadEX3. When using the ET-SCI indices, then please cite the original paper (Dunn et al, 2020) as well.

Dunn, R.J.H, Alexander, L.V., Donat, M.G. 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, https://doi.org/10.1029/2019JD032263

Dunn, R.J.H, Herold, N., Alexander, L.V., Donat, M.G. et al, (2024), Observed Global Changes in Sector-Relevant Climate Extremes Indices - An Extension to HadEX3, Earth and Space Science, https://doi.org/10.1029/2023EA003279

For further citation information, for e.g. the precessor (HadEX2) [3], the uncertainty investigation on HadEX2 [6] or other cross comparisons using these kinds of dataset [5, 15], these references may be useful. Finally, a cross comparison with recent reanalysis products may also be of interest [8].

Further information and contact

For further help please read the rest of the document. The paper describing the data set is the best place to find the technical details. Updated diagnostics are available from http://www.metoffice.gov.uk/hadobs/hadex3/ or http://www.climdex.org/. For further enquiries contact robert.dunn@metoffice.gov.uk. Data set updates will be tweeted from @metofficeHadOBS.

Contents

1 Getting started with HadEX.3.0.4

This section describes some of the basic information about the HadEX.3.0.4 dataset. HadEX.3.0.4 is a dataset of moderate temperature and precipitation extremes indices, with the primary product being gridded fields. Alongside these, some of the station time series are also available.

1.1 How do I get the data?

The gridded data are available from http://www.metoffice.gov.uk/hadobs/hadex3/download.html and http://www.climdex.org/ in NetCDF format. The station indices in `csv` format are available from http://www.climdex.org/ only.

All versions will be archived to the CEDA (Centre for Environmental Data Analysis) Archive at https://catalogue.ceda.ac.uk/.

1.1.1 Gridded data

- Grid for one index (TX90p as an example), annual and monthly, using 1961-90 reference period, 1901-2018, in "HadEX" format (matching earlier HadEX products). link

- Grid for one index (TX90p as an example), number of stations within each grid box for each time point, link

- Grid for one index (TX90p as an example), number of stations within the DLS[1] for each grid box for each time point, link

- Grid for one index (TX90p as an example), annual (monthly available in a separate file), using 1961-90 reference period, 1901-2018, in Climdex format link

1.1.2 Time series data

Data were collected from over 37,000 stations across the globe, but not all of these were included when creating the gridded product. The station time series are available through the http://www.climdex.org/ portal where you can select them by index and location.

1.2 How do I use the data?

The data are in NetCDF format, which is a standard data format for climate data. Furthermore, the data files are almost CF compliant which means that the metadata in the files is in a standardised format. The "almost" is because of both the limitations of the CF standard which does not yet include the ETCCDI[2] climate extremes indices nor those from the ET-SCI[3], and also how we have structured the files to retain consistency with earlier versions of HadEX.

The structure of the data files is described in more detail in Section 2. Some example code for reading and processing the data is found in Section 4. The gridded data are stored in a 3-dimensional array, with dimensions of time × longitude × latitude. There are 192 longitude points and 144 latitude points. The 192 longitude points represent grid cells with centres running from $+0.9372°$E to $+359.0625°$E (the longitudinal bounds being $0° \rightarrow 1.875°$E and $359.125°$E $\rightarrow 360°$E respectively. The 144 latitude points represent grid cells with centres running from $-89.375°$N to $+89.375°$N ($-90°$N $\rightarrow -88.75°$N and $88.75°$N $\rightarrow 90°$N respectively).

---

[1]Decorrelation Length Scale

[2]Former WMO Expert Team on Climate Change Detection and Indices

[3]Former WMO Expert Team on Sector-specific Climate Indices

1.2.1 Dos and Donts of using the data

Do - be aware that there are gaps in the data in space and time. Furthermore, the gridding method interpolates into areas with few station observations (to differing extents depending on the index). Auxiliary products like the number of stations within a grid box or the decorrelation length scale (DLS) can provide useful additional information for identifying grid cells in which the uncertainty is likely to be large.

Do - check the Changelog to see what has been updated since the dataset was released.

Don't - compare HadEX.3.0.4 with other datasets of the ETCCDI indices if the reference periods do not match [for the indices which use a reference period].

Don't - compare HadEX.3.0.4 to other datasets of ETCCDI extremes indices without taking into account the spatio-temporal coverage. Data coverage can affect the comparability of two data sets. See e.g. [2] for an example on the issues this can cause.

Do - read the Open Government Licence v3.

Do - send us feedback when you use the data to robert.dunn@metoffice.gov.uk.

1.3 Basic data statistics

Data from around 37,000 stations was provided for use within HadEX3. After selecting for quality and length of record, up to 7,000 were used in the temperature indices and 17,000 for the precipitation indices. The gridding method used interpolates into regions which have few or no stations. The early part of the dataset has poorer spatial coverage than the latter part of the 20th century (Figure 1). However we encourage users check the spatio-temporal coverage of the dataset when investigating changes over the whole 20th century, especially for the early period.

Alongside the netCDF grids of the data, on http://www.metoffice.gov.uk/hadobs/hadex3/ we have also made available netCDF files containing two sets of station counts. The first show the number of stations available within the confines of the grid box at each time point. As there is no stipulation on the minimum number for this, it can be zero, especially in data sparse parts of the world.

Secondly, we also provide the number of stations within a decorrelation length scale (DLS) of the grid box centre, which must be at least three in order for an average value to be calculated and stored for that grid box at that time point.

1.4 Contact us

If you have further questions about the HadEX.3.0.4 data set, please contact us at robert.dunn@metoffice.gov.uk.

1.5 FAQ

1.5.1 Is HadEX.3.0.4 the data set for me?

There are a number of quasi-global datasets of the ETCCDI indices and many other regional analyses (which we do not cover here). Table 1 provides an overview comparison of these datasets with HadEX.3.0.4.
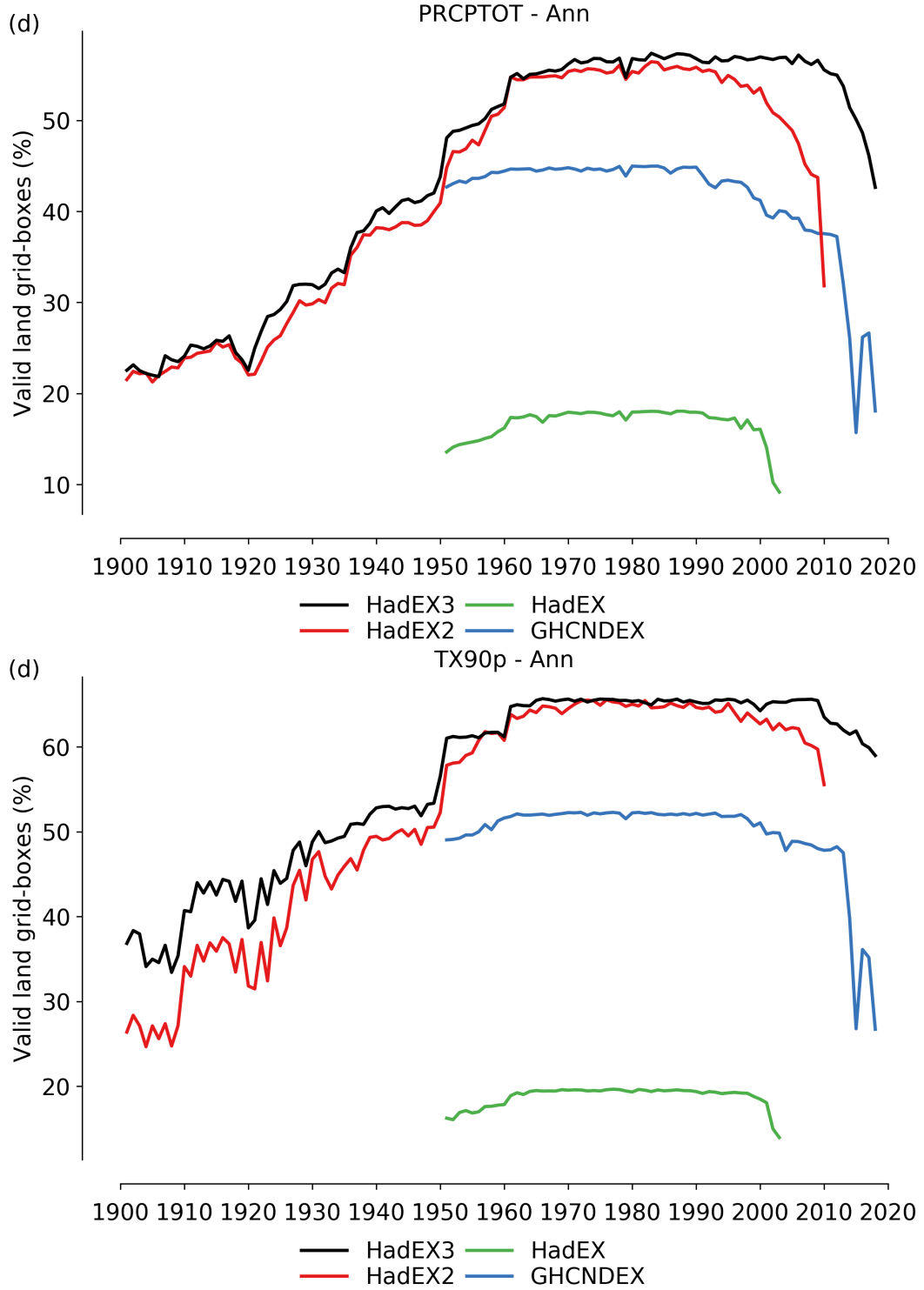
Figure 1: Spatial coverage over time for PRCPTOT and TX90p (1961-90) for HadEX3 in comparison with similar datasets (HadEX2, HadEX and GHCNDEX). Fraction of land grid boxes is shown.

Dataset	Gridded?	Resolution	Updates?	Start	End	Citation
HadEX.3.0.4	Y	$1.25° \times 1.875°$	-	1901	2018	[7]
HadEX2	Y	$2.5° \times 3.75°$	-	1901	2010	[3]
HadEX	Y	$2.5° \times 3.75°$	-	1951	2003	[1]
GHCNDEX	Y	$2.5° \times 2.5°$	Y	1951	-	[4]

Table 1: Dataset characteristic comparison table

1.5.2 What reference period have you used?

A subset of the indices (for ETCCDI:TX90p, TX10p, TN90p, TN10p, WSDI, CSDI, R95p, R99p, R95pTOT, R99pTOT; for ET-SCI: WSDI3, CSDI3, TX2TN2, TXb2TNb2, TXgt50p, TX95t, and heat-wave indices [ECF, EHF, TX90, TN90]) require a reference period in order to determine the numerical thresholds used. For these indices, we provide two versions, one using a reference period of 1961-90 and one of 1981-2010. For other indices the choice of a reference period has no effect.

As some data was submitted as pre-calculated indices we could not adjust the reference period, and so using both reference periods allowed us to make use of all data that had been submitted. In the end, most data providers were able to make available versions using both reference periods, which resulted in comparable spatio-temporal coverage.

By keeping 1961-90, we allow consistency with HadEX2 and other similar products and analyses. But using 1981-2010 looks forward, as there are now more stations which have coverage over this period, especially in parts of the world with a less-dense network available. We were also able to make the comparison between both reference periods within the paper, which ensured we could clearly communicate to users the care needed when using these indices from different reference periods.

We do note that the quality control checks use the data completeness over this reference period to ensure that there are sufficient observations present to set the numerical thresholds. For stations where we received the underlying observations, a similar check was also carried out by the Climpact software. There is also a knock on effect in that the Climpact software calculates no indices when there are no observations during the reference period (rather than only excluding those indices which use the reference period). Hence the station selections for other indices do have a slight dependancy on the reference period, but for clarity we do not make files for these indices available for the second reference period.

1.5.3 Where do the observations come from?

The full list of observation sources is shown in the paper, along with references where these were available. These are from publicly facing archives of global, regional and national collections as well as via individuals co-authors who contributed the data. Some arrived as pre-calculated indices whereas for the daily observations, we used the Climpact2 software.

1.5.4 How does HadEX.3.0.4 differ from HadEX2 and GHCNDEX?

The key differences between HadEX.3.0.4 and other similar datasets are shown in Table 1. Note that HadEX.3.0.4 supercedes HadEX2 and HadEX.

1.5.5 How to cite the data?

Dunn, R.J.H, Alexander, L.V., Donat, M.G. 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, https://doi.org/10.1029/2019JD032263

Dunn, R.J.H, Herold, N., Alexander, L.V., Donat, M.G. et al, (2024), Observed Global Changes in Sector-Relevant Climate Extremes Indices - An Extension to HadEX3, Earth and Space Science, https://doi.org/10.1029/2023EA003279

1.5.6 I want data after 2018, what can I use?

HadEX.3.0.4 is a static data product. It is not operational and so will not be updated. It is highly likely that the next project to release a future version (HadEX4.0.0) will tie into the IPCC[4] process and so will occur a number of years after HadEX.3.0.4's release.

The GHCNDEX [4] product, also available at http://www.climdex.org/, is an updated dataset of gridded ETCCDI indices using the freely available GHCND (Global Historical Climate Network Daily, https://www.ncdc.noaa.gov/ghcn-daily-description, 12). This is updated roughly once a year, and so allows for climate monitoring.

1.5.7 What are the definitions of the indices

We show in Table 2 the details of the ETCCDI indices which are present in HadEX3. This table has been adapted from Table 1 of [7]. We indicate which indices are available on a monthly basis.

---

[4]UNFCCC Intergovernmental Panel on Climate Change

Index	Name	Description	Units
TXx	Hottest day	Monthly and annual highest value of daily max temperature	°C
TNx	Warmest night	Monthly and annual highest value of daily min temperature	°C
TXn	Coldest day	Monthly and annual lowest value of daily max temperature	°C
TNn	Coldest night	Monthly and annual lowest value of daily min temperature	°C
TN10p	Cool nights	Percentage of time when daily min temperature < 10th percentile	%
TX10p	Cool days	Percentage of time when daily max temperature < 10th percentile	%
TN90p	Warm nights	Percentage of time when daily min temperature > 90th percentile	%
TX90p	Warm days	Percentage of time when daily max temperature > 90th percentile	%
DTR	Diurnal temperature range	Annual mean difference between daily max and min temperature	°C
GSL	Growing season length	Annual (1st Jan to 31st Dec in Northern Hemisphere, 1st July to 30th June in Southern Hemisphere) count between first span of at least 6 days with TG > 5°C and first span after July 1 (January 1 in SH) of 6 days with TG < 5°C (where TG is daily mean temperature)	days
ID	Ice days	Annual count when daily maximum temperature < 0°C	days
FD	Frost days	Annual count when daily minimum temperature < 0°C	days
SU	Summer days	Annual count when daily max temperature > 25°C	days
TR	Tropical nights	Annual count when daily min temperature > 20°C	days
WSDI	Warm spell duration index	Annual count when at least six consecutive days of max temperature > 90th percentile	days
CSDI	Cold spell duration index	Annual count when at least six consecutive days of min temperature < 10th percentile	days
* *ETR*	Extreme temperature range	TXx - TNn	°C

Table 2: Details of the ETCCDI indices and the numbers of stations available for each one that were taken forward into the gridded product. The indices labelled 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. (Adapted from Table 1 in [7].)

Index	Name	Description	Units
Rx1day	Max 1 day precipitation amount	Monthly and annual maximum 1 day precipitation	mm
Rx5day	Max 5 day precipitation amount	Monthly and annual maximum consecutive 5 day precipitation	mm
SDII	Simple daily intensity index	The ratio of annual total precipitation to the number of wet days ($\geq$ 1 mm)	mm/day
R10mm	Number of heavy precipitation days	Annual count when precipitation $\geq$ 10 mm	days
R20mm	Number of very heavy precipitation days	Annual count when precipitation $\geq$ 20 mm	days
CDD	Consecutive dry days	Highest number of consecutive days when precipitation $<$ 1 mm	days
CWD	Consecutive wet days	Highest number of consecutive days when precipitation $\geq$ 1 mm	days
R95p	Very wet days	Annual total precipitation from days $>$ 95th percentile	mm
R99p	Extremely wet days	Annual total precipitation from days $>$ 99th percentile	mm
PRCPTOT	Annual total wet day precipitation	Annual total precipitation from days $\geq$ 1 mm	mm
* R95pTOT	Contribution from very wet days	100 * R95p / PRCPTOT	%
* R99pTOT	Contribution from extremely wet days	100 * R99p / PRCPTOT	%

Table 2: (Continued)

Table 3: Details of the ET-SCI indices and the numbers of stations available for each one that were taken forward into the gridded product after selection for record length and quality control. The temperatures (n) for CDDcold$n$, HDDheat$n$ and GDDgrow$n$, and day count (d) for CSDId and WSDId are user defined. The indices labeled in italics are available on a monthly as well as annual basis. The sectors to which the indices are most relevant are listed, using the information as given in the Climpact2 User Guide, Appendix A, where H-Health, A-Agriculture and Food Security, W-Water Resources. This is also available at https://www.climdex.org/learn/indices/sector/. (Adapted from Table 1 in [9].)

Index	Name	Description	Units	Sector(s)
CDDcold$n$	Cooling degree days	The annual sum of $T_{\mathrm{mean}} - n$, where n is a user defined temperature and $T_{\mathrm{mean}} > n$	degree-days	H
HDDheat$n$	Heating degree days	The annual sum of $n - T_{\mathrm{mean}}$, where n is a user defined temperature and $T_{\mathrm{mean}} < n$	degree-days	H
GDDgrow$n$	Growing degree days	The annual sum of $T_{\mathrm{mean}} - n$, where n is a user defined temperature and $T_{\mathrm{mean}} > n$	degree-days	H, A
CSDId	User defined CSDI	Cold spell duration index, annual count of days where d or more consecutive days experience $T_{\mathrm{max}} < 10^{\mathrm{th}}$ percentile	days	H, A, W
WSDId	User defined WSDI	Warm spell duration index, annual count of days where d or more consecutive days experience $T_{\mathrm{max}} > 90^{\mathrm{th}}$ percentile	days	H, A, W
TNlt2	T_{min} below 2°C	Number of days when $T_{\mathrm{min}} < 2$°C	days	A
TNltm2	T_{min} below −2°C	Number of days when $T_{\mathrm{min}} < -2$°C	days	A
TNltm20	T_{min} below −20°C	Number of days when $T_{\mathrm{min}} < -20$°C	days	H, A
TXge30	T_{max} at least 30°C	Number of days when $T_{\mathrm{max}} \geq 30$°C	days	H, A
TXge35	T_{max} at least 35°C	Number of days when $T_{\mathrm{max}} \geq 35$°C	days	H, A
TNm	Mean T_{min}	Mean daily minimum temperature	°C	
TXm	Mean T_{max}	Mean daily maximum temperature	°C	
TMm	Mean T_{mean}	Mean daily mean temperature	°C	
TMlt10	T_{mean} below 10°C	Number of days when $T_{\mathrm{mean}} < 10$°C	days	A
TMlt5	T_{mean} below 5°C	Number of days when $T_{\mathrm{mean}} < 5$°C	days	A
TMge5	T_{mean} at least 5°C	Number of days when $T_{\mathrm{mean}} \geq 5$°C	days	A
TMge10	T_{mean} at least 10°C	Number of days when $T_{\mathrm{mean}} \geq 10$°C	days	A

In the following table, we include those indices from the ET-SCI extension of HadEX3. This table has been adapted from Table 1 of [9]. We indicate which indices are available on a monthly basis.

In last two tables, we present the heat wave metrics and measures which, when combined, form a set of heat wave indices the ET-SCI extension of HadEX3. These tables have been adapted from Tables 2 and 3 of [9]. The available indices are then of the fomr EHF_HWF, or TN90_HWM.

2 Using the gridded data files

Just to confuse matters, there are two sets of files available on http://www.metoffice.gov.uk/hadobs/hadex3/. One set are in what we call "HadEX" format and these are very similar to those produced for HadEX2 and HadEX. For these files, a single NetCDF file contains the annual (and if appropriate, monthly) fields. The monthly fields are stored under each calendar month rather than in time-order

Another set ("Climdex" format) have been produced for transparency. HadEX3 is also hosted at

Table 3: (Continued). Consecutive day count (d) for TXdTNd and TXbdTNbd, monthly interval for SPI and SPEI measures (M) are user defined. For more details on SPI see [11, 17] and the SPEI see [16]. † SPI and SPEI indices are available on a monthly basis only. ∗ This index is given as 365 values, one for each day of the year, and so is of a different format to others within the ET-SCI and ETCCDI families. The sectors to which the indices are most relevant are listed, using the information as given in the Climpact2 User Guide, Appendix A, where H-Health, A-Agriculture and Food Security, W-Water Resources. This is also available at https://www.climdex.org/learn/indices/sector/. (Adapted from Table 1 in [9].)

Index	Name	Description	Units	Sector(s)
TX95t∗	Very warm day threshold	Value of the 95$^{\text{th}}$ percentile of $T_{\max}$	°C	H, A
TXgt50p	Percentage of days with maximum temperature above the median	Percentage of days with $T_{\max} > 50^{\text{th}}$ percentile	%	H, A, W
TXdTNd	User-defined consecutive number of hot days and nights	Annual count of d consecutive days where both $T_{\max} > 95^{\text{th}}$ percentile and $T_{\min} > 95^{\text{th}}$ percentile, where $10 \geq d \geq 2$	days	H, A, W
TXbdTNbd	User-defined consecutive number of cold days and nights	Annual count of d consecutive days where both $T_{\max} < 5^{\text{th}}$ percentile and $T_{\min} < 5^{\text{th}}$ percentile, where $10 \geq d \geq 2$	days	H, A, W
Rxdday	Monthly maximum d-day precipitation	Maximum rainfall occurring within d consecutive days	mm	H, A, W
Mmonth_SPI†	Standardised Precipitation Index	SPI drought measure using precipitation over M months		H, A, W
12month_SPI†	Standardised Precipitation Index	SPI drought measure using precipitation over 12 months		H, A, W
6month_SPI †	Standardised Precipitation Index	SPI drought measure using precipitation over 6 months		H, A, W
3month_SPI†	Standardised Precipitation Index	SPI drought measure using precipitation over 3 months		H, A, W
Mmonth_SPEI†	Standardised Precipitation Evapotranspiratiomn Index	SPEI drought measure using precipitation and evaporation over M months		H, A, W
12month_SPEI†	Standardised Precipitation Evapotranspiration Index	SPEI drought measure using precipitation and evaporation over 12 months		H, A, W
6month_SPEI†	Standardised Precipitation Evapotranspiration Index	SPEI drought measure using precipitation and evaporation over 6 months		H, A, W
3month_SPEI†	Standardised Precipitation Evapotranspiration Index	SPEI drought measure using precipitation and evaporation over 3 months		H, A, W

Table 4: (Continued)

14

Table 5: The various heat wave characteristics defined by [14] as used with the heat wave metrics in Table 6 to create 20 ET-SCI indices, e.g. EHF_HWF. These can also be used for cold waves. A heatwave is defined as being at least 3 days long, using the metrics from Table 6. The sectors to which the indices are most relevant are listed, using the information as given in the Climpact2 User Guide, Appendix A, where H-Health, A-Agriculture and Food Security, W-Water Resources. This is also available at https://www.climdex.org/learn/indices/sector/.(Adapted from Table 2 in [8].)

Abbreviation	Name	Description	Units	Sector(s)
HWA	Heat wave amplitude	Peak value of the hottest heat wave/coldest cool wave	°C	H, A, W
HWD	Heat wave duration	Length of the longest heat wave in a season as identified by HWN	days	H, A, W
HWF	Heat wave frequency	Sum of the number of days which contribute to heat waves as defined by HWN	days	H, A, W
HWM	Heat wave magnitude	Mean temperature of all heat waves identified by HWN	°C	H, A, W
HWN	Heat wave number	Number of individual heat waves each summer season (NovMar in southern hemisphere and MaySep in northern hemisphere)	-	H, A, W

Table 6: The various heat (and cold) wave metrics used with the heat wave characteristics in Table 5 to create 20 ET-SCI indices, e.g. EHF_HWF, TX90_HWA. The ECF and EHF metrics are defined in from [13, 14]. The percentiles are calculated over the 1981-2010 period. Note that TN90 and TX90 are not the same as the TN90p and TX90p indices from the ETCCDI set. The sectors to which the indices are most relevant are listed, using the information as given in the Climpact2 User Guide, Appendix A, where H-Health, A-Agriculture and Food Security, W-Water Resources. This is also available at https://www.climdex.org/learn/indices/sector/. (Adapted from Table 3 in [9].)

Heat/Cold wave definition	Name	Heat/Cold wave criteria	Units	Sector(s)
ECF	Excess Cold Factor	A cold wave occurs when at least 3 or more days have positive ECF	$°C^2$	H, A, W
EHF	Excess Heat Factor	A heat wave occurs when at least 3 or more days have positive EHF	$°C^2$	H, A, W
TN90		A heat wave occurs when at least 3 or more days have $T_{\min} > 90^{\text{th}}$ percentile		H, A, W
TX90		A heat wave occurs when at least 3 or more days have $T_{\max} > 90^{\text{th}}$ percentile		H, A, W

http://www.climdex.org/, which also allows for subsetting of the data. However, in order to do this, the files were split and re-ordered, with separate annual and monthly files (and the latter have all the months in sequential order). We provide both these on http://www.metoffice.gov.uk/hadobs/hadex3/ but focus on the "HadEX" format here.

2.1 Tools that can be used to work with data files

A list of software tools that work with NetCDF files is maintained by UCAR https://www.unidata.ucar.edu/software/netcdf/software.html). Some simple tools for viewing and manipulating NetCDF files in Linux include:

- ncdump: provided with the NetCDF library, produces a text rendering of a NetCDF file (Unidata at UCAR).

- Climate Data Operators (CDO)s: a set of command line utilities for performing operations on NetCDF files including concatenation, editing and mathematics (https://code.mpimet.mpg.de/projects/cdo).

- ncview: a program to produce graphical displays of the contents of NetCDF files. More information can be found here. A more complete list can be found here.

In addition, packages are available in most commonly-used scientific programming languages for reading and working with NetCDF files. For example, in Python there are numerous packages including:

- netCDF4 - this basic package provides functionality to read NetCDF files and extract metadata.

- Iris - developed by the Met Office, Iris provides functionality to read, write and process files in a variety of formats including NetCDF. Some example snippets of code using Iris to process HadEX.3.0.4 are provided in Section 4.

There are a number of packages in R that can be used to process NetCDF files:

- ncdf4: https://cran.r-project.org/web/packages/ncdf4/index.html

- raster: https://cran.r-project.org/web/packages/raster/index.html

- rcdo: https://github.com/r4ecology/rcdo

- RNetCDF: https://cran.r-project.org/web/packages/RNetCDF/index.html

- CM SAF R tools: https://www.mdpi.com/2220-9964/8/3/109

2.2 File names (HadEX format)

The filenames for the gridded data available on http://www.metoffice.gov.uk/hadobs/hadex3/ in "HadEX" format follows a simple pattern:

`HadexVersion_Index_Start-End_Gridding_ReferencePeriod_GridSize.nc`.

We'll walk through this using the example below:

`HadEX3_TXx_1901-2018_ADW_61-90_1.25x1.875deg.nc`

HadEX	Dataset identifier (in line with other HadOBS products)	
3	Major Version	
1901	Start of dataset (January 1, 1931 at 00h00)	
2018	End of dataset (December 31, 2018 at 00h00)	
ADW	Gridding method (Angular Distance Weighting)	
61-90	Reference period used (not relevant for this index, but included for consistency across all indices)	
1.25x1.875deg	Grid box dimensions	

A number of entries in the file names allow for future variation (e.g. gridding method) or are not applicable to all indices (e.g. reference period) but have been kept to future proof these filenames.

2.3   Contents of data files (HadEX format)

We show the field information for the primary NetCDF files in Table 7. These metadata are all available through the NetCDF header information. We show the example for an index with monthly fields. These are obviously not available for indices with annual data only.

Field Name	Size	Description
longitude	[192]	Grid longitudes (degrees)
latitude	[144]	Grid latitudes (degrees)
time	[118]	Time (years)
Ann	[118,144,192]	Annual fields
Jan	[118,144,192]	January fields
Feb	[118,144,192]	February fields
Mar	[118,144,192]	March fields
Apr	[118,144,192]	April fields
May	[118,144,192]	May fields
Jun	[118,144,192]	June fields
Jul	[118,144,192]	July fields
Aug	[118,144,192]	August fields
Sep	[118,144,192]	September fields
Oct	[118,144,192]	October fields
Nov	[118,144,192]	November fields
Dec	[118,144,192]	December fields

Table 7:   Fields present in the primary NetCDF files.

2.4   File names (Climdex format)

The filenames for the gridded data available on http://www.metoffice.gov.uk/hadobs/hadex3/ in "Climdex" format follows a simple pattern:

    HadexVersion_[ReferencePeriod]_Index_Timescale.nc.

We'll walk through this using the examples below:

    HadEX3_61-90_TX90p_ANN.nc
    HadEX3_R10mm_MON.nc

HadEX	Dataset identifier (in line with other HadOBS products)
3	Major Version
61-90	Reference period used (only present in indices where relevant)
TX90p	Index
ANN	Timescale - either ANN-Annual, or MON-Monthly

The change in the file names and format has been done to allow for subsetting by the tools available on the http://www.climdex.org/ server. The monthly data are now stored in separate files to the annual data where these are available.

2.5 Contents of data files (Climdex format)

We show the field information for the primary NetCDF files in Tables 8 and 9. These metadata are all available through the NetCDF header information. We show examples for both the annual and monthly files.

Field Name	Size	Description
longitude	[192]	Grid longitudes (degrees)
latitude	[144]	Grid latitudes (degrees)
time	[118]	Time (years)
Index	[118,144,192]	Annual fields

Table 8: Fields present in the Climdex Annual NetCDF files.

Field Name	Size	Description
longitude	[192]	Grid longitudes (degrees)
latitude	[144]	Grid latitudes (degrees)
time	[1416]	Time (months)
Index	[1416,144,192]	Monthly fields

Table 9: Fields present in the Climdex Monthly NetCDF files.

3 Using the time series files

Although the primary product of HadEX3 is a set of gridded files, these have been built from individual station records. For a subset of the stations we are allowed to share these records. These are available at http://www.climdex.org/.

4 Worked examples

4.1 Calculate global mean index time series

Read data and plot the Annual values for 2018 using the Iris package in Python.

```
import os
import matplotlib.pyplot as plt
import iris
import iris.quickplot as qplt

DATALOC = "/data/local/"
```

```
cubes = iris.load(os.path.join(DATALOC,
                 'HadEX3_TXx_1901 -2018_ADW_61 -90_1.25x1.875deg.nc'))

# extract the Annual data
for cube in cubes:
    if cube.var_name == "Ann":
        annual_cube = cube
        break

# plot the final field (2018)
qplt.pcolor(annual_cube[-1])
plt.show()
```

To produce a simple global average timeseries, using the 1961-90 period to make anomalies and only selecting grid boxes which have ≥90% completeness.

```
import os
import numpy as np
import matplotlib.pyplot as plt
import iris
import iris.quickplot as qplt

DATALOC = "/data/local/"

cubes = iris.load(os.path.join(DATALOC,
                 'HadEX3_TXx_1901 -2018_ADW_61 -90_1.25x1.875deg.nc'))

# extract the Annual data (month identifier stored in {\texttt var_name}).
for cube in cubes:
    if cube.var_name == "Ann":
        annual_cube = cube
        break

# need to convert to anomalies
time_constraint = iris.Constraint(time=lambda cell: 1961 <= cell <= 1990)
climatology_period = annual_cube.extract(time_constraint)
climatology = climatology_period.collapsed(['time'],
                            iris.analysis.MEAN)

anomaly_cube = annual_cube - climatology

# check for 90\% completeness
year_count = np.ma.count(anomaly_cube.data, axis=0)
year_count = np.tile(year_count, [118, 1, 1])

# and mask the data
anomaly_cube.data.mask[year_count < 0.9*118] = True


grid_areas = iris.analysis.cartography.cosine_latitude_weights(annual_cube)
timeseries = anomaly_cube.collapsed(['grid_longitude', 'grid_latitude'],
                            iris.analysis.MEAN, weights=grid_areas)

qplt.plot(timeseries)
plt.show()
```
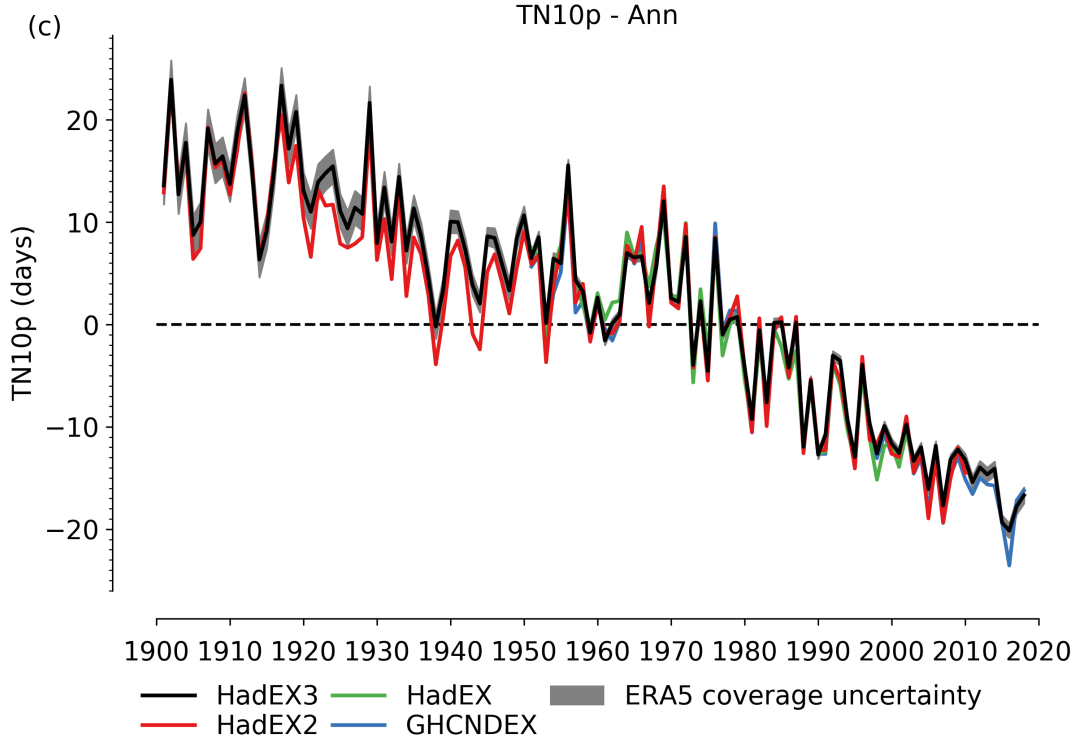
Figure 2: Quasi-global timeseries for TN10p (1961-90) comparing to HadEX2, HadEX and GHCNDEX. The uncertainties are those stemming from the incomplete spatial coverage and have been estimated from ERA5 [10].

5   Dataset Characteristics

5.1   Intercomparisons with other ETCCDI index datasets

At this moment in time we show timeseries plots on http://www.metoffice.gov.uk/hadobs/hadex3/ for each of the indices which compares HadEX.3.0.4 to its predecessors and GHCNDEX in terms of "quasi-"global land averages and spatial coverage. As an example we repeat in Figure 2 the global timeseries for TN10p ("cool nights") in comparison to other ETCCDI global datasets (HadEX2, HadEX and GHCN-DEX).

We have also performed an intercomparison with a number of recent reanalysis products [8], and included many of these in the paper on the sector-specific extension to HadEX3 [9]. For full details, please see these works, though we show an example in Fig. **??**.

5.2   Effects of a chosen reference period

As we note elsewhere in this Product User Guide (Sections 1.5.2) and also the HadEX3 paper [7] a number of indices use the data available over a reference period to determine threshold values which are then used to calculate exceedance counts. For example, the amount of rain which falls very wet days (R95p), the 95th percentile of the daily rainfall amount values is used as the threshold above which the accumulations are summed to obtain the value of the index.

In HadEX and HadEX2, this reference period was 1961-90, corresponding to the recommended WMO practice at the time. For HadEX3, we received some data which had already been converted into indices, using 1981-2010. In some cases this index period had been chosen as it was a better match to the period
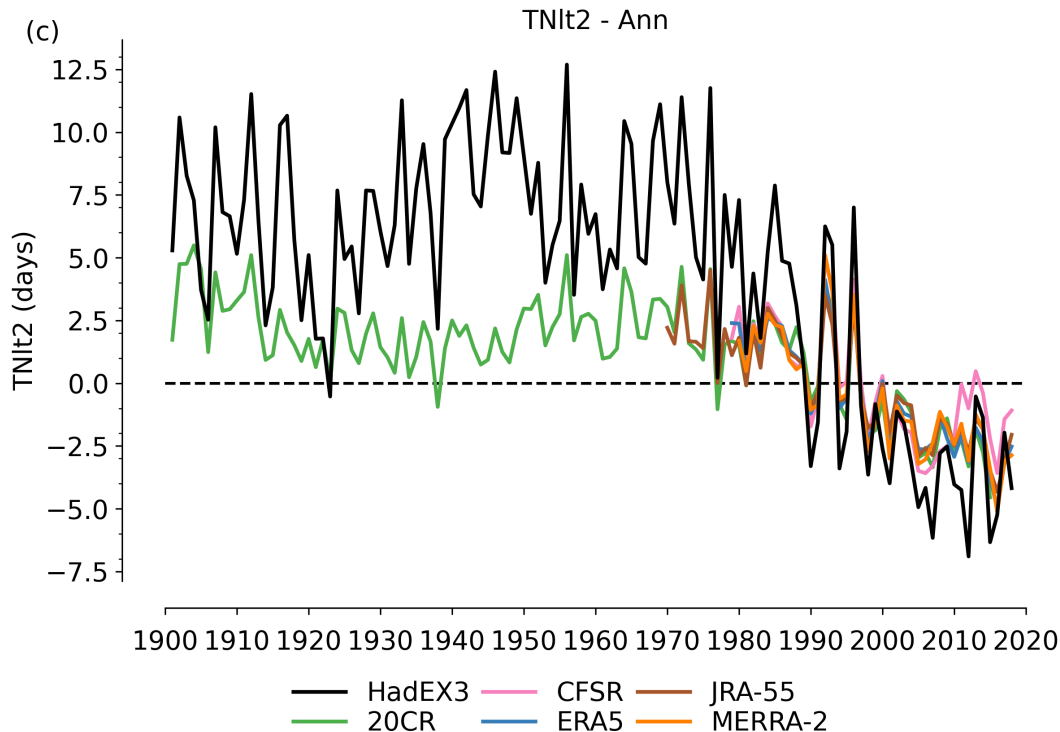
Figure 3: Quasi-global timeseries for TNlt2 (1981-2010) comparing to 20CR, CFSR, ERA5, JRA-55, and MERRA-2.

of record of the majority of the stations. It also matches well with the current recommended reference period from the WMO. However, comparing two versions when different reference periods have been used is non-trivial (see dicussion in the HadEX3 paper as well as e.g. [18]).

5.3 Expected climatological values in the reference period

The percentile indices (TX10p, R95p etc) use a reference period to determine the numerical thresholds set by the 10th or 95th percentile values. Therefore, it would seem reasonable at first glance to expect that the climatological average of the gridded fields would be identically equal to 10 or 5%. However, this is not the case; although we expect these values to be close to the theoretically expected quantities, in fact they may be different by a few percent either way.

Firstly, there are a number of methods available to calculate the percentile values for these quantities, interms of how the discrete distribution is handled to extract the threshold value, and so there will be some variation from that alone. The handling of any missing daily data during this reference period can be quite varied betweeen routines. We have tried to mitigate this to use a single routine where we have done the conversion from daily data to ETCCDI indices. But there are a number of codes available, and we do not know which was used when indices have been submitted for use in HadEX3. Also, specifically for the ETCCDI indices, our calculations include work o specifically avoid inhomogeneities for these indices which mean the calculations are more complex than the simplest version of extracting the percentiles [19].

The percentile thresholds determined for doing the counts above/below are determined from the daily data. However, the monthly and annual index values used in the dataset (on a station basis) are only available if this underlying data are sufficiently complete. If too many days within a specific calendar month or year are missing, then the index value cannot be accurately determined, even if the threshold

value (10th/90th percentile) has been calculated. Therefore when calculating the climatologies from the gridded fields, these missing months and years will affect the average value, reducing the likelihood of it equalling 10% exactly.

There could also be the impact of quality control algorithms. For the HadEX3 scripts, the QC processes flag entire stations, rather than individual years or months, so this should not have an effect, but any upstream QC performed on the indices may have flagged individual years or months and so further affect the completeness, and hence the climatological values.

Finally, these indices are measured in percent. However, for regions of the world which have comparatively low annual temperature variability, this makes the threshold exceedances very sensitive to any impact from the issues outlined above. And of course, the gridding process blends together stations, and so if e.g. missing years in one station affect the back-calculated climatology value, then this will be drawn through into each grid box this station contributes to.

5.4   Effects of station network changes

Despite our best efforts at selecting stations with long, contiguous records, some stations in HadEX3 will have missing years or months of data, start later or end earlier than their neighbours. In many cases this "flutter" in the station network will not have an effect as for lots of the indices, neighbouring stations are recording very similar weather, and hence moderate extremes. There are some indices where this may not be the case, and there are some more subtle effects from a varying network of stations contributing to a grid box.

The temperature indices providing annual (or monthly) actual temperature extremes (TXx, TXn, TNx & TNn) can be affected by a changing station network because the gridding process is blending and interpolating actual point measurements. It is clear that that the elevation influences the temperatures recorded at each station. Therefore if a mountain-top station drops out of the network at the end of the series or there is a missing year, then the weighted average value of all remaining stations will increase as a result, and therefore this grid box may appear to contain a temporal step-change.

The CDD duration index, which can reach large values in desert regions, and only 0.1mm of rainfall is enough to reset that index. Unsurprisingly, the station networks in desert regions are also sparse, and the correlation scales for this index relatively long (around 500 km or more). Therefore distant stations may contribute to a grid box, which could have very different values for CDD, resulting in relatively large changes if one drops out of the network.

During the construction of HadEX3, we attempt to mitigate the effect for a fluctuating station network with the station selection criteria, but in order to retain a reasonable spatio-temporal coverage, we do allow some short gaps when selecting the stations.

5.5   Utility of some indices in certain climates

Some of the indices, being summaries of the underlying daily observations, require careful interpretation, especially when gridded or expressed as trends. For example, in the case of indices using a fixed threshold (SU, TR, FD, ID, R10mm & 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.

Other indices (CDD and maybe CWD) may not be helpful in desert regions, where the number of consecutive dry days may stretch over years. For the reasons noted in the previous section, there may be inhomogeneities in these indices for these regions resulting from an interplay between the fluctuating station network and the gridding method chosen. We encourage users who wish to use these indices over those sorts of regions to think carefully and get in touch.

6 How was HadEX.3.0.4 produced?

At this moment in time, we refer you to the [7] and the code which is available at `DOI:10.5281/zenodo.3886492`. There are also some musings at the blog (https://hadex-extremes.blogspot.com/). Over time, we may add more information here.

6.1 Data sources and data collection

For a full list of the data sources, please see the paper [7]. In summary, we obtained data from publicly facing repositories (GHCND, ECAD & EOBS family, GHCNDEX etc.) as well as contributions from many co-authors. Some data arrived in the form of the daily data (precipitation, maximum and minimum temperatures) whereas others provided pre-calculated indices. In some cases we know which code was used to calculate these indices, but not always.

Unfortunately, the diverse set of sources and individual contacts we used mean that HadEX is not a regularly update dataset. For a product more suitable to monitoring, then use GHCNDEX.

6.2 Quality Control

For more details on the Quality Control, please see the paper [7] or the code. We note that not all indices have specific quality control tests applied - though most have at least some consistency checks.

6.3 Gridding method and Decorrelation Length Scale [DLS]

For more details on the gridding and DLS, please see the paper [7], as well as other papers on HadEX2, GHCNDEX and references therein.

Acknowledgements

References

[1] L. V. Alexander, X. Zhang, T. C. Peterson, J. Caesar, B. Gleason, A. M. G. Klein Tank, M. Haylock, D. Collins, B. Trewin, F. Rahimzadeh, A. Tagipour, K. Rupa Kumar, J. Revadekar, G. Griffiths, L. Vincent, D. B. Stephenson, J. Burn, E. Aguilar, M. Brunet, M. Taylor, M. New, P. Zhai, M. Rusticucci, and J. L. Vazquez-Aguirre. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, 111(D5), 2006.

[2] Russell Blackport, Michael Sigmond, and James A Screen. Models and observations agree on fewer and milder midlatitude cold extremes even over recent decades of rapid arctic warming. *Science Advances*, 10(40):eadp1346, 2024.

[3] M. G. Donat, L. V. Alexander, H. Yang, I. Durre, R. Vose, R. J. H. Dunn, K. M. Willett, E. Aguilar, M. Brunet, J. Caesar, B. Hewitson, C. Jack, A. M. G. Klein Tank, A. C. Kruger, J. Marengo, T. C. Peterson, M. Renom, C. Oria Rojas, M. Rusticucci, J. Salinger, A. S. Elrayah, S. S. Sekele, A. K. Srivastava, B. Trewin, C. Villarroel, L. A. Vincent, P. Zhai, X. Zhang, and S. Kitching. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The hadex2 dataset. *Journal of Geophysical Research: Atmospheres*, 118(5):2098–2118, 2013.

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

[5] Markus G Donat, Jana Sillmann, Simon Wild, Lisa V Alexander, Tanya Lippmann, and Francis W Zwiers. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, 27(13):5019–5035, 2014.

[6] R. J. H. Dunn, MG Donat, and LV Alexander. Investigating uncertainties in global gridded datasets of climate extremes. *Climate of the Past*, 10(6):2171–2199, 2014.

[7] Robert JH Dunn, Lisa V Alexander, Markus G Donat, Xuebin Zhang, Margot Bador, Nicholas Herold, Tanya Lippmann, Rob Allan, Enric Aguilar, Abdoul Aziz Barry, et al. Development of an updated global land in situ-based data set of temperature and precipitation extremes: Hadex3. *Journal of Geophysical Research: Atmospheres*, 125(16):e2019JD032263, 2020.

[8] Robert JH Dunn, Markus G Donat, and Lisa V Alexander. Comparing extremes indices in recent observational and reanalysis products. *Frontiers in Climate*, 4:989505, 2022.

[9] Robert JH Dunn, Nicholas Herold, Lisa V Alexander, Markus G Donat, Rob Allan, Margot Bador, Manola Brunet, Vincent Cheng, Wan Maisarah Wan Ibadullah, Muhammad Khairul Izzat Bin Ibrahim, et al. Observed global changes in sector-relevant climate extremes indicesan extension to hadex3. *Earth and Space Science*, 11(4):e2023EA003279, 2024.

[10] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2020.

[11] Thomas B McKee, Nolan J Doesken, John Kleist, et al. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, 22, pages 179–183. California, 1993.

[12] Matthew J Menne, Imke Durre, Russell S Vose, Byron E Gleason, and Tamara G Houston. An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29(7):897–910, 2012.

[13] John R Nairn and Robert JB Fawcett. The excess heat factor: a metric for heatwave intensity and its use in classifying heatwave severity. *International journal of environmental research and public health*, 12(1):227–253, 2015.

[14] Sarah E Perkins and Lisa V Alexander. On the measurement of heat waves. *Journal of climate*, 26(13):4500–4517, 2013.

[15] J Sillmann, VV Kharin, X Zhang, FW Zwiers, and D Bronaugh. Climate extremes indices in the cmip5 multimodel ensemble: Part 1. model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, 118(4):1716–1733, 2013.

[16] Sergio M Vicente-Serrano, Santiago Beguería, and Juan I López-Moreno. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate*, 23(7):1696–1718, 2010.

[17] World Meteorological Organisation. *Standardized Precipitation Index User Guide*. WMO, 2012.

[18] Yizhak Yosef, Enric Aguilar, and Pinhas Alpert. Is it possible to fit extreme climate change indices together seamlessly in the era of accelerated warming? *International Journal of Climatology*, 2022.

[19] Xuebin Zhang, Gabriele Hegerl, Francis W Zwiers, and Jesse Kenyon. Avoiding inhomogeneity in percentile-based indices of temperature extremes. *Journal of Climate*, 18(11):1641–1651, 2005.