

**Uncertainty of  
gridded extremes  
datasets**

R. J. H. Dunn et al.

This discussion paper is/has been under review for the journal Climate of the Past (CP).  
Please refer to the corresponding final paper in CP if available.

# Investigating uncertainties in global gridded datasets of climate extremes

R. J. H. Dunn<sup>1</sup>, M. G. Donat<sup>2</sup>, and L. V. Alexander<sup>2</sup>

<sup>1</sup>Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK

<sup>2</sup>ARC Centre of Excellence for Climate System Science and Climate Change Research Centre, UNSW, Sydney, NSW 2052, Australia

Received: 3 April 2014 – Accepted: 22 April 2014 – Published: 13 May 2014

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



## Abstract

We assess the effects of different methodological choices made during the construction of gridded datasets of climate extremes, focusing primarily on HadEX2. Using global timeseries of the indices and their coverage, as well as uncertainty maps, we show that the choices which have the greatest effect are those relating to the station network used or which drastically change the values for individual grid boxes. The latter are most affected by the number of stations required in or around a grid box and the gridding method used. Most parametric changes have a small impact, on global and on grid box scales, whereas structural changes to the methods or input station networks may have large effects. On grid box scales, trends in temperature indices are very robust to most choices, especially in areas which have high station density (e.g. North America, Europe and Asia). Precipitation trends, being less spatially coherent, can be more susceptible to methodological changes, but are still clear in regions of high station density. Regional trends from all indices derived from areas with few stations should be treated with care. On a global scale, the linear trends over 1951–2010 from almost all choices fall within the statistical range of trends from HadEX2. This demonstrates the robust nature of HadEX2 and related datasets to choices in the creation method.

## 1 Introduction

In recent years a number of gridded datasets of climate extremes indices have been released. The first global dataset to contain all 27 indices recommended by the World Meteorological Organization (WMO)/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) was HadEX (Alexander et al., 2006), which was based on work by Frich et al. (2002). HadEX covered the period 1951–2003 and was static i.e. not updated. So over recent years an international effort aimed to bring the dataset up to date. This work resulted in HadEX2 (Donat et al., 2013a), which substantially increases the time span covered to 1901–2010 and also has a greater surface coverage

CPD

10, 2105–2161, 2014

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



of the globe. A number of partner datasets also exist which follow a similar methodology. GHCNDEX (Donat et al., 2013b) uses only the Global Historical Climate Network (GHCN)-Daily dataset as input, rather than the mix of large collections of data and country-based inputs to HadEX2. GHCNDEX can also be used for a monitoring as it is

continually updated in near real-time. Both HadEX2 and GHCNDEX calculate the extremes indices for each station, and then average the stations to form a gridded dataset. To assess whether the order of these two processes causes a large change in the global trends, HadGHCNDEX was created from the gridded HadGHCND dataset of Caesar et al. (2006). This dataset gridded the daily maximum and minimum temperature observations in GHCN-Daily first, and then calculated the extremes indices on the daily grid box values. No precipitation grids were calculated in the creation of HadGHCND, and so none of the precipitation based indices are available in HadGNCDEX.

On the whole, for the temperature based indices, all three of these datasets agree, both for global averages and regional trends. The best agreement is between the percentile based indices (see Sect. 2 and Table 1) and for the trends of the absolute indices. Some systematic differences were found for the values of the absolute indices: those calculated from the daily gridded data (HadGHCND) were less extreme (lower maxima and higher minima, Donat et al., 2013b). However for the precipitation indices, the agreement between HadEX and GHCNDEX is less robust (Donat et al., 2013b), though the agreement improves when the datasets are masked to have the same coverage. This indicates that the spatial coverage of the datasets is important, especially for the precipitation indices. The precipitation indices have more complex spatial changes, and hence the global averages are more sensitive to changes in coverage than the temperature indices.

It is therefore important to assess whether global or regional results are unduly sensitive to any choices in these methods. An assessment of this nature probes the parametric and structural uncertainties of the dataset. The Intergovernmental Panel on Climate Change (IPCC) define the parametric uncertainty as that coming from choices of

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



parameters within the analysis scheme and the structural uncertainty as from choices of the scheme itself (Hartmann et al., 2013, Box 2.1). For a complex method, choices in the order of a calculation (e.g. GHCNDEX vs. HadGHCNDEX above) or values of a threshold parameter may have unexpected consequences for the final outcomes.

5 Recently, Sillmann et al. (2013) compared extremes indices calculated from state-of-the-art global climate models participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) to four reanalysis datasets; and Donat et al. (2014) compared the three in-situ datasets described above to five reanalysis products. For these types of studies it is important to know if areas of disagreement between the models, reanalyses and observational data can be reduced by more fully understanding the uncertainties associated with the observational datasets. Also, with the advent of coordinated efforts to provide climate services, policy and planning decisions are increasingly being made using insight from observational datasets combined with model analyses. Therefore it is also important to highlight where results from datasets are robust and, moreover, where they are not.

15 Here we will focus specifically on HadEX2, but as noted above, the results are applicable to all datasets which follow a similar calculation method. We first outline the HadEX2 methods relevant to this work in Sect. 2 and the effect of the completeness requirement used when plotting global average time series (Sect. 3). The individual methodological choices are presented in Sect. 4. We discuss our findings in Sect. 5 and Sect. 6 summarizes the study.

## 2 HadEX2 methods

In this section we will outline in particular, the methodological choices (parametric and structural) which will be assessed below. For a full description of the methods used to create HadEX2 see Alexander et al. (2006) and Donat et al. (2013a).

25 To perform the gridding of the station data, HadEX2 uses a modified form of Shepard's angular distance weighting (ADW) scheme (Shepard, 1968). It was initially cho-

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



sen when creating HadEX because it had been shown to be the most appropriate method for gridding irregularly spaced data (New et al., 2000). For each index, the correlation coefficients between all station pairs are calculated and are plotted against the distance between the stations. The correlation coefficients decay with distance, and averaging over 100 km bins, this decay is fitted with second order polynomial (see Fig. A1 of Alexander et al., 2006). It is assumed that the bin at zero distance has perfect correlation. Using this polynomial fit, the distance at which the correlation has fallen by a factor of  $1/e$  is obtained. This distance is the decorrelation length scale (DLS, also known as the correlation decay distance). Furthermore, as part of the ADW scheme, there is a weighting parameter,  $m$ , which determines the steepness of the decay with distance. In HadEX2 (and the other related datasets), this weighting function has been chosen to be  $m = 4$  (it is the effect of parametric choices like this which are investigated in the course of this study). A DLS is calculated for each index individually, and for five separate latitude bands: four  $30^\circ$  bands between  $30^\circ$  S and  $90^\circ$  N and one  $60^\circ$  band between  $30^\circ$  S and  $90^\circ$  S where there are few stations. The DLS values are then linearly interpolated to avoid discontinuities at the band boundaries.

A  $3.75^\circ \times 2.5^\circ$  grid is used for HadEX2. For each grid box centre, all the stations within a DLS are combined using ADW to obtain the value for the grid box in HadEX2. There have to be a minimum of three stations within a radius of one DLS for a grid box value to be calculated. This means that there exist grid boxes which themselves do not contain any stations, but are just in sufficient proximity to three stations to have a value assigned. These annual (and monthly in the case of some indices) grid box values make up the final dataset.

HadEX2 contains a total of 29 extremes indices, 17 temperature based and 12 precipitation based, which are defined in Table 1. Some indices are calculated in a similar way to others, and so fall naturally into categories – there are percentile based ones (TX10p, TX90p, TN10p, TN90p), block maxima/minima (TXx, TXn, TNx, TNn), duration based indices (CSDI, WSDI, CDD, CWD, GSL), ones based on fixed threshold values (SU, TR, FD, ID), precipitation frequency (R10mm, R20mm), precipitation

totals (PRCPTOT, R95p, R99p, Rx1day, Rx5day), precipitation intensity (R95pTOT, R99pTOT, SDII) and others which do not fit into any of the above categories (DTR, ETR). For this study we use the version of HadEX2 as of May 2013.

### 3 Grid box completeness

In the calculation of the global average, Donat et al. (2013a) use only those boxes which have at least 90 % of data during the period, i.e. 99 years over 1901–2010. This is to reduce the effect of varying coverage on the global average. However this adds a large restriction onto the areas of the globe which can contribute to the global average. This is not so much a methodological choice as a presentational one, and so we investigate the effect of this separately from the rest of the methods. We note that all of the global time series plots have been normalised over the period 1961–1990 as per Donat et al. (2013a).

As can be seen in Fig. 2 of Donat et al. (2013a), there is a steady climb in the number of grid boxes with data until around 1960, where the curve levels off, before beginning to fall again in the early 2000s. When the number of grid boxes is restricted to those which have 90 % completeness over the 110 year period of the dataset (99 years of valid data) the number of grid boxes is much reduced, but more constant over the period. For TX90p, the number of grid boxes with 90 % completeness remains steady around 700 for most of the period (see Fig. 1a). This number depends on the index used, e.g. CDD (Fig. 2).

When using a percentile based temperature index, e.g. TX90p, there is only a relatively small effect of the completeness requirement on the global trends (see Fig. 1b). The global average time series are highly correlated with HadEX2 for all completeness values, and only for very low completenesses in time (less than around 50 % of years present for a grid box to be included) are there somewhat larger deviations. And even these are only apparent in the most recent decades of the time series where the grid box numbers diverge by a large amount. The correlation coefficients remain above 0.9,

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



indicating that in this case the long term trend is still apparent, which can be clearly seen in Fig. 1b. In general, the warming estimate seems to be more conservative (i.e. less strong) if using a stricter completeness criterion.

However, when using another index, e.g. CDD, see Fig. 2, completeness requirements have clear systematic effects on the results. Although year-to-year variations are very similar, the global averages before 1960 smaller for low completenesses. By including extra-short term grid boxes, the period post 1960 has been biased upwards, which, combined with the normalisation period of 1961 to 1990, results in the apparent low values prior to this. The results for CDD and long term stations (see Sect. 4.4) indicate that stations in central and eastern Asia report mainly in the later period of the dataset and HadEX2 shows high values of CDD in western China (Xinjiang and the Tibetan Plateau), which supports this reasoning.

The correlation coefficients for both CDD and TX90p remain above 0.9 until the completeness drops below 60 % of years, and similar results are seen in the plots for the remaining indices (see Supplement). Therefore if studying the full time span of the HadEX2 dataset (1901–2010) then a completeness criterion of at least 60 per cent would be required. However, the higher the completeness threshold is set, then the more reliable the results will be, albeit for a smaller fraction of the globe. The recommendation for completeness masks when assessing time series of area-averages was also highlighted by Donat et al. (2013b).

However, although the effect of selecting only the grid boxes which have at least 90 % of years with non-missing data is large, we will compare the remaining methodological choices with this restriction in place rather than reducing the requirement down to 60 %. This means that the global average curves labeled as “HadEX2” are exactly the same as in Donat et al. (2013a). Plots showing how the number of grid boxes filled by each methodological choice will be shown *without* this restriction in place, to clearly demonstrate changes in this quantity more clearly. Any differences to this approach will be noted in the text and figure caption.

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 4 Uncertainty investigations

As HadEX2 contains 29 indices, in around 8 categories, and we are studying a range of methodological choices, we only show representative results, likely to be of greatest interest. In the majority of cases, the uncertainties will be very similar for indices in the same category. A full set of results is provided in the Supplement.

We will present the small, parametric changes first as these are likely to have the smallest effect. Then we will focus on larger changes to the source data or methods (structural changes). In all cases, the time series have been anomalised relative to the period 1961–1990.

### 4.1 Weighting function (parametric)

In the ADW scheme used in HadEX, a parameter,  $m$ , determines the steepness of the decay of the weighting function with distance (Eq. A2 in Alexander et al., 2006). The weighting function has been set to  $m = 4$  in both HadEX and HadEX2, as this provided a reasonable compromise between reducing the root mean square error ( $e_{\text{RMS}}$ ) between the gridded and station data and spatial smoothing (Donat et al., 2013a; Caesar et al., 2006). Donat et al. (2013a) note that the results are almost identical when using values between one and 10 for  $m$ .

We vary this weighting parameter from one to eight, and show the results in Fig. 3 for the R99p index. Indeed, the changes are very minor, and are almost imperceptible on the time series plot. For a large  $m$ , the decay of the weighting function is steep which would lead to stronger gradients locally. A small  $m$  results in a slower decay, weaker gradients and hence smoother variation across grid boxes. Hence changes in  $m$  may lead to small local differences, but as seen in Fig. 3, area-averaged continental scale results are almost identical for all values of  $m$ .

For all other indices, there are also only very small changes to the global time series, and the correlation coefficients rarely differ from  $r = 1$ . The coverage (grid boxes with non-missing values, not shown) is also identical to that of HadEX2.

CPD

10, 2105–2161, 2014

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## 4.2 Stations within a DLS (parametric)

In HadEX2, for a grid box to have a value, there have to be at least three stations within a radius of one DLS of the grid box centre. For the percentile based temperature indices this DLS can be in the order of 1000 km, but is much shorter for the precipitation based indices. We vary this parametric choice between one and nine stations within one DLS of the grid box centre. The resulting global average coverage and time series for TX10p are shown in Fig. 4. Again, the changes between versions are very minor, though larger than for the choices of the weighting parameter. The correlation between the different time series remains very high, but as the number of stations required within a grid box rises, so does the root-mean-square error ( $e_{\text{RMS}}$  in Fig. 4b) between each curve and HadEX2. The greatest difference is observed at early times if only one station is required within a DLS. There is also a change in the coverage, from a peak of almost 1600 grid boxes containing a single station or more within a DLS of the centre, to a peak of only 1400 for nine stations.

When masking all of the nine choices by the coverage of the version requiring nine stations within one DLS (the most restrictive choice), then all the global time series curves are identical. This means that the changes in the global time series curves seen in Fig. 4b are driven by the coverage and not by changes in the grid box values. Of course it is very likely that there are small changes in individual grid box values, but on a global average, these are not apparent.

To show how the choices affect results on a regional level we refer to the maps in Fig. 5a. The correlation coefficient ( $r$ ) of the local detrended time series for each of the nine possible choices with HadEX2 is calculated over the 1951–2010 period. The mean of these for each grid box is shown in the left panel in Fig. 5a. We use the correlation coefficient of the detrended time series in order to pick out the short timescale variability rather than any long term trend resulting from climate change that would dominate a non-detrended  $r$ . The linear trend used to detrend the data was determined using a median of pairwise slopes estimator (Sen, 1968; Lanzante, 1996),

CPD

10, 2105–2161, 2014

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



over the period 1951–2010. At least two thirds (40 of 60 years) have to have valid data for a trend to be calculated.

In some grid boxes only one of the possible versions results in a value, and these have been shaded in grey. They arise from the choice which only requires one station per DLS as this is the least restrictive. In the case of TX10p, the correlation is very high for most of North and South America, Europe, Asia and Australia and southern Africa, though in some cases not all of the choices result in a value for a specific grid box. The only regions which have low correlation values or few choices which fill the grid boxes are around central and Saharan Africa and the Amazonian region.

As the representation of long term trends is also important we show in Fig. 5b the standard deviation of the linear trends normalised by the mean of the trends ( $\sigma/\mu$ ) also calculated over 1951–2010. Linear trends for all methods were calculated using the median of pairwise slopes method as above, and from the resulting distribution of trends the standard deviation and mean were obtained. If the value of  $\sigma/\mu$  is small, then there is little variation in the value of the trends compared to the size of the mean trend, and so the trends are robust to the different choices. However, if the value of  $\sigma/\mu$  is large, then there are large variations in the trends, and so they are not robust. This is more likely if the value of the mean trend is small. On the Indian subcontinent and also in South America, although the mean detrended  $r$  were high, there is a higher variation in  $\sigma/\mu$  than compared to North America, Europe and Asia.

The confidence of the sign of a trend is also determined from the median of pairwise slopes method, by requiring that both the 5th and 95th percentiles of the slopes have the same sign. In this way we can be confident that a non-zero trend exists, but make no statement about its magnitude. Grid boxes where this is the case in HadEX2 are highlighted with a solid surround. This is different to the way trend significance was calculated in Donat et al. (2013a) who used the Mann–Kendall test.

For some of the fixed threshold indices (FD, ID, SU, TR) in regions of the globe where these thresholds are always or never exceeded this can lead to some very high correlation values and very low variances. However, it is important to note that in these

**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





actual variances are large because the trends are sensitive to the number of stations within a DLS. Despite this, there is high confidence in the trends for North America, Europe along with parts of southern Africa, Australia, India and eastern Asia.

There is no clear correlation between the location of grid boxes where there is high confidence in a non-zero trend in HadEX2 and those regions where the  $\sigma/\mu$  is small (e.g. South America, Central Asia). In fact some areas with high  $\sigma/\mu$  have high confidence in non-zero trends in HadEX2. Hence, even if there is strong evidence for a non-zero trend, further investigation would be required to be sure of its significance as it may well be extrapolated from (very) distant stations.

### 4.3 Stations within a grid box (parametric)

The ADW gridding method of HadEX2 does not require that a given grid box contains any stations within it, but merely that there are at least three stations within one DLS of the grid box centre (see Sect. 4.2). This means that there could be large numbers of grid boxes which are not close to any actual station, and only have values because of the interpolation performed by the ADW gridding method. We now require that there are stations within the grid box itself (ranging from one to five), with the HadEX2 method requiring zero. The weighting used to calculate the grid box value is the same as in the ADW scheme, but boxes are only filled if there are the requisite number of stations inside them. Requiring even just one station within the grid box has a drastic effect on the coverage (see Fig. 7a).

On the global time series, the requirement that a grid box contains stations has little effect on the overall trend across the entire period of the dataset (Fig. 7b). This is not really surprising as, although HadEX2 has a greater coverage through ADW interpolation, the number of stations that are included are the same. The extra coverage in HadEX2 is extrapolated from data from these stations and so it would be very surprising if there were drastic changes in the trends. This emphasizes the fact that although the coverage in HadEX2 is large, the dataset really only captures the changes in the indices in regions where there are stations, and not elsewhere.

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Although the long time scale trend for all the choices are similar, what is apparent is that some individual short-term spikes become more pronounced as the required number of stations required increases. If these features in the global average arise from small regions, then as the number of stations per grid box increases, there are fewer and fewer grid boxes which have valid values. Therefore these small regions can more easily dominate the global signal.

The uncertainty maps for this set of methodological choices are shown for CSDI in Fig. 8. There are many regions where only one choice gives a value, which are shown in grey, and these will be from the HadEX2 dataset, which does not require any stations to be present within a grid box. Grid boxes which are filled by five or six of the choices, have high mean correlations, whereas those which are only filled by two to four have low correlations.

Where only two of the choices result in grid box having a value, there must be one station in the grid box (the other choice being HadEX2). As this station dominates the value of the grid box, the timeseries can be noisy, and hence the correlation with HadEX2 can be low. If more stations are required, the time series will become less noisy as each station contributes less to the grid box average time series. At the same time, as stations within a grid box are in close proximity, they are likely to correlate well, and so the mean correlation between all the series will improve with an increase in the required number of stations, which is what is observed in Fig. 8a (boxes coloured green are very pale, blue less so, and red ones are the most intense).

For all indices the coverage reduces to North America, Europe and China (and also India for precipitation) when five or more stations are required, with only central Asia being a large area filled in when this restriction is relaxed down to one station (for temperature indices). This highlights the limitations in the available data, and the effect of the DLS in increasing the apparent coverage of HadEX2. For indices where the DLS is large (e.g. TN90p), then although most of the land surface is covered in HadEX2, only those areas listed above survive when the restriction on the number of stations is im-

posed. For indices which have a small DLS (e.g. Rx1day) the effect is less pronounced as the size of the DLS already limits valid grid boxes to those with stations.

#### 4.4 Long-term stations (parametric)

Most stations in HadEX2 do not report for the full 1901–2010 period. Stations dropping-in and -out could cause inhomogeneities and changes in the coverage which may feed through into the global average. As was seen in Sect. 3, selecting grid boxes which have 90 % completeness results in much smoother global average time series, than if all grid boxes are selected.

We therefore select stations which have reported for a long period of time and see how using these effects the coverage and time series. Stations can either be selected requiring that they report for greater than a given number of years or that they have a start date before a given year (and an end date after a different year, if desired), with the latter highlighting areas which only report during more recent times. The difference in which stations were selected has a knock-on effect on the DLS calculated. Hence the coverage can be higher than for HadEX2, especially in the early part of the series. We show the results from both selection methods. In the maps, trends and correlations have been calculated over the 1951–2010 period.

Stations were selected which reported for more than 40 to more than 80 years out of the total of 110, in 10 year increments. As can be seen for TXx in Fig. 9a, this has a large effect on the number of grid boxes available, but there are few differences in the global time series (Fig. 9b) and those that are found, are mainly in the early part of the dataset (which may be partly because the longest records are concentrated in limited regions). This stability is likely due to the grid-box completeness criterion (Sect. 3), though selecting only stations with very long records will not result in the same grid boxes contributing to global average time series as selecting grid boxes directly with long records. The choices with shorter record lengths (40 and 50 years) have higher correlations, lower  $e_{\text{RMS}}$  and more similar variances to HadEX2 than those choices with longer record lengths (70 and 80 years). This is because HadEX2 does not place

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



any restriction on the length of record a station must have for it to be included. But in most cases only long records are used because these dominate the data bases from which HadEX2 and GHCNDEX are drawn.

The uncertainty maps in Fig. 10 show clearly how different parts of the globe have different station record lengths. The regions where all six choices for the length of station record result in filled grid boxes are those where sufficient stations have very long record lengths, e.g. North America, Europe, Australia for all indices; and these have high average correlation values. Regions with 4–5 choices resulting in filled grid boxes, e.g. Eastern Russia, South America for the temperature indices; have shorter records on average, with China standing out as having relatively short records in the HadEX2 dataset. There is some correlation between those grid boxes which have confidence in non-zero trends and those with low  $\sigma/\mu$  values. However, this low  $\sigma/\mu$  does not tie with the number of choices which fill that particular grid box, i.e. highlighted boxes and/or dark shading in Fig. 10b are not always red.

When taking stations which report over a specific number of years we use the following five time periods: 1950–2000, 1940–2000, 1930–2000, 1920–2000 and 1910–2000. Again, the global time series look very similar (see Supplement) for all indices. The uncertainty maps also appear reasonably similar, but with some important differences. Firstly, China, India and parts of central Africa and South America appear in grey, indicating that only one of the choices (presumably the least restrictive one) results in a valid grid box in these locations. Also there is an area in eastern Russia, where all six choices result in a valid grid box values, but that in the station record length was only filled in four or five cases. This indicates that in this region there are some large gaps in the station records.

#### 4.5 DLS fitting methods (parametric)

The method by which the DLS is found in HadEX2 is described in Sect. 2. As the DLS is just an input to the gridding scheme, we have classed this as a parametric uncertainty. A polynomial expansion was used to fit the decay curve and obtain the DLS. Using

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a polynomial is an approximation to an exponential in this case. Therefore a number of different curves were fitted to the binned correlation values to obtain the DLS. By taking the logarithm of the correlations ( $y$  values) a linear function was fitted, which was then converted back to the exponential form (labeled “log\_lin” in Fig. 11). This has the advantage of fitting a straight line, but it places more weight on the correlations at larger distances. A true exponential was fitted, along with one which allowed for a non-zero offset (i.e.  $ae^{bx} + c$ )<sup>1</sup>. All four of these fitting methods are shown in Fig. 11 for the CDD index and the latitude band running from 60° N to 90° N. The DLS is the location where the curve has dropped to  $1/e$  of its value at zero separation.

In most cases, there are rarely large differences between the DLSs calculated using the different methods. For most indices and latitude bands, the exponential methods (with or without offset) result in the closest fit to the data. The “log\_lin” method is worst at capturing the curvature at small separations, especially for precipitation indices where the correlations drop very rapidly over the first few bins. The polynomial approximation is in many cases a good approximation to use, and does not give wildly inaccurate results. The exponential with an offset is the most physically reasonable model to use and in Fig. 11 results in the best fit.

In some cases the decline in the correlation values follows a smooth decline, but in others (and this can be seen in Fig. 11) there are bumps and wiggles as the correlations fall off. There can be quite a range in the DLS obtained, as shown in Fig. 11 with a range of 500 km to 1000 km (we have chosen this index and latitude range to have a large range in DLS values). When the values of the DLS are small, this can make a large difference to which stations contribute to a grid box value. For indices which have a large DLS (most temperature indices), then the change in the DLS value will have a more limited impact on the grid box value.

In the global time series, there are differences between HadEX2 and the alternative fitting methods (Fig. 12). However, the small scale peaks and troughs in the global

<sup>1</sup>The fitting of all methods bar the Polynomial one failed for the latitude band 30° S–0° for ID and FD. This is because no stations measured any ice or frost days in this latitude band.



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



timeseries occur at the same times: only the level differs. Consequently, the correlations between HadEX2 and the exponential model versions are very high (0.97 and 0.98). As was mentioned in the discussion about the number of stations within a grid box (Sect. 4.3), the stations present in HadEX2 remain the same, and as the DLS increases, then the extra coverage is extrapolated from the same data. It would therefore be surprising if there were large changes in the behaviour of the global average time series. A similar result is found for all other indices, with correlations between global averages in many cases being above 0.9. The log-lin model usually has the lowest correlations with HadEX2.

Geographical differences, as shown in Fig. 13, will depend on whether the DLSs determined using the four different methods are similar or very different. We have chosen to show an index (CDD) where the DLS changes by a large amount, especially in the high latitude regions (Fig. 11). Focusing on the Canadian Arctic in Fig. 13 a gradation can be seen, from the areas where all four methods result in filled grid boxes to those where only one method does (Greenland). Where all four methods result in a value are areas which likely have a high station density, and the further from these regions the grid box is, the fewer of the methods result in a value for that box. Also, although the  $\sigma/\mu$  may be small, the correlations are also reduced outside of the areas with high station density. CDD does not have particularly strong trends in these mid-high latitude regions (see Fig. 8 in Donat et al., 2013a). So as the DLS decreases, local variations become more prominent as dense station networks in the vicinity are no longer able to smooth them out. This results in local differences between the time series, and a small correlation coefficient in these high latitude regions (northern Canada and Russia).

Grid boxes in which there is some confidence that a trend is non-zero are predominantly found in areas where all four of the DLS calculations result in a grid box value. However, the converse is not true, especially for precipitation indices. For indices which have large DLS values, there are few areas where one or more methods exclude the grid box (see for example TN10p in the Supplement). The changes in coverage are

affected more by the correlation decay for a given index than by changes in the fitting method used.

## 4.6 Gridding methods (structural)

The gridding method used in HadEX2 is an adapted version of the angular distance weighting scheme (Shepard, 1968). This accounts for the angular distribution of the stations as well as their distance from the grid box centre. It also interpolates into empty grid boxes which are close to stations.

Three alternative gridding methods are outlined below; the Climate Anomaly Method (CAM, Jones, 1994), the Reference Station Method (RSM, Hansen and Lebedeff, 1987) and the First Difference Method (FDM, Peterson et al., 1998).

### 4.6.1 Climate anomaly method

The CAM has a long history of being used in other gridded datasets, for example the HadCRUT series of surface temperature anomalies (see e.g. Jones, 1994; Morice et al., 2012). Anomalies are calculated from a common reference period and then these anomalies are combined. Usually a 30 year reference period is used, with either some requirements on the number of years present within that period or using for example neighbouring stations to estimate the effect of the missing data on the full period value.

In this analysis, the climatological reference period has been chosen to be 1961–1990 to match those used for the calculation of the percentile based temperature indices and the normalisation of all other global average time series. At least 25 out of the 30 years had to have valid data for a climatology to be calculated. A simple mean across all anomalised stations in the grid box resulted in each annual value.

By the nature of this method, for a grid box to have a value, there must be at least one station present within it (with sufficient data in the period 1961–1990). Of course if there is only one station present, this has been assumed to be representative of the entire grid box. As this may not be a valid assumption, should this method be used, it

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



may be prudent to require at least two or three stations per grid box, but this will result in a further decrease in the coverage (see Sect. 4.3).

#### 4.6.2 Reference station method

The RSM described by Hansen and Lebedeff (1987) starts by selecting the station with the longest record within the area of influence of the grid box centre. Hansen and Lebedeff (1987) used a fixed distance of 1200 km, which was derived from the decay of the correlation between stations. In this study we use the DLS appropriate for the latitude of the grid box as in the ADW for HadEX2. Having selected the station with the longest record within one DLS of the grid box centre, successively shorter stations are processed. The new stations' temperature records are adjusted so that their mean over the common period is the same as the composite of all stations that have been processed so far. Then, distance-weighted averages are calculated to obtain the new composite station. This process is repeated until all stations within one DLS of the grid box centre have been included into the composite station.

The advantages of this method are that a common reference period is not required. However, stations are still required to have at least 20 years of overlap with the composite station, and so stations with very short records are still not included. However, stations which report in early times, but not during a specified reference period can be accommodated by this method. Peterson et al. (1998) point out that this method relies on the reference station (the one with the longest record) providing an accurate representation of climatic changes, uncorrupted by biases arising from station moves, instrument changes or other changes in observing procedures.

If there are non-climatic inhomogeneities in the reference station, these will feed through into the final grid box value. In the creation of HadEX2, quality control and homogeneity checks varied from country to country, but in most cases data had been carefully checked by researchers from the country of origin (Donat et al., 2013a). We therefore presume that issues with the quality and inhomogeneity are minor.

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



### 4.6.3 First difference method

The FDM was proposed by Peterson et al. (1998) as an alternative to the previous two methods, as it did not require either a common reference period, or a reference station, and therefore could use all available data. This method does, however, suffer if there are frequent gaps in the data, and is sensitive to outliers if they occur at the beginning or end of the time series.

In this method, the annual values are converted to a series of first differences, with the value for the first year of the series being zero. We average the first differences from different stations within the grid box to obtain the grid box value. The final step is to reconstruct the time series using a cumulative sum resulting in an anomalised series.

Using a simple first difference is probably most suited to temperature indices, whereas something more complex, e.g. a ratio, may be more appropriate for precipitation. For simplicity, we used the simple first difference for all indices, and did not find any blatantly spurious results. Years with missing data within the time series were filled with the average first difference for that station, except years before the start and after the end of the station's reporting period.

### 4.6.4 Gridding methods results

The results from the four gridding methods for PRCPTOT can be seen in Fig. 14. The coverage of the RSM, which uses the DLS to find stations within a region to merge together, is similar to that of the ADW method, and in early times, is actually larger. For the CAM and the FDM methods, the coverage is smaller, as these methods require that stations be present within a grid box. The CAM is more restrictive on which stations it is able to include when calculating grid box average values, and so has the lower coverage of the two.

The correspondence between the different gridding methods on the time series is relatively good. Although the sizes of the short timescale variations do not always match,

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



they occur at the same time and in the same direction (i.e. local year-to-year differences have the same sign, if not the same magnitude). Sometimes the magnitudes of these short timescale variations are larger than in HadEX2. The RSM has the highest correlation coefficients with HadEX2, owing to the extrapolation, smoothing and coverage that this method has in common with the ADW scheme of HadEX2.

Long time scale variations match those from HadEX2 reasonably well, especially later in the period. Large differences are apparent before around 1950, leading to an increase in fitted linear trends (FDM) or decrease (CAM).

To illustrate the regional influences of the gridding methods, we show the corresponding uncertainty maps for PRCPTOT in Fig. 15. Only a few regions have high detrended- $r$  values, on the whole corresponding to those regions with dense station networks. Many grid boxes in high latitudes, South America, Africa, Australia and parts of Asia are only filled by two of the methods (likely to be RSM and ADW). In these regions, as the grid boxes have been interpolated from surrounding stations, the differences in the two methods have resulted in differences in the short time scale variations, and hence low  $r$  values. In regions where the station density is high, and all four methods fill the boxes, the grid box average is driven by stations within the box and so short term variations match between the four methods more often, resulting in higher  $r$  values.

The values of  $\sigma/\mu$  indicate that there is often a large spread in the values of the trend compared to the mean trend. Indices which have strong trends (usually temperature based ones) have smaller relative spreads than those which have weak trends (primarily precipitation indices). However, some grid boxes where only two methods fill the box have a very small spread in the trends compared to regions where all four methods fill the box. These areas are likely to be only filled by the RSM and ADW methods, which both interpolate using neighbouring stations and so have similar sets of stations combining to form a grid box value, resulting in similar trends. When all four methods fill a grid box, then there is a greater range in methodological choices, and so a likely greater range in the trend magnitudes and a larger relative spread.

## 4.7 Jackknife experiment

To assess the effect of the station network on the global and regional trends we perform a jackknife experiment. Jackknifing is a technique that is very similar to bootstrapping. In both cases the parent population (in this case, the stations) is sub-sampled, in the case of bootstrapping, with replacement, and for jackknifing, without. To sample the effect of fewer stations within the network, 100 iterations each were run using random selections of 25, 50 and 75 % of the total station number, for each index. These iterations recalculated the DLS and gridding for each run. Of course, with most stations being found in North America, Europe and East Asia, even random selections will have, on average, similar distributions and coverages to those using the full network.

As can be seen in Fig. 16a for the ETR index, the grid box coverage is unsurprisingly lower for the runs with only 25 per cent of stations, compared to those with 50 or 75 % or the complete set of HadEX2 stations. The scatter in the coverage also increases as the number of stations decreases, as would be expected. It is possible in some cases for the coverage to be better for the 25 per cent runs, through a larger DLS being calculated for these station networks. From the global average time series, it is not clear if a single jackknife run is biased to high or low values. However, most short-timescale features are picked out in most case, especially post 1920 in the example of ETR shown.

In the global average time series, the three different jackknife run-sets form a band around the HadEX2 series (Fig. 16b), but at individual short-timescale peaks and troughs are more extreme in the 25 % runs, especially at early times. The width of the band also increases towards the start of the dataset, which corresponds to the decrease in coverage observed before around 1950.

Some other indices show a similar close relationship with HadEX2. The temperature percentile indices show very little variation. However the precipitation indices are more variable: see Fig. 17. The runs with 75% of stations capture the year-to-year peaks and troughs, whereas those with 25 % of stations are much noisier.

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Most of the Northern Hemisphere grid boxes are filled by over two thirds of the 25 % runs in both the temperature (ETR) and precipitation (PRCPTOT) indices as shown in Fig. 18. This indicates that in the regions of high station density grid boxes are almost always filled by the HadEX2 creation process even with a much reduced station density.

In the Southern Hemisphere, however, there are large areas where less than one third of the runs fill the grid boxes. Only parts of South America and Australia are filled by most runs in both indices.

Although the grid boxes are filled by most of the runs, the range in trends is not uniform. In North America, the ETR shows a consistent small range in the trends, but parts of central Europe and central Asia exhibit a large variance in the trends. Similar patterns are seen in PRCPTOT, where even western and northern North America have a large variance in the trends.

## 5 Discussion

### 5.1 Taylor Diagrams

In order to make the different methodological choices easier to interpret we use a presentation method common in climate model validation analyses, the Taylor Diagram (Taylor, 2001). An example for R10mm is shown in Fig. 19, and one from each of the categories of indices outlined in Sect. 2 are shown in the Appendix.

The x- and y-axes are the standard deviation of the time series, with the reference dataset (HadEX2 in this case) being plotted on the x-axis. The polar axis represents the correlation between the two data sets ranging from zero at 0° to one at 90°. The advantage of this diagram is that it also shows the root-mean-square (RMS) error, shown by the grey semi-circles centered on the reference dataset in Fig. 19. In this way all of the time series shown in the previous sections can be compared to the reference series calculated from HadEX2.

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



This diagram allows the comparison of both the long-term and short-term variation between the parametric and structural choices and is used in the discussion below. In those indices with a strong long-term trend, the correlation coefficient will give a measure of how well this trend is captured, but the magnitude of any deviations around this trend will be represented by the standard deviation. We also show the effect of the completeness criterion, as discussed in Sect. 3, on the diagrams, and this also can have some large effects for some indices.

### 5.2 Parametric Uncertainties

The smallest effects on global trends and variances were observed when investigating the parametric uncertainties. The weighting (scale) parameter of the ADW scheme had almost zero effect. Selecting only long-term stations has an impact via the station coverage, but this is muted because HadEX2 stipulates that  $\gtrsim 90\%$  temporal completeness for each grid box.

Relaxing the criterion for at least three stations within a DLS allows values for each index to be calculated for more land-surface grid boxes, so apparently increasing the coverage. This is unlikely to change any of the results in areas with a high station network density, but decreases the correlation with HadEX2 for the global average. Conversely, increasing the number of stations required within a DLS reduces the coverage to just those areas which have a high station density and also reduced the correlation with HadEX2 for the global average. It is, therefore, through those regions which have few stations that this choice has the greatest effect on the global average.

When the minimum number of stations per grid box is increases, the coverage decreases but the global average trends do not change by much, because the global average is still dominated by data from North America, Europe and Asia, i.e. those areas with high station density. However, the reduced coverage allows the short term changes to differ from HadEX2, reducing the correlation.

Although the DLS fitting method used in HadEX2 is relatively simple, a second order Taylor expansion, it proves to be effective in capturing the geographical coherence of



the data, especially for some of the temperature indices which have very large DLS. The precipitation indices have a smaller DLS and so the analysis may be more sensitive to the fitting methods. However the behaviour of the indices is not coherent on the grid box scale, and so the effects of the DLS fitting method on the global time series are not marked.

### 5.3 Structural uncertainties

The structural uncertainties often had a larger impact than most of the parametric changes. By changing the gridding scheme, the weighting given to different stations changes. This has one of the larger effects on the global trends, and more at a regional level. Two types of scheme were tested, those which interpolated (RSM and ADW) and those which did not (FDM and CAM). Grid boxes mainly had values from only two interpolation methods or from all four methods. The spread of linear trends is likely to have been undersampled for grid boxes where only two methods resulted in values. Hence the trends in these regions have a small spread, but have a larger spread in the areas where all four methods fill a grid box. Also, the similarity between the two pairs of methods resulted in higher correlations and lower normalised variances when only two methods filled a grid box than when all four methods did.

Regions which had a high station density were less susceptible to the changes brought about by the different gridding methods. In these regions, the grid box value depends more on stations within the grid box than those outside. Therefore the effect of interpolation and weighting is less.

In many indices, changing the input station network had the largest effect on the global timeseries. When using only 25% of the stations, the coverage could be larger than when using the full station list because of changes in the DLS. However the small set of stations resulted in large deviations, especially at early times when the stations contributing to the global average are further reduced. On a regional level, however, areas of high station density still have consistent and robust trends for some of the precipitation indices (see Fig. 18).

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 5.4 Combined uncertainties

The extremes indices assessed in this study are regularly used for the monitoring of changes in the occurrence of extremes across the globe. Therefore the overall uncertainties are important to ensure that the reliability of the trends can be assessed. To summarize the results from each of the parametric and structural uncertainties over all the indices we give the linear trends in Table 1. As well as the index definition, we show the linear trend from HadEX2 calculated over 1951–2010 using the median of pairwise slopes estimator. We also show the statistical range in slopes as the 5th to 95th percentile. For each of the methodological choices we show the range in linear trend calculated for each choice, as well as how many of the trends fall within the statistical range of HadEX2. This can only give a global overview. For most of the percentile and block maxima temperature indices, almost all the choices fall within the statistical trend of HadEX2, as do some of the duration based ones (CSDI, WSDI). GSL and ID have the worst agreement, but this is only apparent in a few of the choices. The precipitation indices are on the whole only slightly less robust than the temperature based ones, and this may be in part due to their relatively large trend uncertainties in HadEX2. The heavy precipitation totals and duration indices appear to be the least robust. However, in all indices, for most choices more than half the global trends fall within the statistical range of HadEX2.

The comparison between the gridding scheme results, which will change the values of individual grid boxes, and the minimum number of stations within a grid box, which will change the coverage demonstrates the two main sources of uncertainty within HadEX2 and its related datasets. Most of the stations in HadEX2 and its related datasets are found in North America, Europe, Asia and Australia. Methodological choices which affect coverage generally have small effects on the global averages because these are dominated by those well-observed areas. Also, there are only small effects on grid box values in those areas, as the correlations are high and the spread in trends low (Fig. 8). Changing the gridding method changes the coverage, but also

CPD

10, 2105–2161, 2014

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



affects how the stations are blended together, changing values of the grid boxes. This results in lower correlations and a greater spread in linear trends for individual grid boxes (Fig. 15). These methodological choices tend to have the largest effects as assessed on a global average on the Taylor diagrams.

5 In this analysis we have assessed a number of methodological choices taken during the creation of the HadEX2 dataset to assess the sensitivity of global and regional trends to these parameters. There are now a range of datasets available which follow the HadEX2 method, as mentioned in Sect. 1, and all of these are very likely to have similar sensitivities to the choices assessed here. This family of datasets do, however, probe the effect of completely independent station networks, in a way that the jackknife experiments run in this analysis (Sect. 4.7) cannot do (see Donat et al., 10 2013b). As expected many of the parametric uncertainties have the smallest effect on the results, whereas the structural ones have the largest effect. Methodological choices which drastically change the grid box value or the coverage are those which have the greatest effects: gridding method, station network (jackknife), stations within a grid box.

15 In the course of this study we have not been able to assess all sources of uncertainty. The quality of the station data has not been investigated. Although quality control procedures have been applied by National Meteorological Services and also at regional workshops, these are unlikely to be consistent for all stations over the entire span of the data set. When taking global averages, we have not investigated the effect of the averaging algorithm, nor the latitude weighting. And we have also only used one method for obtaining linear trends (median of pairwise slopes), and in some cases a linear trend is unlikely to be the most applicable. These unassessed sources of uncertainty will also have an impact on conclusions drawn from the datasets.

25 The indices which have strong global trends in HadEX2 continue to have strong global trends under all of the model choices assessed here. On the whole, these are the temperature based indices. Regionally, the areas with a high station density are also more robust to the different methodological choices, with high correlations between the different choices and small variances in the trends for each grid box. Those areas

---

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

---

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

which have a lower station density are more susceptible to local changes in the trends and short timescale behaviour arising from the effect of the methodological choices. Users should therefore be cautious when using these datasets for small regions and be aware of the coverage of the data when doing assessments.

5 The main limitation to the dataset and its cousins is the availability of the data. All the different choices and experiments run in this assessment take larger or smaller fractions of the available station data and process it in similar (albeit slightly different) ways. None of the methods are able to use more data than are available, especially in  
10 parts of the world which are currently under sampled. Therefore the global time series for indices which have a strong trend are very similar, as all grid boxes are interpolated from the same parent data. For indices which have no strong trend, or are inherently more variable, the changes in methods rarely introduce a strong trend or a drastic change in the variability either.

## 6 Summary

15 We have assessed the affects of a number of methodological choices, both parametric and structural, which were made during the creation of the HadEX2 dataset of gridded extremes indices. The largest effects on a global average timeseries come from those methodological choices which make large changes to the final spatial coverage of the dataset or to the grid box values themselves. The main choices which result  
20 in these kinds of changes are changes in the station network (jackknife experiments), the gridding method used and the requirement to have a certain number of stations within the grid box or DLS. When comparing the global average timeseries with those from HadEX2 using a Taylor Diagram, these choices have the lowest correlations and the largest difference in standard deviation. Trends and variances are most robust in  
25 North America, Europe and Asia as well as the southern tip of Africa and eastern Australia. These are also the areas which have the highest station network density. High latitude regions and the majority of South America and Africa have few or no areas of

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



agreement for all indices, as these have low station densities and so are more susceptible to changes in coverage and local grid box values. Temperature indices appear to be more coherent and resistant to changes in the methods than precipitation indices, but this is in part due to the small magnitude of global trends in most of the precipitation indices. In regions with high station density, and for indices which have a clear non-zero trend over 1950–2010, the linear trends from almost all choices fall within the statistical range of trends from HadEX2. This demonstrates the robust nature of HadEX2 and related datasets to choices in the creation method.

## Appendix A

We show the Taylor diagrams for one index in each of the classes outlined in Sect. 4.

**Supplementary material related to this article is available online at <http://www.clim-past-discuss.net/10/2105/2014/cpd-10-2105-2014-supplement.zip>.**

*Acknowledgements.* RJHD was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) and thanks the Australian Research Council's Centre of Excellence for Climate System Science (UNSW, Sydney) for support and hospitality during the final stages of this project. MGD and LVA were supported by Australian Research Council grants CE110001028 and LP100200690. We thank David Parker for suggestions which improved the final manuscript. RJHD thanks Gareth Jones for suggesting the use of a Taylor Diagram and Colin Morice for useful discussions about gridding methods.

The works published in this journal are distributed under the Creative Commons Attribution 3.0 License. This license does not affect the Crown copyright work, which is re-usable under

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the Open Government Licence (OGL). The Creative Commons Attribution 3.0 License and the OGL are interoperable and do not conflict with, reduce or limit each other. © Crown copyright 2014.

## References

- 5 Alexander, L., Zhang, S. X., Peterson, T., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzade, H. F., 2106, 2108, 2109, 2112  
Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and Vazquez-Aguirre, J. L.: Global observed changes in daily climate extremes of temperature and precipi-  
10 tation, *J. Geophys. Res.*, 111, D05109, doi:10.1029/2005JD006290, 2006.
- Caesar, J., Alexander, L., and Vose, R.: Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded data set, *J. Geophys. Res.*, 111, D05101, doi:10.1029/2005JD006280, 2006. 2107, 2112
- 15 Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A. M. G., Kruger, A. C., Marengo, J., Peterson, T. C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Sanhoury Elrayah, A., Sekele, S. S., Srivastava, A. K., Trewin, B., Villarroya, C., Vincent, L. A., Zhai, P., Zhang, X., and Kitching, S.: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset, *J. Geophys. Res.-Atmos.*, 118, 2098–2118, doi:10.1002/jgrd.50150, 2013a. 2106, 2108, 2110, 2111, 2112, 2114, 2121, 2123
- 20 Donat, M. G., Alexander, L. V., Yang, H., Durre, I., Vose, R., and Caesar, J.: Global land-based datasets for monitoring climatic extremes, *B. Am. Meteorol. Soc.*, 94, 997–1006, doi:10.1175/BAMS-D-12-00109.1, 2013b. 2107, 2111, 2131
- 25 Donat, M. G., Sillmann, J., Wild, S., Alexander, L. V., Lippmann, T., and Zwiers, F. W.: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis data sets, *J. Climate*, in press, doi:10.1175/JCLI-D-13-00405.1, 2014. 2108

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

- Frich, P., Alexander, L., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A., and Peterson, T.: Observed coherent changes in climatic extremes during the second half of the twentieth century, *Clim. Res.*, 19, 193–212, 2002. 2106
- Hansen, J. and Lebedeff, S.: Global trends of measured surface air temperature, *J. Geophys. Res.-Atmos.*, 92, 13345–13372, 1987. 2122, 2123
- Hartmann, D. L., Tank, A. M. G. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P. M.: Observations: atmosphere and surface, in: *Climate Change 2013: The Physical Science Basis, Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013. 2108
- Jones, P. D.: Hemispheric surface air temperature variations: a reanalysis and an update to 1993, *J. Climate*, 7, 1794–1802, 1994. 2122
- Lanzante, J. R.: Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, 16, 1197–1226, 1996. 2113, 2137
- Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set, *J. Geophys. Res.-Atmos.*, 117, D08101, doi:10.1029/2011JD017187, 2012. 2122
- New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability, Part II: Development of 1901–1996 monthly grids of terrestrial surface climate, *J. Climate*, 13, 2217–2238, 2000. 2109
- Peterson, T. C., Karl, T. R., Jamason, P. F., Knight, R., and Easterling, D. R.: First difference method: maximizing station density for the calculation of long-term global temperature change, *J. Geophys. Res.-Atmos.*, 103, 25967–25974, 1998. 2122, 2123, 2124
- Sen, P. K.: Estimates of the regression coefficient based on Kendall's Tau, *J. Am. Stat. Assoc.*, 63, 1379–1389, 1968. 2113, 2137
- Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data, in: *Proceedings of the 1968 23rd ACM National Conference*, 27–29 August, ACM, 517–524, 1968. 2108, 2122

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, *J. Geophys. Res.-Atmos.*, 118, 1716–1733, doi:10.1002/jgrd.50203, 2013. 2108

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.-Atmos.*, 106, 7183–7192, 2001. 2127

5

## CPD

10, 2105–2161, 2014

### Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)





## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

**Table 1.** The definition of all the indices and a summary of the uncertainties assessed in this work. For HadEX2, the linear trends and their uncertainties (5th to 95th percentile range) have been calculated over 1951–2010 using the Median of Pairwise Slopes estimator Sen (1968); Lanzante (1996) and are per decade. For the six different choices investigated in this study, the range obtained is shown as well as the number of choices which fall within the statistical range of the HadEX2 slopes.

| Index       | Name                          | Definition   | Unit | Range of Linear Trend of Global Average Timeseries |                        |                          |                        |                        |                        |                        |
|-------------|-------------------------------|--|------|--|------------------------|--------------------------|------------------------|------------------------|------------------------|------------------------|
|             |                               |  |      | HadEX2   | Weighting              | Stations/DLS             | Stations/Grid Box      | Long stations          | DLS Methods            | Gridding               |
| Temperature |                               |  |      |  |                        |                          |                        |                        |                        |                        |
| TXx         | Max $T_{max}$                 | Warmest daily maximum temperature  | °C   | 0.11<br>(0.04 → 0.17)                              | 0.11 → 0.11<br>(8/8)   | 0.10 → 0.22<br>(9/10)    | 0.03 → 0.11<br>(5/6)   | 0.10 → 0.11<br>(6/6)   | 0.08 → 0.11<br>(4/4)   | 0.08 → 0.12<br>(4/4)   |
| TXn         | Min $T_{max}$                 | Coldest daily maximum temperature  | °C   | 0.33<br>(0.21 → 0.45)                              | 0.33 → 0.33<br>(8/8)   | 0.29 → 0.39<br>(10/10)   | 0.13 → 0.33<br>(3/6)   | 0.29 → 0.38<br>(6/6)   | 0.27 → 0.33<br>(4/4)   | 0.24 → 0.35<br>(4/4)   |
| TNx         | Max $T_{min}$                 | Warmest daily minimum temperature  | °C   | 0.17<br>(0.11 → 0.23)                              | 0.17 → 0.17<br>(8/8)   | 0.17 → 0.27<br>(9/10)    | 0.11 → 0.17<br>(6/6)   | 0.16 → 0.18<br>(6/6)   | 0.12 → 0.18<br>(4/4)   | 0.17 → 0.23<br>(3/4)   |
| TNn         | Min $T_{min}$                 | Coldest daily minimum temperature  | °C   | 0.43<br>(0.32 → 0.53)                              | 0.42 → 0.43<br>(8/8)   | 0.33 → 0.48<br>(10/10)   | 0.37 → 0.43<br>(6/6)   | 0.40 → 0.47<br>(6/6)   | 0.29 → 0.43<br>(3/4)   | 0.38 → 0.50<br>(4/4)   |
| DTR         | Diurnal temperature range     | Mean difference between daily maximum and daily minimum temperature  | °C   | -0.07<br>(-0.09 → -0.06)                           | -0.07 → -0.07<br>(8/8) | -0.07 → -0.06<br>(9/10)  | -0.09 → -0.07<br>(6/6) | -0.09 → -0.05<br>(5/6) | -0.07 → -0.06<br>(4/4) | -0.13 → -0.07<br>(1/4) |
| ETR         | Extreme temperature range     | Difference between monthly maximum and minimum temperature   | °C   | -0.12<br>(-0.20 → 0.04)                            | -0.13 → -0.12<br>(8/8) | -0.13 → 0.05<br>(9/10)   | -0.27 → -0.12<br>(2/6) | -0.20 → -0.12<br>(5/6) | -0.14 → 0.09<br>(4/4)  | -0.30 → -0.12<br>(2/4) |
| GSL         | Growing season length         | Annual number of days between the first occurrence of 6 consecutive days with $T_{mean} > 5^{\circ}\text{C}$ and the first occurrence of 6 consecutive days with $T_{mean} < 5^{\circ}\text{C}$ . For the Northern (Southern) Hemisphere this is calculated between 1 January and 31 December (1 July to 30 June). | Days | 0.83<br>(0.34 → 1.29)                              | 0.79 → 0.84<br>(8/8)   | 0.83 → 1.81<br>(6/10)    | 0.83 → 1.31<br>(5/6)   | 0.83 → 1.47<br>(3/6)   | 0.78 → 1.08<br>(4/4)   | 0.83 → 1.88<br>(1/4)   |
| CSDI        | Cold spell duration indicator | Annual number of days with at least 6 consecutive day when $T_{min} < 10$ th percentile  | Days | -0.66<br>(-0.84 → -0.47)                           | -0.67 → -0.64<br>(8/8) | -0.70 → -0.50<br>(10/10) | -0.66 → -0.52<br>(6/6) | -0.71 → -0.51<br>(6/6) | -0.69 → -0.54<br>(4/4) | -0.81 → -0.66<br>(4/4) |
| WSDI        | Warm spell duration indicator | Annual number of days with at least 6 consecutive day when $T_{max} > 90$ th percentile  | Days | 1.32<br>(0.85 → 1.69)                              | 1.27 → 1.32<br>(8/8)   | 1.16 → 1.62<br>(10/10)   | 0.89 → 1.32<br>(6/6)   | 0.89 → 1.32<br>(6/6)   | 1.16 → 1.72<br>(3/4)   | 1.02 → 1.32<br>(4/4)   |

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

**Table 1.** Continued.

| Index                   | Name                              | Definition  | Unit                | Range of Linear Trend of Global Average Timeseries |                        |                          |                        |                        |                        |                        |
|-------------------------|-----------------------------------|---|---------------------|--|------------------------|--------------------------|------------------------|------------------------|------------------------|------------------------|
|                         |                                   |   |                     | HadEX2   | Weighting              | Stations/DLS             | Stations/Grid Box      | Long stations          | DLS Methods            | Gridding               |
| Temperature (continued) |                                   |   |                     |  |                        |                          |                        |                        |                        |                        |
| TX10p                   | Cool days                         | Percentage of days when $T_{\max} < 10\text{th percentile}$   | % of days           | -2.49<br>(-3.10 → -1.95)                           | -2.51 → -2.48<br>(8/8) | -2.56 → -2.46<br>(10/10) | -2.49 → -1.61<br>(4/6) | -2.56 → -2.49<br>(6/6) | -2.57 → -2.49<br>(4/4) | -2.49 → -1.62<br>(3/4) |
| TX90p                   | Warm days                         | Percentage of days when $T_{\max} > 90\text{th percentile}$   | % of days           | 3.17<br>(2.08 → 4.27)                              | 3.15 → 3.19<br>(8/8)   | 2.87 → 3.47<br>(10/10)   | 2.04 → 3.17<br>(5/6)   | 2.66 → 3.20<br>(6/6)   | 3.17 → 3.91<br>(4/4)   | 2.83 → 3.23<br>(4/4)   |
| TN10p                   | Cool nights                       | Percentage of days when $T_{\min} < 10\text{th percentile}$   | % of days           | -4.12<br>(-4.78 → -3.49)                           | -4.13 → -4.11<br>(8/8) | -4.19 → -4.00<br>(10/10) | -4.12 → -3.55<br>(6/6) | -4.18 → -3.64<br>(6/6) | -4.17 → -4.12<br>(4/4) | -4.38 → -3.37<br>(3/4) |
| TN90p                   | Warm nights                       | Percentage of days when $T_{\min} > 90\text{th percentile}$   | % of days           | 5.87<br>(4.81 → 7.05)                              | 5.80 → 5.98<br>(8/8)   | 5.24 → 6.02<br>(10/10)   | 4.11 → 5.87<br>(3/6)   | 4.16 → 5.87<br>(3/6)   | 5.87 → 6.02<br>(4/4)   | 3.08 → 5.87<br>(2/4)   |
| FD                      | Frost days                        | Annual number of days when $T_{\min} < 0^\circ\text{C}$   | Days                | -1.75<br>(-2.23 → -1.30)                           | -1.77 → -1.73<br>(8/8) | -2.09 → -1.14<br>(9/10)  | -1.75 → -1.43<br>(6/6) | -1.75 → -1.34<br>(6/6) | -1.82 → -1.07<br>(3/4) | -2.09 → -1.75<br>(4/4) |
| ID                      | Ice days                          | Annual number of days when $T_{\max} < 0^\circ\text{C}$   | days                | -0.70<br>(-1.00 → -0.42)                           | -0.71 → -0.68<br>(8/8) | -1.14 → -0.56<br>(5/10)  | -0.95 → -0.66<br>(6/6) | -1.61 → -0.70<br>(4/6) | -0.93 → -0.70<br>(4/4) | -1.35 → -0.70<br>(1/4) |
| SU                      | Summer days                       | Annual number of days when $T_{\max} > 25^\circ\text{C}$  | days                | 1.07<br>(0.69 → 1.42)                              | 1.06 → 1.11<br>(8/8)   | 0.90 → 1.39<br>(10/10)   | 0.36 → 1.07<br>(3/6)   | 0.85 → 1.30<br>(6/6)   | 0.51 → 1.07<br>(3/4)   | 1.02 → 1.31<br>(4/4)   |
| TR                      | Tropical nights                   | Annual number of days when $T_{\min} > 20^\circ\text{C}$  | days                | 1.24<br>(0.95 → 1.52)                              | 1.16 → 1.27<br>(8/8)   | 0.96 → 1.94<br>(9/10)    | 0.69 → 1.24<br>(2/6)   | 0.78 → 1.24<br>(5/6)   | 1.12 → 1.74<br>(2/4)   | 1.08 → 1.45<br>(4/4)   |
| Rx1day                  | Maximum 1 day precipitation       | Maximum 1 day precipitation total   | mm                  | 0.42<br>(0.18 → 0.69)                              | 0.40 → 0.46<br>(8/8)   | 0.33 → 0.82<br>(7/10)    | 0.42 → 0.62<br>(6/6)   | 0.42 → 0.63<br>(6/6)   | 0.11 → 0.47<br>(3/4)   | 0.13 → 0.48<br>(3/4)   |
| Rx5day                  | Maximum 5 day precipitation       | Maximum 5 day precipitation total   | mm                  | 0.49<br>(-0.03 → 1.03)                             | 0.40 → 0.54<br>(8/8)   | 0.39 → 1.27<br>(7/10)    | 0.48 → 0.76<br>(6/6)   | 0.49 → 0.82<br>(6/6)   | 0.23 → 0.49<br>(4/4)   | 0.25 → 0.81<br>(4/4)   |
| PRCPTOT                 | Annual contribution from wet days | Annual sum of daily precipitation $\geq 1.0\text{ mm}$  | mm                  | 4.50<br>(1.66 → 7.19)                              | 4.30 → 4.68<br>(8/8)   | 2.10 → 5.29<br>(10/10)   | 3.51 → 5.70<br>(6/6)   | 4.50 → 9.58<br>(3/6)   | 4.50 → 6.22<br>(4/4)   | 4.50 → 8.85<br>(3/4)   |
| SDII                    | Simple daily index                | Annual total precipitation divided by the number of wet days (when total precipitation $\geq 1.0\text{ mm}$ ) | mmday <sup>-1</sup> | 0.06<br>(0.04 → 0.08)                              | 0.06 → 0.06<br>(8/8)   | 0.04 → 0.07<br>(8/10)    | 0.04 → 0.06<br>(6/6)   | 0.04 → 0.06<br>(5/6)   | 0.05 → 0.07<br>(4/4)   | 0.04 → 0.07<br>(4/4)   |

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.

**Table 1.** Continued.

| Index         | Name  | Definition  | Unit | Linear Trend Uncertainty |                        |                        |                       |                       |                       |                       |
|---------------|---|---|------|--------------------------|------------------------|------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|               |   |   |      | HadEX2                   | Weighting              | Stations/DLS           | Stations/Grid Box     | Long stations         | DLS Methods           | Gridding              |
| Precipitation |   |   |      |                          |                        |                        |                       |                       |                       |                       |
| R95p          | Annual contribution from very wet days      | Annual sum of daily precipitation > 95th percentile                         | mm   | 3.29<br>(2.08 → 4.66)    | 3.08 → 3.37<br>(8/8)   | 2.69 → 5.24<br>(6/10)  | 3.29 → 5.20<br>(4/6)  | 3.18 → 4.31<br>(6/6)  | 2.99 → 3.36<br>(4/4)  | 2.92 → 4.94<br>(3/4)  |
| R95pTOT       | Fraction from very wet days                 | R95p × 100/PRCPTOT  | %    | 0.30<br>(0.18 → 0.42)    | 0.29 → 0.31<br>(8/8)   | 0.26 → 0.45<br>(6/10)  | 0.30 → 0.45<br>(4/6)  | 0.28 → 0.34<br>(6/6)  | 0.23 → 0.31<br>(4/4)  | 0.21 → 0.30<br>(4/4)  |
| R99p          | Annual contribution from extremely wet days | Annual sum of daily precipitation > 99th percentile                         | mm   | 1.60<br>(0.81 → 2.44)    | 1.50 → 1.74<br>(8/8)   | 1.38 → 2.73<br>(6/10)  | 1.60 → 2.39<br>(6/6)  | 1.60 → 1.97<br>(6/6)  | 1.37 → 1.60<br>(4/4)  | 1.54 → 2.33<br>(4/4)  |
| R99pTOT       | Fraction from extremely wet days            | R99p × 100/PRCPTOT  | %    | 0.14<br>(0.06 → 0.22)    | 0.13 → 0.15<br>(8/8)   | 0.12 → 0.23<br>(8/10)  | 0.14 → 0.19<br>(6/6)  | 0.12 → 0.14<br>(6/6)  | 0.10 → 0.14<br>(4/4)  | 0.12 → 0.14<br>(4/4)  |
| CWD           | Consecutive wet days                        | Maximum annual number of consecutive wet days (when precipitation ≥ 1.0 mm) | Days | -0.01<br>(-0.03 → 0.01)  | -0.01 → -0.01<br>(8/8) | -0.01 → 0.02<br>(6/10) | -0.01 → 0.01<br>(6/6) | -0.01 → 0.02<br>(3/6) | -0.01 → 0.03<br>(3/4) | -0.01 → 0.02<br>(3/4) |
| CDD           | Consecutive dry days                        | Maximum annual number of consecutive dry days (when precipitation < 1.0 mm) | Days | 0.24<br>(-0.10 → 0.59)   | 0.21 → 0.31<br>(8/8)   | -0.54 → 0.28<br>(6/10) | -0.12 → 0.24<br>(4/6) | -0.61 → 0.24<br>(1/6) | -0.08 → 0.32<br>(4/4) | -0.42 → 0.24<br>(2/4) |
| R10mm         | Heavy precipitation days                    | Annual number of days when precipitation > 10 mm                            | Days | 0.14<br>(0.05 → 0.20)    | 0.13 → 0.15<br>(8/8)   | 0.04 → 0.14<br>(9/10)  | 0.12 → 0.19<br>(6/6)  | 0.13 → 0.27<br>(3/6)  | 0.10 → 0.14<br>(4/4)  | 0.13 → 0.23<br>(2/4)  |
| R20mm         | Very heavy precipitation days               | Annual number of days when precipitation > 20 mm                            | Days | 0.04<br>(-0.01 → 0.10)   | 0.04 → 0.05<br>(8/8)   | 0.04 → 0.15<br>(6/10)  | 0.04 → 0.10<br>(5/6)  | 0.04 → 0.13<br>(3/6)  | 0.04 → 0.08<br>(4/4)  | 0.04 → 0.11<br>(3/4)  |

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

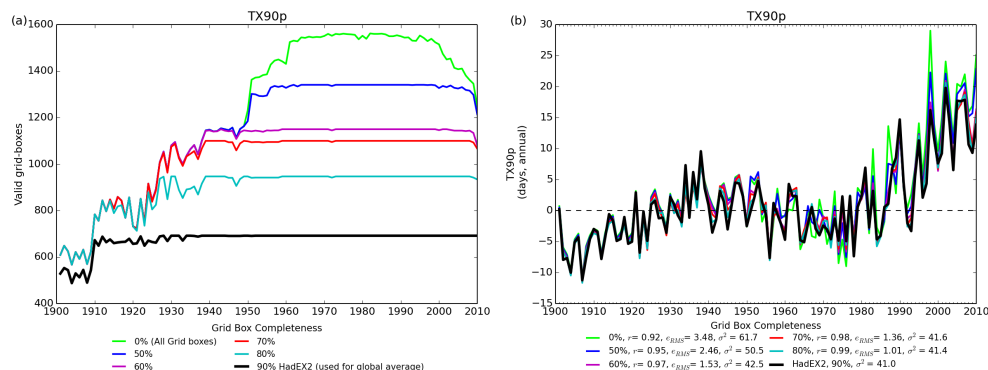
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 1.** (a) Numbers of grid boxes covered given different completeness requirements, for TX90p. The total coverage is shown in light green, whereas the coverage used when calculating global averages is in black, with a range of other completeness percentages shown in between. (b) The time series for global average of TX90p in HadEX2 (black). Also shown are the comparison of other completenesses to that used in HadEX2 (90 %) using the correlation coefficient,  $r$ , the root mean square error  $e_{RMS}$  and the variance,  $\sigma^2$ .

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

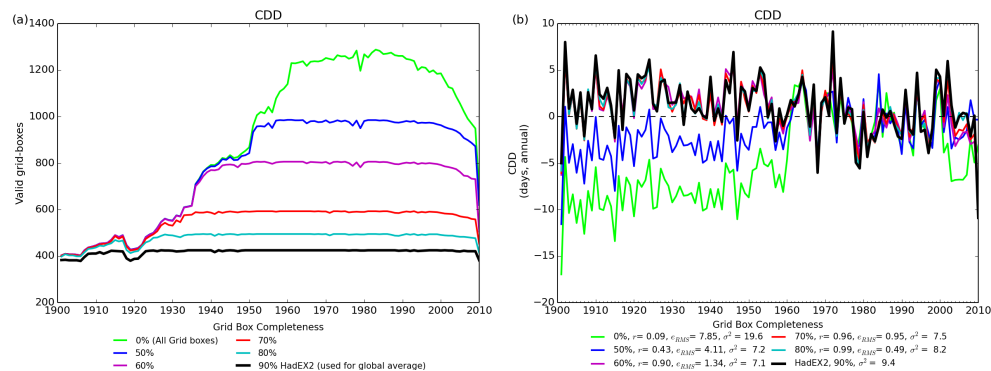
Printer-friendly Version

Interactive Discussion



**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.



**Fig. 2.** As for Fig. 1 but for CDD.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

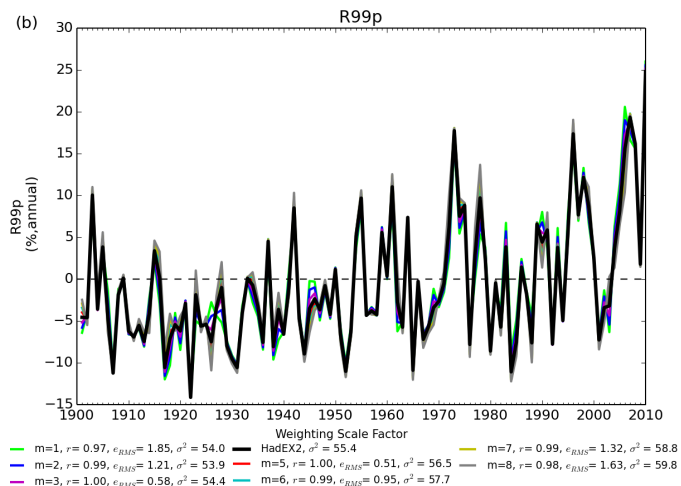
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 3.** The time series for global average R99p showing the different curves for the different weighting values (colours are indicated below the plot). The figure also shows the correlation coefficient ( $r$ ), variance ( $\sigma^2$ ) and root-mean-square error ( $e_{RMS}$ ) between each series and HadEX2. All choices have been normalised relative to the 1961–1990 period.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

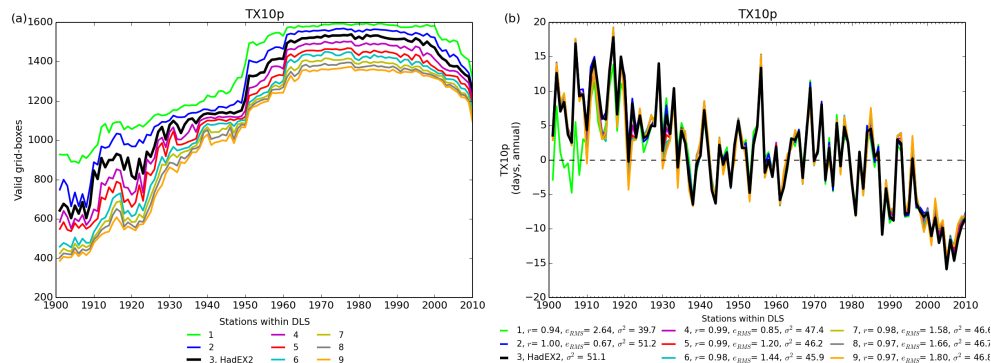
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 4.** (a) The number of non-missing grid boxes for TX10p for the different numbers of stations required within a DLS of the grid box centre. All grid boxes with sufficient stations are shown in the coverage series. (b) The time series for global average TX10p for the different numbers of stations required within a DLS of the grid box centre, given the overall 90 % completeness requirement.

[Title Page](#)

[Abstract](#) | [Introduction](#)

[Conclusions](#) | [References](#)

[Tables](#) | [Figures](#)

[◀](#) | [▶](#)

[◀](#) | [▶](#)

[Back](#) | [Close](#)

[Full Screen / Esc](#)

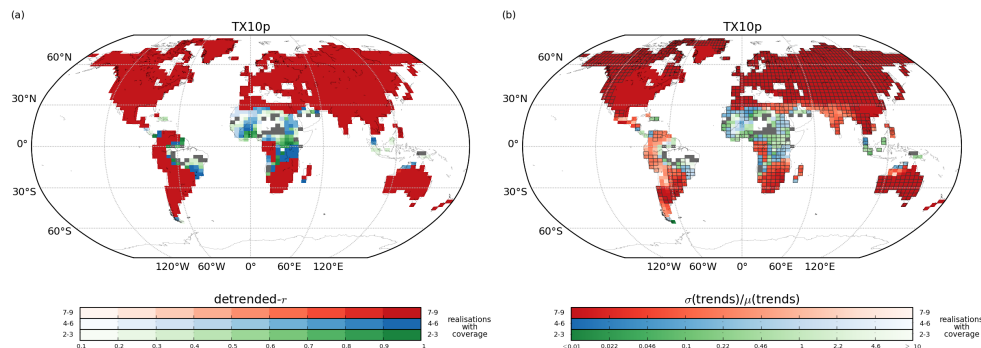
[Printer-friendly Version](#)

[Interactive Discussion](#)



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



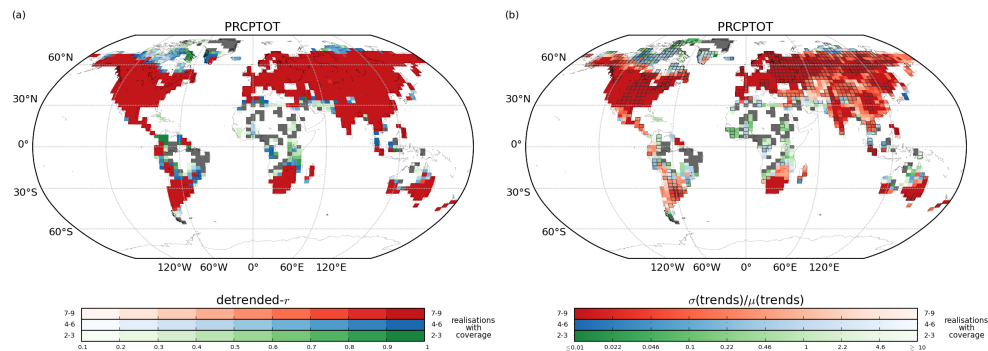
**Fig. 5.** (a) The mean detrended correlation coefficient,  $r$  with HadEX2 for all the grid boxes and (b) the standard deviation of the trends normalised by the mean trend ( $\sigma/\mu$ ) calculated for the period 1951 to 2010. Grey grid boxes are those where only one of the 9 possible options for the number of stations within the DLS results in a value. The grid boxes are green when 2–3, blue when 4–6 and red when 7–9 of the possible choices result in a value. In the right panel, boxes which have been outlined are those where there is high confidence in a non-zero trend in HadEX2. For details on the calculation method, see text.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)




## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 6.** As for Fig. 5 but for PRCPTOT.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

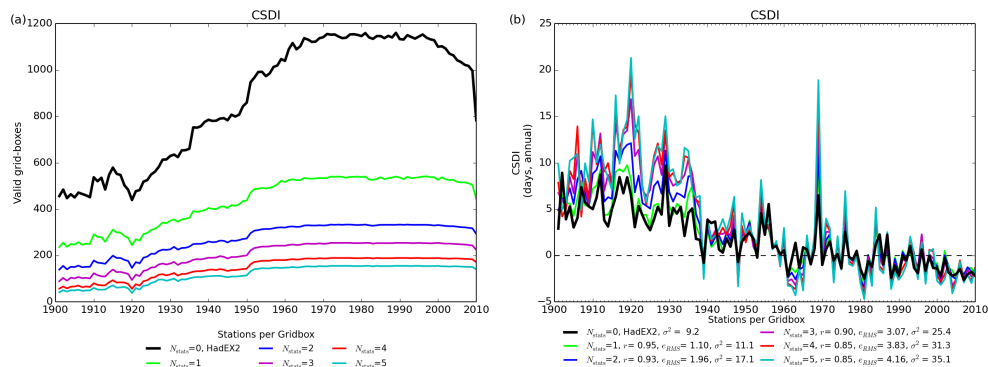
Printer-friendly Version

Interactive Discussion



**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.



**Fig. 7.** As for Fig. 4 but for the CSDI index and for different numbers of stations within a grid box. Note that the global average only takes grid boxes which have 90 % completeness or more, whereas all grid boxes are shown in the coverage series.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

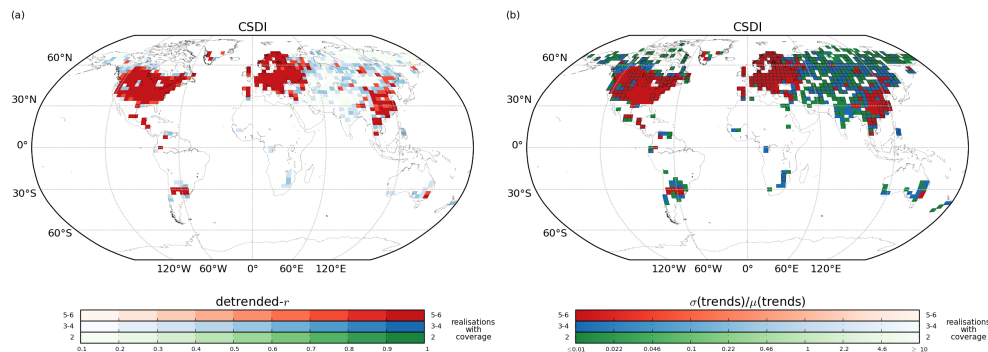
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 8.** As for Fig. 5 but for CSDI. Green grid boxes are those where two of the 6 possible stations per grid box choices results in a value, blue when 3–4 and red when 5–6 of the possible choices result in a value. Note that the HadEX2 method does not require any stations to be present within a grid box for a value to be calculated.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

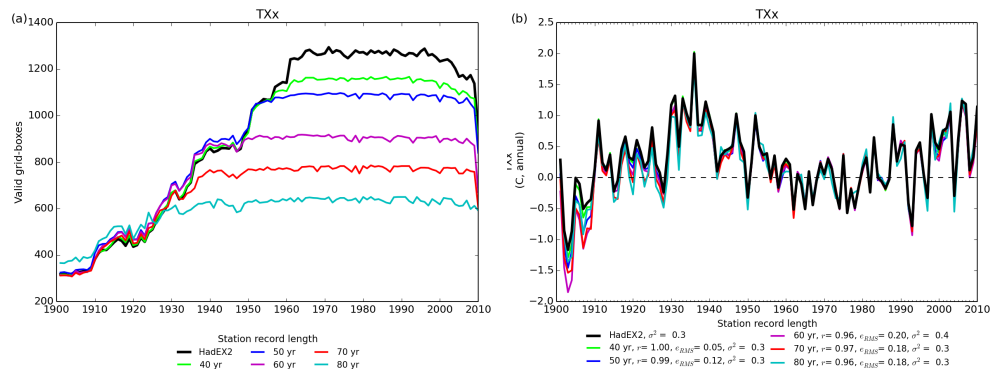
Printer-friendly Version

Interactive Discussion



**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.



**Fig. 9.** As for Fig. 4 but for TXx and long-term stations.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

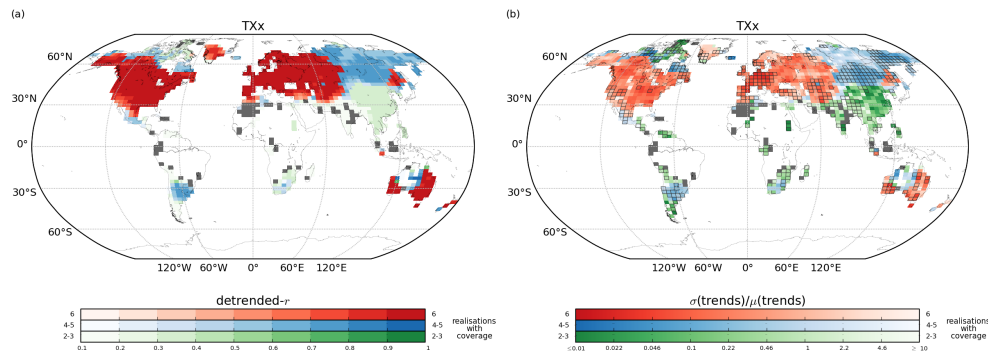
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

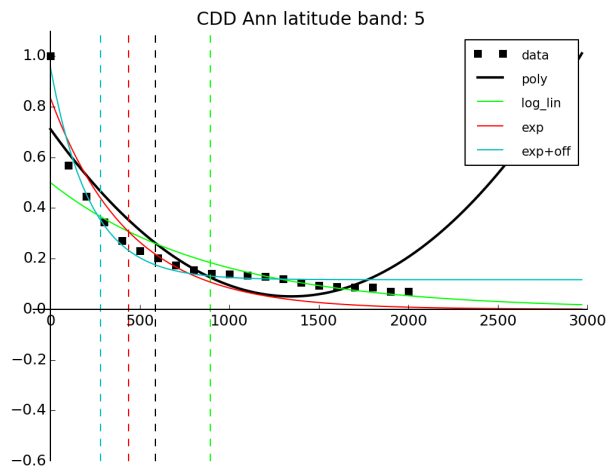


**Fig. 10.** As for Fig. 5 but for TXx and long-term stations. Grey grid boxes are those where only one of the 6 possible station reporting length choices results in a value, green when 2–3, blue when 4–5 and red when 6 of the possible choices result in a value. In the right-hand plot, boxes which have been outlined are ones where the trend was significant in the HadEX2 version. For details on the calculation method, see text.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

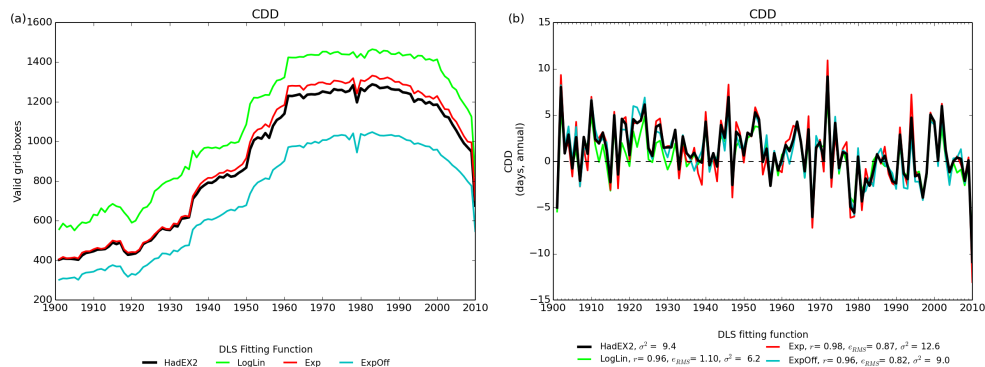


**Fig. 11.** The binned inter-station correlation coefficients against distance. All four fitting methods are shown. The vertical dashed lines are the derived decorrelation length scales.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.



**Fig. 12.** As for Fig. 4 but for CDD and choices in the DLS calculation method.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

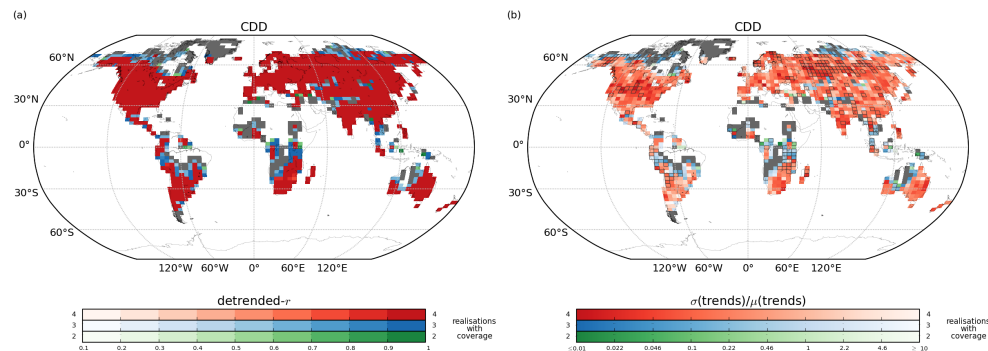
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 13.** As for Fig. 5 but for CDD and for choices of the DLS fitting method. Grey grid boxes are those where only one of the four possible DLS calculation choices results in a value, green when 2, blue when 3 and red when all 4 of the possible choices result in a value. In the right-hand plot, boxes which have been outlined are ones where the trend was significant in the HadEX2 version. For details on the calculation method, see text.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.

[Title Page](#)

[Abstract](#)   [Introduction](#)

[Conclusions](#)   [References](#)

[Tables](#)   [Figures](#)

[◀](#)   [▶](#)

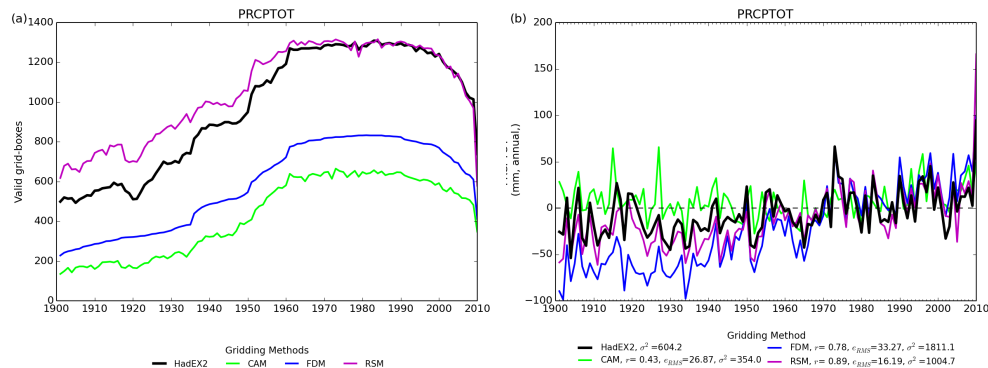
[◀](#)   [▶](#)

[Back](#)   [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

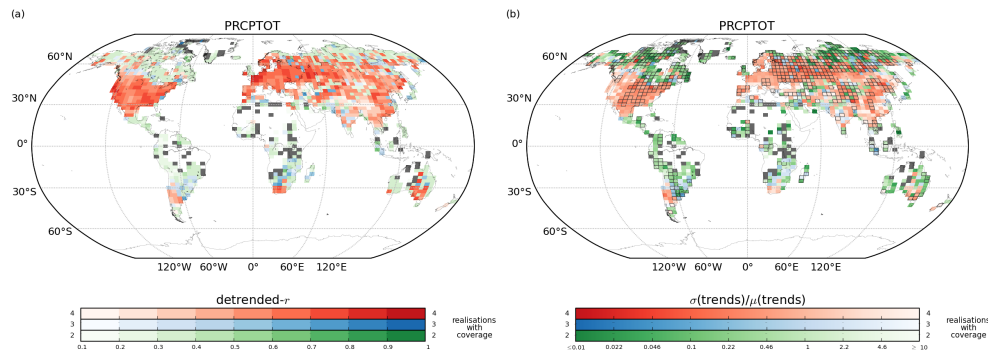
[Interactive Discussion](#)



**Fig. 14.** As for Fig. 4 but for PRCPTOT and for the four different gridding methods.

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 15.** As for Fig. 5 but for PRCPTOT and the four different gridding methods. Grey grid boxes are those where only one of the four possible gridding methods results in a value, green when 2, blue when 3 and red when all 4 of the possible choices result in a value.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

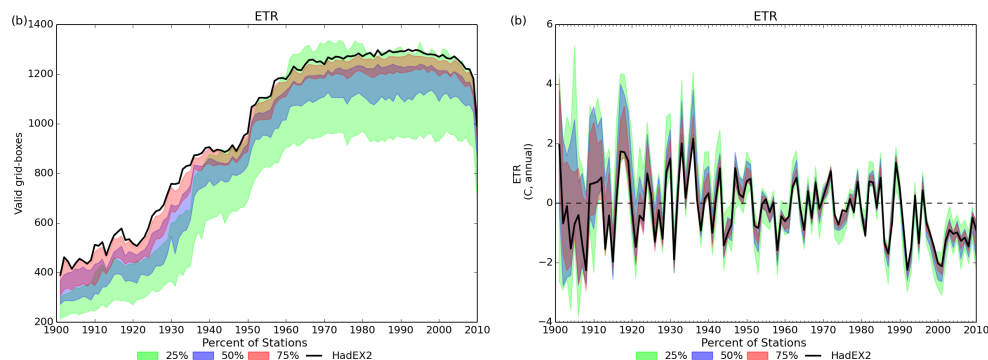
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

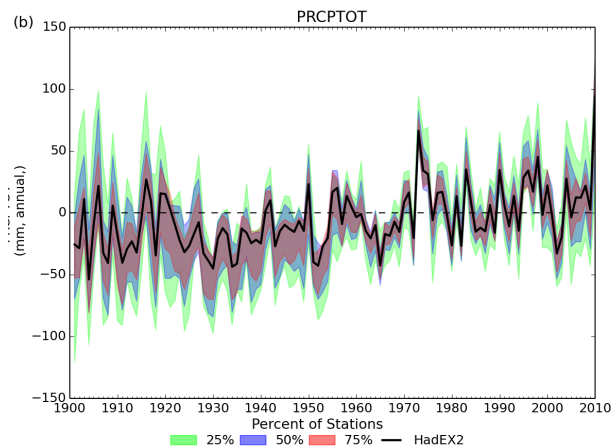


**Fig. 16.** (a) The number of non-missing grid boxes and (b) the time series for the global average of ETR for the jackknife runs. Each colour shows the maximum range of the time series for each of the three sets of runs, 25% of stations in green, 50% in blue and 75% in purple respectively. The solid black line shows the HadEX2 results in both panels. Note that the global average only takes grid boxes which have 90% completeness or more, whereas all grid boxes are shown in the coverage series.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

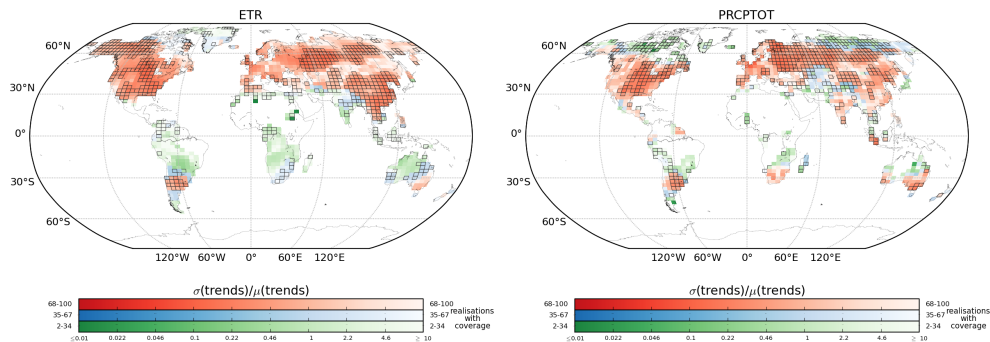


**Fig. 17.** As for Fig. 16b but showing the time series for the global average of PRCPTOT for the jackknife runs.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Uncertainty of gridded extremes datasets**

R. J. H. Dunn et al.



**Fig. 18.** As for Fig. 5b but for ETR and PRCPTOT and the 25 % runs of the jackknifing analysis different gridding methods.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

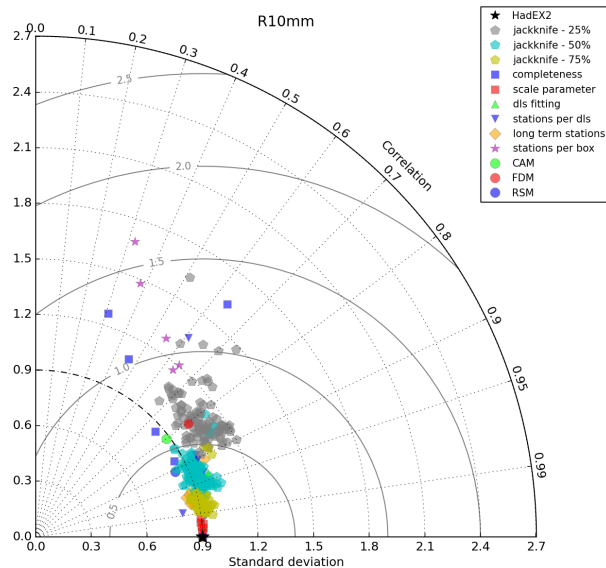
Printer-friendly Version

Interactive Discussion



## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.



**Fig. 19.** The Taylor Diagram for R10mm calculated using the global average timeseries for each choice. Each of the different methodological choices in the previous sections are shown using a different symbol and colour as indicated in the legend.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

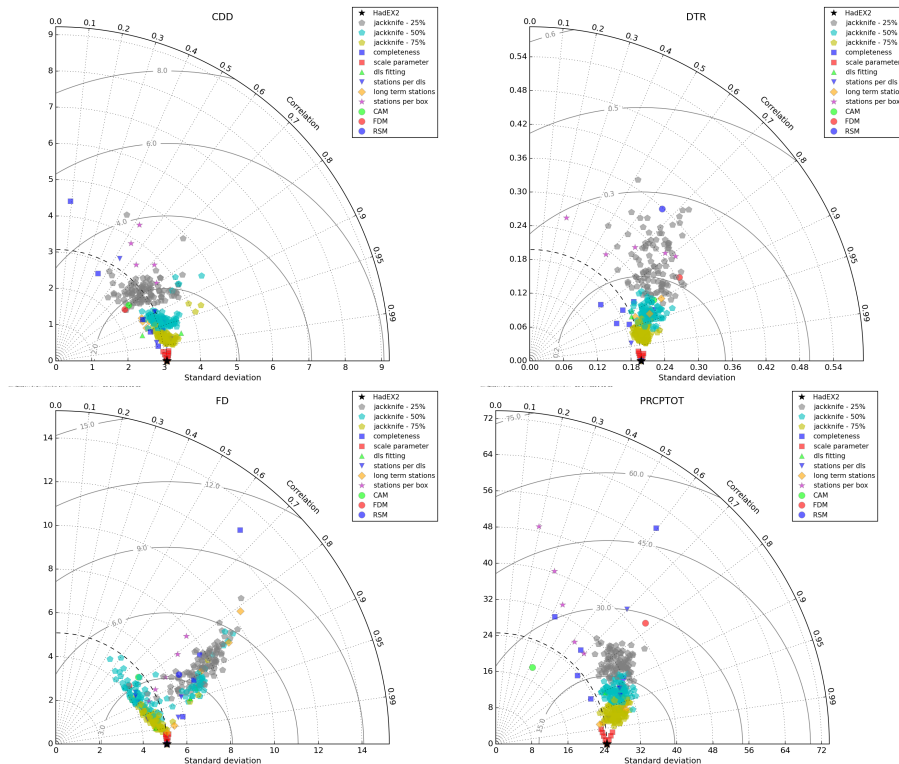


Fig. 20. Taylor diagrams for each of the classes of extremes indices analysed in this study.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

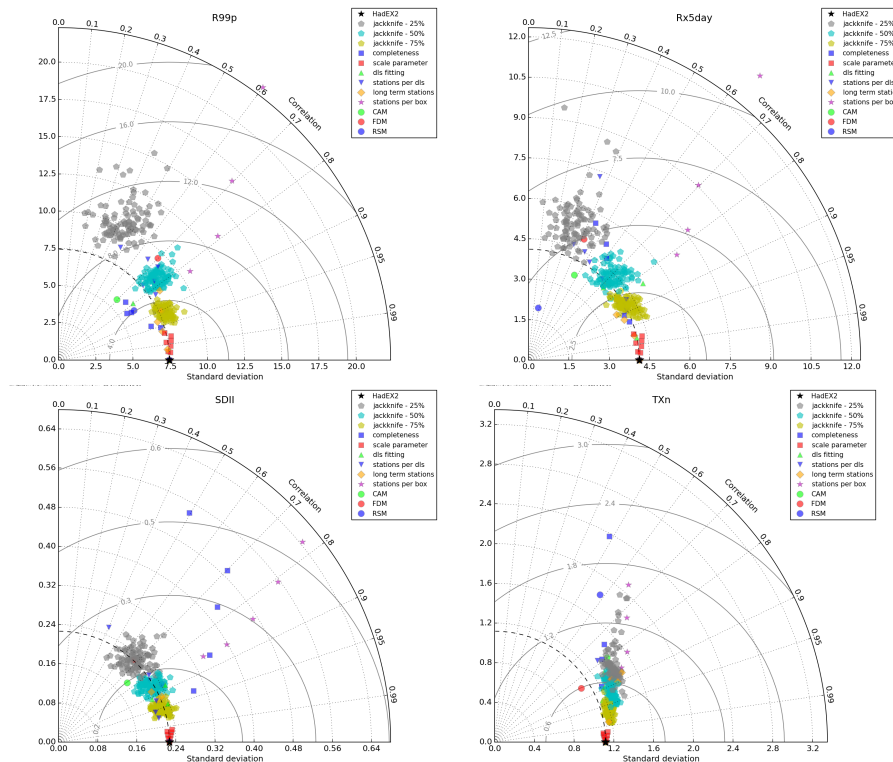


Fig. 20. Continued.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## Uncertainty of gridded extremes datasets

R. J. H. Dunn et al.

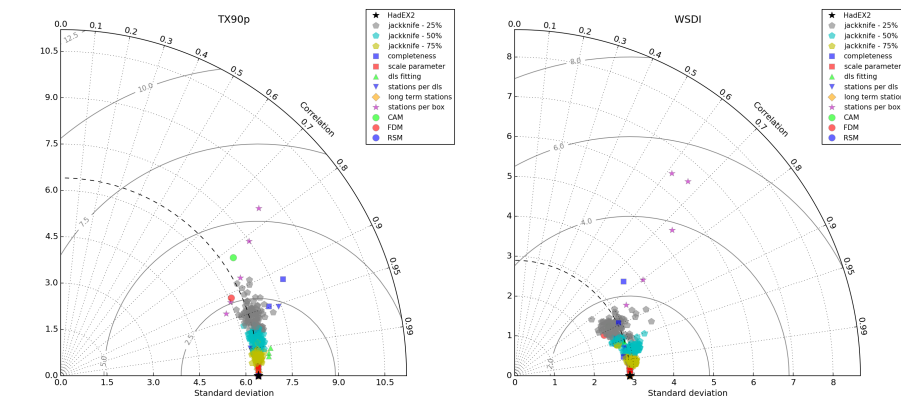


Fig. 20. Continued.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

