Systematic errors in temperature extreme definitions and their impacts

Lukas Brunner* and Aiko Voigt
*University of Vienna (soon Universität Hamburg)

13. Deutsche Klimatagung | 13.3. 2024

Ramsau am Dachstein (1.100msl) in January 2023.

Overview

- Part 1: A bias in defining temperature extremes
- Part 2: Pitfalls in diagnosing temperature extremes
- Part 3: Eliminating the bias
- Summary and conclusions

Overview

- Part 1: A bias in defining temperature extremes
- Part 2: Pitfalls in diagnosing temperature extremes
- Part 3: Eliminating the bias
- Summary and conclusions

Reference:

Brunner and Voigt (in press): Pitfalls in diagnosing temperature extremes. *Nature Communications*, DOI: https://doi.org/10.1038/s41467-024-46349-x

Paper embargoed until Monday May 18th, 11.00 CET!
You can take pictures but please don't share on social media!

Part 1: A bias in defining temperature extremes

universität wien

Definition of temperature extremes

Index shorthand	Characteristic measured & timescales	Index definition
TN10p	Frequency; monthly & annual	Occurrence of cold nights (daily minimum temperature) below
TN90p	Frequency; monthly & annual	Occurrence of warm nights above the 90th percentile
TX10p	Frequency; monthly & annual	Occurrence of cold days (daily maximum temperature) below t
TX90p	Frequency; monthly & annual	Occurrence of warm days above the 90th percentile.
TXx	Intensity; monthly & annual	Maximum daily maximum temperature
TNx	Intensity; monthly & annual	Maximum daily minimum temperature
TXn	Intensity; monthly & annual	Minimum daily maximum temperature
TNn	Intensity; monthly & annual	Minimum daily minimum temperature

Contents lists available at ScienceDirect
Atmospheric Research
journal homepage: www.elsevier.com/locate/atmos

Invited review article

A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale

Sarah E. Perkins *

JGR Atmospheres
RESEARCH ARTICLE
10.1029/2023JD038906

Detecting Extreme Temperature Events Using Gaussian Mixture Models

Aytaç Paçal[1,2], Birgit Hassler[1], Katja Weigel[1,2], M. Levent Kurnaz[3], Michael F. Wehner[4], and Veronika Eyring[1,2]
[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany,

Key Points:

- Extreme temperature events are detected with Gaussian Mixture Models to follow a multimodal rather than a unimodal distribution
- 10-year temperature extremes will occur 13.6 times more frequently under 3.0°C future warming
- Colder days are getting warmer faster

VOLUME 34 JOURNAL OF CLIMATE 1 OCTOBER 2021

A New Framework for Identifying and Investigating Seasonal Climate Extremes

MATTHIAS RÖTHLISBERGER,[a] MAURO HERMANN,[a] CHRISTOPH FREI,[b] FLAVIO LEHNER,[c,a,d] ERICH M. FISCHER,[a] ...RNLI[a]
...ürich, Zürich, Switzerland
...Swiss, Zürich, Switzerland
...University, Ithaca, New York
...oulder, Colorado

Extreme metrics from large ensembles: investigating the effects of ensemble size on their estimates

Claudia Tebaldi[1,2], Kalyn Dorheim[1], Michael Wehner[2], and Ruby Leung[3]
[1]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA
[2]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

DEFINING SINGLE EXTREME WEATHER EVENTS IN A CLIMATE PERSPECTIVE

JULIEN CATTIAUX AND AURÉLIEN RIBES

Geophysical Research Letters
RESEARCH LETTER
10.1029/2023GL103540

Increasing Intensity of Extreme Heatwaves: The Crucial Role of Metrics

Emmanuele Russo[1] and Daniela I. V. Domeisen[1,2]
[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zürich, Switzerland, [2]Université de Lausanne, Lausanne, Switzerland

Key Points:

- The most intense heatwaves of 1950–2021 considerably change if considering intensity indices either based on cumulative or averaged values
- An appropriate measure of heatwave

The effect of a short observational record on the statistics of temperature extremes

Joel Zeder[1], Sebastian Sippel[1], Olivier C. Pasche[2], Sebastian Engelke[2] and Erich M. Fischer[1]
[1]Institute for Atmospheric and Climate Science, ETH Zurich, 8092 Zurich, Switzerland
[2]Research Center for Statistics, University of Geneva, 1205 Geneva, Switzerland

Extremes are often defined relative to the local temperature distribution

Various approaches are used to define extremes. These are generally based on the determination of **relative** (e.g., 90th percentile) or absolute (e.g., 35°C for a hot day) **thresholds.** IPCC AR6 WG1 CH11

Extremes are often defined relative to the local temperature distribution

Various approaches are used to define extremes. These are generally based on the determination of **relative** (e.g., 90th percentile) or absolute (e.g., 35°C for a hot day) **thresholds.** IPCC AR6 WG1 CH11

For percentile-based definitions the **Expert Team on Climate Change Detection and Indices (ETCCDI)** recommends a threshold based on

- the **90th percentile** relative to daily maximum temperature,
- the **30 year period** 1961-1990, and
- **5 day running window** across the seasonal cycle.

Properties of relative extreme definitions

When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*

*in sample

Properties of relative extreme definitions

When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*

- **independent of the season**
  since the threshold follows the annual cycle, Tank and Können 2003; Fischer and Schär 2010; Hirsch et al. 2021

*in sample

Properties of relative extreme definitions

When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*

- **independent of the season**
  since the threshold follows the annual cycle, Tank and Können 2003; Fischer and Schär 2010; Hirsch et al. 2021
- **independent of the location**
  since the threshold follows the spatial temperature distribution, Zhang et al. 2011; Schoetter et al. 2015

*in sample

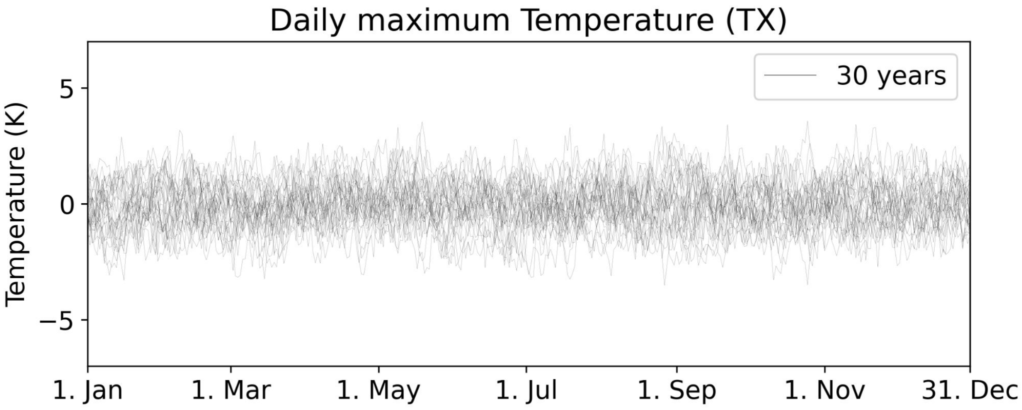
Properties of relative extreme definitions

When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*

- **independent of the season**
  since the threshold follows the annual cycle, Tank and Können 2003; Fischer and Schär 2010; Hirsch et al. 2021
- **independent of the location**
  since the threshold follows the spatial temperature distribution, Zhang et al. 2011; Schoetter et al. 2015
- **independent of the dataset**
  since the threshold provides an implicit bias correction. Freychet et al. 2021; Schoetter et al. 2015

*in sample

Creation of a synthetic temperature time series

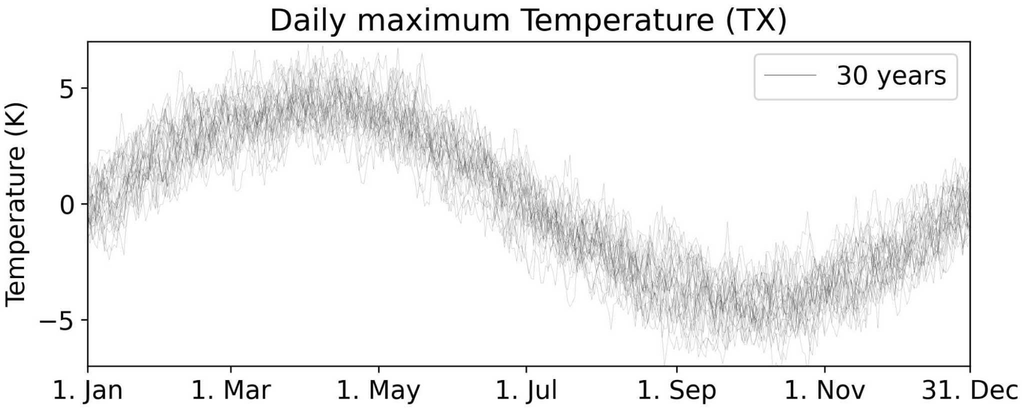


Synthetic temperature

- white noise with standard deviation 1K
- 30 years with 365 days
- lag 1 day autocorrelation: 0.8

Following Zhang et al. 2005

Creation of a synthetic temperature time series



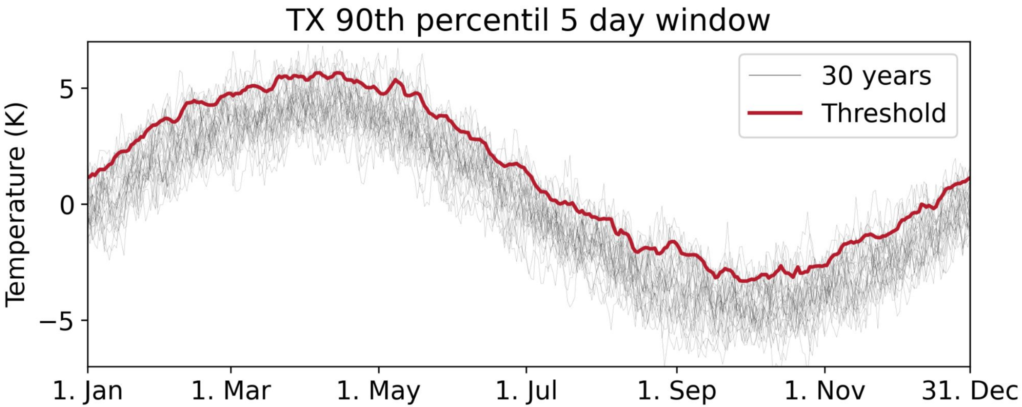
Synthetic temperature

- white noise with standard deviation 1K
- 30 years with 365 days
- lag 1 day autocorrelation: 0.8

Following Zhang et al. 2005

- sine with amplitude 3K

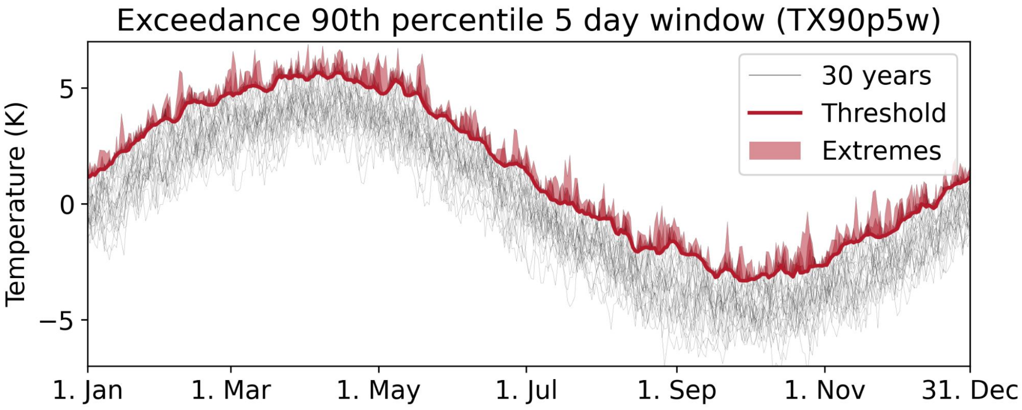
Definition of relative temperature extremes



ETCCDI threshold:

- 90th percentile
- 30 year
- 5 day running window

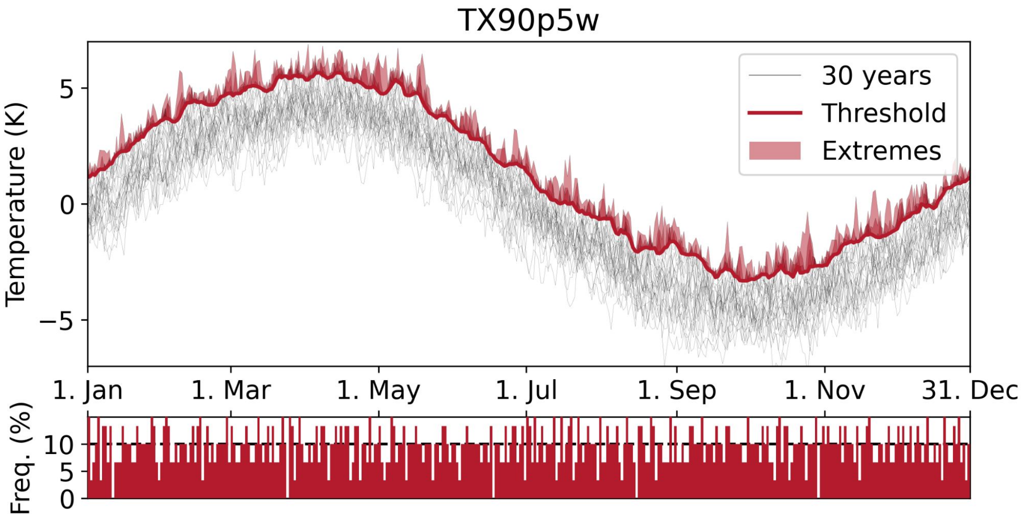
Definition of relative temperature extremes



ETCCDI threshold:

- 90th percentile
- 30 year
- 5 day running window

Definition of relative temperature extremes

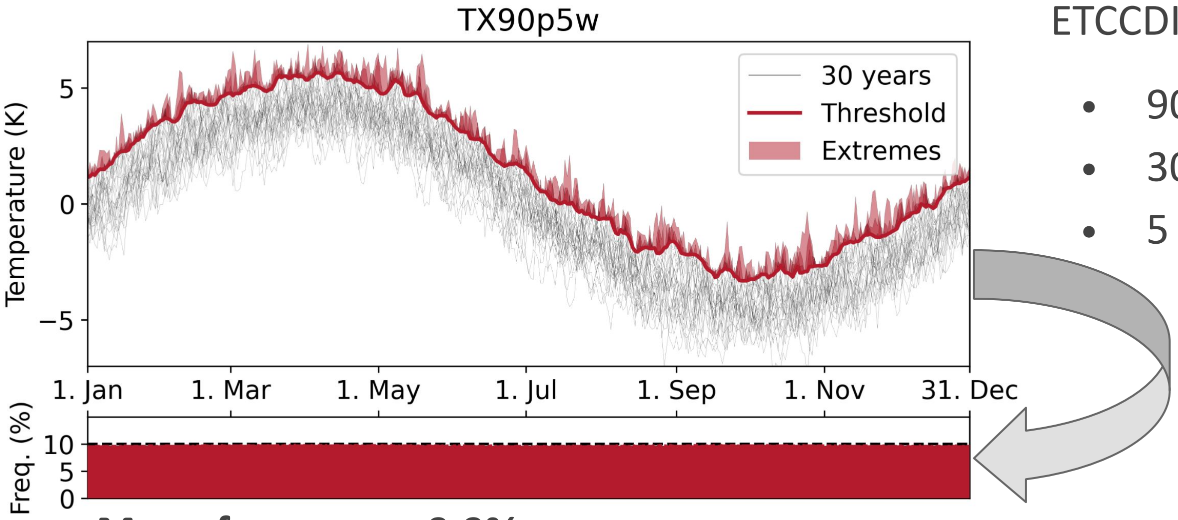


ETCCDI threshold:

- 90th percentile
- 30 year
- 5 day running window

- reference frequency: 10%

Definition of relative temperature extremes

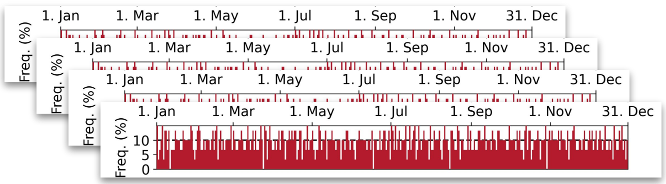


Mean frequency: 9.9%

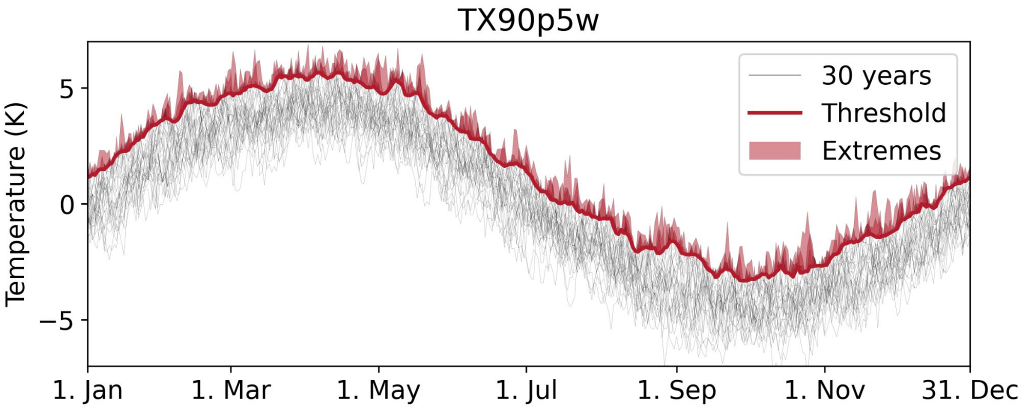
ETCCDI threshold:

- 90th percentile
- 30 year
- 5 day running window

repeat 5’000 times & average



Many studies do not follow the ETCCDI recommendation and use longer running window sizes

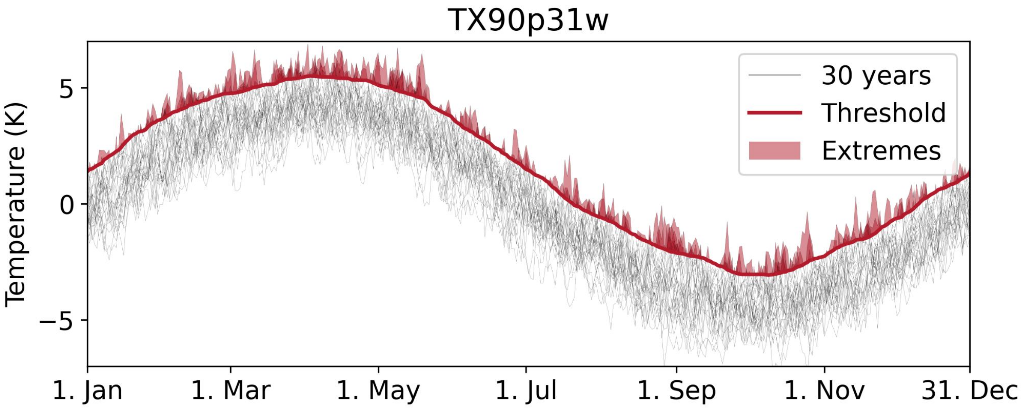


ETCCDI threshold:

- 90th percentile
- 30 year
- 5 day running window

Given the relatively short historical period used, daily ***percentile values can fluctuate*** *up and down somewhat from one day to the next,* ***an undesired result of sampling variability*** *rather than changes in seasonally varying climate.* Lyon et al. 2019

Many studies do not follow the ETCCDI recommendation and use longer running window sizes



ETCCDI threshold:

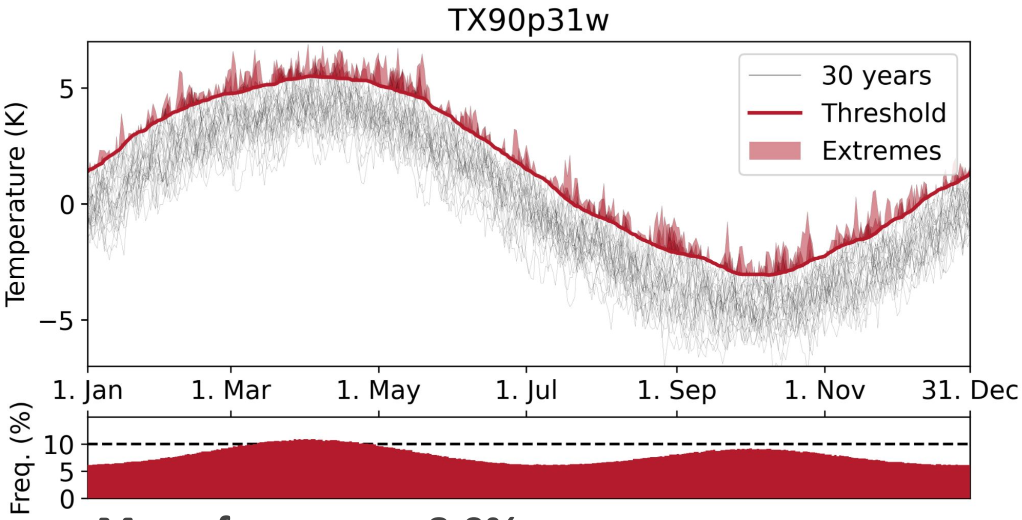
- 90th percentile
- 30 year
- **31 day running window**

Russo et al. 2015; Ceccherini et al. 2016; Russo et al. 2016; Sun et al. 2017; Brunner et al. 2018; Dosio et al. 2018; Zschenderlein et al. 2018; Spensberger et al. 2020; Vogel et al. 2020; Freychet et al. 2021; Schielicke et al. 2022; Aadhar et al. 2023; Russo et al. 2023

- (15 day running window)

Della-Marta et al. 2007; Fischer et al. 2010; Perkins et al. 2012; Perkins et al. 2013; Spinoni et al. 2015; Perkins-Kirkpatrick et al. 2017; Lyon et al. 2019; Perkins-Kirkpatrick et al. 2020; Engdaw et al. 2021; Hirsch et al. 2021; Reddy et al. 2021; Wu et al. 2023

Many studies do not follow the ETCCDI recommendation and use longer running window sizes which leads to a bias



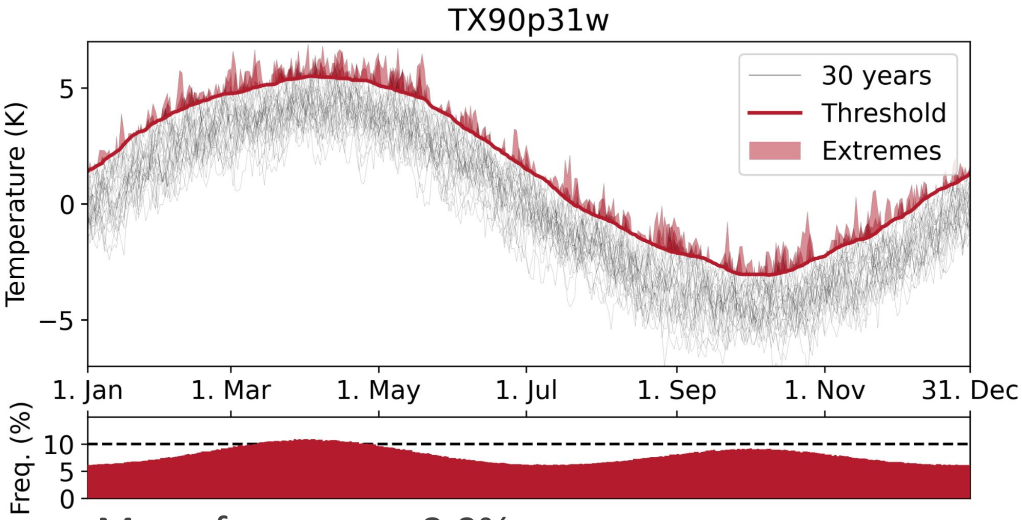
Mean frequency: 8.0%

ETCCDI threshold:

- 90th percentile
- 30 year
- **31 day running window**

Russo et al. 2015; Ceccherini et al. 2016; Russo et al. 2016;
Sun et al. 2017; Brunner et al. 2018; Dosio et al. 2018;
Zschenderlein et al. 2018; Spensberger et al. 2020;
Vogel et al. 2020; Freychet et al. 2021;
Schielicke et al. 2022; Aadhar et al. 2023; Russo et al. 2023

Many studies do not follow the ETCCDI recommendation and use longer running window sizes which leads to a bias



Mean frequency: 8.0%

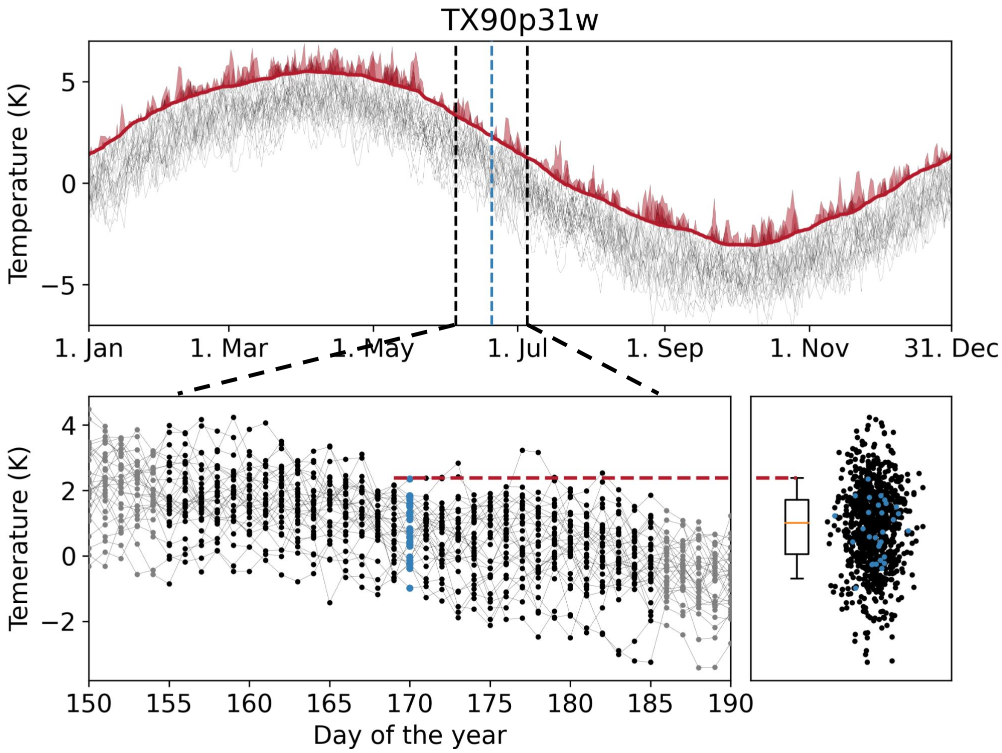
Mean bias: -20%

Definition. Frequency bias

Relative deviation from the expected extreme frequency

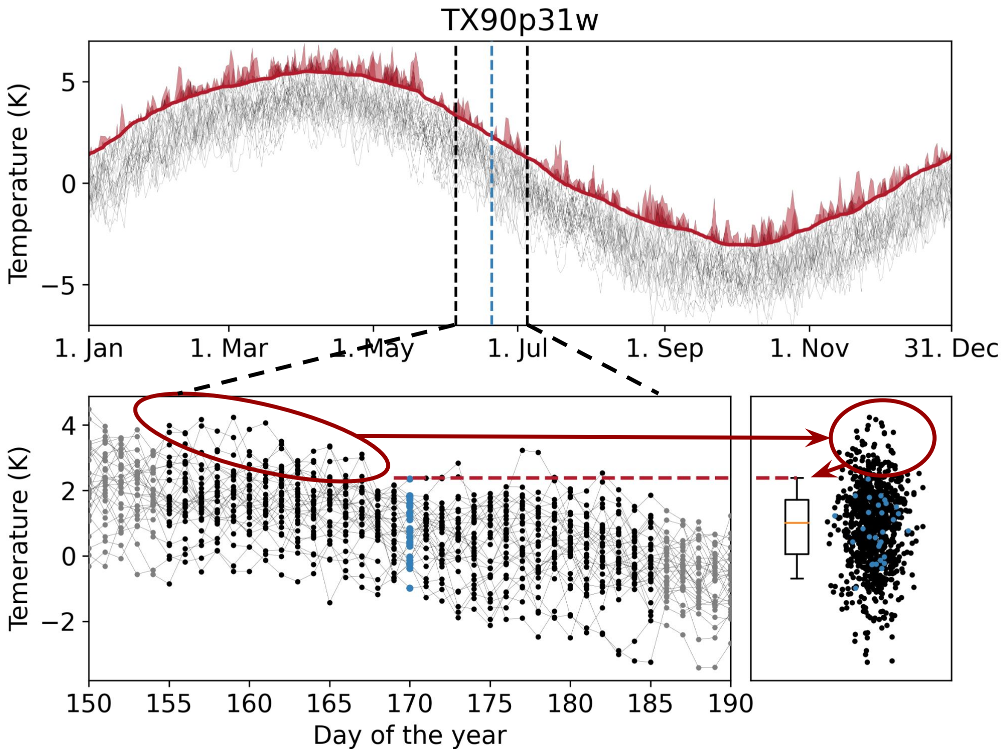
$$f'(p, w) = \frac{f(p, w) - f_{\mathrm{exp}}(p)}{f_{\mathrm{exp}}(p)} \times 100\,\%$$

Seasonally warmer periods dominate the extreme threshold when using long windows



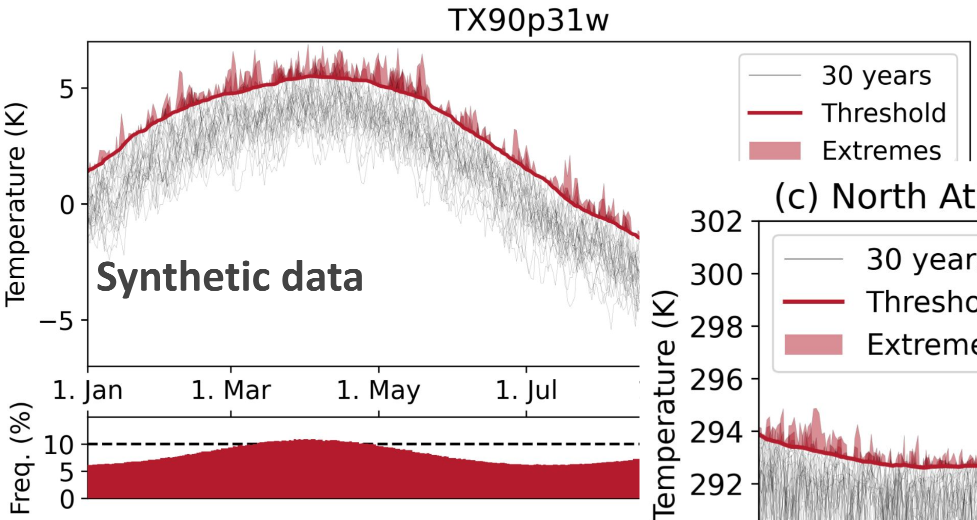
The strongest bias occurs in periods of strong seasonal gradients.

Seasonally warmer periods dominate the extreme threshold when using long windows



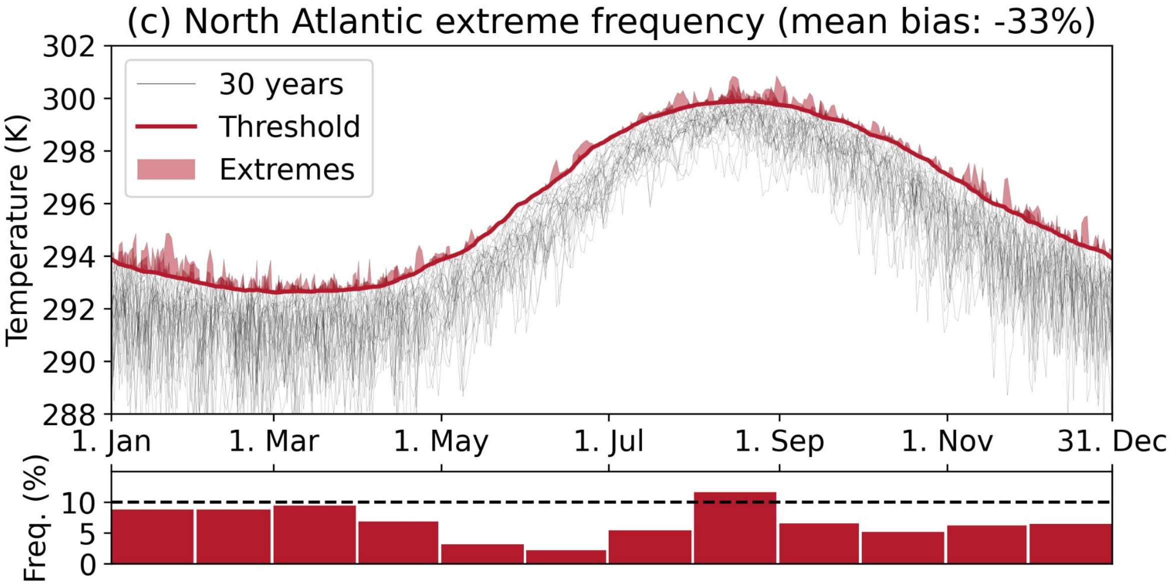
The strongest bias occurs in periods of strong seasonal gradients.

Time for real data: daily maximum temperatures from ERA5

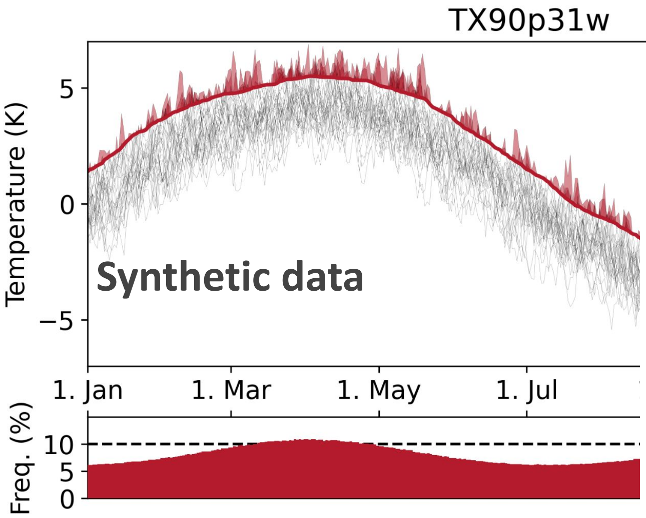


ERA5 1961-1990

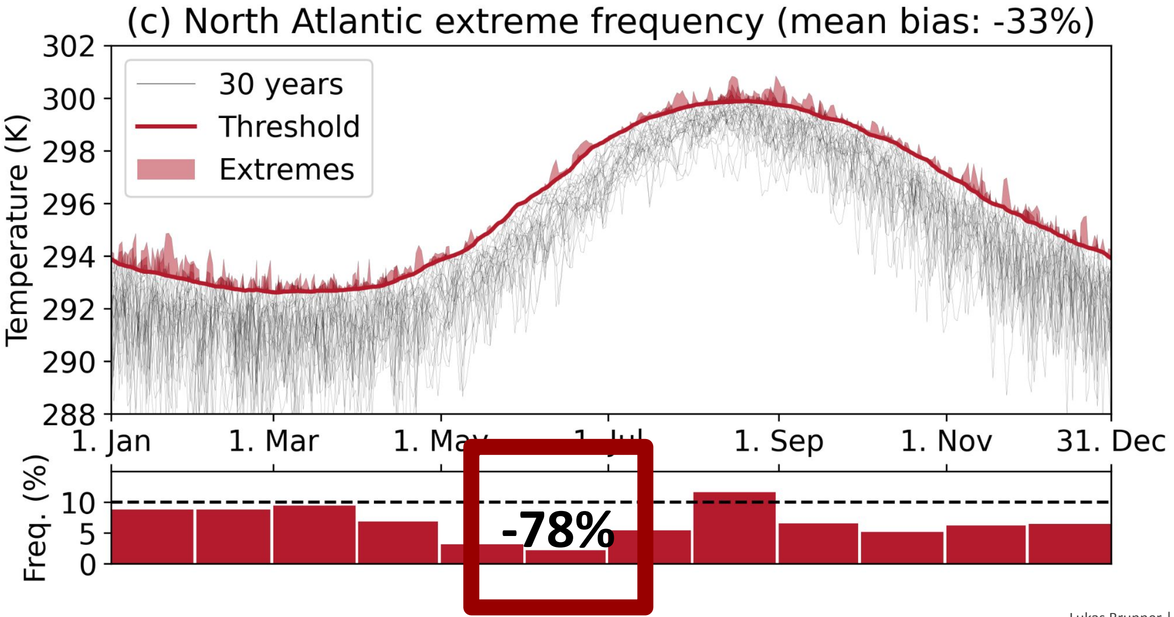
grid cell in the North Atlantic



Time for real data: daily maximum temperatures from ERA5



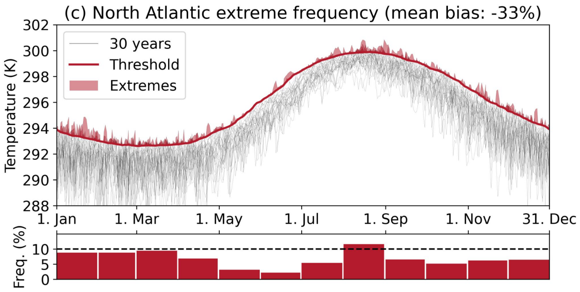
ERA5 1961-1990
grid cell in the North Atlantic



Part 2: Pitfalls in diagnosing temperature extremes

Properties of relative extreme definitions

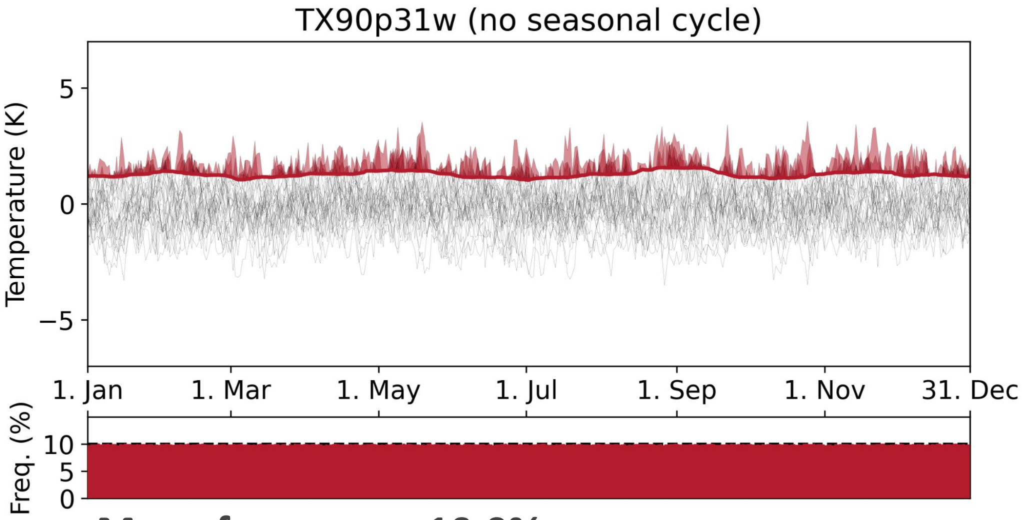
When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*



- ~~**independent of the season**~~ **Bias depends on the season**
  since the threshold follows the annual cycle, Tank and Können 2003; Fischer and Schär 2010; Hirsch et al. 2021
- **independent of the location**
  since the threshold follows the spatial temperature distribution, Zhang et al. 2011; Schoetter et al. 2015
- **independent of the dataset**
  since the threshold provides an implicit bias correction. Freychet et al. 2021; Schoetter et al. 2015

*in sample

The bias depends on the strength of the seasonal cycle

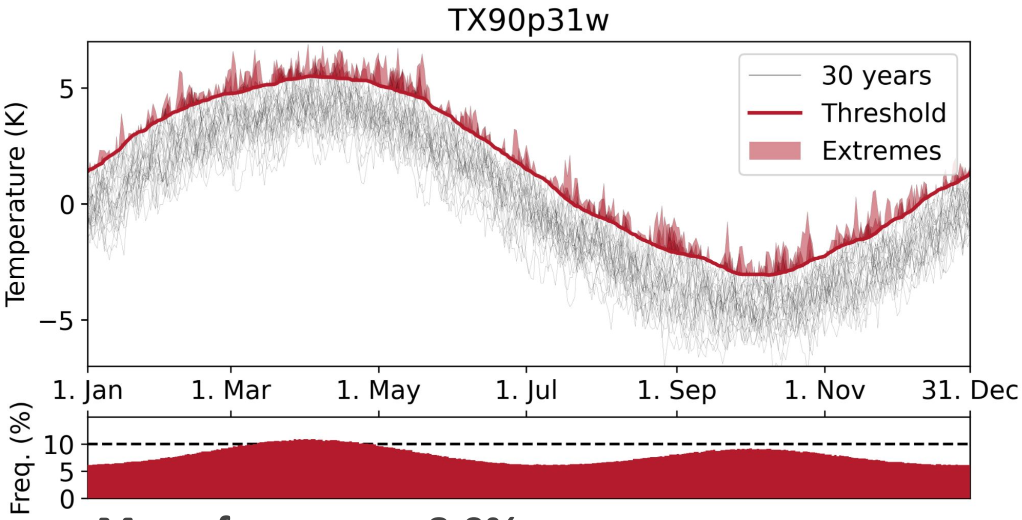


Mean frequency: 10.0%
Mean bias: 0%

Synthetic temperature

- white noise with standard deviation 1K
- 30 years with 365 days
- lag 1 day autocorrelation: 0.8
- **sine with amplitude 0K**

The bias depends on the strength of the seasonal cycle

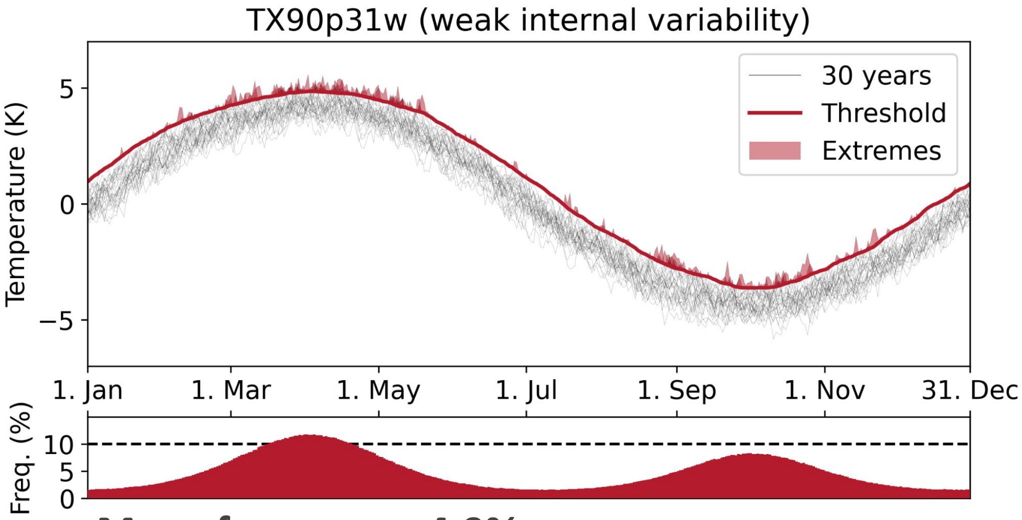


Mean frequency: 8.0%
Mean bias: -20%

Synthetic temperature

- white noise with standard deviation 1K
- 30 years with 365 days
- lag 1 day autocorrelation: 0.8
- **sine with amplitude 3K**

The bias depends on the strength of the seasonal cycle relative to the amplitude of the internal variability



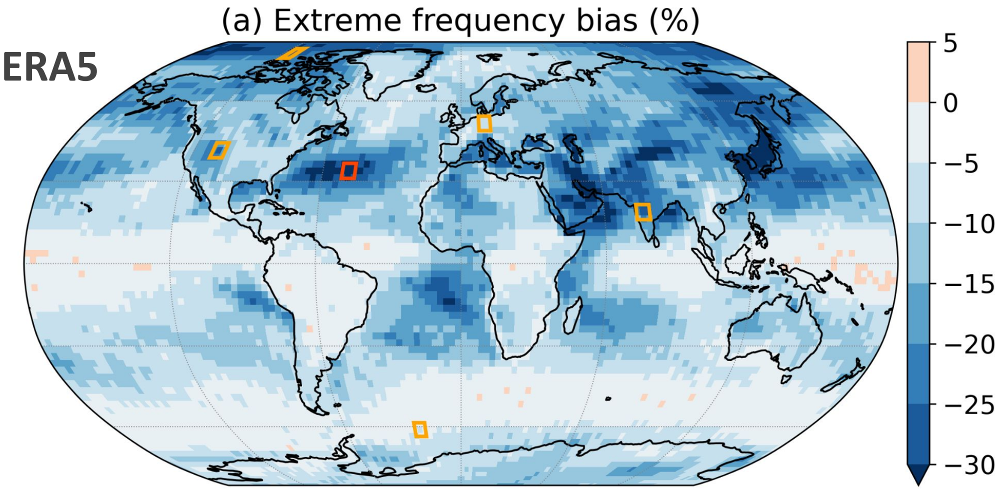
Mean frequency: 4.9%
Mean bias: -51%

Synthetic temperature

- white noise with **standard deviation 0.5K**
- 30 years with 365 days
- lag 1 day autocorrelation: 0.8
- **sine with amplitude 3K**

The amplitude of the seasonal cycle varies regionally and with it the strength of the bias

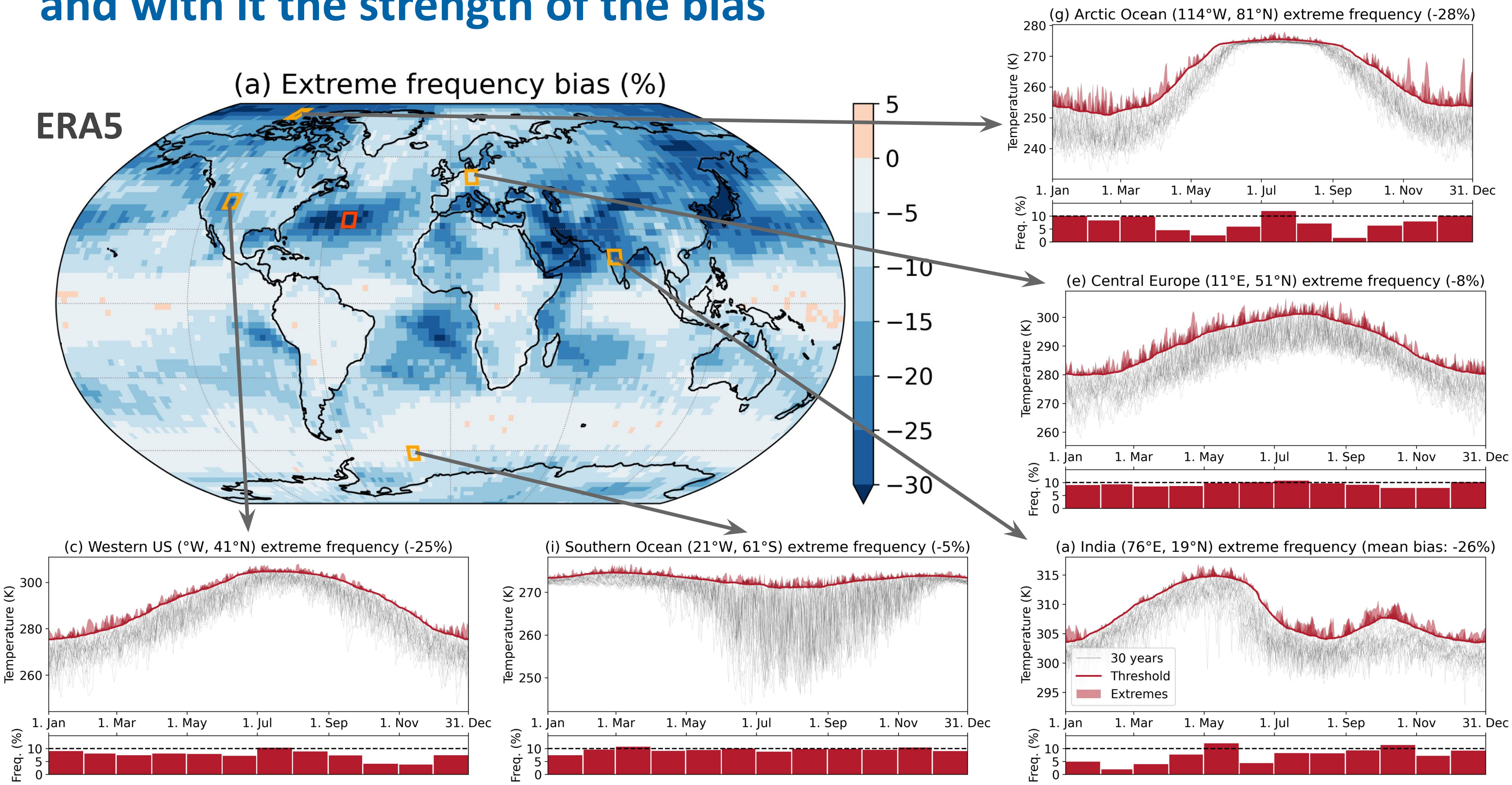
ERA5



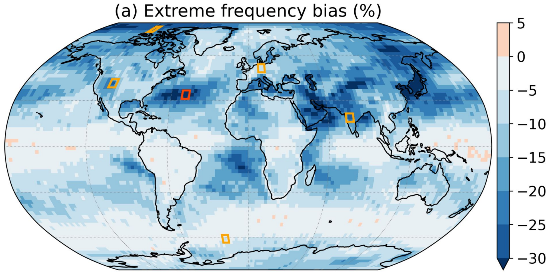
The **global mean bias** in the 30 year period 1961-1990 in ERA5 is **-10%**

Regionally the bias can exceed **-30%**

The amplitude of the seasonal cycle varies regionally and with it the strength of the bias



Properties of relative extreme definitions



When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*

- **~~independent of the season~~** Bias depends on the season
  since the threshold follows the annual cycle, Tank and Können 2003; Fischer and Schär 2010; Hirsch et al. 2021
- **~~independent of the location~~** Bias depends on the region
  since the threshold follows the spatial temperature distribution, Zhang et al. 2011; Schoetter et al. 2015
- **independent of the dataset**
  since the threshold provides an implicit bias correction. Freychet et al. 2021; Schoetter et al. 2015

*in sample

Relative temperature extreme definitions are used as implicit bias correction

The choice of a percentile-based threshold instead of a fixed threshold allows for an ***implicit bias correction of the climate model results****.* Schoetter et al. 2015

Relative temperature extreme definitions are used as implicit bias correction

*The choice of a percentile-based threshold instead of a fixed threshold allows for an **implicit bias correction of the climate model results**.* Schoetter et al. 2015

The use of separate thresholds for each dataset (e.g., observations and climate models) is intended to account for

- offsets in absolute temperature and
- differences in the temperature distribution.

Remaining differences in derived metrics such as cumulative heat and heatwave area or duration a are then attributed to non-linear model errors.

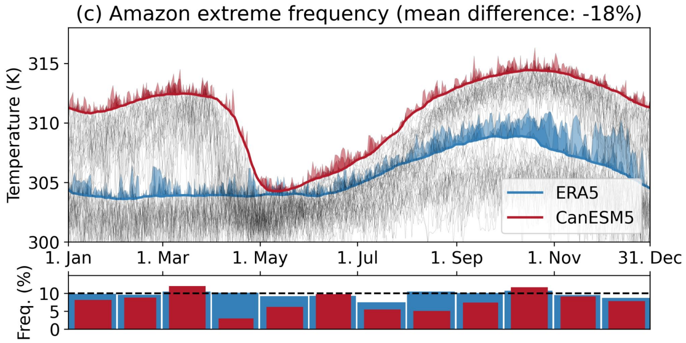
Relative temperature extreme definitions are used as implicit bias correction

*The choice of a percentile-based threshold instead of a fixed threshold allows for an **implicit bias correction of the climate model results.*** Schoetter et al. 2015

The use of separate thresholds for each dataset (e.g., observations and climate models) is intended to account for

- offsets in absolute temperature and
- differences in the temperature distribution.

Remaining differences in derived metrics such as cumulative heat and heatwave area or duration a are then attributed to non-linear model errors.



TX90p31w difference for one grid cell in the Amazon between CanESM5 and ERA5 due to differences in the mean seasonal cycle.

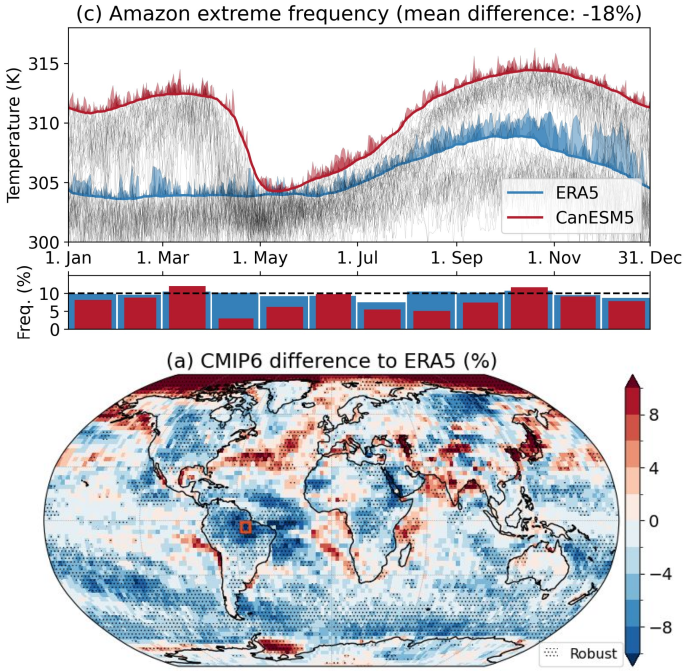
Relative temperature extreme definitions are used as implicit bias correction

*The choice of a percentile-based threshold instead of a fixed threshold allows for an **implicit bias correction of the climate model results.*** Schoetter et al. 2015

The use of separate thresholds for each dataset (e.g., observations and climate models) is intended to account for

- offsets in absolute temperature and
- differences in the temperature distribution.

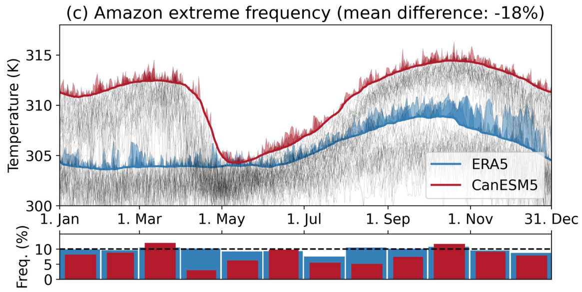
Remaining differences in derived metrics such as cumulative heat and heatwave area or duration a are then attributed to non-linear model errors.



(top) TX90p31w difference for one grid cell in the Amazon between CanESM5 and ERA5 due to differences in the mean seasonal cycle. (bottom) Mean difference over 26 CMIP6 models.

Properties of relative extreme definitions

When defining relative extremes based on a 90th percentile threshold we can expect *on average* 10% extreme frequency*



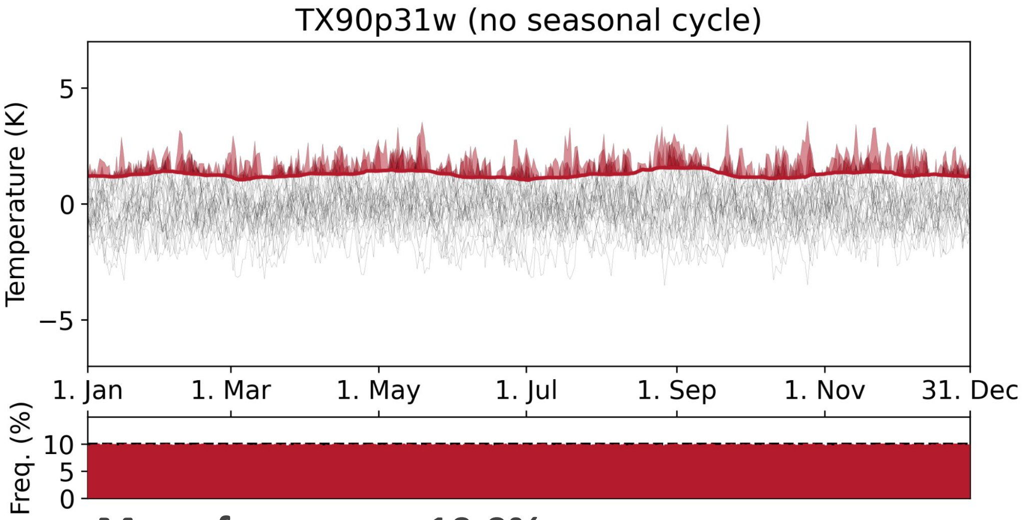
- ~~**independent of the season**~~ **Bias depends on the season**
  since the threshold follows the annual cycle, Tank and Können 2003; Fischer and Schär 2010; Hirsch et al. 2021
- ~~**independent of the location**~~ **Bias depends on the region**
  since the threshold follows the spatial temperature distribution, Zhang et al. 2011; Schoetter et al. 2015
- ~~**independent of the dataset**~~ **Bias depends on the dataset**
  since the threshold provides an implicit bias correction. Freychet et al. 2021; Schoetter et al. 2015

*in sample

Part 3: Eliminating the bias

The solution:
No seasonal cycle – no problem

The solution:
No seasonal cycle – no problem



Mean frequency: 10.0%
Mean bias: 0%

Without a seasonal cycle in the data, the bias disappears.

The solution:
No seasonal cycle – no problem

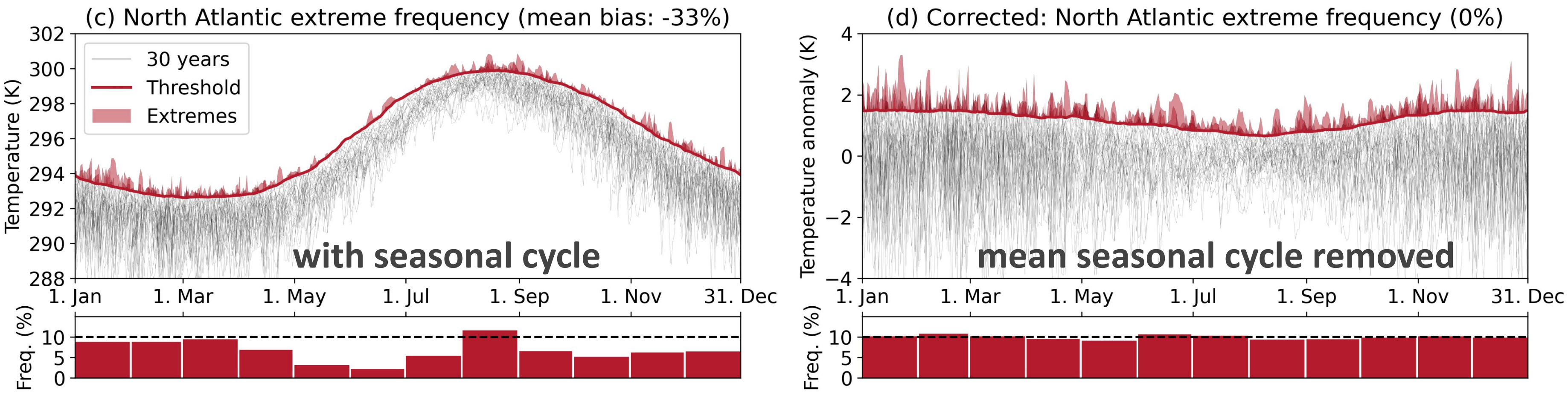
THE USE OF INDICES TO IDENTIFY CHANGES IN CLIMATIC EXTREMES

P.D. JONES[1], E.B. HORTON[2], C.K. FOLLAND[2], M. HULME[1], D.E. PARKER[2] and T.A. BASNETT[2]

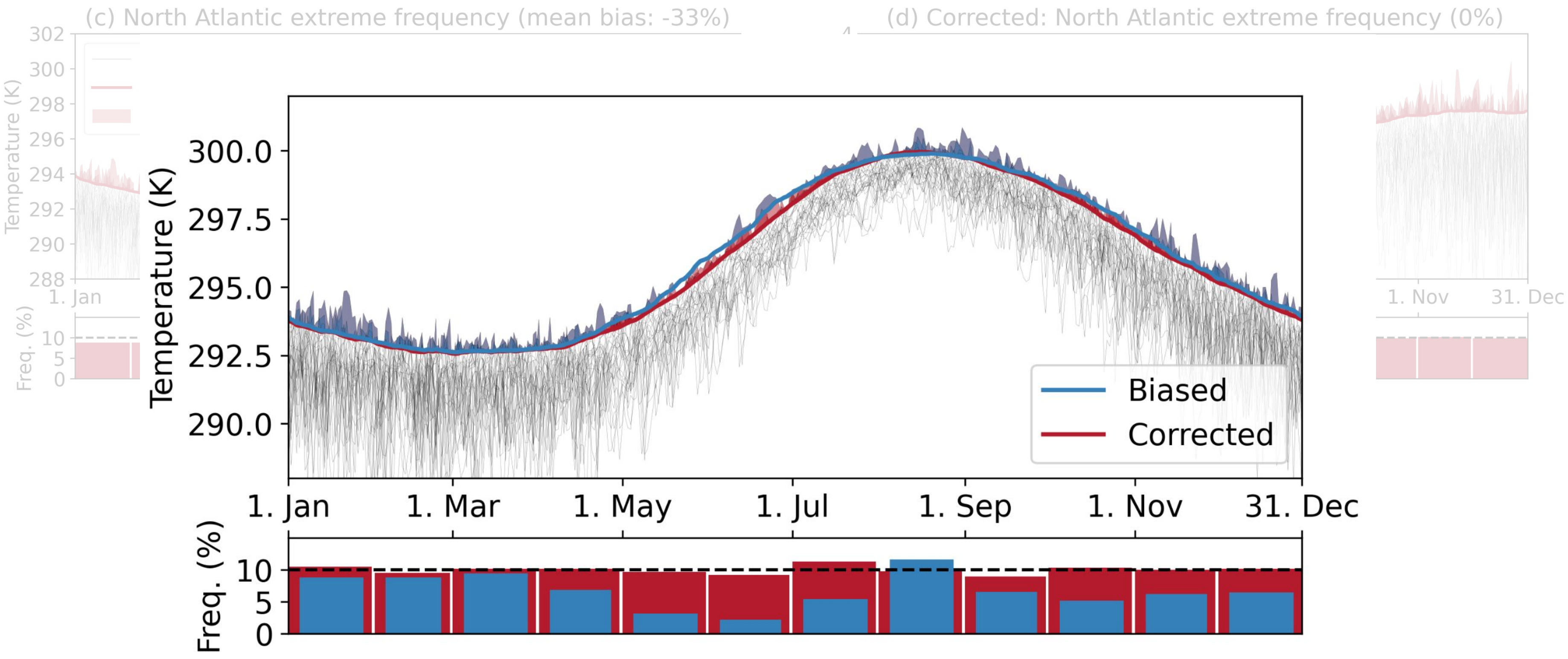
[1]*Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, U.K.*
[2]*Hadley Centre, Meteorological Office, Bracknell, RG12 2SY, U.K.*

As a first step, an average daily temperature value for each day of the year is derived from the period 1961-1990. [...] In the second step [a ***percentile*** *or return period]* ***is fitted*** *[...]* ***to the daily anomaly values*** *relative to the smoothed daily mean.* Jones et al. 1999

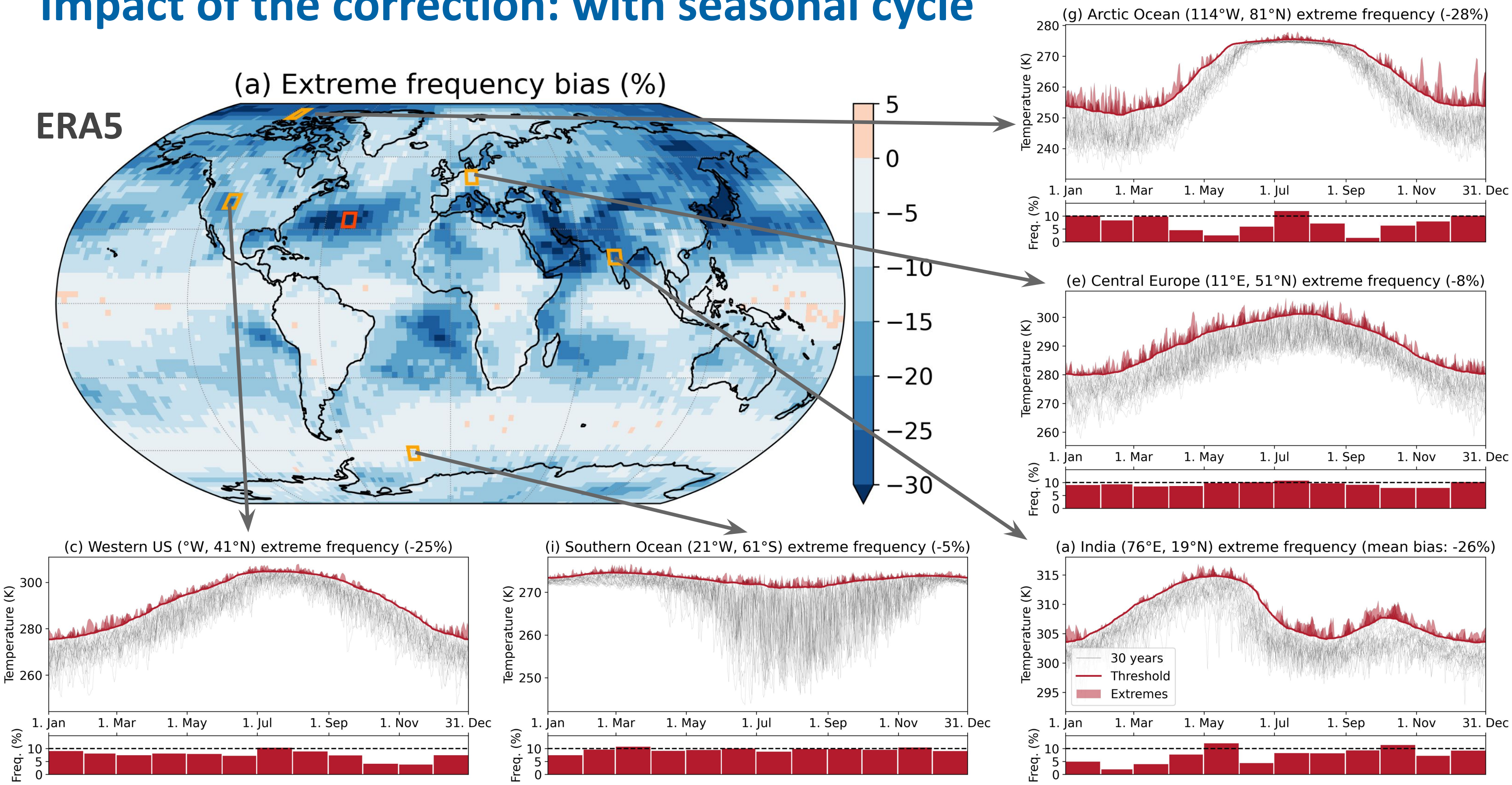
The solution:
No seasonal cycle – no problem



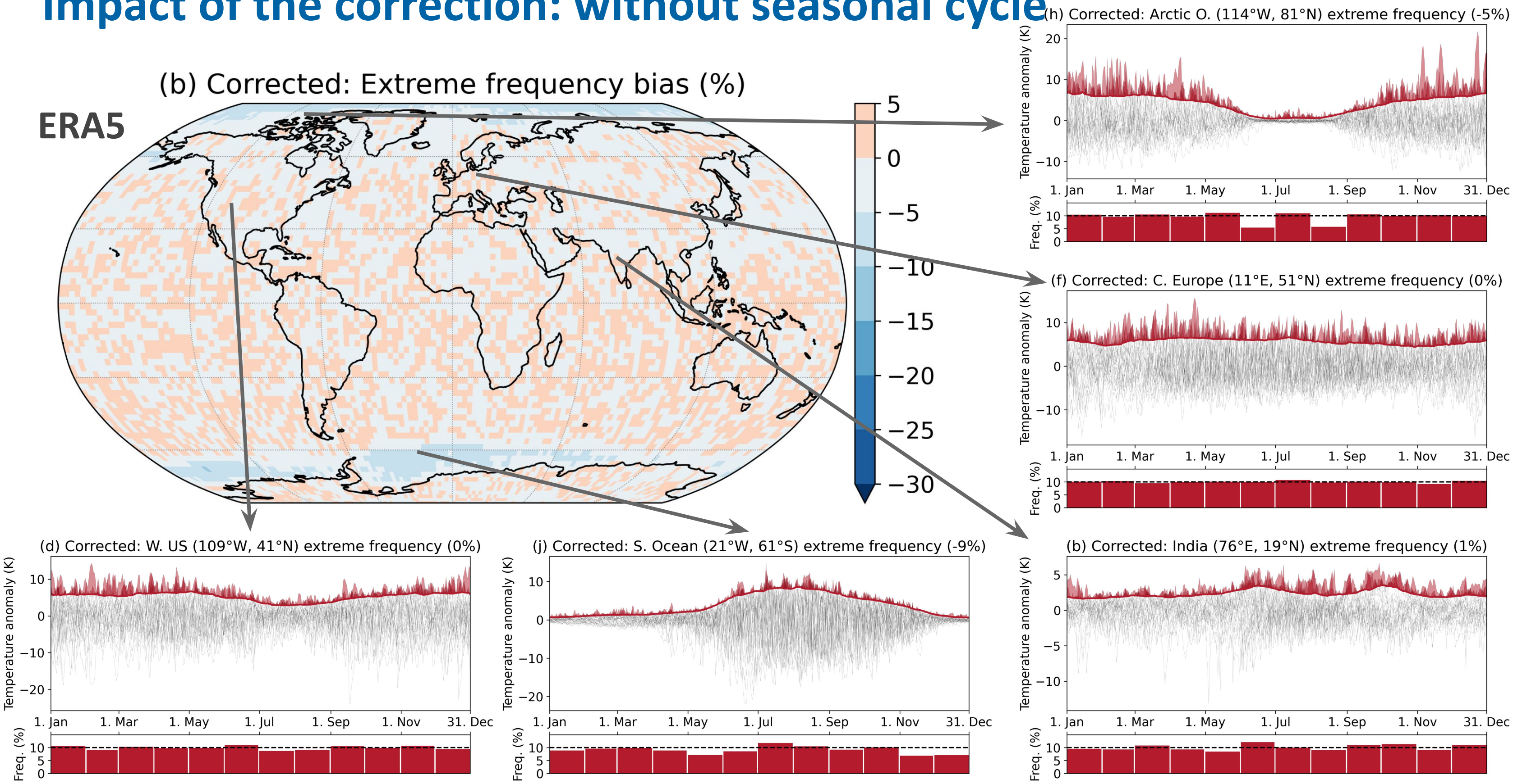
The solution:
No seasonal cycle (during threshold calculation) – no problem



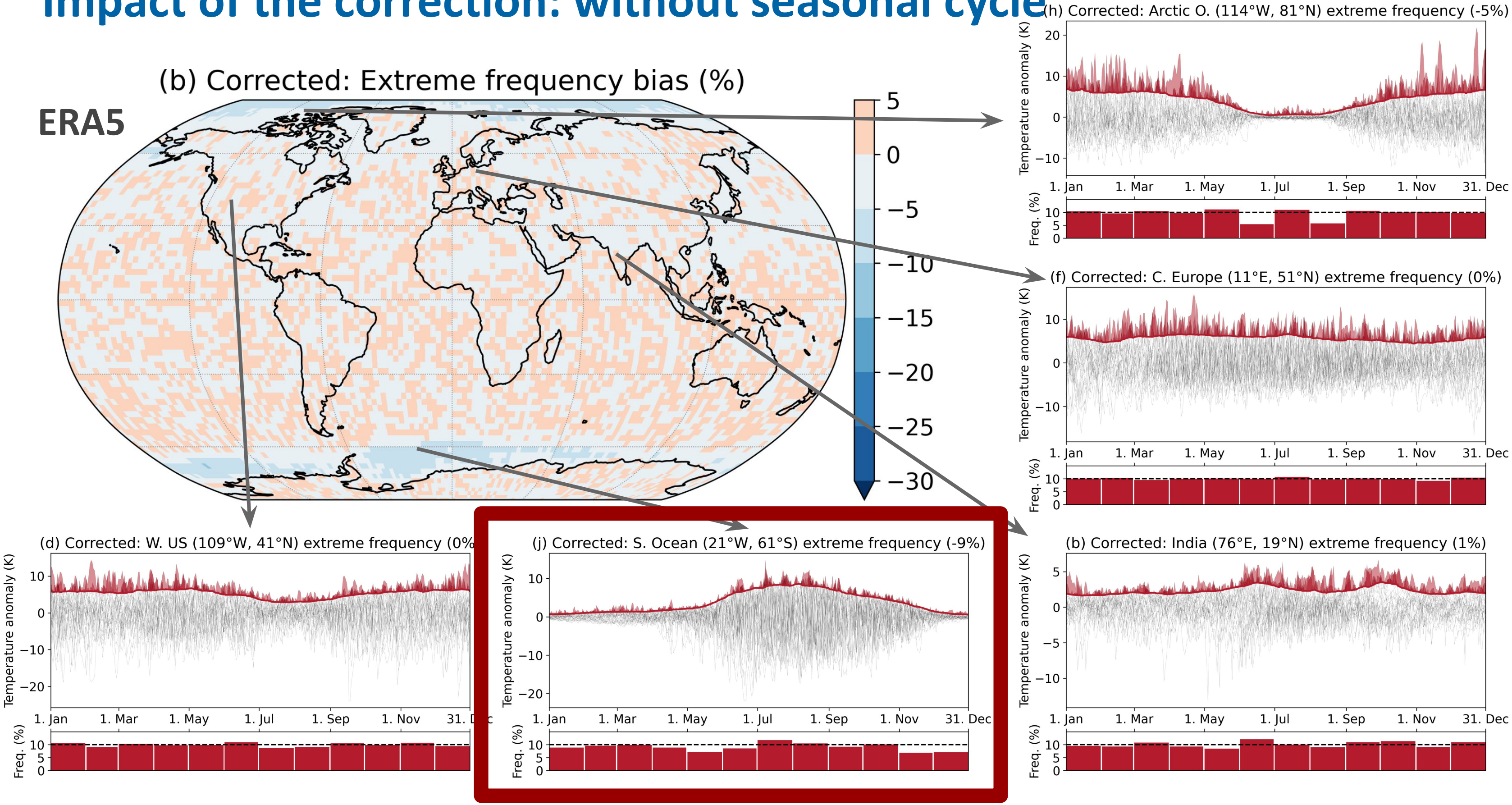
Impact of the correction: with seasonal cycle



Impact of the correction: without seasonal cycle

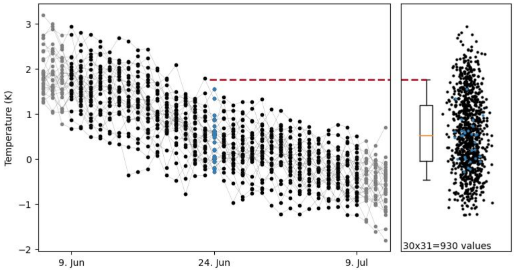


Impact of the correction: without seasonal cycle



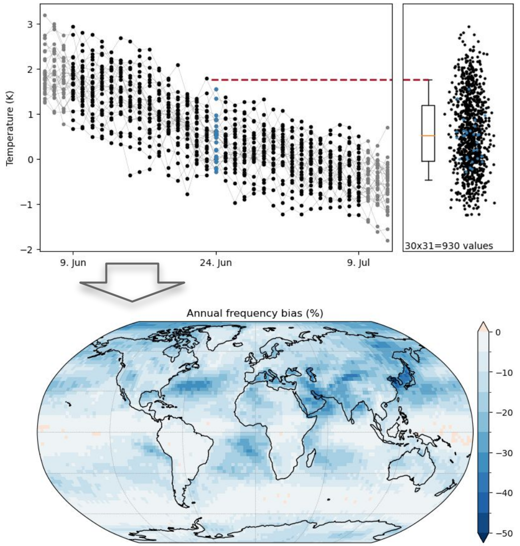
Summary and conclusions

- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**



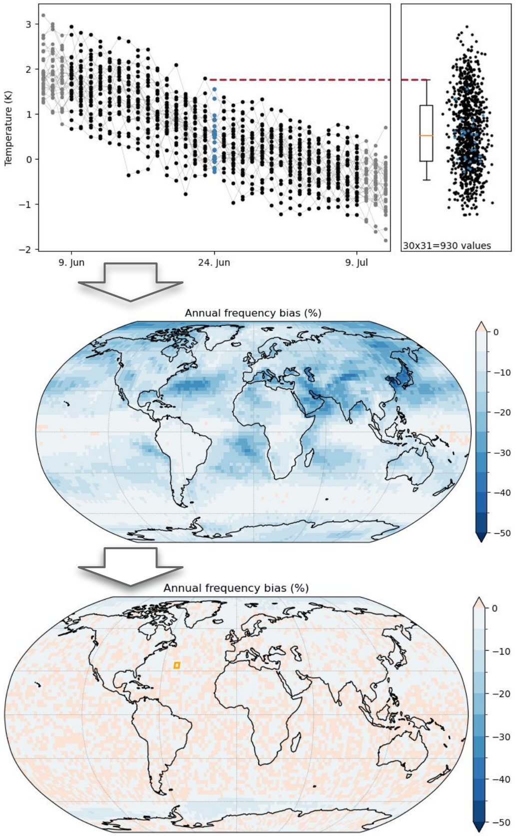
Summary and conclusions

- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**
- The bias varies across seasons, regions, datasets, and climatic states, **violating assumptions about properties of relative extreme definitions**



Summary and conclusions

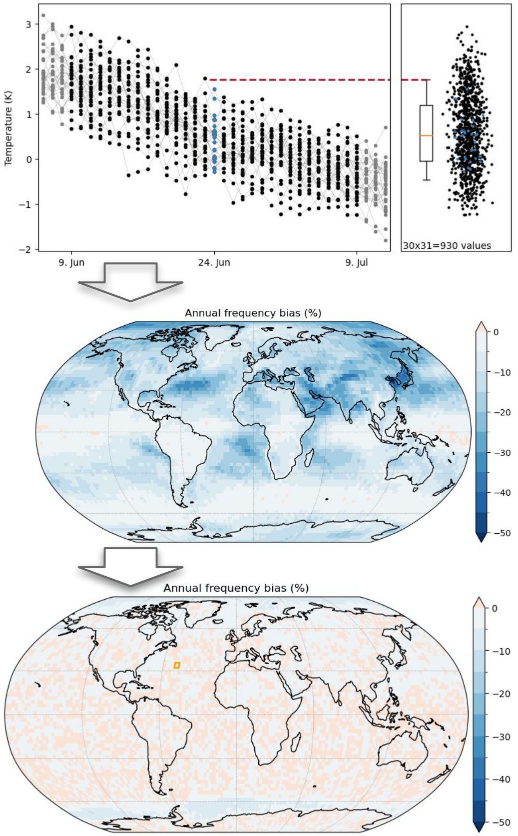
- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**
- The bias varies across seasons, regions, datasets, and climatic states, **violating assumptions about properties of relative extreme definitions**
- It is mostly eliminated by removing the mean seasonal cycle before calculating the extreme threshold



Summary and conclusions

- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**
- The bias varies across seasons, regions, datasets, and climatic states, **violating assumptions about properties of relative extreme definitions**
- It is mostly eliminated by removing the mean seasonal cycle before calculating the extreme threshold

*We strongly **warn against the use of long running windows without correction** when calculating extreme thresholds. The use of such a biased method is never advisable, even though the impacts on derived metrics might not always be strong or immediately apparent.* Brunner and Voigt (in press)



Bonus slides: More pitfalls

Bias impact on future change signals using a fixed 1961-1990 threshold

Question: How does the bias affect estimates of future extreme changes?

Problem: We don’t know what extreme frequency to expect in the future (out-of-base).

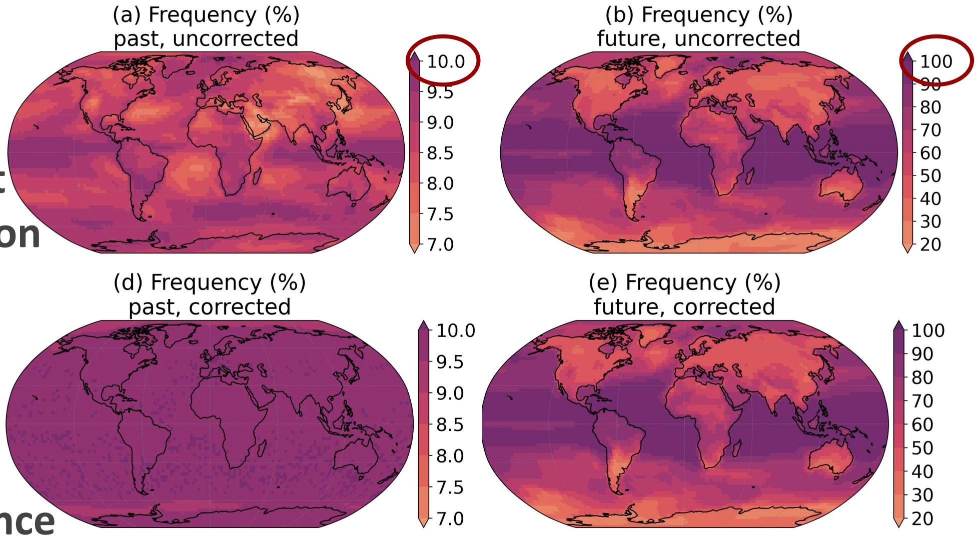
Bias impact on future change signals using a fixed 1961-1990 threshold

universität wien

Base period (1961-1990) | **Future period (2071-2100)**

Without correction

Reference (with correction)



Question: How does the bias affect estimates of future extreme changes?

Problem: We don’t know what extreme frequency to expect in the future (out-of-base).

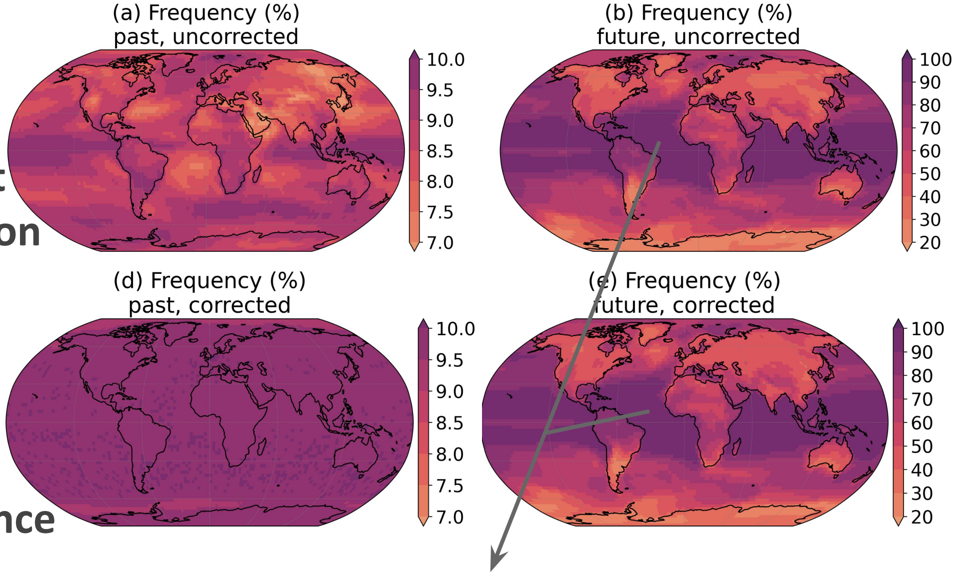
Solution: Use the corrected frequency as reference which is also available in the future.

Bias impact on future change signals using a fixed 1961-1990 threshold

Base period (1961-1990)

Future period (2071-2100)

Without correction



Reference

Question: How does the bias affect estimates of future extreme changes?

Problem: We don’t know what extreme frequency to expect in the future (out-of-base).

Solution: Use the corrected frequency as reference which is also available in the future.

In the future some regions have 100% extreme frequency → **bias must be 0%**

→ **The bias generally decreases with increasing extreme frequency!**

Bias impact on future change signals using a fixed 1961-1990 threshold

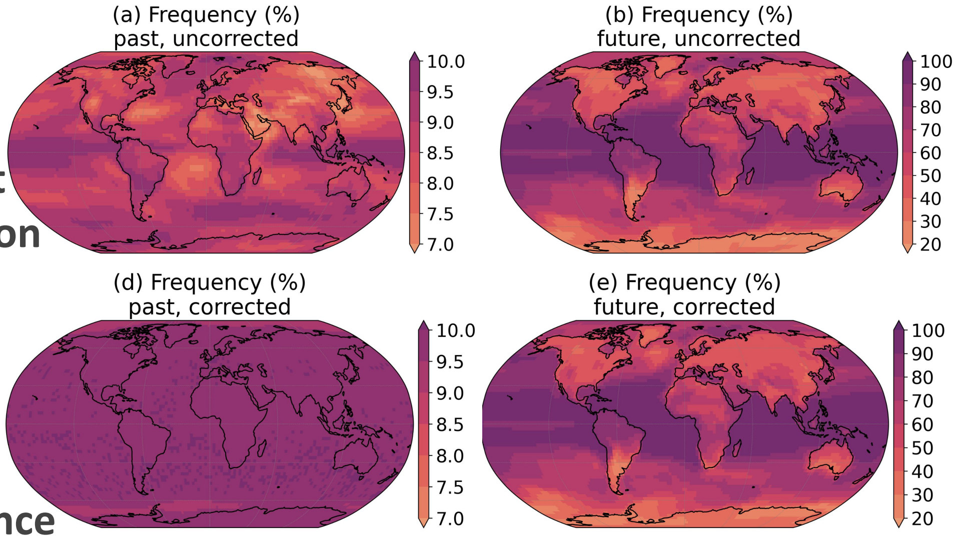
Change

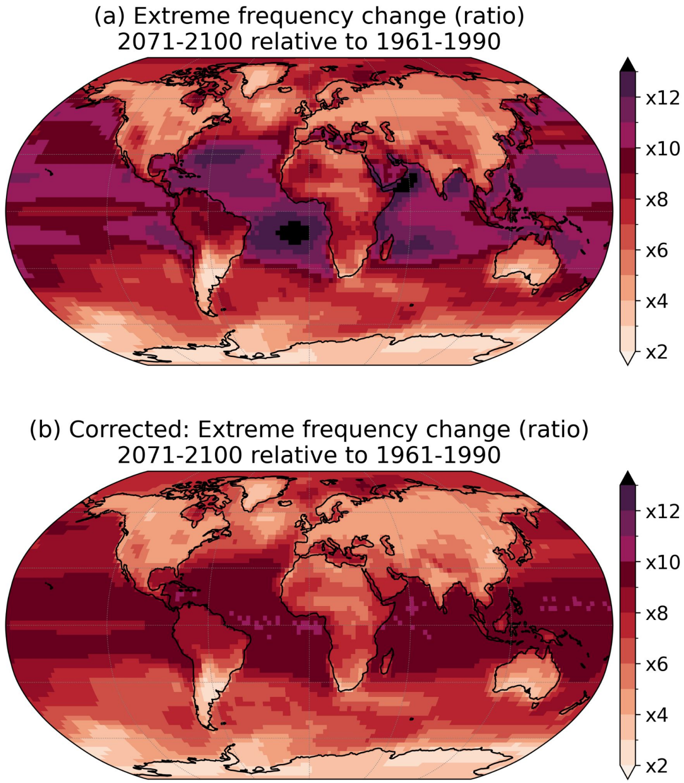
Base period (1961-1990)

Future period (2071-2100)

Without correction

Reference

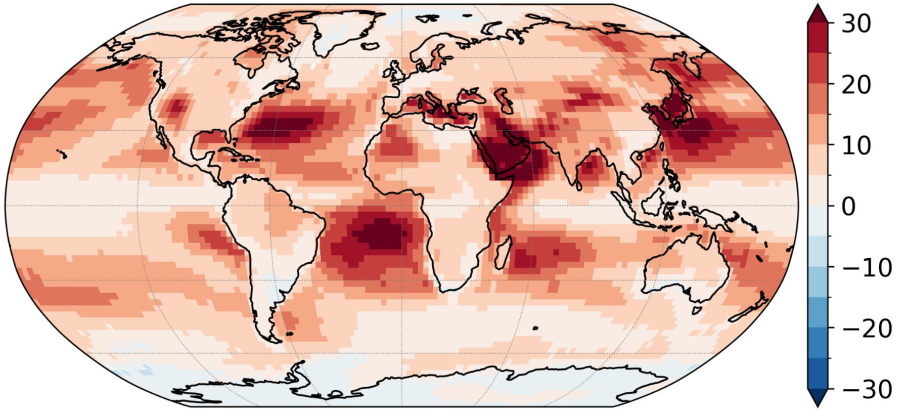




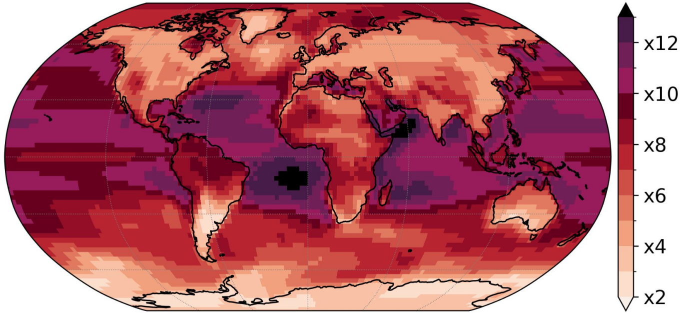
Bias impact on future change signals using a fixed 1961-1990 threshold

Change

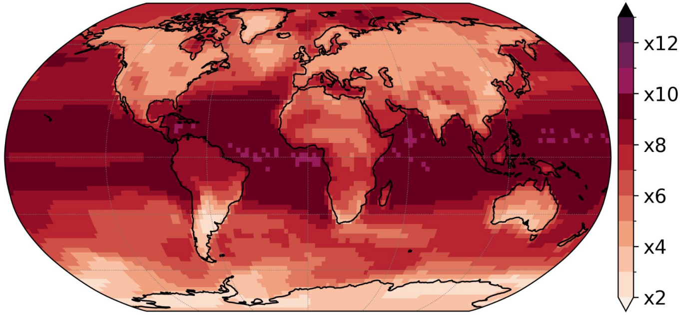
(c) Extreme frequency change bias (%)



(a) Extreme frequency change (ratio)
2071-2100 relative to 1961-1990



(b) Corrected: Extreme frequency change (ratio)
2071-2100 relative to 1961-1990

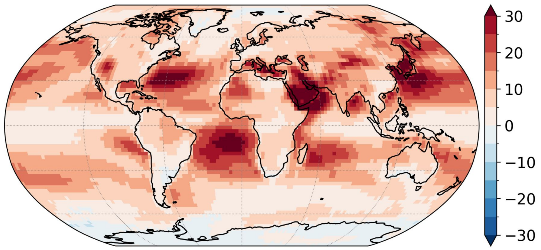


→ **the bias leads to an overestimation of extreme changes by up to 30%!**

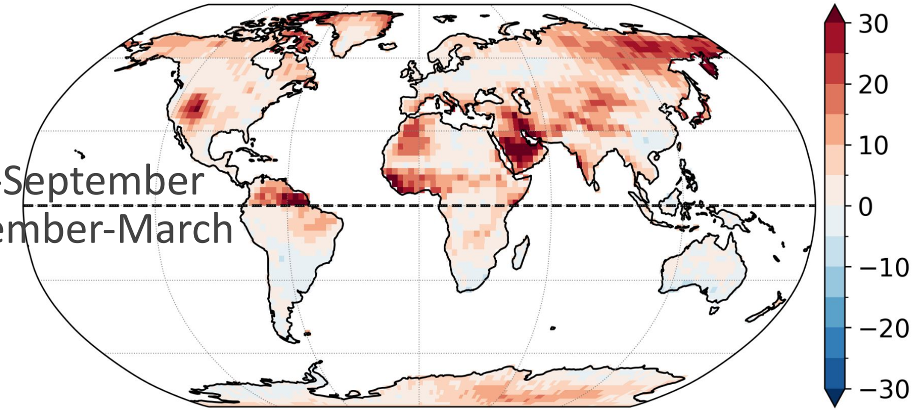
universität wien

Bias impact on summer heatwaves changes

(c) Extreme frequency change bias (%)



(d) Heatwave frequency change bias (%)
Extended summer land surface



Definition heatwave: At least 3 consecutive TX90p31w days.

Backup Slides

Effect of the window size

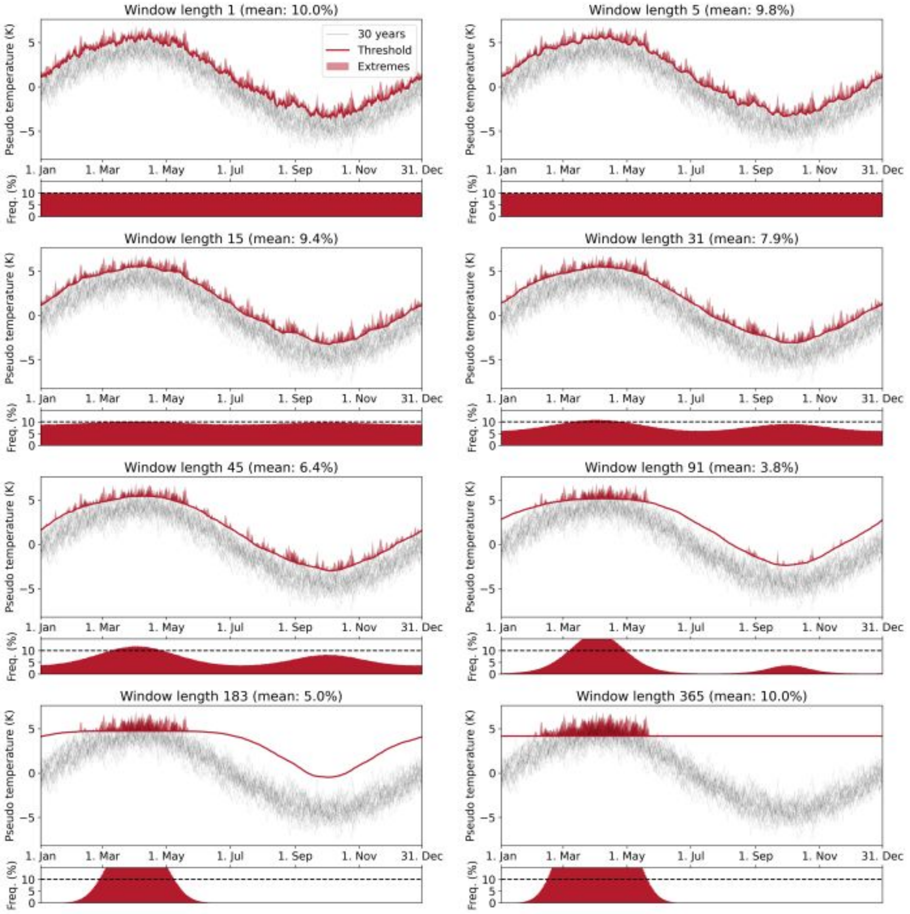
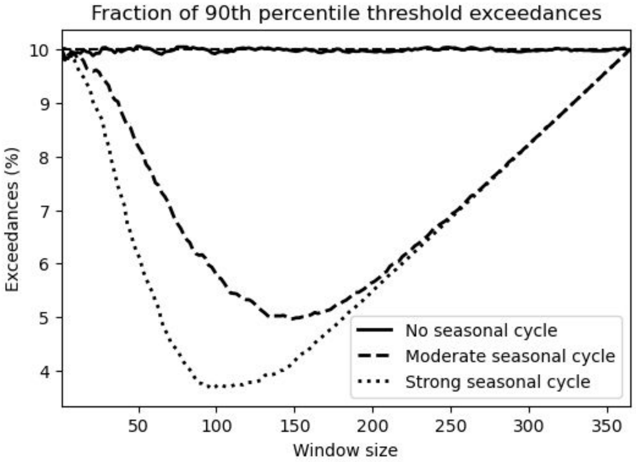
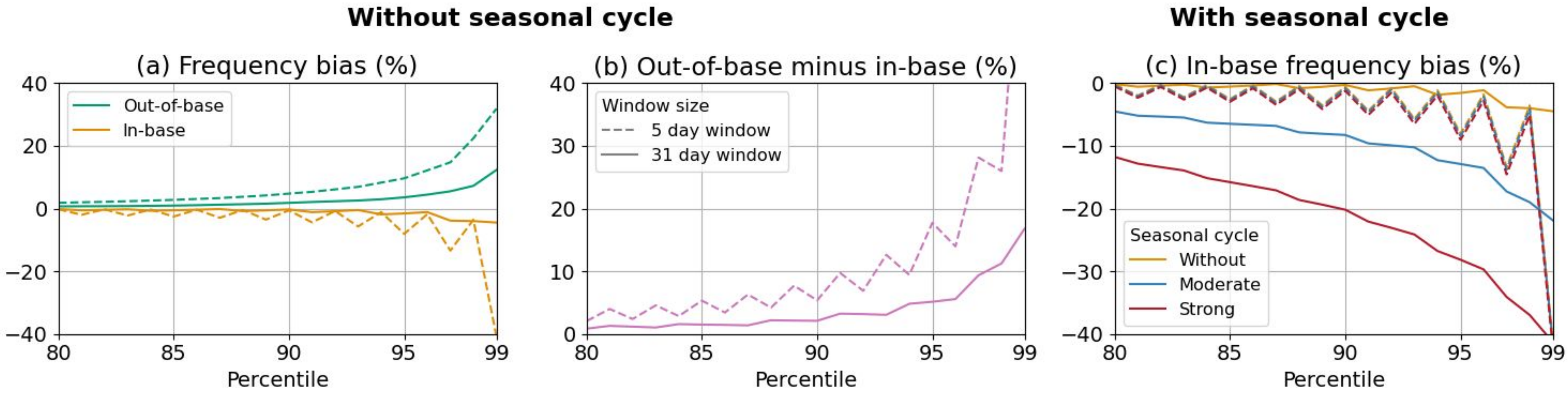


Figure S3: Threshold exceedances for different window sizes in synthetic data. Effect of different window sizes on the frequency of 90th percentile exceedances using the synthetic data with a strong seasonal cycle from figure 2 in the main manuscript. The respective top panels show threshold and exceedances for 30 seasonal cycles. The smaller bottom panels show exceedances for each day of the year averaged over all 5000 bootstrap samples.



The running window bias exceeds the well know in-base/out-of-base jump



Zhang et al. 2005

Brunner and Voigt (accepted)

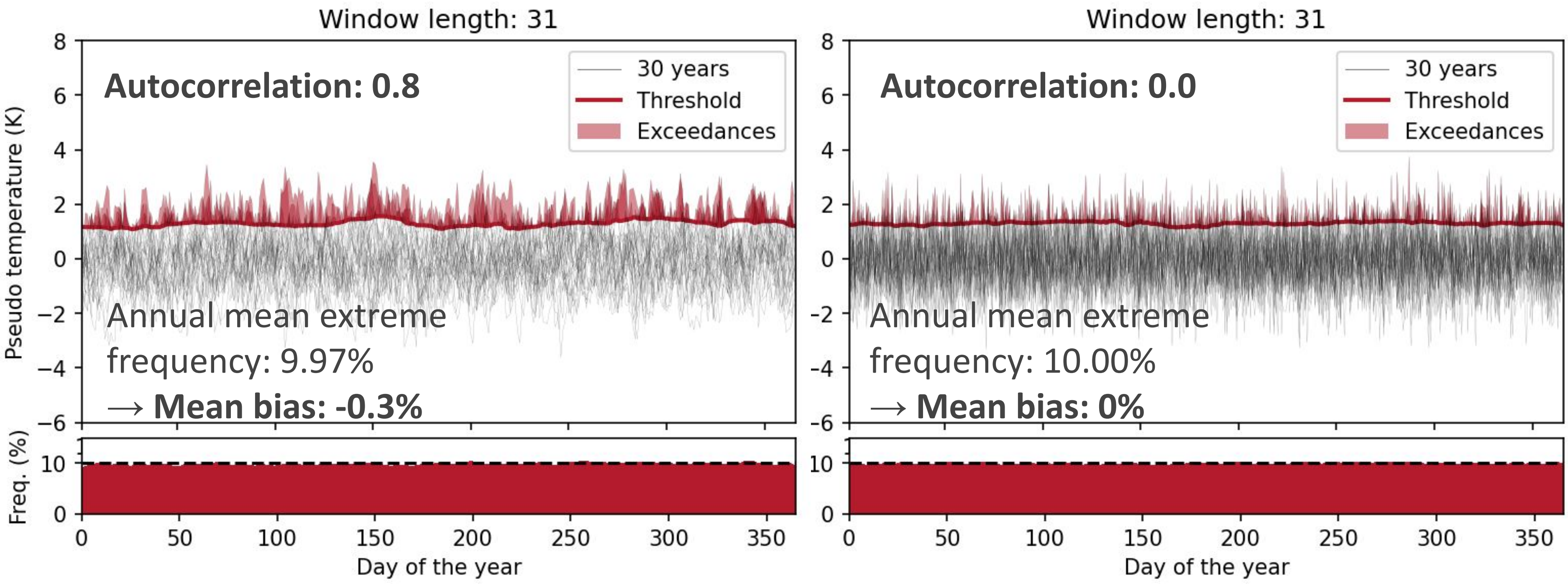
The extreme frequency difference between regions with high and low bias can reach about 25%

universität wien

(a) Spatial bias inhomogeneity

Running window size (day) \ Percentile	50	90	95	98	99
1	**Inhomogeneity** 5th/95th perc Mean	**0.0%** 0.0%/0.0% 0.0%	**0.0%** 33.3%/33.3% 33.3%	**0.0%** 66.7%/66.7% 66.7%	**0.0%** 233.3%/233.3% 233.3%
5	**0.6%** -0.3%/0.3% 0.0%	**2.6%** -2.2%/0.4% -0.8%	**4.6%** 2.6%/7.2% 5.0%	**9.6%** -8.2%/1.4% -3.3%	**19.2%** 13.2%/32.4% 23.1%
15	**1.1%** -0.6%/0.5% 0.0%	**7.9%** -7.9%/-0.1% -3.1%	**12.1%** -10.0%/2.1% -2.7%	**18.7%** -18.3%/0.5% -7.5%	**24.7%** -14.2%/10.5% 0.1%
31	**2.7%** -1.7%/1.0% -0.1%	**23.5%** -24.7%/-1.2% -10.4%	**32.0%** -32.6%/-0.6% -13.8%	**41.6%** -41.6%/0.0% -18.1%	**47.5%** -44.3%/3.2% -18.1%
45	**4.3%** -2.8%/1.6% -0.2%	**36.6%** -39.0%/-2.4% -18.0%	**45.5%** -48.1%/-2.6% -23.3%	**53.0%** -57.5%/-4.6% -30.4%	**57.5%** -60.7%/-3.2% -32.5%

Autocorrelation leads to a small bias even without seasonal cycle

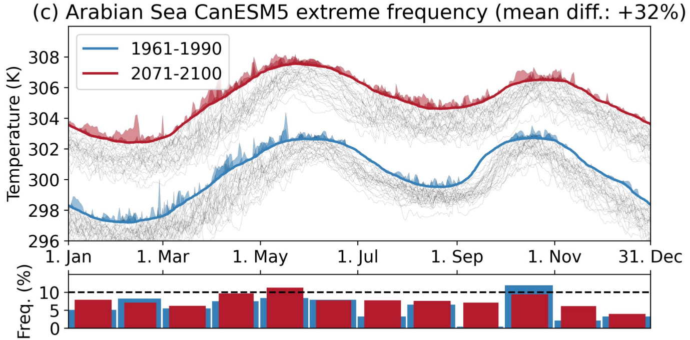


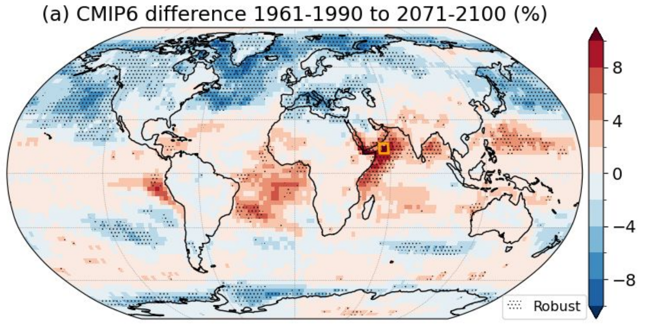
Relative temperature extreme definitions are intended to offset distributional shifts due to climate change

*[Relative thresholds with shifting base-periods] can be seen as a proxy for full adaptation to the respective prevailing future climate. [...] **Changes in [heatwave] duration** with [such] thresholds would be **related to** physical drivers of heatwaves such as **circulation changes or land-atmosphere feedbacks**.* Vogel et al. (2020)

When using two base-periods, with separate thresholds frequencies in both periods are assumed to be about 10%.

Changes in the shape of the seasonal cycle under warming can lead to a shift in the bias and, hence to differences.





(top) TX90p31w difference for one grid cell in the Arabian Sea between CanESM5 in the period 1961-1990 and 2071-2100. (bottom) Mean difference over 26 CMIP6 models. Brunner und Voigt (in review)

Full disclosure:
We are not the first to come up with this

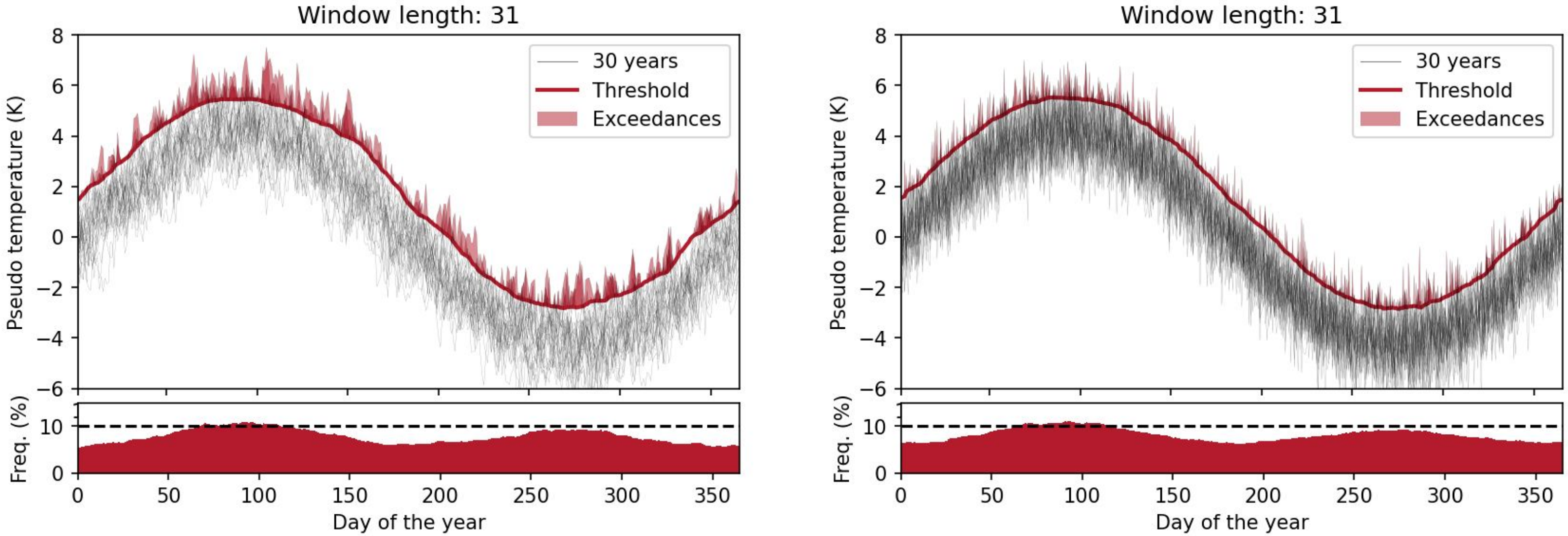
THE USE OF INDICES TO IDENTIFY CHANGES IN CLIMATIC EXTREMES

P.D. JONES[1], E.B. HORTON[2], C.K. FOLLAND[2], M. HULME[1], D.E. PARKER[2] and T.A. BASNETT[2]

[1]Climatic Research Unit, University of East Anglia, Norwich, NR4 7TJ, U.K.
[2]Hadley Centre, Meteorological Office, Bracknell, RG12 2SY, U.K.

As a first step, an average daily temperature value for each day of the year is derived from the period 1961-1990. [...] In the second step [a ***percentile*** *or return period]* ***is fitted*** *[...]* ***to the daily anomaly values*** *relative to the smoothed daily mean*

Jones et al. 1999



White noise with lag 1 day autocorrelation 0.8
→ 7.9%

White noise no autocorrelation
→ 8.3%