Revealing a systematic bias in percentile-based temperature extremes

Lukas Brunner and Aiko Voigt

Department of Meteorology and Geophysics, University of Vienna

Definition of percentile-based temperature extremes

Percentile-based extremes are typically defined as exceedance of a relative threshold

- for each day of the year and
- grid cell or region

based on

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

Expert Team on Climate Change Detection and Indices (ETCCDI)

Properties of percentile-based temperature extremes

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

*in sample

Properties of percentile-based temperature extremes

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

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

*in sample

Properties of percentile-based temperature extremes

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

- **independent of the season**
  since the threshold follows the seasonal 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

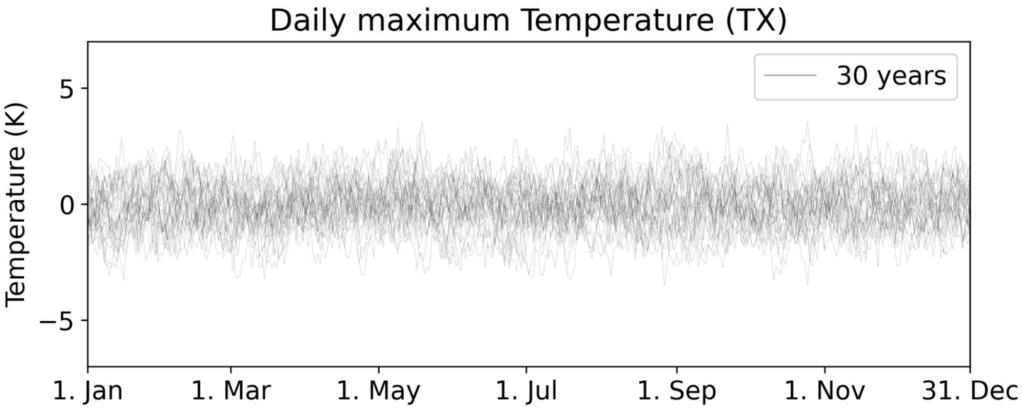
Properties of percentile-based temperature extremes

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

- **independent of the season**
  since the threshold follows the seasonal 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

Temperature extremes in a synthetic time series

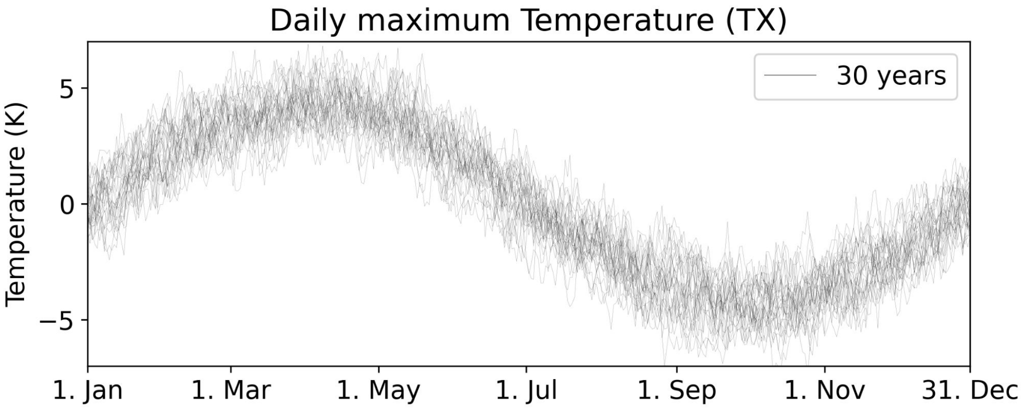


Synthetic temperature

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

Following Zhang et al. 2005

Temperature extremes in a synthetic time series



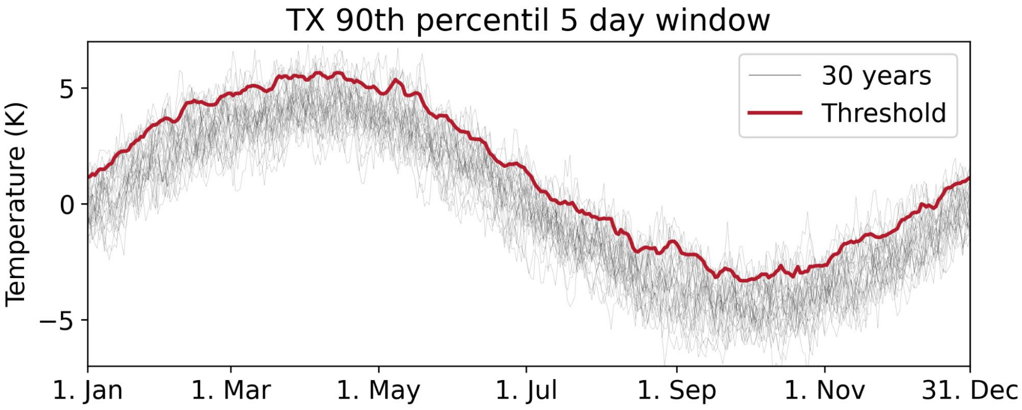
Synthetic temperature

- 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

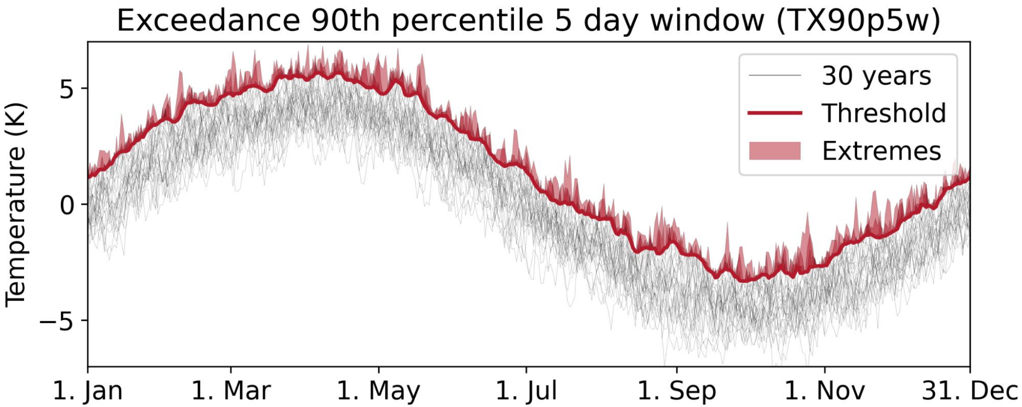
Temperature extremes in a synthetic time series



ETCCDI threshold:

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

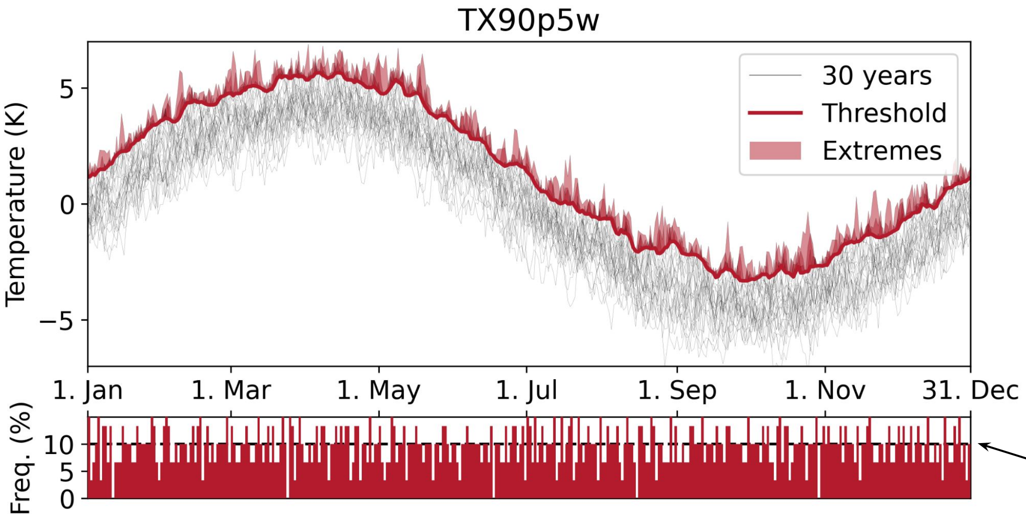
Temperature extremes in a synthetic time series



ETCCDI threshold:

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

Temperature extremes in a synthetic time series



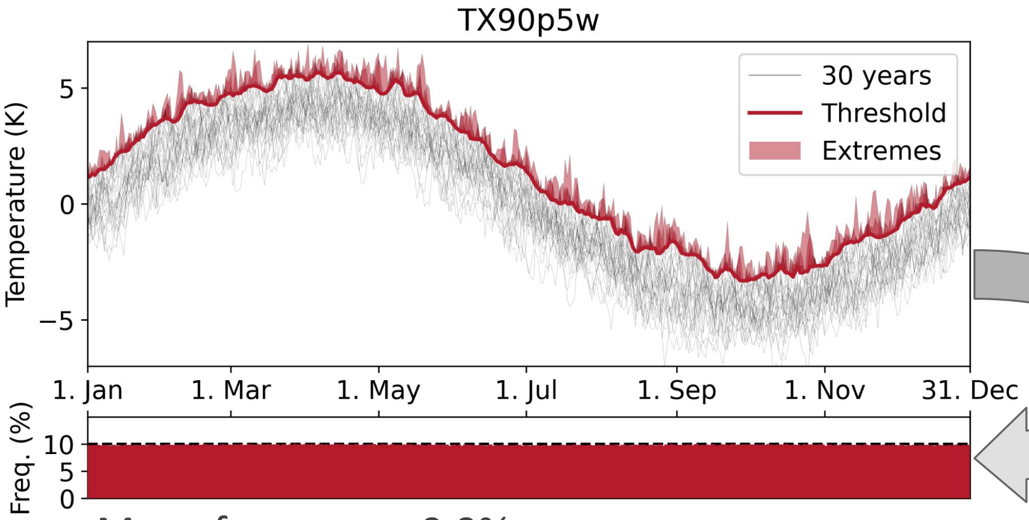
ETCCDI threshold:

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

expected frequency: 10%

Temperature extremes in a synthetic time series

universität wien

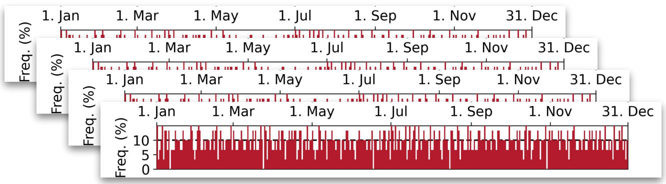


Mean frequency: 9.9%

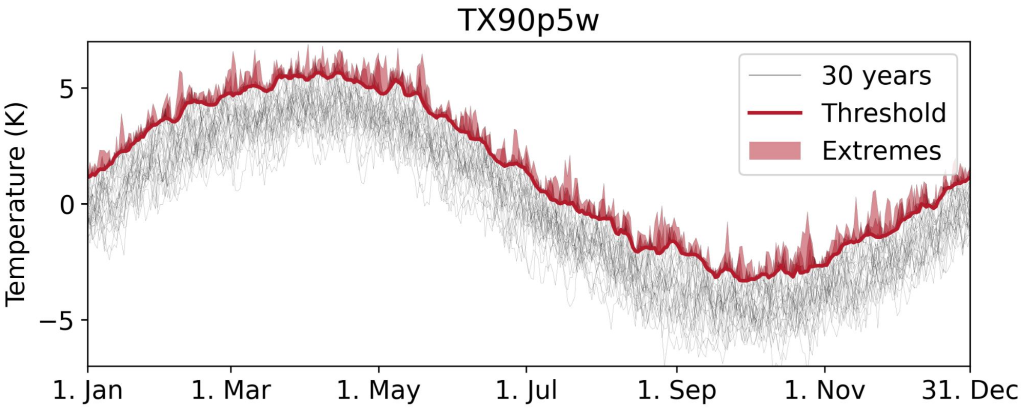
ETCCDI threshold:

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

repeat 5’000 times & average



Problems with the ETCCDI extreme threshold

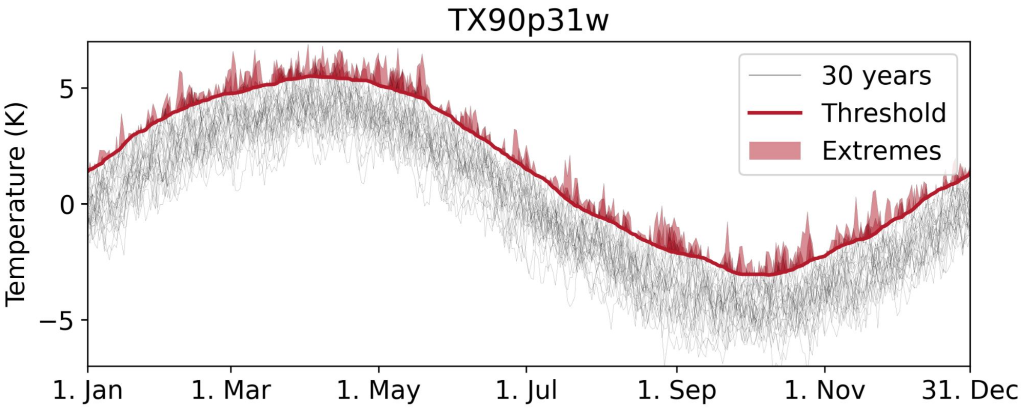


ETCCDI threshold:

- 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*** *[...].* Lyon et al. 2019

Many studies use longer windows to smooth the threshold



ETCCDI threshold:

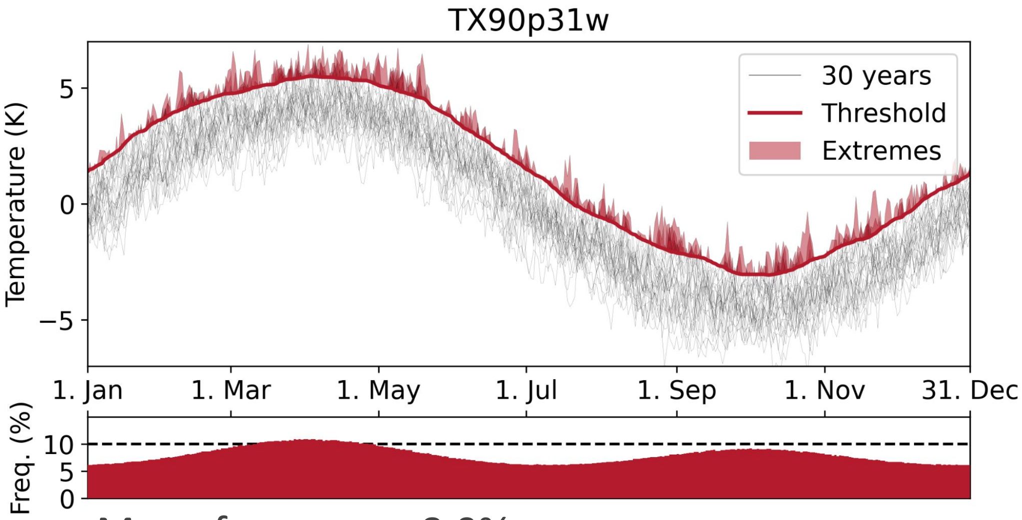
- 90th percentile
- 30 year
- **15 day running window** (not shown here)

E.g., 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; Luo et al. 2024; Perkins-Kirkpatrick et al. 2024

- **31 day running window**

E.g., 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 use longer windows to smooth the threshold and introduce a bias into the extreme frequency



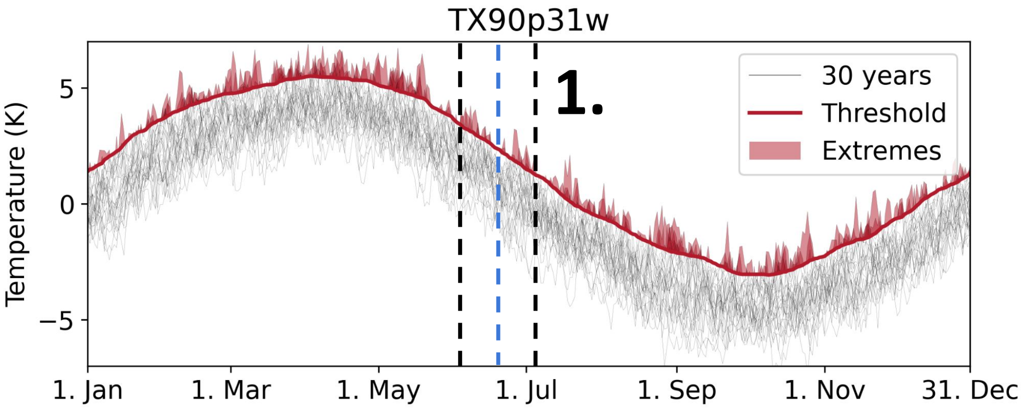
Mean frequency: 8.0%

Relative bias: -20%

ETCCDI threshold:

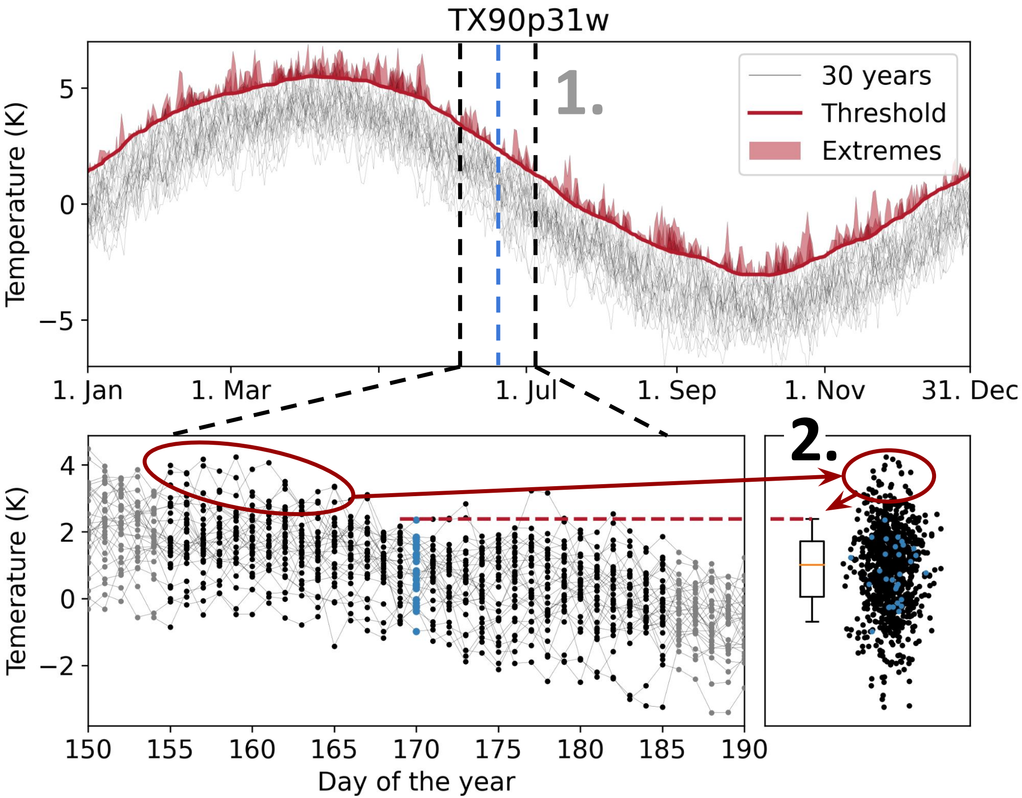
- 90th percentile
- 30 year
- 31 day running window

The bias originates from too long windows mixing the seasonal cycle into the extreme threshold



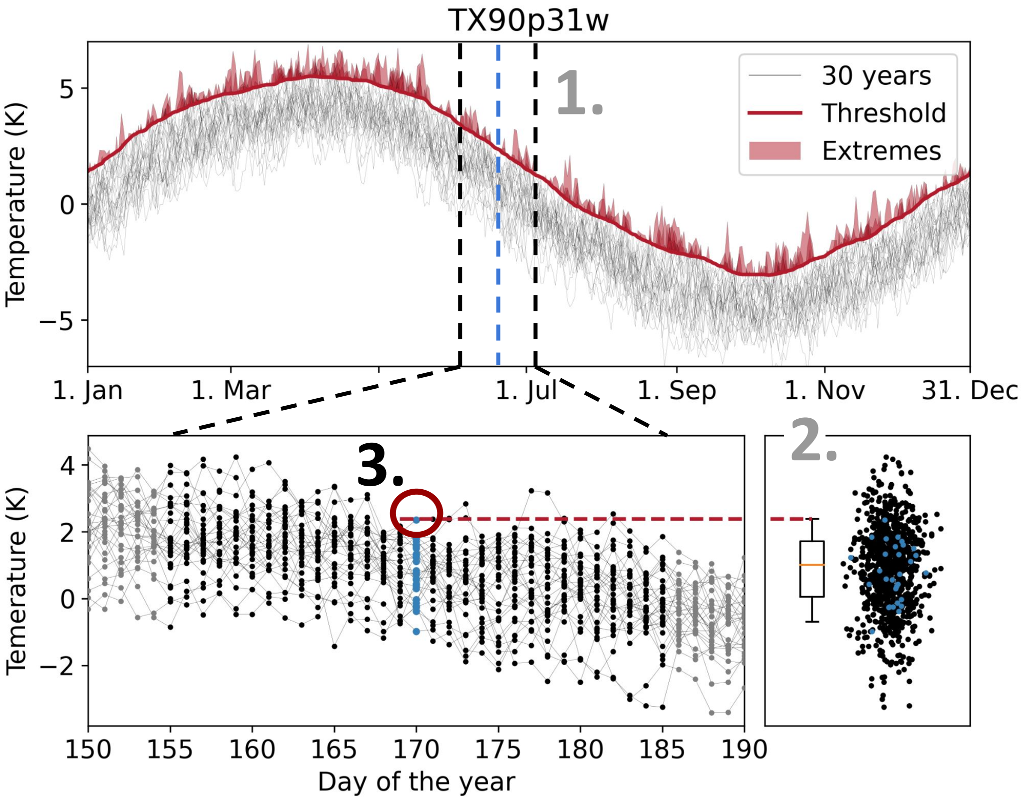
1. The threshold for the **day of the year 170** is calculated using ±15 days (= 31 day window)

The bias originates from too long windows mixing the seasonal cycle into the extreme threshold



1. The threshold for the **day of the year 170** is calculated using ±15 days (= 31 day window)
2. **Seasonally warmer days** within the window **dominate the 90th percentile** threshold

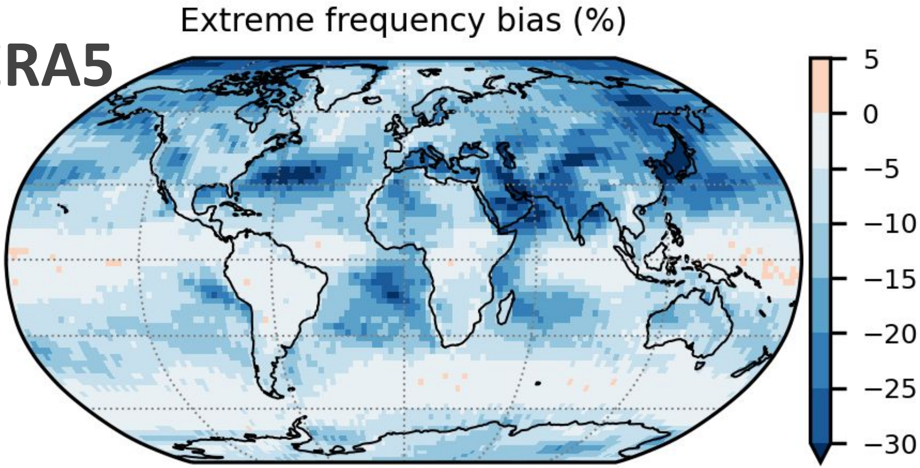
The bias originates from too long windows mixing the seasonal cycle into the extreme threshold



1. The threshold for the **day of the year 170** is calculated using ±15 days (= 31 day window)
2. **Seasonally warmer days** within the window **dominate the 90th percentile** threshold
3. Only **1 out of 30 (3%) samples exceeds the threshold** for this example

The bias varies regionally

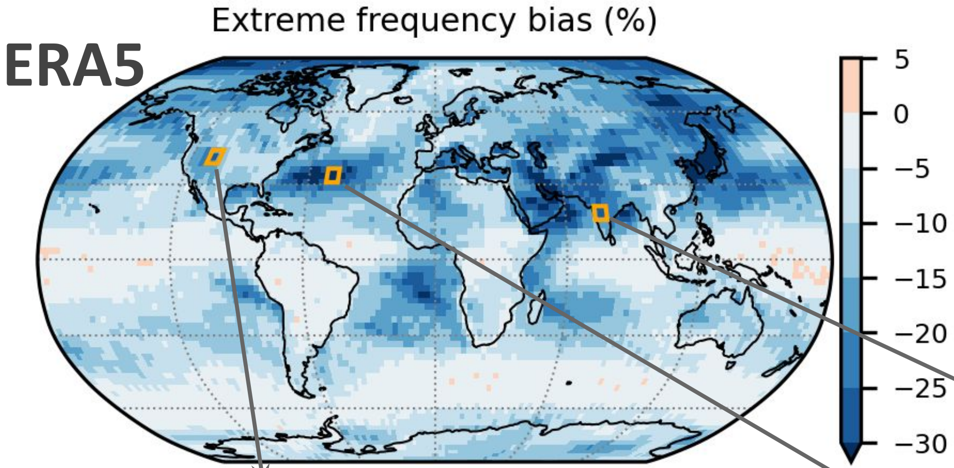
ERA5



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

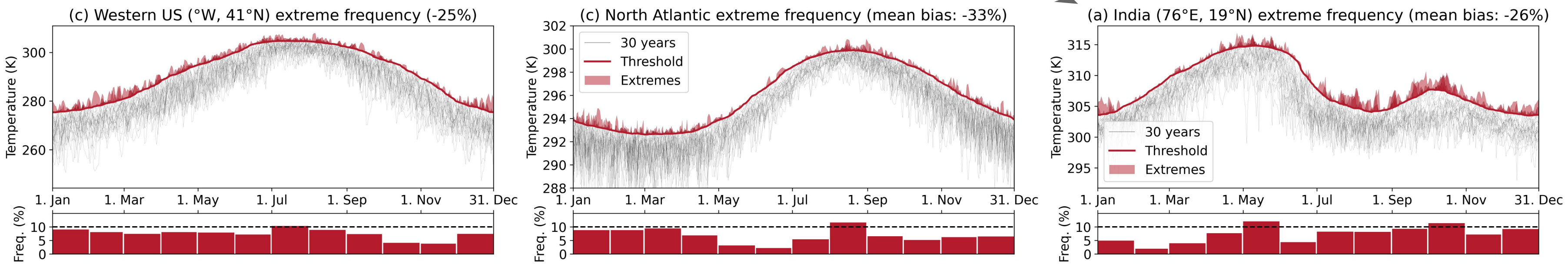
Regionally the bias can exceed **-30%**

The bias varies regionally and seasonally



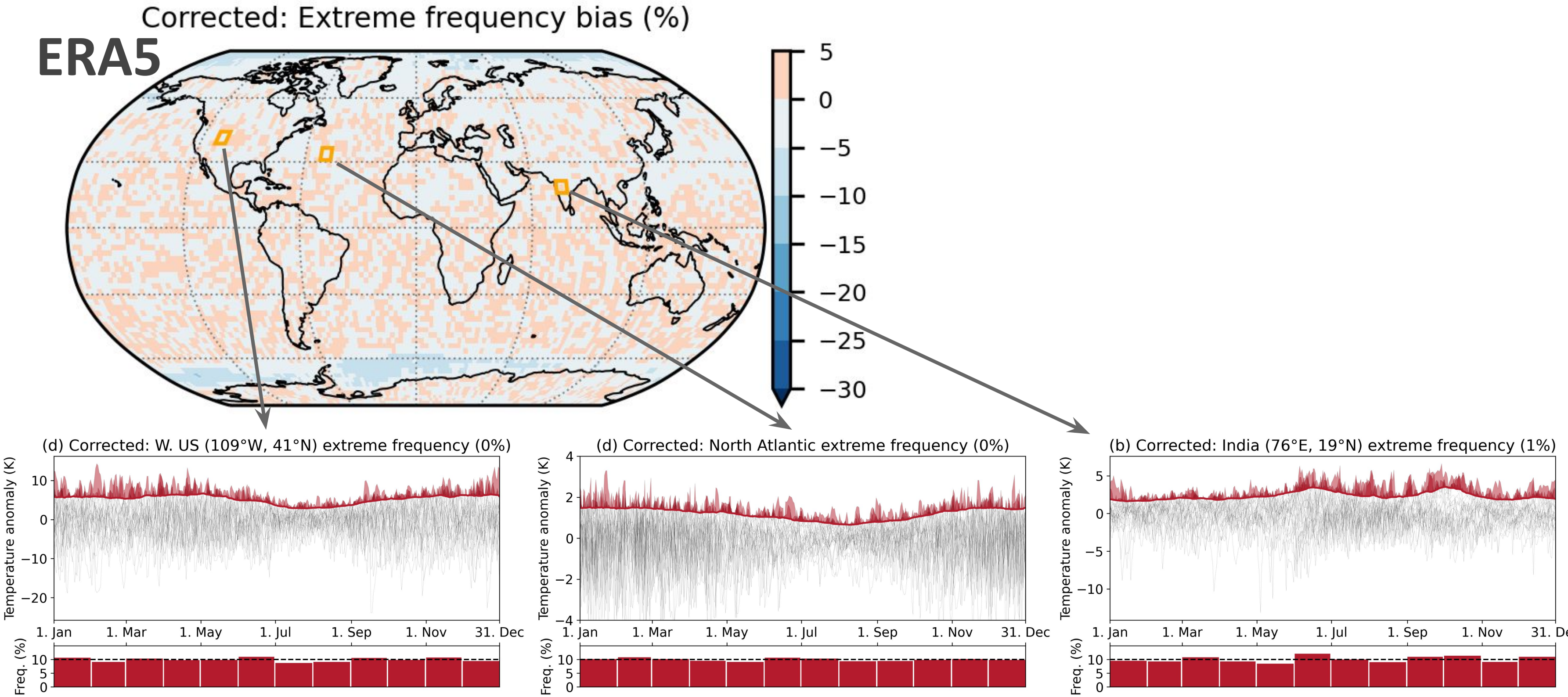
The **global mean bias** in the 30 year period 1961-1990 in ERA5 is **-10%**

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



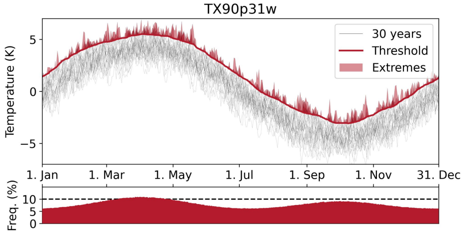
The bias varies regionally and seasonally and can be corrected by removing the seasonal cycle

ERA5



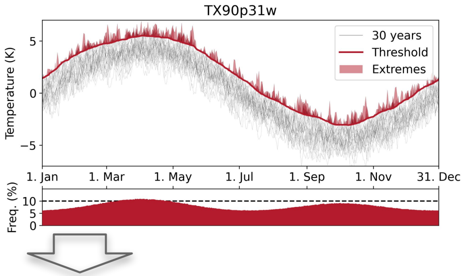
Summary and conclusions

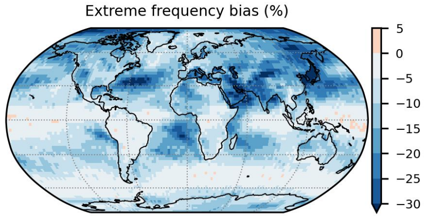
- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**



Summary and conclusions

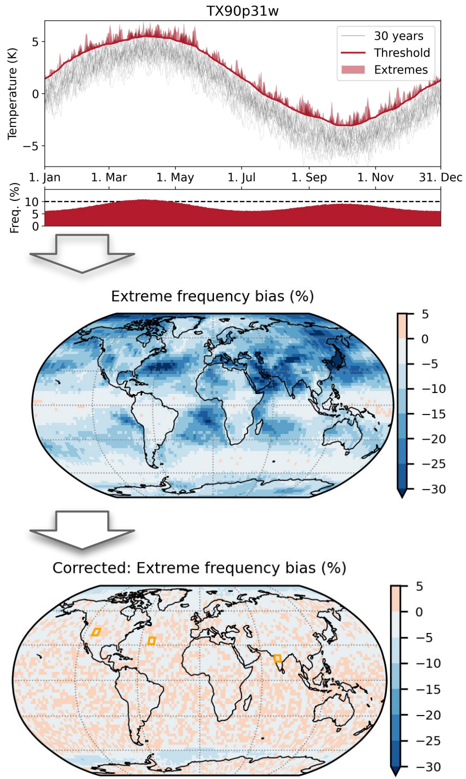
- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**
- The bias varies across seasons, regions, datasets, and climatic states, **violating assumptions about properties of relative extreme definitions**





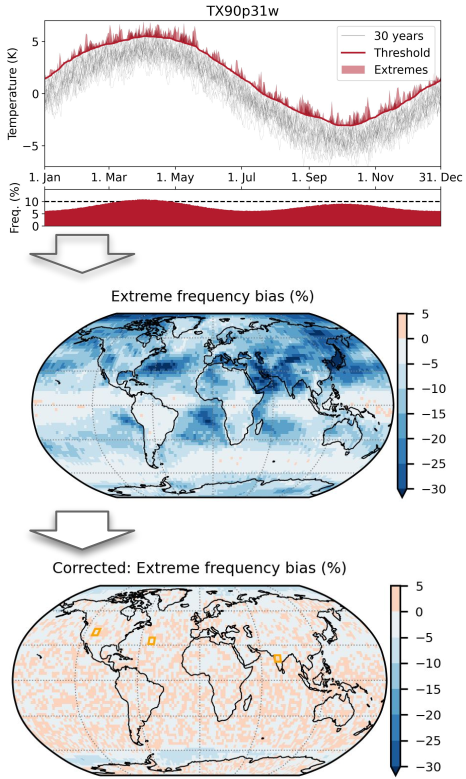
Summary and conclusions

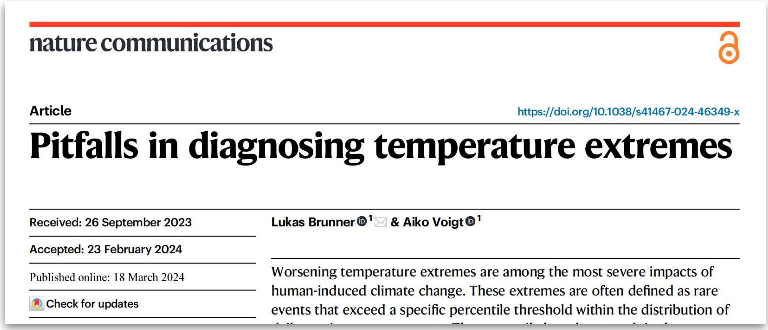
- An interaction between running windows and the seasonal cycle leads to a considerable **bias in temperature extremes**
- The bias varies across seasons, regions, datasets, and climatic states, **violating assumptions about properties of relative extreme definitions**
- It is mostly eliminated by removing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





nature communications

Article https://doi.org/10.1038/s41467-024-46349-x

Pitfalls in diagnosing temperature extremes

Received: 26 September 2023

Accepted: 23 February 2024

Published online: 18 March 2024

Check for updates

Lukas Brunner[1] & Aiko Voigt[1]

Worsening temperature extremes are among the most severe impacts of human-induced climate change. These extremes are often defined as rare events that exceed a specific percentile threshold within the distribution of

Bonus slides

Effect of the window size

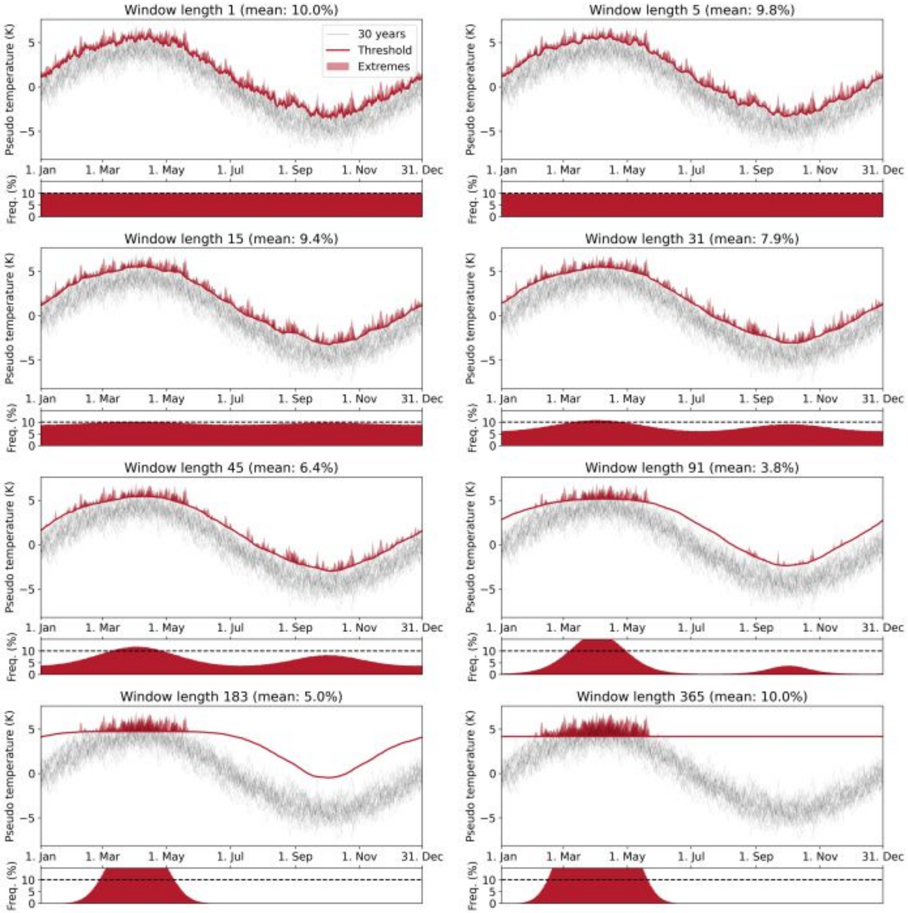
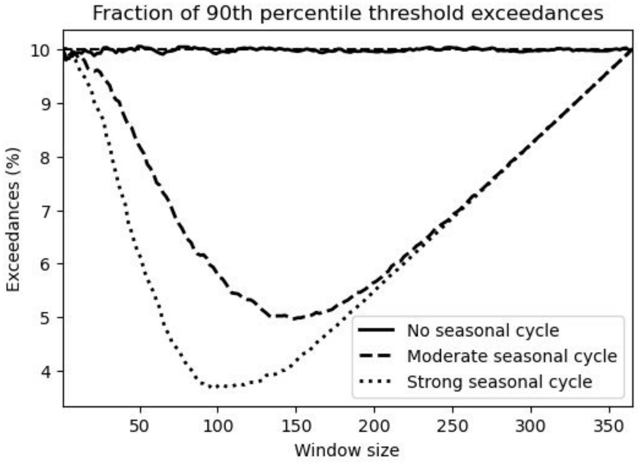
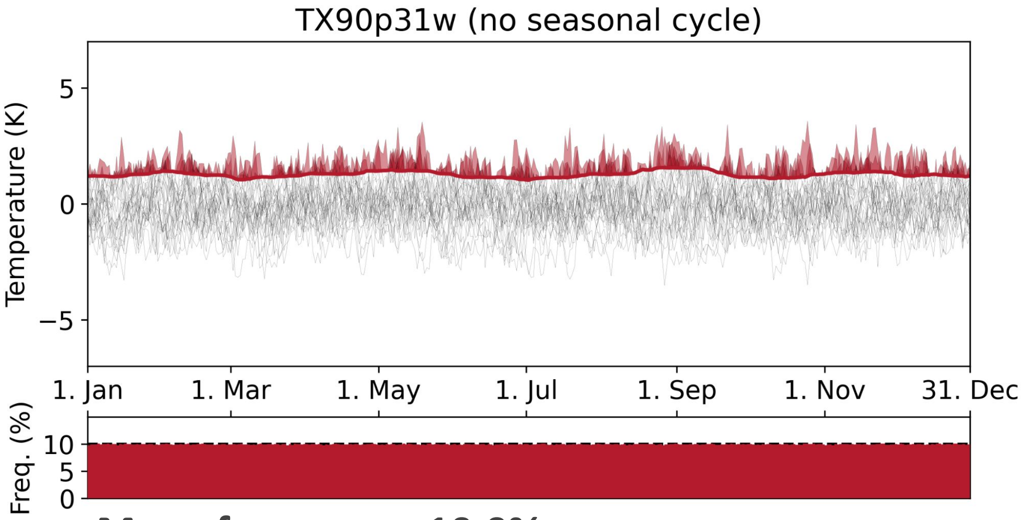


Figure S3: Threshold exceedances for different window sizes in synthetic data. Effect of different window sizes on the frequency of 90th percentile exceedances using the synthetic data with a strong seasonal cycle from figure 2 in the main manuscript. The respective top panels show threshold and exceedances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



The bias depends on the strength of the seasonal cycle

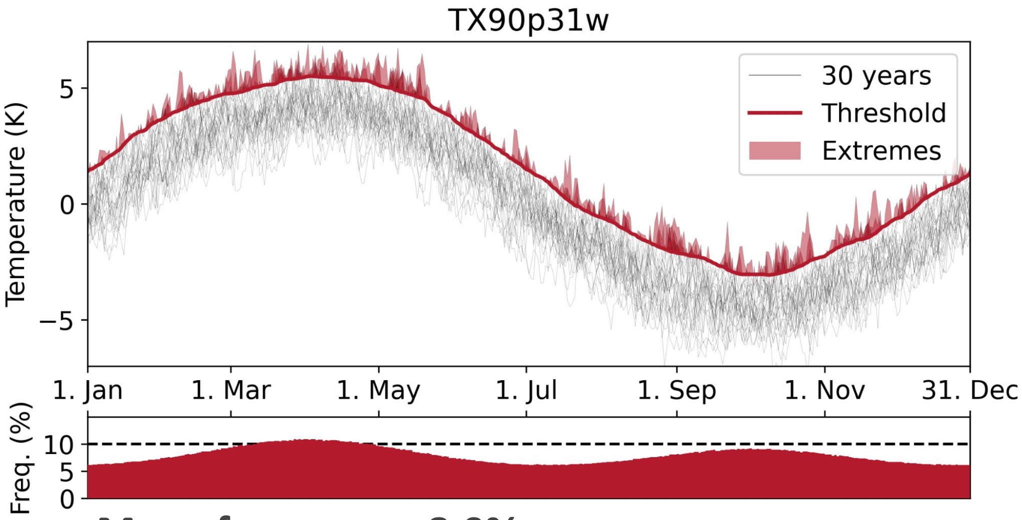


Mean frequency: 10.0%
Mean bias: 0%

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



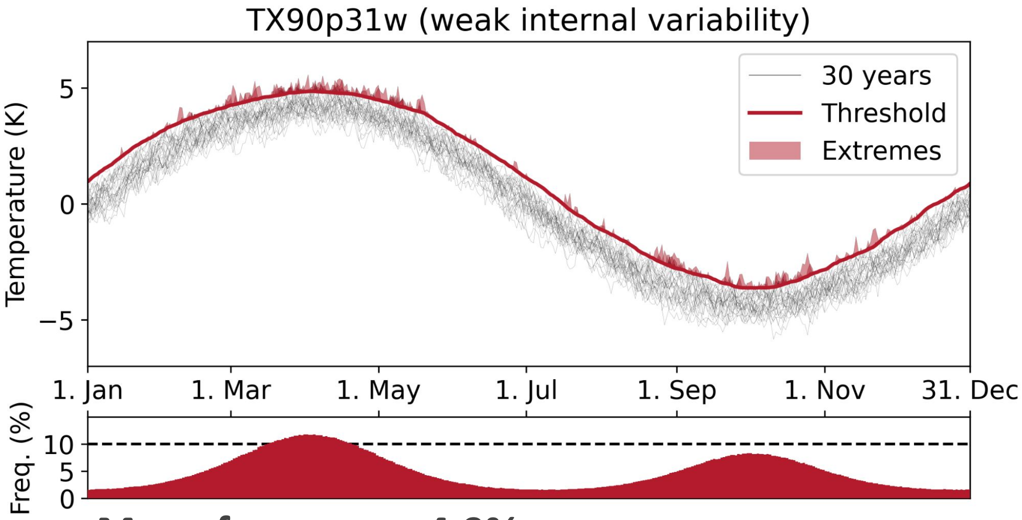
Mean frequency: 8.0%

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**

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

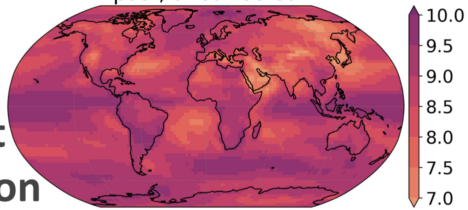
Base period (1961-1990)

Future period (2071-2100)

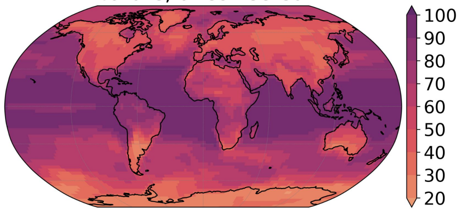
Change

Without correction

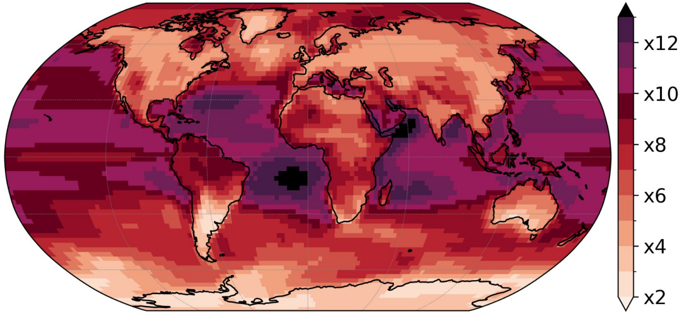
(a) Frequency (%)
past, uncorrected



(b) Frequency (%)
future, uncorrected

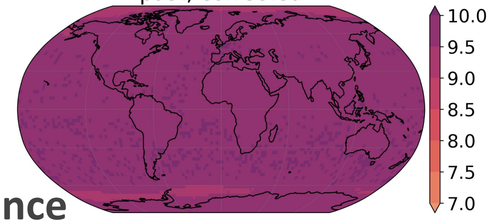


(a) Extreme frequency change (ratio)
2071-2100 relative to 1961-1990

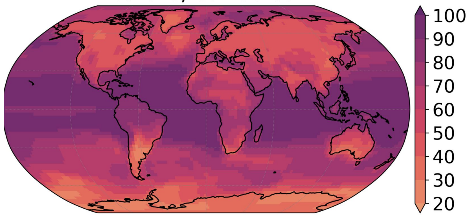


Reference

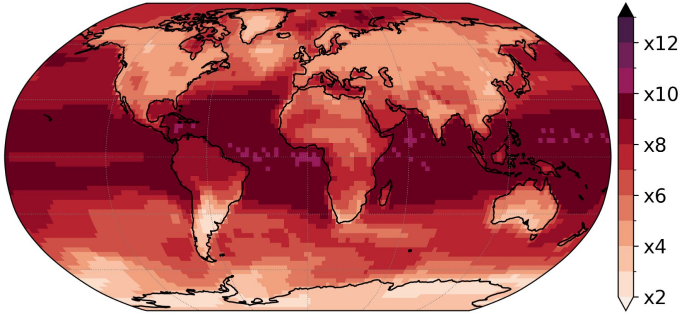
(d) Frequency (%)
past, corrected



(e) Frequency (%)
future, corrected



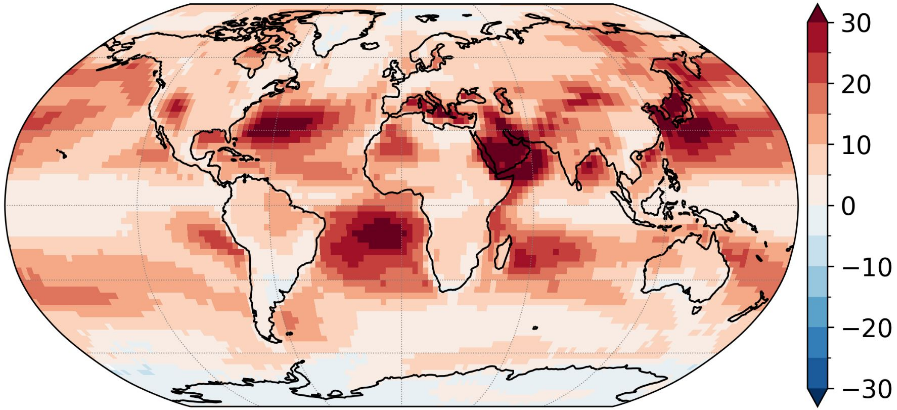
(b) Corrected: Extreme frequency change (ratio)
2071-2100 relative to 1961-1990



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

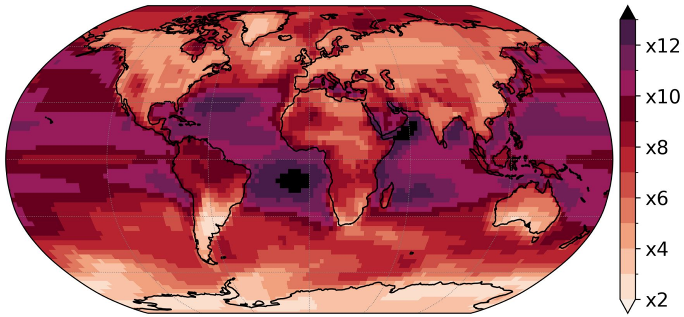
Change

(c) Extreme frequency change bias (%)

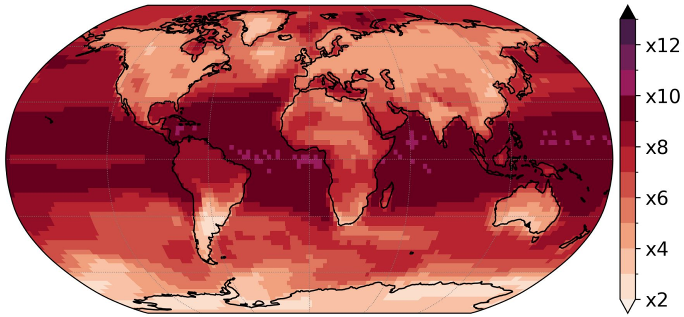


→ **the bias leads to an overestimation of extreme changes by up to 30%!**

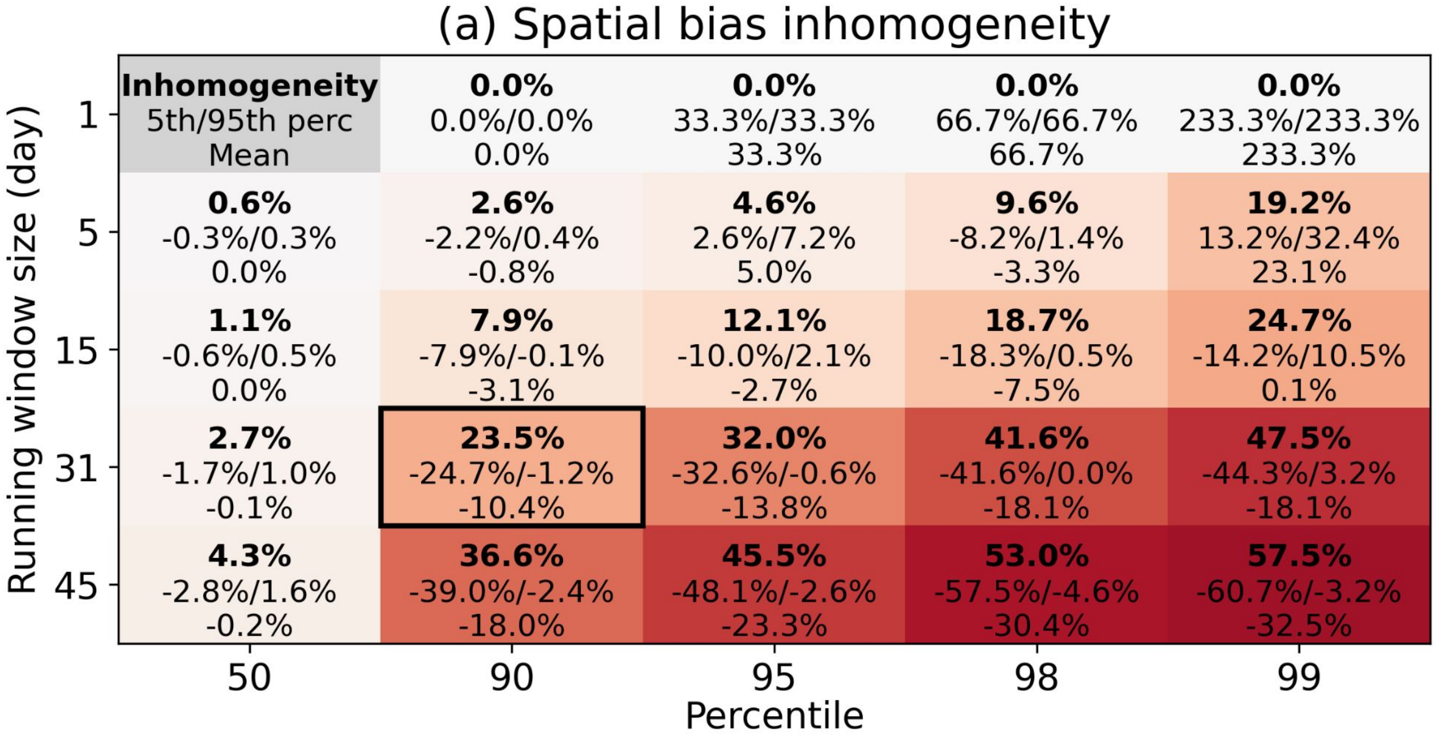
(a) Extreme frequency change (ratio)
2071-2100 relative to 1961-1990



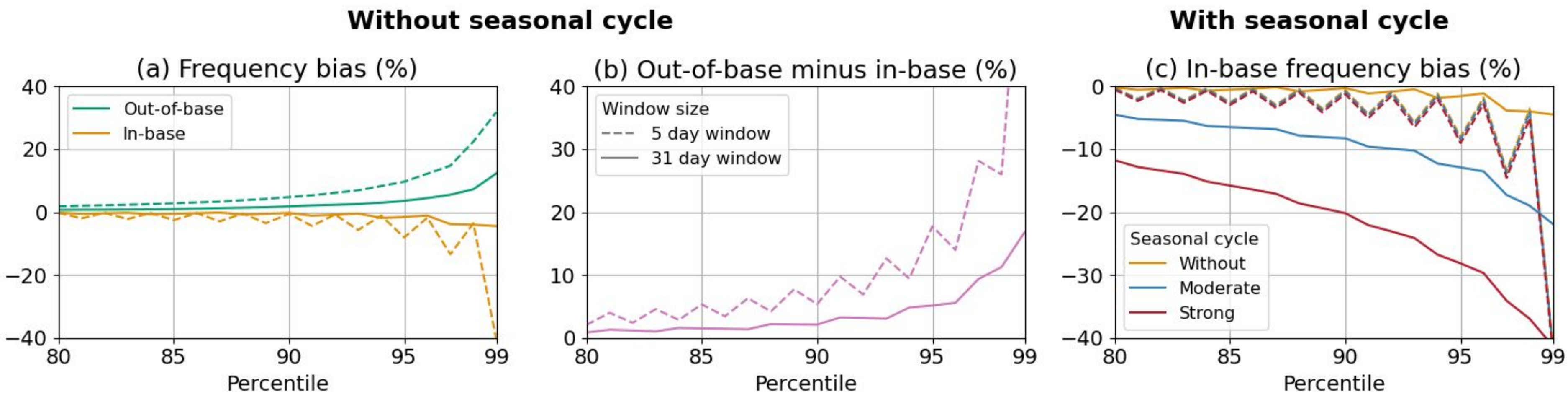
(b) Corrected: Extreme frequency change (ratio)
2071-2100 relative to 1961-1990



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



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



Zhang et al. 2005

Brunner and Voigt (in review)

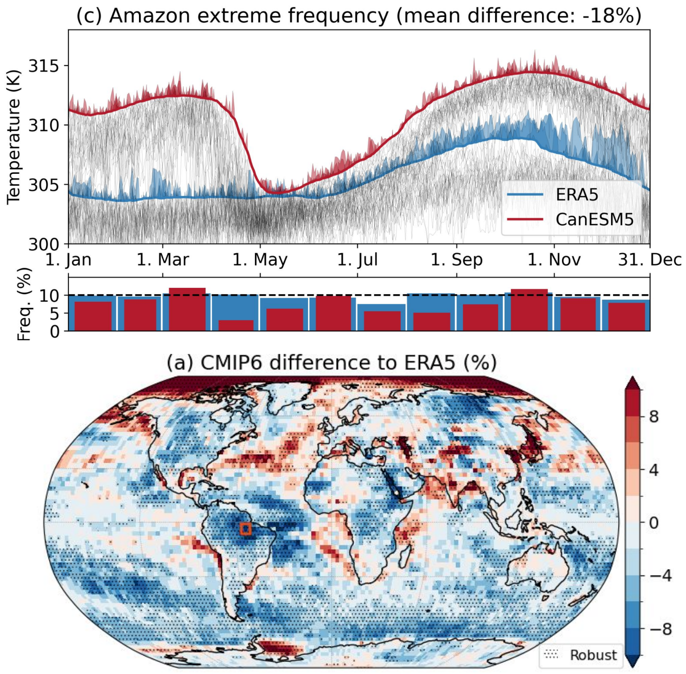
Relative temperature extreme definitions are used as implicit bias correction

*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. Brunner und Voigt (in review)