# Claire the Climate Scientist

## What is she doing? (everyday work of a climatologist)



- Claire lives in a world with a CESM 1.2.2 (2° Resolution) climate and no global change
- She studies the frequency and severity of heat extremes
- She relies on extreme value theory – or more specifically the block-maxima approach – to estimate quantities like return levels or return periods
- Currently, she investigates a 100yr long observational time series of daily maximum temperature data


CC BY

# Claire the Climate Scientist
## What is her concern?



How would her estimates change if she …

- … had more data to investigate?
  (if only one had measured temperature for more than just 100 years)
- ... would choose a larger than annual block size?
  (are annual maxima extreme enough to be analysed with EVT?)
- ... knew the effects of internal climate variability?
  (what if she did the analysis 100 years ago?)


CC BY

# Claire the Climate Scientist

## What would she find?

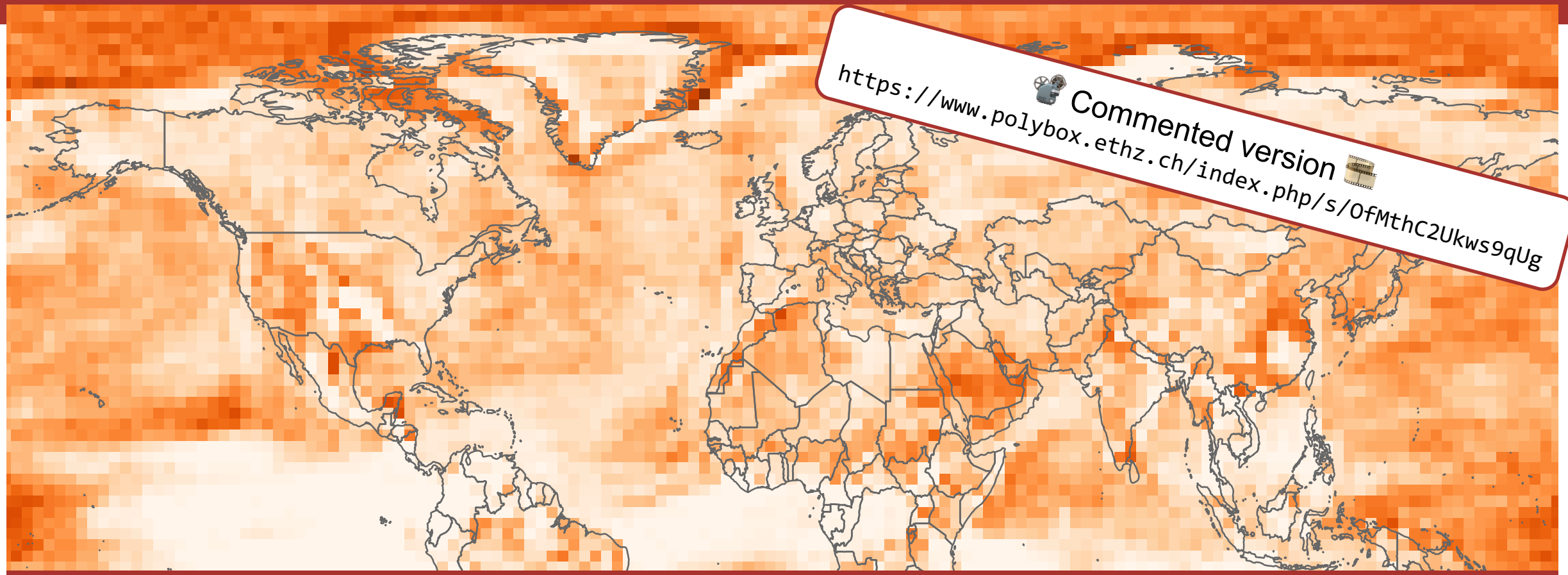


Sometimes at night she dreams about having thousands of years of data available to answer these questions. She would find that …

- … return levels of heat extremes are slightly underestimated when using a limited amount of data.
- … confidence intervals as measures of uncertainty should be regarded as lower bound estimates of the total uncertainty.
- … the (true) frequency of extreme events is very often massively underestimated.


CC BY

ETH zürich




# The Challenge of Estimating High Return Levels with Short Records under Large Internal Variability

EGU Session on Large Ensembles

IAC Institute for Atmospheric and Climate Science

CC BY

# Claire the Climate Scientist
## Motivation



Model:

- CESM 1.2.2 (2° Resolution)
- Pre-industrial control run
- and further CMIP6 models

Target variable:

- Tx7d (Maximum 7d temperature mean)
- and Rx1d & Rx5d (max 1-/5-day precipitation)

Research focus:

- Return level and return period estimates for Maxima under different
  - time series durations $t_{sub}$ or
  - block sizes $m$

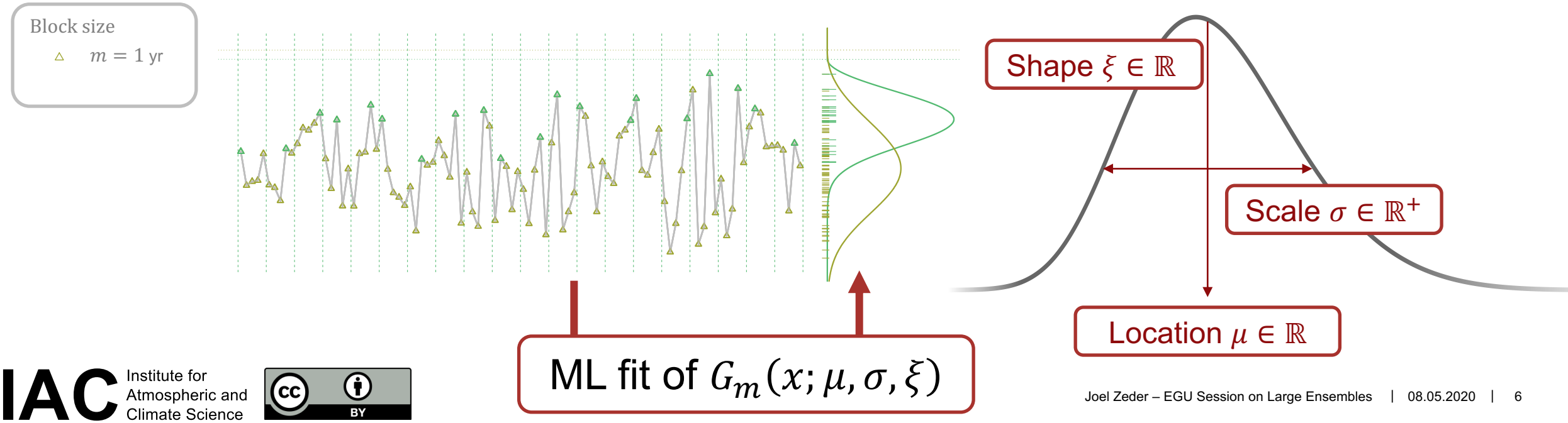

CC BY



# Generalised Extreme Value Distribution $G(x; \mu, \sigma, \xi)$

## Recap in Extreme Value Theory

- Statistical distribution of maxima taken from blocks (length $m$) of $iid$ data
- E.g. highest daily temperature in consecutive five year blocks ($m = 5\text{yr}$) from 1901 to 2000 ($t_{sub} = 100\text{yr}$), i.e. 1901-1905, 1906-1910, …, 1996-2000





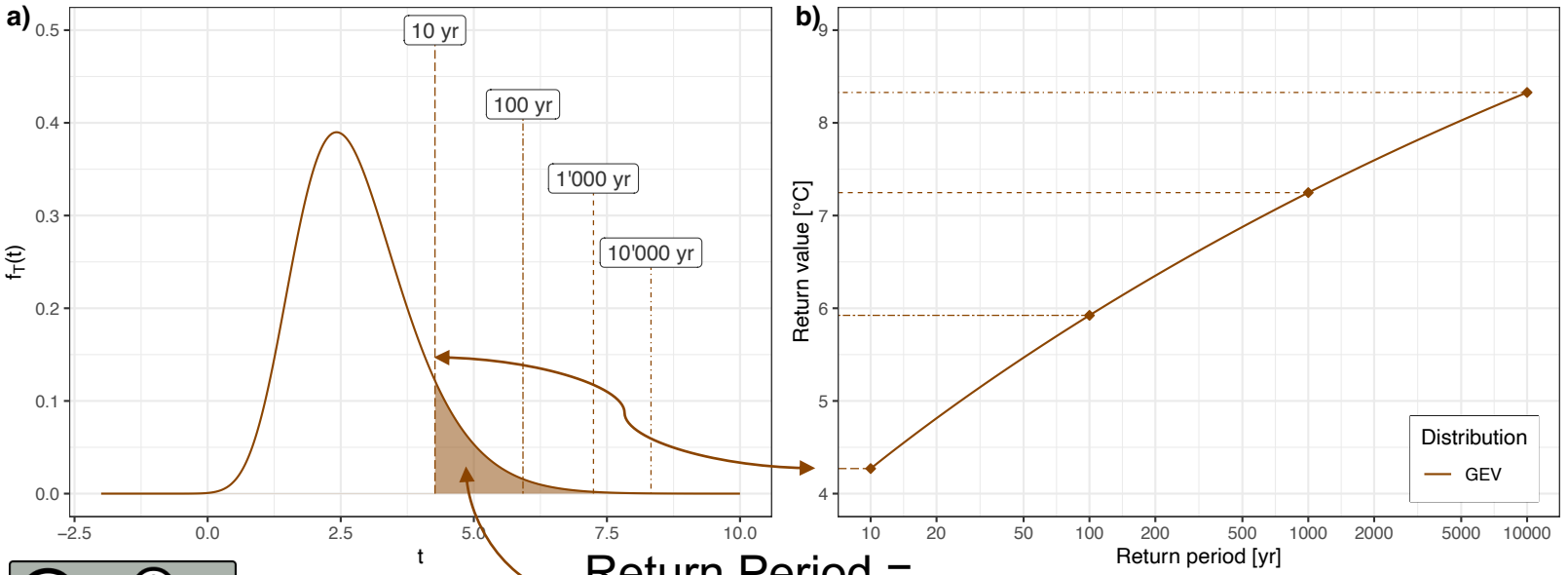

CC BY

# Generalised Extreme Value Distribution $G(x; \mu, \sigma, \xi)$

## Recap in Extreme Value Theory

- Having estimated the model parameters $\hat{\mu}, \hat{\sigma}, \hat{\xi}$, we can infer the respective return levels $z_p$ for specified return periods $1/p$:

$$\hat{z}_p(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}[1 - (-\log(1-p))^{-\hat{\xi}}]$$




Return Period = 1/Exceedance Probability

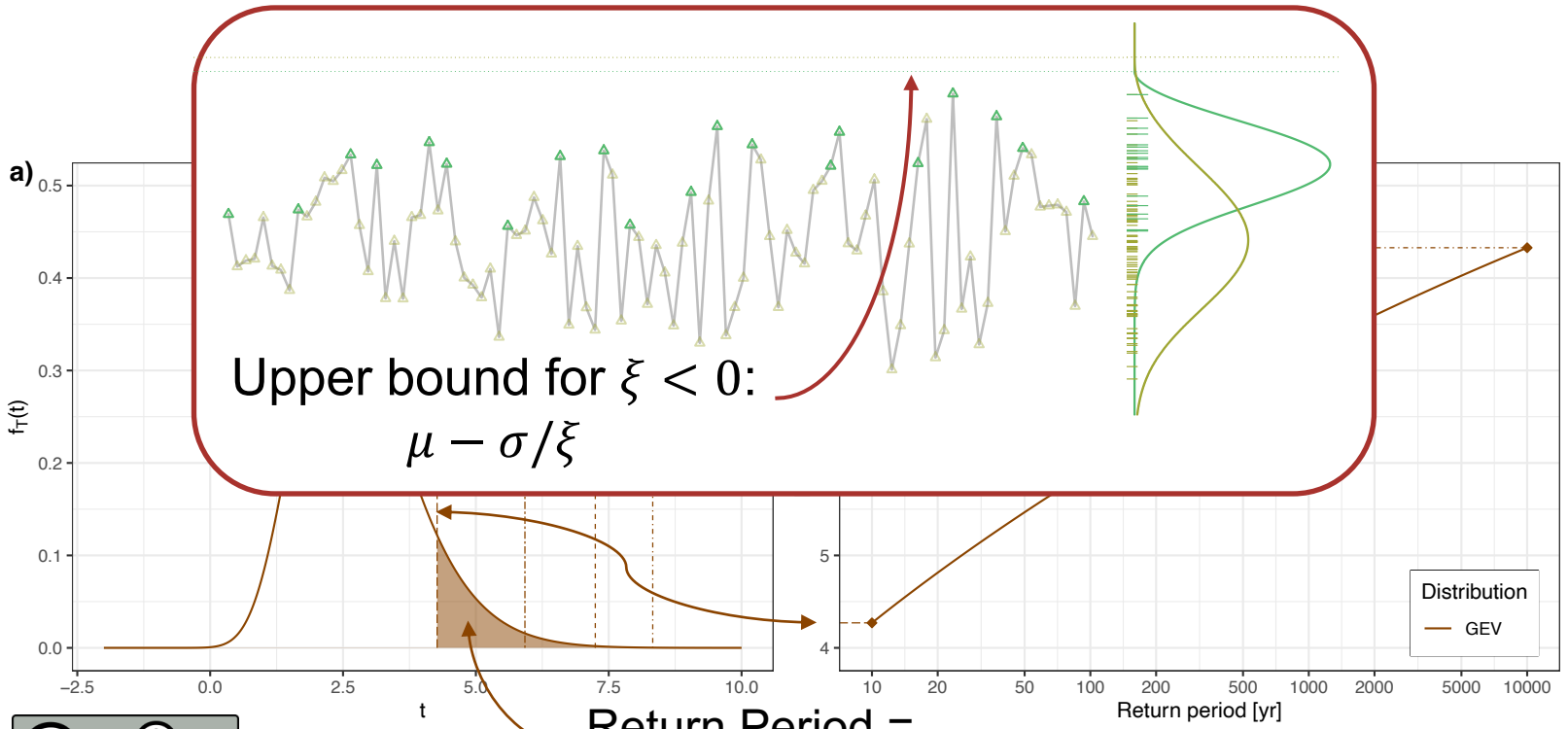


CC BY

# Generalised Extreme Value Distribution $G(x; \mu, \sigma, \xi)$

## Recap in Extreme Value Theory

- Having estimated the model parameters $\hat{\mu}, \hat{\sigma}, \hat{\xi}$, we can infer the respective return levels $z_p$ for specified return periods $1/p$:





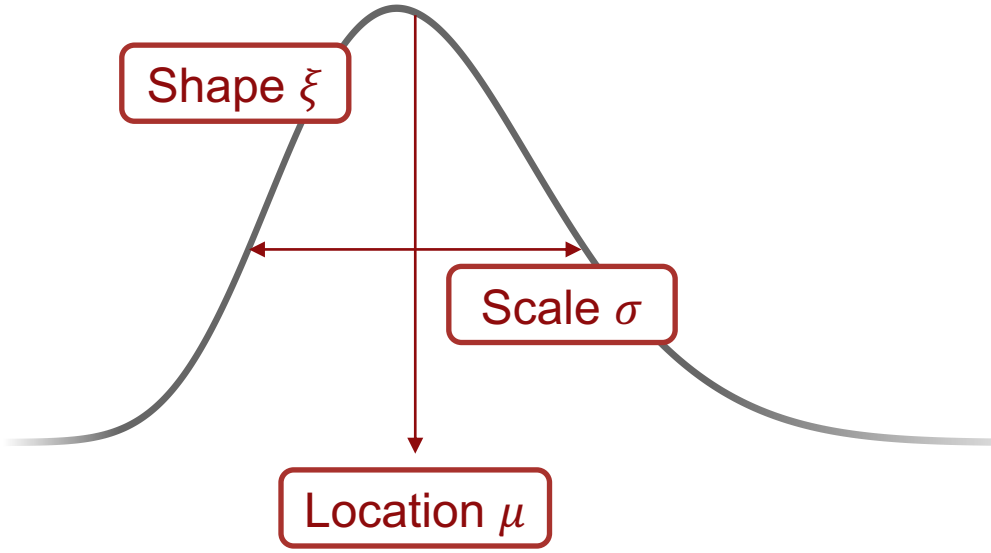
CC BY

# Research Objective

How do estimates of model parameters $(\mu, \sigma, \xi)$, return values $(100/200/500/1000\mathrm{yr})$ and exceedance probabilities (i.e. return periods) depend on

- block size $(m = 1, 5, 20\mathrm{yr})$ and
- time series length $(t_{sub} = 50, 100, 500, 1000\mathrm{yr})$

under the influence of internal variability?





CC BY

# Extremes in a Stationary (Model) Climate

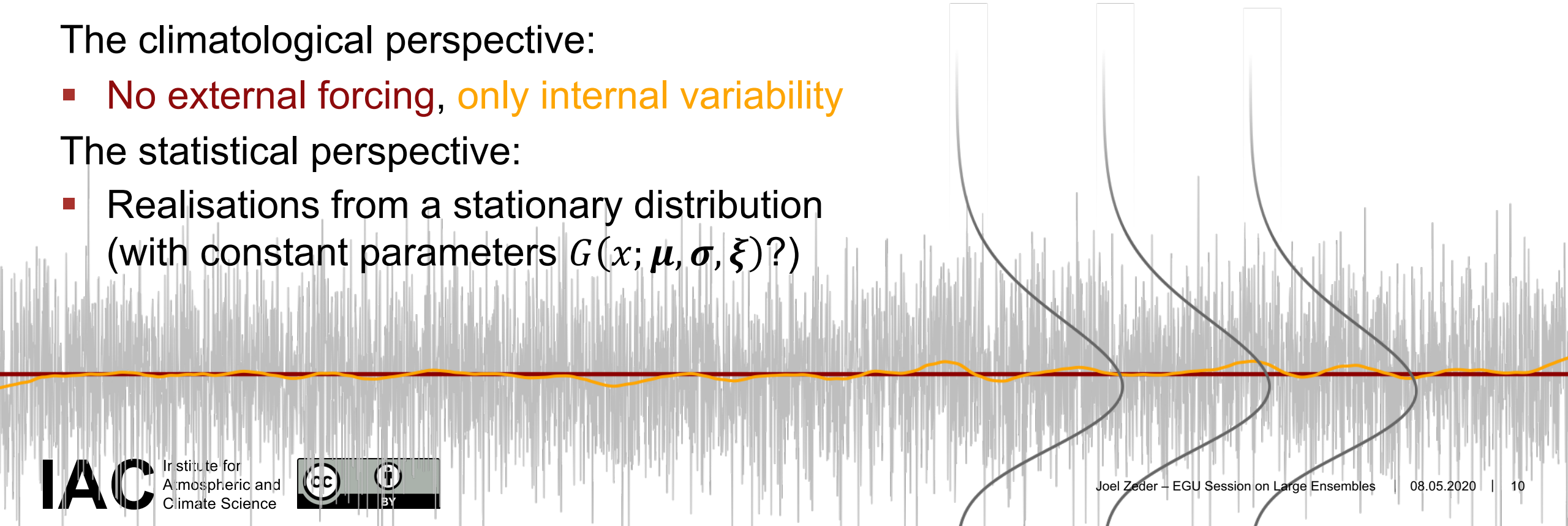
Focus on internal variability → Analyse extremes in stationary model climate

The climatological perspective:

- No external forcing, only internal variability

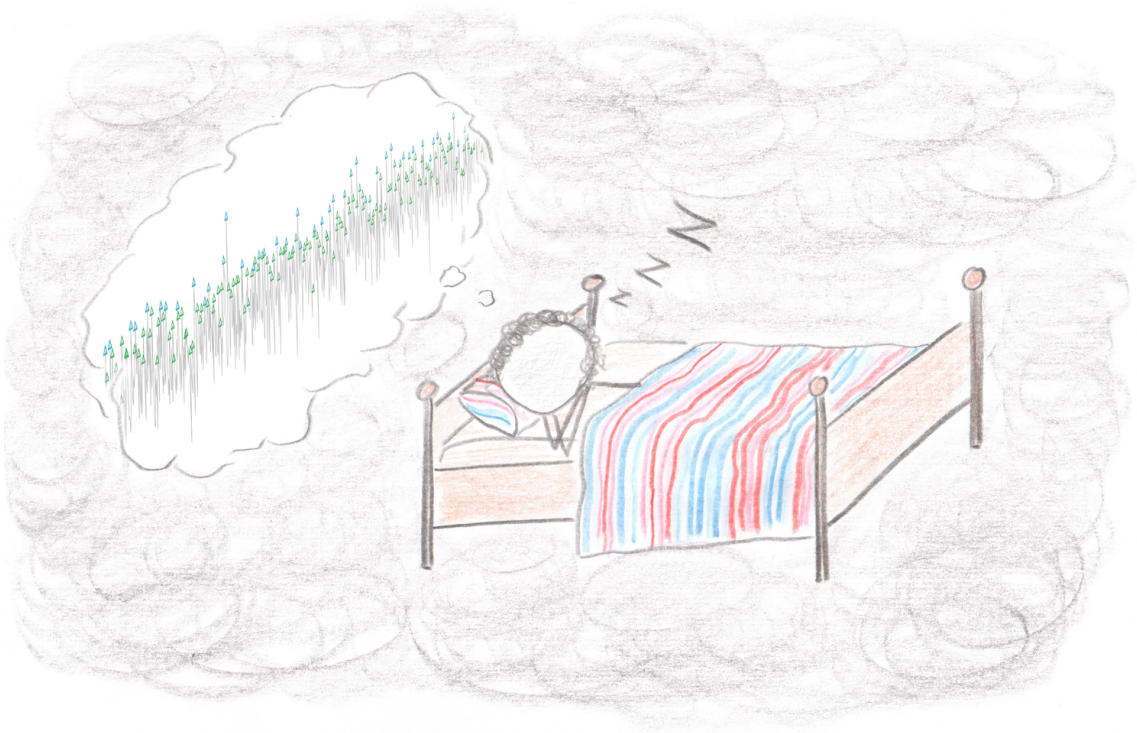
The statistical perspective:

- Realisations from a stationary distribution (with constant parameters $G(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi})$?)



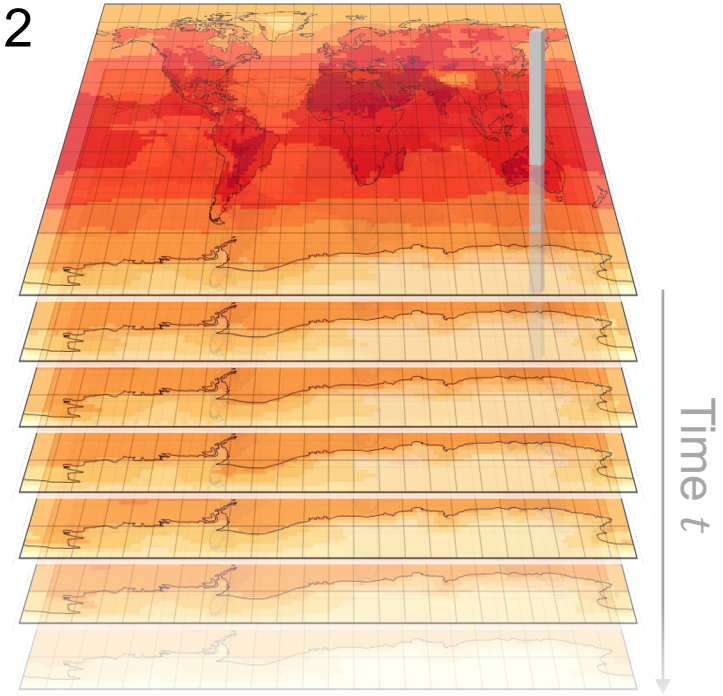

CC BY

# Claire the Climate Scientist
## Motivation



Data form CESM12 pre-industrial control run:

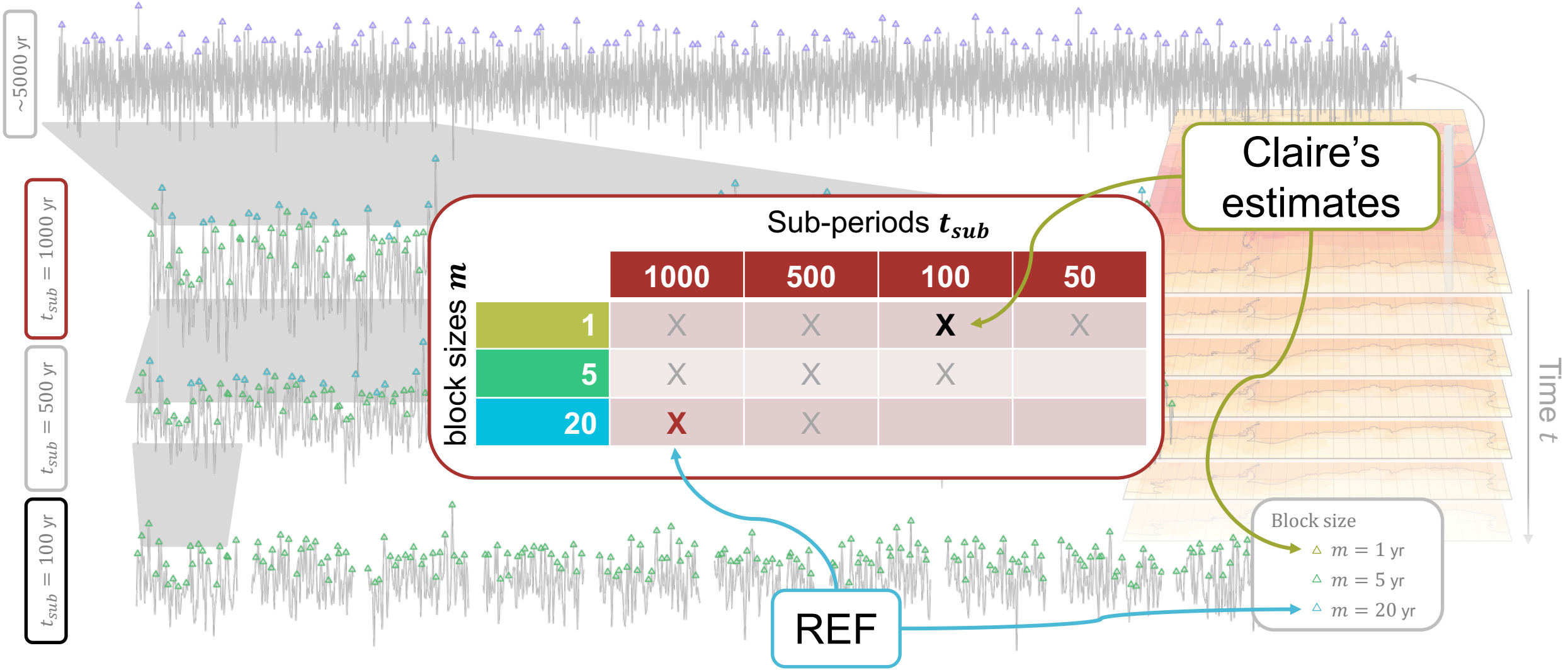
- 7day maxima of surface temperature
- at each grid point
- for ~5000yr model years





CC BY

# Sub-periods $t_{sub}$ and block sizes $m$



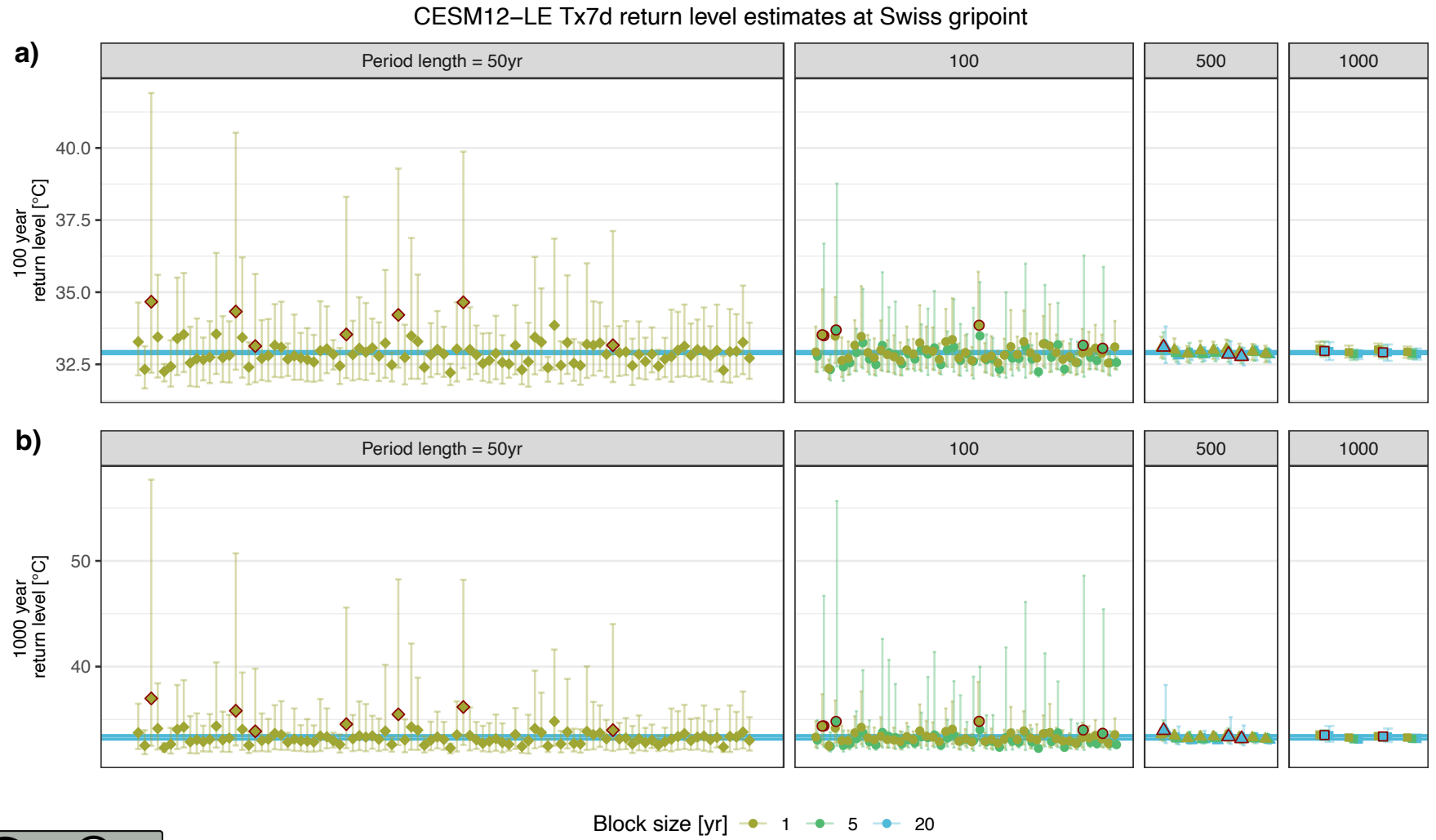


CC BY

# Estimated Tx7d GEV return levels at Swiss grid point

100 year
return level [°C]

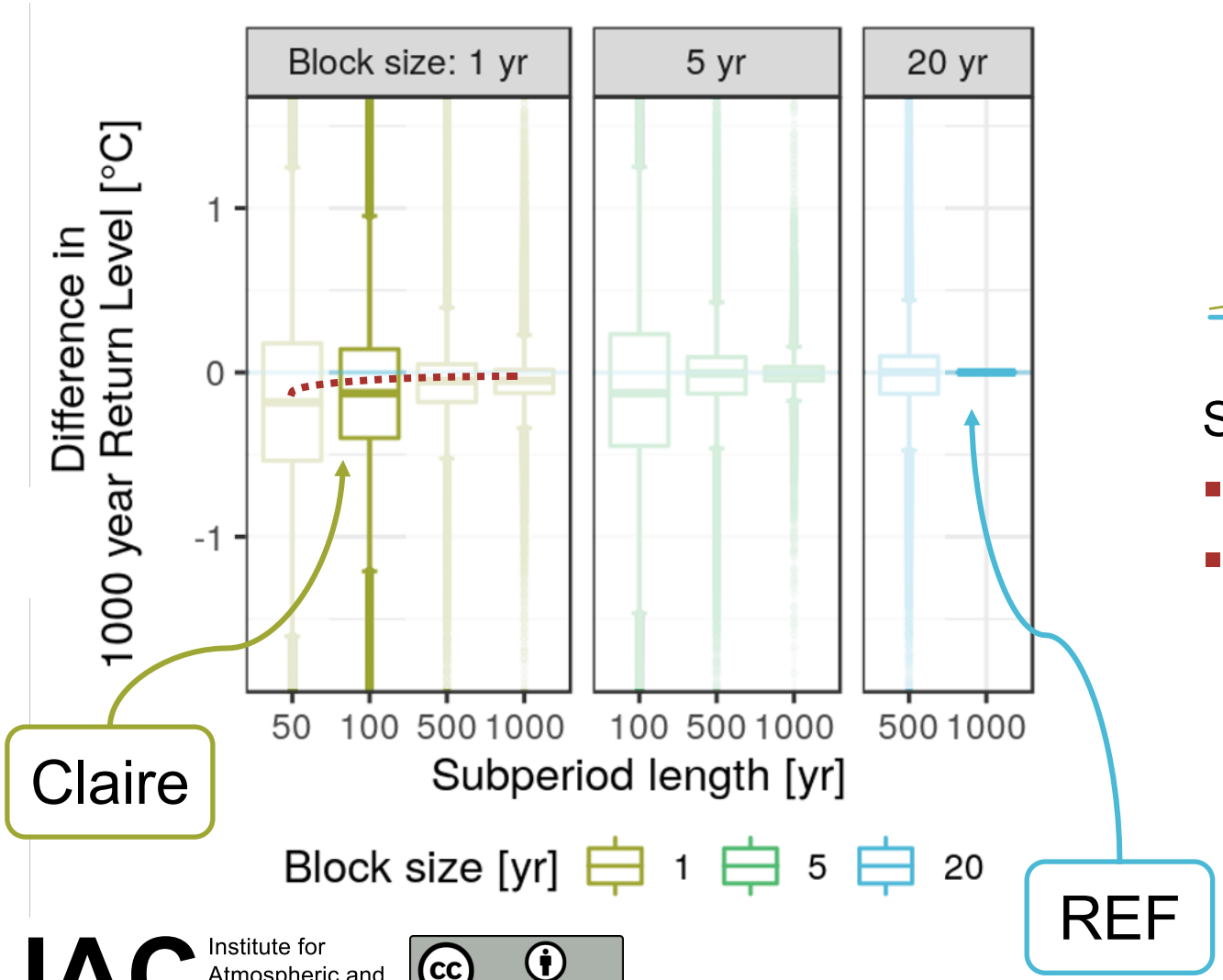
1000 year
return level [°C]

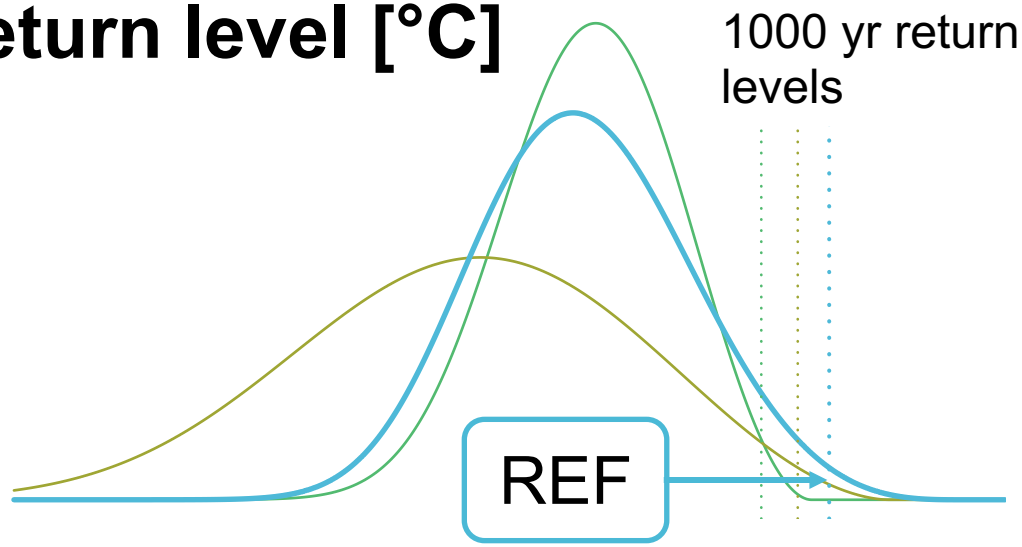





CC BY

# 1. Systematic underestimation of return level [°C]




Small systemic underestimation of return levels

- observable for shorter sub-periods $t_{sub}$
- not so much for smaller block sizes $m$

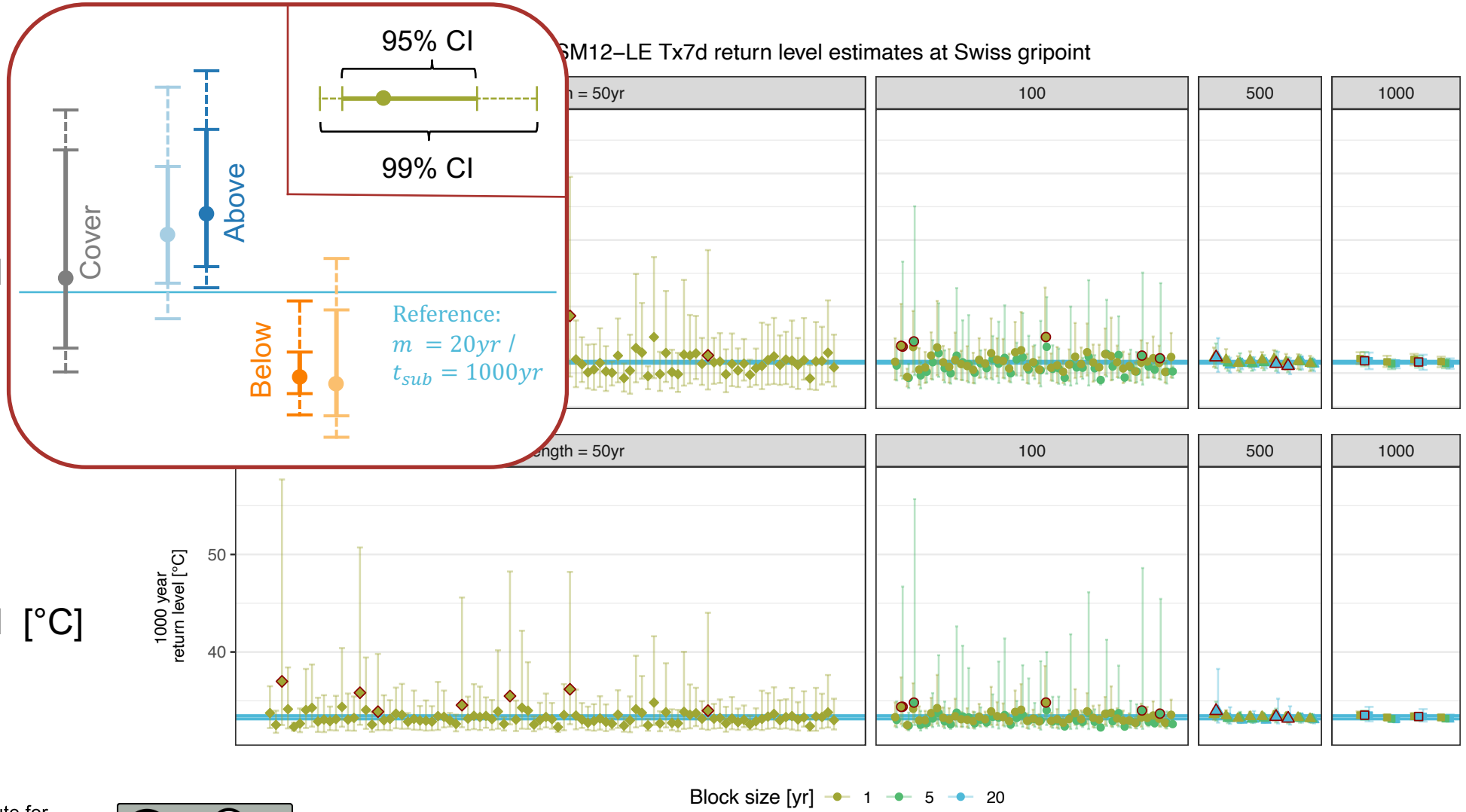
IAC Institute for Atmospheric and Climate Science

CC BY

# CI of estimated Tx7d GEV return levels at Swiss grid point

100 year return level

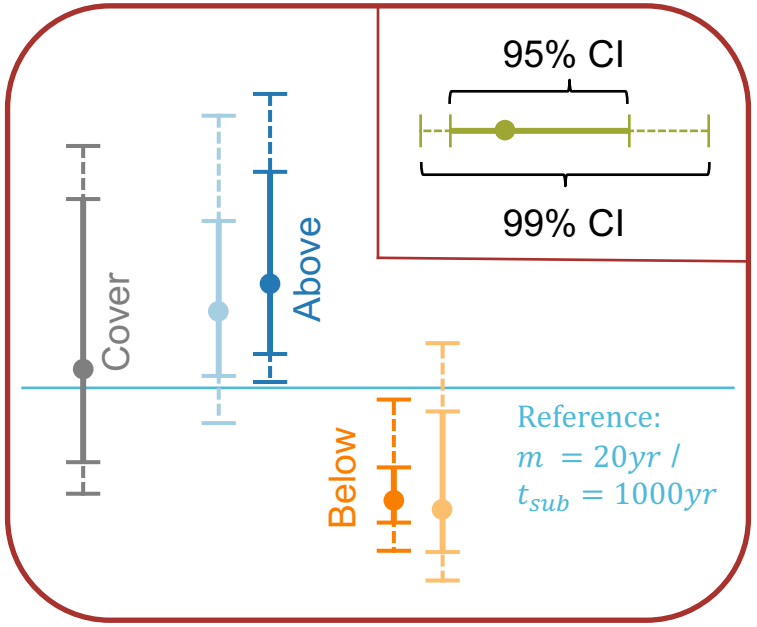
1000 year return level [°C]



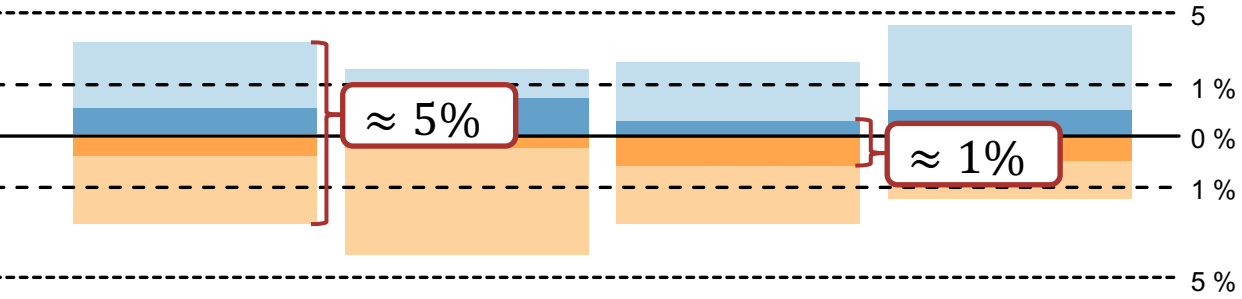


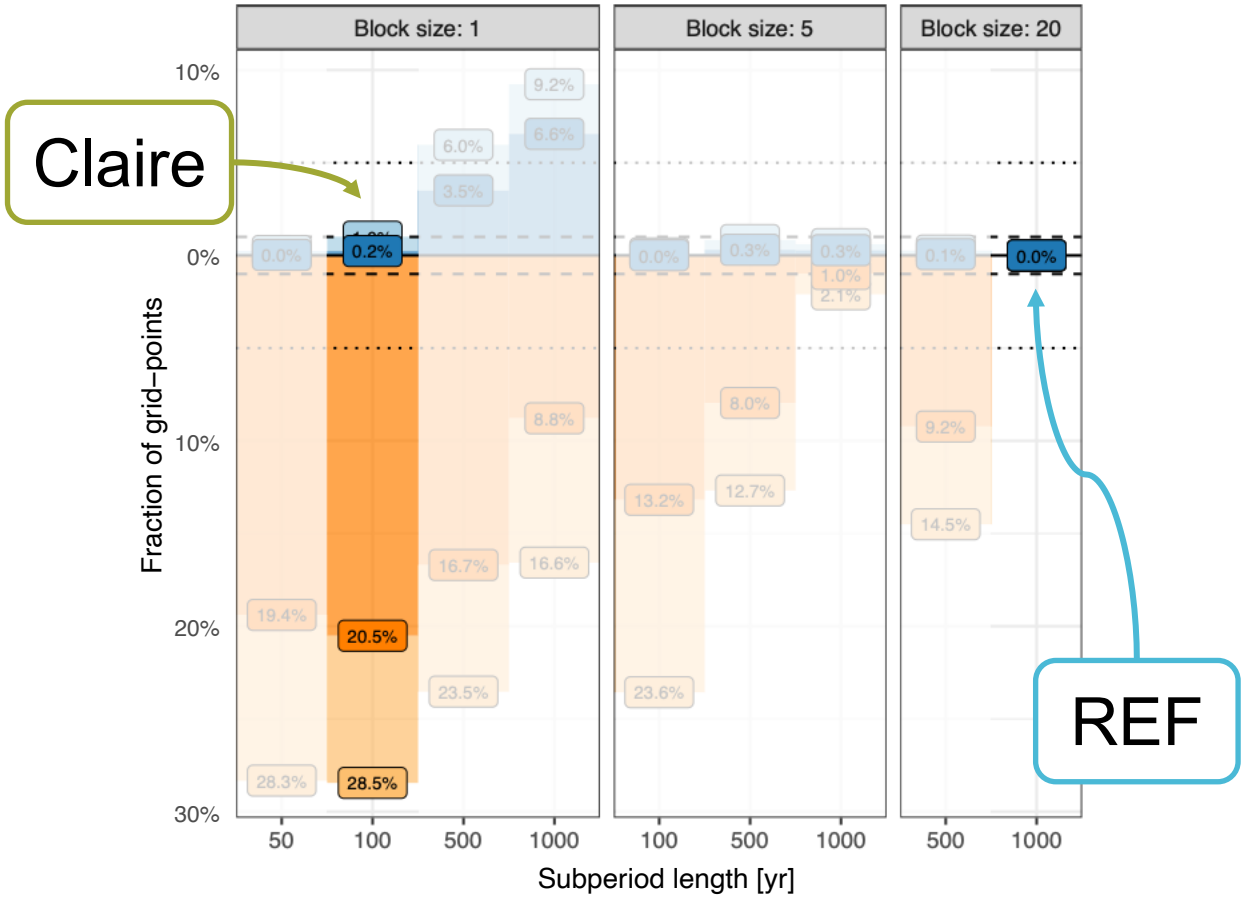

CC BY

# 2. CI coverage of 1000yr return level reference estimate




Expectation from theory:



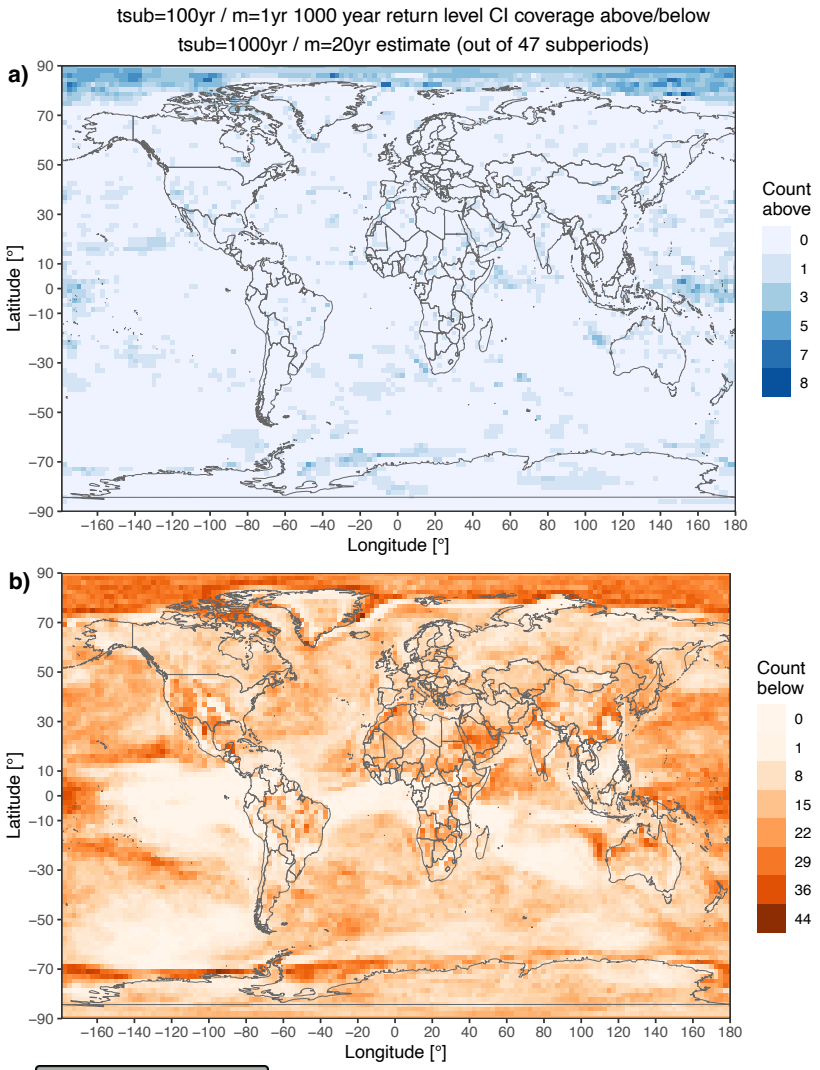

1000 year return level CI comparison with tsub=1000yr / m=20yr parameter:

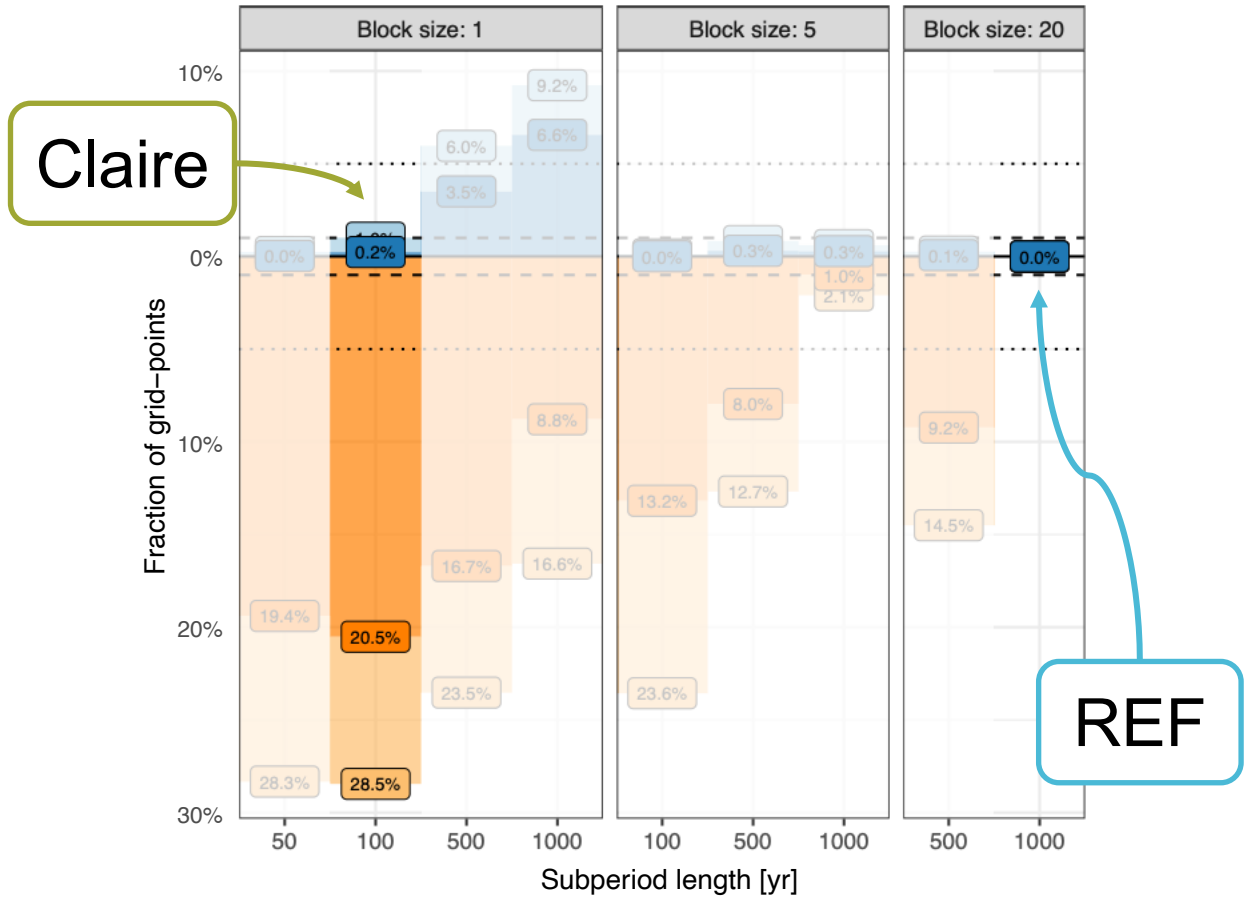
Above parameter (95% CI)
Above parameter (99% CI)
Below parameter (95% CI)
Below parameter (99% CI)

IAC Institute for Atmospheric and Climate Science

CC BY

# 2. CI coverage of 1000yr return level reference estimate




1000 year return level CI comparison with tsub=1000yr / m=20yr parameter:

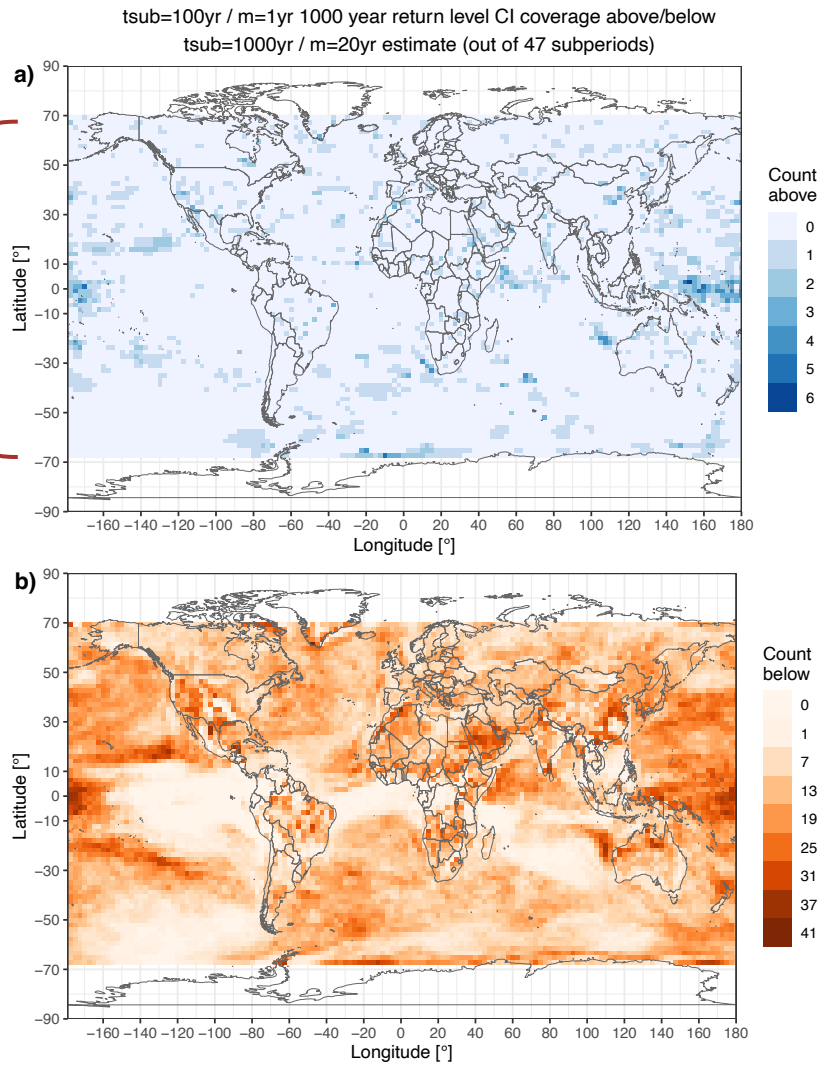
Above parameter (95% CI) | Below parameter (95% CI)
Above parameter (99% CI) | Below parameter (99% CI)

IAC Institute for Atmospheric and Climate Science

CC BY

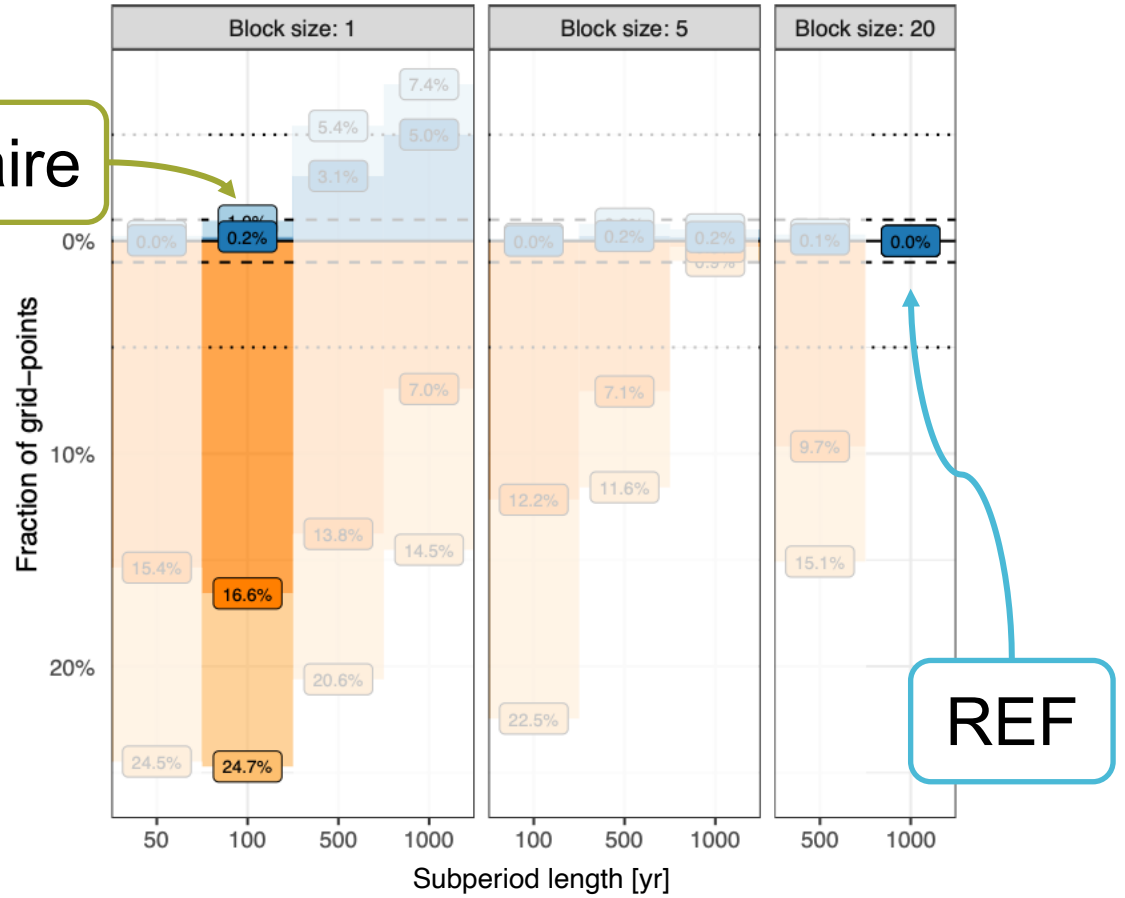
# 2. CI coverage of 1000yr return level reference estimate

Only low latitudes




Claire

REF



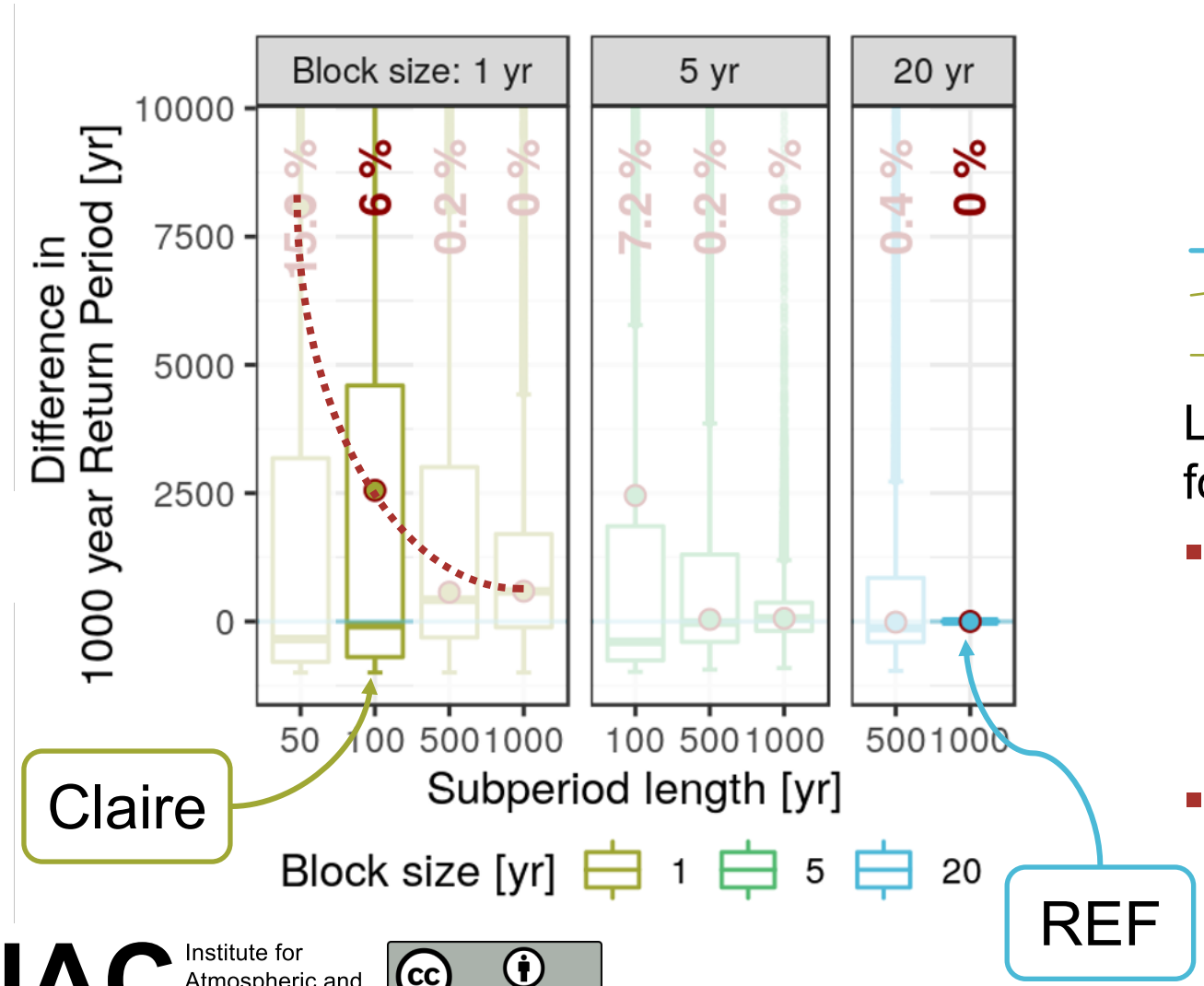

1000 year return level CI comparison with tsub=1000yr / m=20yr parameter:

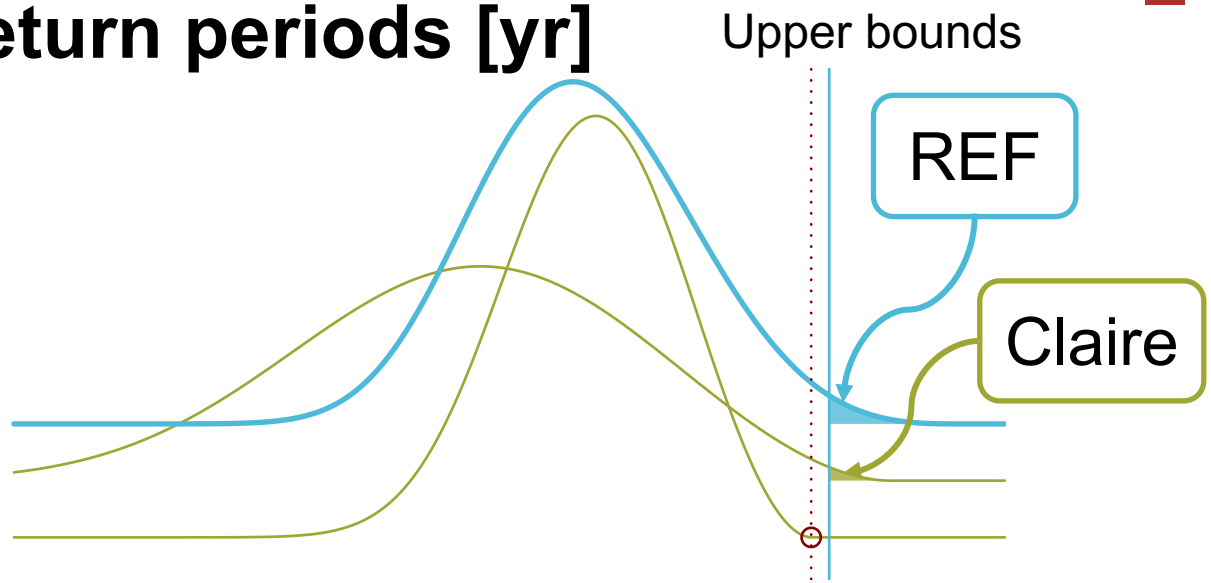
Above parameter (95% CI)
Above parameter (99% CI)
Below parameter (95% CI)
Below parameter (99% CI)



CC BY

# 3. Systematic underestimation of return periods [yr]




Large systemic overestimation of return periods for high return levels in the reference data:

- The median return period of a 1000 $yr$ reference return level is estimated to be a ~3500 $yr$ return level for $t_{sub} = 100yr$ and $m = 1yr$ fits.
- In 6% of $t_{sub} = 100yr$ and $m = 1yr$ fits, the 1000 $yr$ reference return level has a return period of infinity.

IAC Institute for Atmospheric and Climate Science

CC BY

# Conclusions and Outlook

Even under stationary conditions,

- ✓ high return values are often underestimated and return periods are overestimated by fitting on small block sizes and short time series.
- ✓ Confidence intervals are often not covering the refence value and thus underestimate uncertainty of the estimate.

The same findings can be concluded for transient LE CESM12 data, other CMIP6 control run simulations and using $m = \mathbf{5}\mathrm{yr}$ / $t_{sub} = 1000\mathrm{yr}$ estimates as reference.

Outlook:

- ❑ Extend analysis to non-stationary **transient** LE simulations.


CC BY

# Thank you very much for your attention


- Joel Zeder, PhD Candidate
  joel.zeder@env.ethz.ch / @JoelZeder
- Dr. Erich Fischer, Senior Scientist
  erich.fischer@env.ethz.ch / @erichfischer

ETH Zurich
Climate Physics Group of Prof. Reto Knutti
Institute for Atmospheric and Climate Science
Universitätstrasse 16
8092 Zurich
Switzerland






Also have a look at the presentation of Claudia Gessner, working on a similar project:

https://meetingorganizer.copernicus.org/EGU2020/EGU2020-1628.html



CC BY