universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Changes in heatwave properties under climate change
A systematic bias in future heatwave diagnostics throughout the seasonal cycle

verfasst von | submitted by

Maximilian Meindl BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 614
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Meteorology
Betreut von | Supervisor:	Univ.-Prof. Dr. Aiko Voigt
Mitbetreut von | Co-Supervisor:	Dr. Lukas Brunner BSc MSc

Abstract

Human-induced climate change is leading to a warming Earth, resulting in more frequent and intense temperature extremes. Daily temperature extremes can be defined following various approaches, with relative percentile-based thresholds being a common method. In this master thesis I am exploring spatiotemporal heatwaves across the seasonal cycle derived from daily temperature extremes, emphasizing the critical role of the extreme threshold chosen in their definition.

To investigate the sensitivity of heatwave characteristics to the extreme threshold definition, I focus on the approach utilizing a so-called moving threshold. This method involves a 31-day running window to increase the sample size for percentile calculations as well as an additional 31-year running window to account for the impact of global warming. It is recognized that the usage of a seasonal running window may introduce biases in threshold exceedances. To address this issue, Brunner et al., 2024 proposed a simple bias-correction method, involving the removal of the mean seasonal cycle before percentile threshold calculation, which I will also use here to explore effects on downstream impact metrics.

I focus on the 99th percentile as threshold and show the potential for a significant bias in the extreme frequency, exceeding 50% in certain regions according to 5 selected CMIP6 models. My findings further reveal that without bias-correction, this also leads to a substantial underestimation of derived heatwave properties, in particular area, duration, and magnitude. For one ensemble member of the ACCESS-CM2 model, the difference in heatwave area can reach up to 40%, when comparing biased and corrected results for the 100 biggest events in the period 1960-1990.

A statistical analysis of 5 ensemble members of the ACCESS-CM2 model shows that changes in heatwave properties are statistically significant when the bias-correction method is not applied. When defined relative to a future climate, simulated heatwaves show no significant changes unless no proper bias-correction method is employed.

The results of my master thesis contribute to a better understanding of the implications of using a seasonally running window on heatwave characteristics, providing valuable insights for future climate projections. I emphasize the importance of adopting appropriate methods and bias-correction techniques to enhance the accuracy of temperature extremes in the context of ongoing climate change.

Kurzfassung

Der vom Menschen verursachte Klimawandel führt zu einer globalen Erwärmung und damit zu häufigeren und intensiveren Temperaturextremen. Tägliche Temperaturextreme können nach verschiedenen Ansätzen definiert werden, wobei sogenannte Schwellenwerte auf der Grundlage relativer Perzentile eine gängige Methode sind. In dieser Masterarbeit werde ich Hitzewellen, die in Raum und Zeit verbunden sind, im Verlauf des saisonalen Zyklus untersuchen, wobei sich diese aus täglichen Temperaturextremen ableiten lassen. Die genaue Definition der Schwellenwerte spielt dabei eine wesentliche Rolle.

Um die Sensitivität auf abgeleitete Metriken von Hitzewellen gegenüber der Definition des Extremschwellenwertes zu untersuchen, verwende ich den Ansatz von sogenannten beweglichen Schwellenwerten. Diese Methode umfasst ein gleitendes Fenster von 31 Tagen, um die Stichprobengröße für Perzentilberechnungen zu erhöhen, sowie ein zusätzliches gleitendes Fenster von 31 Jahren, um den Einfluss der globalen Erwärmung zu berücksichtigen. Es wird gezeigt, dass die Einführung eines gleitenden Fensters über den saisonalen Zyklus zu einem Bias bei der Überschreitung von Schwellenwerten führen kann. Um dieses Problem zu lösen, schlugen Brunner et al., 2024 eine einfache Bias-Korrekturmethode vor, die die Entfernung des mittleren saisonalen Zyklus vor der Berechnung der Schwellenwerte beinhaltet. Diese Methode verwende ich im Zuge meiner Masterarbeit, um Auswirkungen auf abgeleitete Metriken von Hitzewellen zu untersuchen.

Ich verwende das 99. Perzentil als Schwellenwert und illustriere das Potenzial für einen signifikanten Bias in der Häufigkeit von Extremen, der in bestimmten Regionen laut 5 ausgewählten CMIP6-Modellen über 50% liegt. Meine Ergebnisse zeigen außerdem, dass ohne Anwendung der Biaskorrektur eine erhebliche Unterschätzung abgeleiteter Eigenschaften von Hitzewellen auftritt, insbesondere in Bezug auf Fläche, Dauer und Intensität. Für ein Ensemblemember des globalen Klimamodells ACCESS-CM2 kann die Differenz in der Fläche von Hitzewellen bei Vergleich der biaskorrigierten und nicht biaskorrigierten Ergebnisse für die 100 größten Ereignisse im Zeitraum 1960-1990 bis zu 40% betragen.

Eine statistische Analyse von 5 Ensemblemitgliedern des ACCESS-CM2 Modells zeigt, dass Änderungen in den Eigenschaften von Hitzewellen statistisch signifikant sind, wenn die Biaskorrektur nicht angewendet wird. Nur wenn eine geeignete Methode zur Biaskorrektur verwendet wird gibt es keine signifikanten Veränderungen von projizierten Hitzewellen, wenn sie relativ zur Zukunft definiert sind.

Kurzfassung

Die Ergebnisse meiner Masterarbeit tragen dazu bei, ein besseres Verständnis der Auswirkungen saisonaler gleitender Fenster auf die Eigenschaften von Hitzewellen zu erlangen und liefern wertvolle Erkenntnisse für zukünftige Klimaprojektionen. Die Existenz eines systematischen Bias in der Diagnostik zukünftiger Hitzewellen über den saisonalen Zyklus hinweg wurde bereits von Brunner et al., 2024 nachgewiesen, wobei meine Arbeit insbesondere die Auswirkungen des Bias auf zukünftige Veränderungen von Hitzewellen zeigt. Zudem betone ich die Bedeutung der Anwendung geeigneter Methoden zur Biaskorrektur, um die Genauigkeit von Temperaturextremen im Zusammenhang mit dem aktuellen Klimawandel zu verbessern.

Acknowledgements

I would like to take this opportunity to thank all those who supported and motivated me during my Master's project. First of all, I would like to thank Aiko Voigt and Lukas Brunner, who supervised and reviewed my Master's thesis and always took the time to answer my questions. Lukas Brunner in particular proposed the topic in February 2023, as this research field is a hot topic in the truest sense of the word, given its significance and relevance. The work resulted in many interesting discoveries, whereby my two supervisors have recorded parts of it in a scientific publication (Brunner et al., 2024). At this point, I would like to express my sincere thanks once again for all suggestions, technical aspects, and helpful additions. Moreover, I would also like to thank all members within the Climate Dynamics and Modeling group for providing valuable feedback on my findings. Finally, I would also like to express my gratitude to the Department of Meteorology and Geophysics at the University of Vienna, where I had the privilege of working part-time in IT during my master's studies. Their support and flexibility allowed me to successfully navigate the demands of both work and study.

Ich möchte diese Gelegenheit nutzen, um allen zu danken, die mich während meines Masterprojekts unterstützt und motiviert haben. Zunächst möchte ich Aiko Voigt und Lukas Brunner danken, die meine Masterarbeit betreut und begutachtet haben und sich stets die Zeit genommen haben, um meine Fragen zu beantworten. Insbesondere Lukas Brunner schlug das Thema im Februar 2023 vor, da dieser Bereich aufgrund seiner Bedeutung und Relevanz im wahrsten Sinne des Wortes ein heiß diskutiertes Thema ist. Die Arbeit führte zu vielen interessanten Erkenntnissen, wobei meine beiden Betreuer Teile davon in einer wissenschaftlichen Publikation festgehalten haben (Brunner et al., 2024). An dieser Stelle möchte ich mich noch einmal für alle Anregungen, fachlichen Aspekte und hilfreichen Ergänzungen bedanken. Darüber hinaus möchte ich auch allen Mitgliedern der Gruppe Klimadynamik und Modellierung für das wertvolle Feedback zu meinen Ergebnissen danken. Schließlich möchte ich mich auch beim Institut für Meteorologie und Geophysik der Universität Wien bedanken, wo ich während meines Masterstudiums im IT-Bereich arbeiten durfte. Die Unterstützung und Flexibilität, die mir entgegengebracht wurde, ermöglichte es mir, die Anforderungen von Arbeit und Studium erfolgreich zu bewältigen.

Contents

1. Introduction

1.1. Heatwaves and their driving mechanisms

The term "extreme" denotes the outermost (lat. extremus) manifestations under certain conditions. An extreme weather/climate, as defined by the Intergovernmental Panel on Climate Change (IPCC), is an event that is rare at a specific location and season. These extremes can manifest in various ways including extreme temperature events, intense precipitation events, droughts, hurricanes and other forms of extreme weather. One type of extreme temperature events are heatwaves, which are generally defined as several consecutive days with temperatures exceeding a certain threshold (Perkins et al., 2012), affecting a specific area. In this context heatwaves have a special role, because such events can range from few days up to several weeks, thus acting both as weather and climate extreme (Kirch et al., 2005).

This period with persistent hot conditions is often related to the large-scale dynamics, corresponding to a quasi-stationary high-pressure system. Even though the dynamical drivers can vary locally, typically clear sky conditions related to the subsiding motions connected to the high-pressure system are contributing to an increased amount of shortwave radiation reaching the surface (Pfahl et al., 2012). Beside land-atmosphere feedbacks, like feedbacks with soil moisture, which are also an important factor for the amplification of heatwaves (Sillmann et al., 2017), anomalies in the atmospheric circulation are the primary driving factors. When talking about circulation anomalies, specifically in the extratropical region, the literature often refers to the term "atmospheric blocking" or "blocking highs" (Zschenderlein et al., 2018), characterized by persistent anticyclones that interrupt the usual west-to-east flow of the atmosphere. Apart from the impact on near-surface-temperatures trough anomalies in the surface radiation budget, the anticyclonic circulation also affects temperature advection. This connection between warm temperature extremes and blocking patterns should not be neglected, especially with respect to global warming, even if other physical processes, or even their combination can also play a crucial role.

The combination of multiple drivers leading to significant impacts is referred to as a compound event (Zscheischler et al., 2018). Although we do not fully understand some single driver processes in our climate system, it is essential to better understand the interplay of physical mechanisms in generating heatwaves. Improving the understanding of these processes will contribute to more accurate representations of weather and climate extremes simulated by climate models (Sillmann et al., 2017).

1.2. Heatwaves and their linked impacts

Nowadays an increasing number of people are experiencing warmer temperatures, for this reason heatwaves have drawn considerable attention from a variety of research disciplines, as well as from the general public. Since heatwaves develop jointly in space and time it is particularly important to investigate different characteristics of heatwaves, because such events can have severe implications across various sectors. For example, the duration of heatwaves strongly influences the impact on natural ecosystems and human health, because it determines if an affected system can recover or not. In combination with the spatial component, heatwaves also cause adverse effects on agriculture and infrastructure, leading to strong economic losses. A representative example is the high mortality rate during the 2003 European heatwave, exceeding 70.000 (Black et al., 2004).

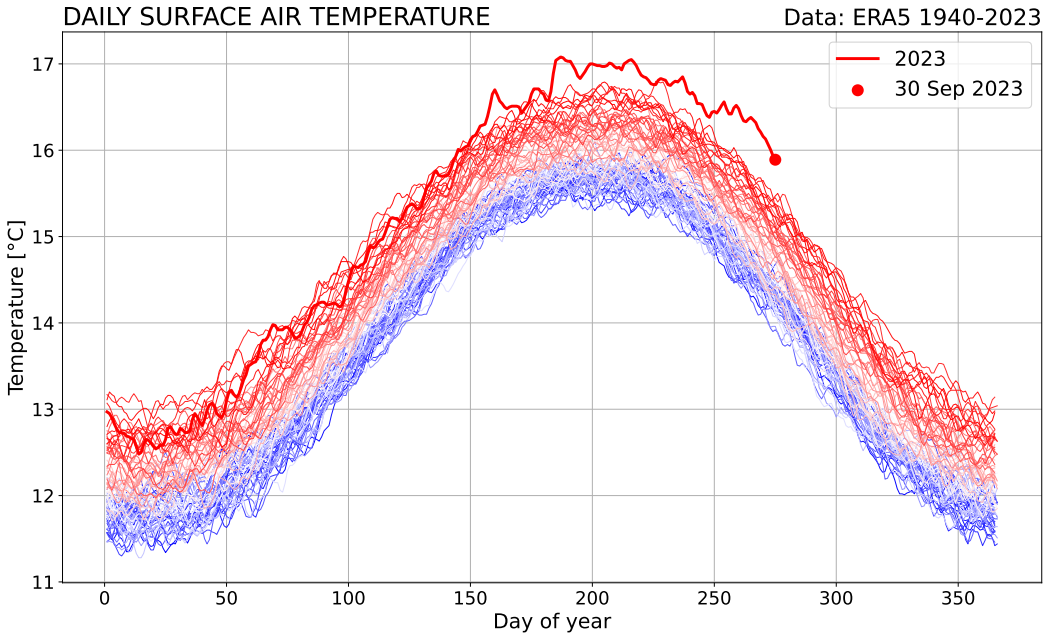

Figure 1.: Global daily surface air temperature (°C) from 1 January 1940 to 30 September 2023, plotted as time series for each year. 2023 is shown as thick red line, other years are shaded according to the respective year, from blue (1940) to red (2023).

More recent heatwaves as those that occurred in 2023 (Schäfer et al., 2023) have again set new temperature records with July being the hottest month ever recorded, as shown in Fig. 1. Additionally, 2023 was the hottest year on record, 1.48°C warmer than the pre-industrial level (`https://climate.copernicus.eu/global-climate-highlights-2023`). Beside global warming, climate change may also cause changes in the jet stream probably influencing the persistence of heatwaves. The development of these time series provides information about the future course of global surface air temperatures, implying that future warming strongly depends on future emissions, outlined in section 1.6.

1.3. Changes in heatwave properties

As described in section 1.2, human-induced climate change leads to a warming Earth which could consequently also lead to more intense and frequent heatwaves affecting the environment and society. This assumption has been investigated many times, as for example by Perkins-Kirkpatrick et al., 2020, who found that since the 1950s heatwaves have increased in their frequency, duration and intensity. This study highlights not only a positive, but also an accelerating trend in the presence of anthropogenic climate change. More recently Russo et al., 2023 showed that heatwaves from 1950-1985 intensified, becoming 10 times more frequent and three times more intense in 1986-2021. Moreover, they emphasize the crucial role of metrics, advocating for a heatwave intensity measure based on cumulative indices to facilitate comparisons across events of varying durations.

The changes in these properties have their origin in changes in the distribution of temperature extremes. Fig. 2 demonstrates that extreme temperatures, represented by the two ends of the distribution, can change due to a shift in mean temperatures or a change in the variability. A shift in the mean implies more hot extremes at the right end of the distribution, while a change in the variability can lead to more extremes in both directions.

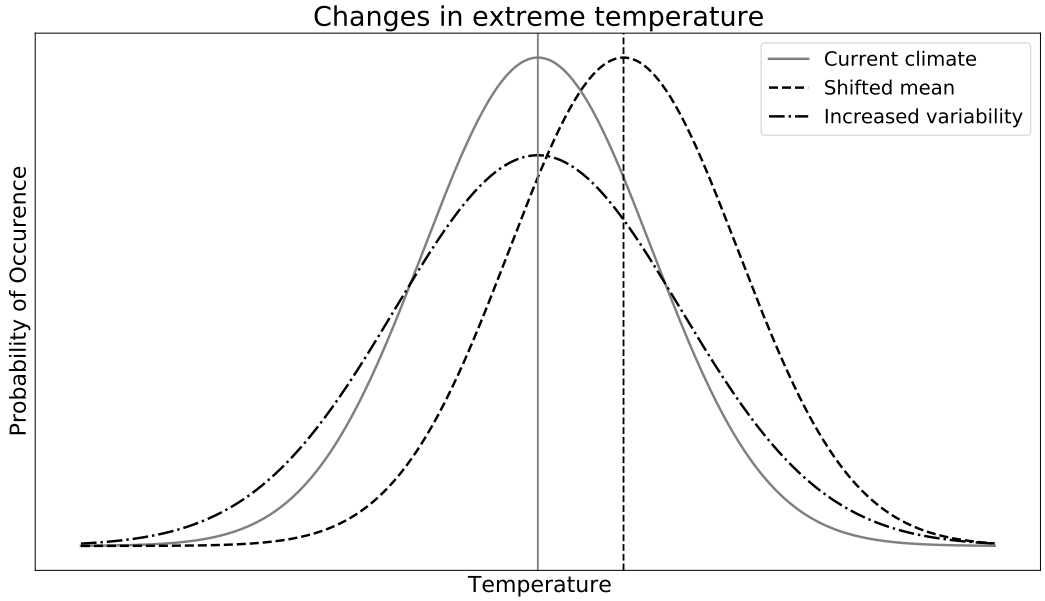

Figure 2.: Schematic showing changes in extreme temperature. The gray curve represents the current climate, while the black dashed/dashed-dotted lines illustrate a respective shift. Adapted from Figure SPM.3 of IPCC (2012).

Another study by Perkins et al., 2012 presented global changes in heatwaves using three different definitions and characteristics. Based on observational records from 1950-2011 (HadGHCND), all definitions show positive trends per characteristic, despite different quantitative results. Nevertheless, given the inconsistency in the definition of heatwaves, it has been difficult to document their changes at the global scale. Most publications studying heatwaves state that there is no universal definition to measure these events, making an exact comparison of existing results quite difficult (Perkins, 2015).

1.4. Defining heatwave events

Indeed, there is a strong sensitivity of heatwave characteristics to the extreme threshold definition. In the absence of a universal definition of a heatwave, various approaches are used to define such temperature 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. Because absolute thresholds are purely fixed, they do not account for regional variations in the temperature distribution. Although percentile-based extreme definitions allow temperature extremes throughout the year, most studies focus on the warmest period. In order to set the rarity of an extreme event, exceedances of a chosen percentile value, typically on a daily bases, are calculated. Selecting for example the 90th percentile means that, on average, 10% of days in the evaluated period are expected to exceed the calculated temperature threshold. As base-period, the period used to calculate extreme thresholds, the Expert Team on Climate Change Detection and Indices (ETCCDI) recommends to use a 30-year period, for example 1961-1990. There are some studies using longer base-periods or even shifting periods like Vogel et al., 2020. These shifting periods are used to understand changes in the climate system and to evaluate heatwave properties based on extreme events that still happen rarely, even as the climate warms up. This should answer the question how to define heatwaves in a changing climate. If heatwaves are defined using a fixed period this results in a strong increase in heatwave area, duration and magnitude in a warming climate due to the mean warming alone (Zhang et al., 2011). But if they are defined in a way that accounts for global warming using temporally varying thresholds, changes are relatively small according to the findings by Vogel et al., 2020. Therefore, changes in heatwave characteristics strongly depend on the for the analysis chosen threshold, highlighting the sensitivity of future heatwaves to the extreme threshold definition.

To increase the sample size used for the percentile calculation the ETCCDI suggests to use a running window of 5 days. However, many studies on heatwaves, as well as the often-used Heat Wave Magnitude Index (HWMI) by Russo et al., 2014, use longer windows. This index describes the intensity of an event by summing up the daily magnitudes for each day with the duration of a particular heatwave.

Recently, Brunner et al., 2024 showed that using such a seasonal running window leads to large and systematic biases in threshold exceedances and therefore in the frequency of temperature extremes. This running window bias originates from seasonally warmer days dominating the extreme threshold and can vary between seasons, regions, and datasets. Their results highlight that is particularly important to account for those biases to correctly interpret the effect on downstream impact metrics like the HWMI. Disregarding this bias in the temperature extreme frequency could in this case probably lead to a misinterpretation in the actual magnitude of heatwave events.

1.5. Simulating heatwave events

To gain insights into future changes of heatwaves and other climate extremes, simulations of global climate models (GCMs) can be a helpful tool. These numerical models use a discrete representation of the governing equations to simulate and predict the interactions of various components of the Earth's climate system including atmosphere, oceans, land surface and ice.

Due to the inherent variability in the outputs of different climate models, the comparison of these models becomes essential to improve our understanding of the climate system. Therefore, the objective of the Coupled Model Intercomparison Project (CMIP) is to evaluate the performance of climate models and identify uncertainties in GCMs. Since the release of the CMIP Phase 3 in 2005, many GCMs have provided daily data, which is key in the measurement of heatwaves. Throughout the evolving phases of the Coupled Model Intercomparison Project, GCMs have become more complex to better represent physicals processes and moved towards higher resolution (Eyring et al., 2016). Thus, in addition to observational and reanalysis data, GCMs represent an important source of robust and reliable climate information. It is essential to evaluate the model output of CMIP simulations to measure the representativeness regarding climate extremes in general. Existing literature has already evaluated CMIP model performance for a range of climate extremes. Di Luca et al., 2020, for example, showed systematic improvements in daily temperature extreme errors in CMIP6 compared to CMIP5, linked to increased horizontal resolution and model biases in land-atmosphere feedbacks (Sippel et al., 2017).

However, the heatwave frequency is systematically underestimated in the past two CMIP cycles, while heatwave duration and magnitude are both over- and underestimated depending on the region studied. Furthermore, comparisons between CMIP5 and CMIP6 show that improvements in skill for the heatwave metrics evaluated by Hirsch et al., 2021 are marginal.

One way to possibly achieve better skills in simulating heatwaves is the usage of high-resolution models on a km-scale. Because current global climate models use horizonal grid spacings of 50-100 km, processes that occur on scales smaller than the resolution need to be parameterized (i.e. mathematically represented, not simulated by the model). These parameterizations are the root source of uncertainties in current GCMs (Palmer et al., 2019). This could affect how heatwaves are represented because important small-scale processes like surface energy fluxes or soil moisture may not be accurately simulated, causing systematic biases (Vautard et al., 2013).

Because heatwaves are quite sensitive to internal climate variability, Perkins-Kirkpatrick et al., 2020 recommend to use an analysis period of at least 3-4 decades to robustly assess changes in heatwaves. This in turn makes the use of high-resolution models on a km-scale challenging, as a simulation over such a period requires enormous capacities for data

storage and computer resources. Despite these facts, climate models are still the key tool for understanding how climate changes with increased anthropogenic activity. Considering the noted limitations and evaluating the model's ability in reproducing historical heatwave statistics can give us insights into how heatwaves might respond to future projections.

1.6. Future heatwave diagnostics

The current state of research shows, that many studies, such as Rousi et al., 2023, primarily focus on analyzing historical heatwaves events using various data sources, including models, reanalysis, or observational data. The few existing studies dedicated to investigate future changes, consistently show that as greenhouse gas levels rise in future climate scenarios, all temperature extremes, particularly heatwaves, become more intense, frequent and prolonged. As described in section 1.4, this positive trend highly depends on the exact definition of temperature extremes. It indeed appears that nearly every climatological study examining heatwaves utilizes a different metric. The use of fixed thresholds in the context of future heatwaves seems somewhat eccentric, as the rarity of events loses its actual meaning.

This highlights the importance of studying future changes of heatwave properties. In connection with the points already mentioned, the primary objective of my thesis is to enhance our understanding of how global warming affects heatwaves. In order to investigate these future changes, I will use a relative, percentile-based threshold that is adapted to future climate conditions. This makes it possible to examine how selected CMIP models simulate heatwaves on a global scale. Since the GCMs are free running, meaning that they are not driven with observational data, the goal of this master thesis is not to exactly reproduce certain patterns of historical heatwaves, but rather a statistical analysis of the changes in heatwave properties.

While the impacts of heatwaves are to a certain degree well known, their joint temporal and spatial dimensions are rarely studied together. Unlike observed heatwaves, projected changes from climate models face fewer spatial and temporal issues, which simplifies the 3-dimensional analysis. To identify heatwave objects that are contiguous in space and time a clustering algorithm can be applied. These so-called connected components are used to label connected regions in binary images, where in this case the binarity is determined by exceeding a calculated threshold value on a grid point basis.

To study future changes in heatwaves beyond mean global warming, moving thresholds that are adapted to the respective climatic conditions can be used. An analysis of future changes by Schielicke et al., 2022, investigated the processes associated with European heatwaves. In this study, they dealt with the question whether changes in thermodynamic or dynamic processes lead to an amplification of heatwaves beyond mean warming. Their main finding was an amplification of heatwaves due to increased diabatic heating in the boundary layer. Moreover, increased heating is linked to stronger adiabatic warming trough enhanced descent of air masses. This is in contrast with the findings of Vogel et al., 2020 who suggested that changes in dynamics should be less important. More investigation is needed to assess the findings' robustness, including other climate models as demonstrated in this thesis.

2. Research questions

The overall goal of my thesis is to improve the understanding of changes in heatwave properties under climate change. This is accomplished by constructing a Python-based framework for detecting heatwaves in CMIP6 data. Within this work I address the following two research questions:

1. What is the effect of a running window when using moving thresholds for defining heatwaves? The method chosen for calculating thresholds is crucial, as it influences the number of hot days. The literature review has shown that many existing studies use relatively large running windows of 15- or even 31-days. When using the 99th percentile as threshold, one would expect that the overall exceedance frequency is 1%. However, recently, Brunner et al., 2024 showed that the use of such long running windows introduces a systematic bias that leads to a striking underestimation in the expected extreme frequency. Beside that finding they also proposed a simple bias-correction method, which is also used in this thesis to explore effects on downstream impact metrics.

2. How do different heatwave characteristics respond to global warming in CMIP6 models? In order to investigate future changes two time periods, one in the past and one in the future, are compared to each other. Using moving thresholds, the effect of global warming is considered by defining projected heatwaves relative to the future. Therefore, it is hypothesized that future heatwaves do not change much. If changes in heatwave characteristics with fully moving thresholds still occur, they could be related to physical drivers such as circulation changes or land-atmosphere feedbacks. Depending on the used definition, I will quantify this potential trend trough some statistical analysis.

Compared to other previous studies, which mostly use only the warmest months for their analysis, I am going to evaluate the full seasonal cycle allowing heatwaves to occur throughout the whole year. While heatwaves are often associated with summer months, in some regions, deviations in the seasonal temperature distribution can take place at any time of the year, including also winter months. The use of global climate models, as described in section 1.5, allows a global analysis of heatwaves, whereby this work is not limited to any specific region. In combination with the first research question, the corrected and biased results are also compared, whereby it is expected that without applying the bias-correction, we obtain an underestimation of heatwave properties.

3. Data and Methods

The first part of this chapter describes the used dataset as well as the justification for the models I use. Afterwards, the steps for calculating percentile-based thresholds, extreme frequency biases, and the bias-correction method are explained. The last two sections include the explanation of the computation of spatiotemporal heatwaves as well as the analysis of future changes through statistical heatwave properties.

3.1. Climate Models: CMIP6

The analysis is based on simulations of the daily maximum 2m surface air temperatures (variable tasmax) from the Coupled Model Intercomparison Project phase 6 (Eyring et al., 2016). The actual data, providing daily maximum temperature for historical and SSP370 scenarios, is available from the ETH Zurich CMIP6 next generation archive (Brunner et al., 2020) on a common 2.5°×2.5° longitude-latitude grid. The SSP370 scenario represents a specific socio-economic development pathway defined by a prescribed trajectory of atmospheric greenhouse gas concentrations, designed to achieve a radiative forcing of 7 W/m^2 by the year 2100 (Meinshausen et al., 2020).

The data used for the analysis covers two distinct periods. A historical period that spans from 1850-01-01 to 2014-12-31 and a projection period that extends from 2015-01-01 to 2100-12-31. Combining these two periods results in a continuous, comprehensive dataset spanning 251 years across 27 distinct climate models on a global scale.

3.1.1. Model selection

Since an analysis of all CMIP6 models would be too time consuming for the scope of my thesis, I will only use simulations from 5 selected global climate models based on a simulation skill metric from the International Land Model Benchmarking (ILAMB) system (Collier et al., 2018), also used by Hirsch et al., 2021. This metric evaluates the simulation skill between some observational data (O) and the model output (M) for different heatwave characteristics. The relative bias is calculated as the difference between the modeled and observed time mean, weighted by a form of the standard deviation of the observations (for big enough samples):

$$\epsilon_{bias} = (\overline{M} - \overline{O}) / \sqrt{\frac{1}{N} \sum_{i=1}^{N} (O_i - \overline{O})^2} \tag{1}$$

The subscript i respectively N in Eq. 1 denotes the number of the heatwave season over which the calculation is made. For the evaluation of the GCMs, heatwave characteristics are compared to those derived from observed daily maximum temperature from the Berkeley Earth dataset (Rohde et al., 2013). This dataset is available for the years 1950-2014, which results in a total of $N = 65$ heatwave seasons.

Within the CMIP6 ensemble, the relative skill score evaluates how individual models rank across several heatwave characteristics and in specific climate reference regions. The BCC-CSM2-MR model consistently exhibits the lowest skill ranking across multiple regions. Conversely, the INM-CM5-0 model often has the highest skill for several characteristics particularly in Eurasian regions. The ACCESS-CM2 model usually falls within a middle ranking across different regions and characteristics.

To determine the impact of the model resolution, I will also consider a model that was run at two different resolutions, namely MPI-ESM1-2-HR and MPI-ESM1-2-LR where HR stands for high-resolution and LR for low-resolution. Finally, I have chosen a set of five distinct GCMs, ranging in the simulation skill from high to low.

MODEL-ID	**Native Resolution [lat x lon]**	**Modeling center**
MPI-ESM1-2-HR	192 x 384	Max-Planck-Institute for Meteorology, Germany
MPI-ESM1-2-LR	96 x 192	Max-Planck-Institute for Meteorology, Germany
INM-CM5-0	120 x 180	Institute for Numerical Mathematics, Russian Academy of Sciences
ACCESS-CM2	144 x 192	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia
BCC-CSM2-MR	160 x 320	Beijing Climate Center, China Meteorological Administration

Table 1.: Overview of selected climate models. Note that the skill metrics were not calculated within this thesis, but the results from the study by Hirsch et al., 2021 were used.

One ensemble member is used per model, except for the ACCESS-CM2 model where 5 ensemble members are used. These different realizations of a certain model were generated by different initial conditions and are used to examine the sensitivity to internal variability and to test the statistical significance of the calculated results (see section 4.4).

3.2. Calculation of percentile-based extreme thresholds

For a (relative) percentile-based approach the Expert Team on Climate Change Detection and Indices (ETCCDI) recommends to use the 90th percentile of a 30-year period (e.g. 1961-1990) with a 5-day running window (`https://etccdi.pacificclimate.org/list_27_indices.shtml`). The percentile method for extreme events is a statistical approach used to determine thresholds beyond which an event is considered as extreme.

Lyon et al., 2019 found, that using a 5-day window leads to a quite noisy extreme threshold, with the result that most studies (like Russo et al., 2023) are using longer windows. Therefore, I adapt the recommended approach by computing the 99th percentile temperature distribution of each day based on the 31 neighboring days and 31 neighboring years. Using the 99th percentile, for example, instead of the 90th percentile provides a more conservative estimate capturing events that are even more extreme, as the higher the percentile chosen, the more extreme the values it captures. The 31-day running window is used to increase the sample size for percentile calculations to get smooth extreme thresholds, while the 31-year running window accounts for the impact of global warming. This method is also known as moving threshold, because of using a shifted baseline instead of a fixed baseline as described in section 1.4. As illustrated in Fig. 3 the calculation involves 31 neighboring days and 31 neighboring years resulting in a total of $31 \times 31 = 961$ data points for each grid cell and each day separately.

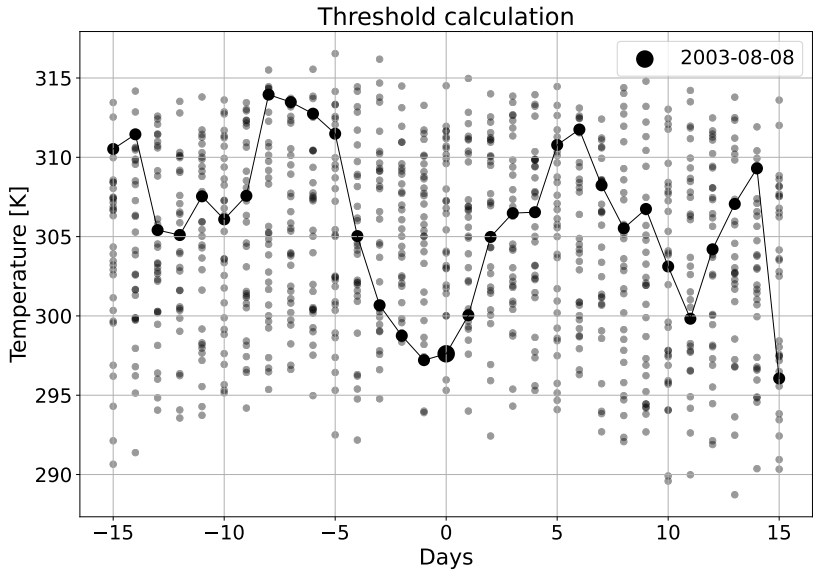

Figure 3.: Sample illustration of the calculation for the moving threshold for a grid cell close to the Austrian-Czech border from INM-CM5-0. The data points shown are the 961 used to calculate the extreme threshold for 2003-08-08 (thick black dot). The x-axis represents the 31-day running window, while the small gray dots represent the corresponding climatology over 31 years.

In the following the calculation of the 99th percentile using a 31-day window is described as $TX99p31w$. If one adds an additional running window, as already described before, the calculation of the daily temperature thresholds can be described by the following formula:

$$TX99p31w(d) = Percentile(\{T_{i,j}(d)\}_{i,j=-15}^{+15}, 99\%) \tag{2}$$

In Eq. 2 subscript i represents the previous and following 15 days, while subscript j describes the previous and following 15 years. The moving thresholds are then calculated for every grid point, every date in the timeseries and for each of the model selected. It should be noted, that the first and last 15 years of the original timeseries cannot be used for further analysis, because there is simply no data before 1850 and after 2100 needed for the threshold calculation.

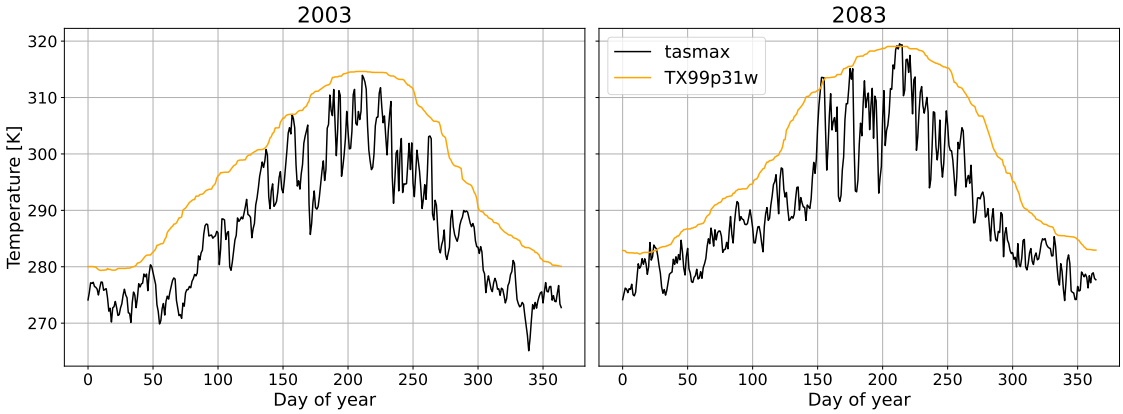

Figure 4.: Moving thresholds to compute hot days (orange line). Daily maximum temperature (black line) for 2003 and 2083 from INM-CM5-0 for a grid cell close to the Austrian-Czech border.

A hot day is then defined as a day that exceeds the 99th percentile distribution as calculated by Eq. 2. Thereby the threshold of the 220th day of the year 2003 in Fig. 4 (left) corresponds to the 99th percentile from the data basis of Fig 3. The right-hand side of Fig. 4 is intended for demonstration purposes, whereby the amplitude in the year 2083 is much more pronounced due to global warming simulated by the climate model. By selecting the 99th percentile one would expect that on average 1% of the data is exceeding the threshold, representing the extreme end of the distribution of the maximum temperature. However, this assumption is likely wrong, as will be shown later in this thesis.

3.3. Calculation of extreme frequencies and biases

Throughout the research phase of this master thesis it was found that a 31-day window leads to a smoother threshold, but the distribution of extremes is no longer uniform. Consequently, introducing a seasonal running window may result in a bias in the threshold exceedances, depending on the used window size. To investigate this possible bias, one can as a first step calculate the extreme frequency as the fraction of days exceeding the extreme threshold in a given period in percent:

$$f(p,w) = \frac{x_{exceed(p,w)}}{x_{base}} \times 100\% \tag{3}$$

with p indicating the percentile and w the window size. In my case $p = 99$ and $w = 31$. For example, the period 1960-1990 can be used as a baseline. Based on Brunner et al., 2024 results, a bias in the extreme frequency is expected, therefore I introduce the expected frequency which is per definition 1% for the 99th percentile. However, by using a 31-year window the data is never fully in-sample, resulting in the fact that the actual expected frequency is never exactly 1%.

Based on the extreme frequency one can calculate the relative frequency bias, which is defined as the actual extreme frequency (as defined by Eq. 3) minus the expected frequency, divided by the expected frequency in percent:

$$f'(p,w) = \frac{f(p,w) - (100-p)}{100-p} \times 100\% \tag{4}$$

with p and w again indicating the percentile respectively the window since. An analysis on grid point basis makes it also possible to investigate the spatial distribution of relative frequency biases, discussed in section 4.2. Note that Eq. 4 is one of the key points of the study by Brunner et al., 2024, whereby I am going to use this approach to show effects on downstream impact metrics such as the area of heatwaves.

3.4. Bias correction: Removing the mean seasonal cycle

To address the issue described in section 3.3, Brunner et al., 2024 also proposed a simple bias-correction method, involving the removal of the mean seasonal cycle before percentile threshold calculation, which is also used here. The mean seasonal cycle is calculated using a 31-year running window on a day of the year basis for each grid cell:

$$\overline{T}(d, y) = \frac{1}{31} \sum_{i=-15}^{i=15} T(d, y_i) \tag{5}$$

After calculating the mean season cycle for each day of the historical and future period, the mean seasonal cycle, as calculated in Eq. 5, is subtracted from the daily maximum temperatures at each grid cell resulting in daily anomalies:

$$T_{anom}(d, y) = T(d, y) - \overline{T}(d, y) \tag{6}$$

Based on the anomalies, calculated by Eq. 6, the extreme thresholds are computed again by using the method already explained in section 3.2.

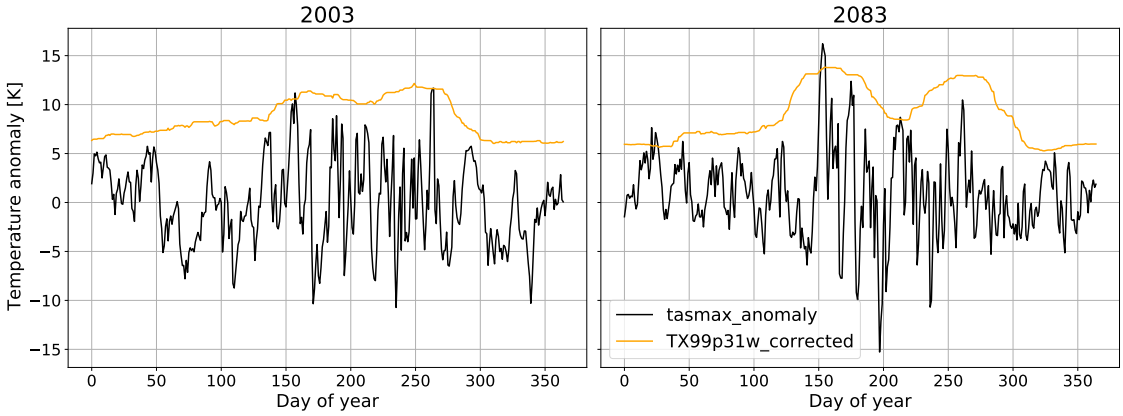

Figure 5.: Applying the bias-correction to calculate moving thresholds (orange line). Anomalies of the daily maximum temperature (black line) for 2003 and 2083 from INM-CM5-0 for a grid cell close to the Austrian-Czech border.

This solution has already been suggested by Folland et al., 1999, although they did not explain the reason for their recommendation. The interaction between the running window and the seasonal cycle may explain why grid cells with a pronounced seasonal cycle tend to exhibit a relatively high bias. By taking seasonal variations into account, I allow the use of longer window sizes without introducing substantial biases into the extreme frequency. Since Brunner et al., 2024 have shown a general reduction in the relative frequency bias using the ERA5 and CMIP6 dataset, it can generally be assumed that this approach can also be used in my case.

3.5. Spatiotemporal heatwaves

To compute spatiotemporal heatwaves, I will apply a three-dimensional clustering algorithm to identify events that are continuous in space and time over both the land and the ocean. This method is essential for visualizing contiguous regions with temperatures above a defined threshold, making it easier to quantify the spatial distribution of certain grid cells that are influenced by a heatwave. The clusters are computed using the connected components function cc3d in Python (`https://pypi.org/project/connected-components-3d/`). The core of the package uses a variant of the two-pass method proposed by Rosenfeld et al., 1966, which was originally developed for image processing to identify and label connected regions in a 2D binary image. The algorithm is extended to work in 3D, which allows me to consider the temporal evolution of already labeled patterns, supporting the exploration of spatiotemporal heatwaves. Additionally, my co-supervisor (Lukas Brunner), has added a component to the version of cc3d that I use, addressing the issue where the original algorihm cannot correctly handle edge crossings on the latitude (e.g connected components that extent over $180°W/180°E$).

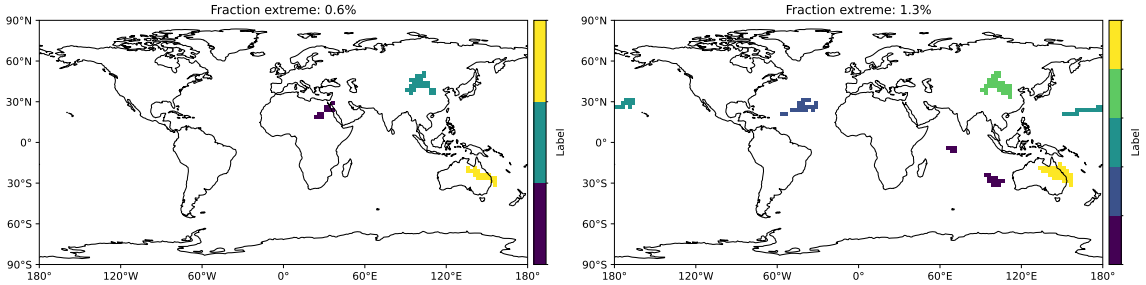

Figure 6.: Connected components for the ACCESS-CM2 model for 10-07-1960. Biased case (left) and corrected case (right).

The percentile-based thresholds as calculated in section 3.2 are not directly used for the labeling routine, but rather the deviation from the threshold, here defined as the so-called exceedance. This exceedance can in principle be interpreted as a binary code: If the daily maximum temperature is greater than the computed grid cell threshold the value is set to 1, indicating the occurrence of a heatwave, otherwise it is zero. Each label in Fig. 6 is representing a heatwave, whereby in the corrected case 1.3% of all grid cells are marked by a heatwave event.

After clustering single hot days into spatiotemporal patterns, only the 100 largest events in a past (1960-1990) and a future period (2050-2080) are used for further analysis. Thus, by studying the 100 biggest events, I target the largest and potentially most impactful heatwaves, but can also already explain a large part of the original distribution containing several ten thousand single events.

3.6. Statistical heatwave characteristics

After calculating the spatiotemporal patterns as described in section 3.5, the heatwave events are characterized by their area, duration and magnitude. To estimate the heatwave area (HWA) I am counting the grid cells that are affected by a heatwave over the entire duration of the event. Due to the longitudinal convergence towards the poles, the grid cells vary in size, which leads me to consider area weighting. By using the regridded cell area available from the model output, I can directly calculate the heatwave area in km^2 if the size and number of the grid cells influenced by a heatwave are known. Since I am considering the summed area over the duration in days, HWA consequently receives the unit km^2 day.

The heatwave duration (HWD) is the time period of all cells contributing to the heatwave. The duration starts when a label appears for the first time and ends when the label has completely disappeared again. In addition, each grid cell at each timestep has a certain threshold exceedance. The area weighted sum of this exceedances over the entire duration of a heatwave event results in the heatwave magnitude (HWM), describing the intensity or severity of an event.

To investigate future changes and to study the impact of the bias-correction, for each selected model two different time periods and the difference in the biased and corrected data are analyzed. The characteristics of heatwaves are presented using their distribution function in form of histograms based on the 100 largest evenzs, while the changes can be investigated using various statistical approaches.

3.6.1. Calculating the correlation coefficient

The Person correlation coefficient (PCC) is a correlation coefficient that measures the linear correlation between two sets of data. It relies on the assumption of a normal distribution for both variables considered. Since a non-normal distribution could lead to an overestimation or underestimation of the relationship between the variables, I assume that the data is at least close to a normal distribution. The Pearson correlation coefficient is calculated as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7}$$

Where n is the sample size, x_i and y_i are the individual sample points of the two variables and $\bar{x}$ and $\bar{y}$ are the corresponding sample means. The right-hand side of Eq. 7 equals the ratio between the covariance of the two variables and the product of their standard deviations. The correlation coefficient ranges from -1 to 1, thus offering various interpretations, as outlined in Table 2.

Correlation	Interpretation
0.90 – 1.00 (-0.90 – -1.00)	Very high positive (negative) correlation
0.70 – 0.90 (-0.70 – -0.90)	High positive (negative) correlation
0.50 – 0.70 (-0.50 – -0.70)	Moderate positive (negative) correlation
0.30 – 0.50 (-0.30 – -0.50)	Low positive (negative) correlation
0.00 – 0.30 (0.00 – -0.30)	Negligible correlation

Table 2.: Interpretation of the correlation coefficient (Mukaka, 2012).

This statistical parameter is used to better understand the correlation between the heatwave duration (HWD) and the heatwave area (HWA) as well as their three-dimensional relationship. By considering two different time periods, as well as a corrected and a biased version of the results, I compute the following four correlations for each model selected:

1. $r(HWD_{corrected-past}, HWA_{corrected-past})$
2. $r(HWD_{biased-past}, HWA_{biased-past})$
3. $r(HWD_{corrected-future}, HWA_{corrected-future})$
4. $r(HWD_{biased-future}, HWA_{biased-future})$

3.6.2. Permutation tests

To quantify significant changes of heatwave characteristics I perform a permutation test and compare the biased and the corrected case. This permutation test is a non-parametric statistical test that assesses the significance of an observed effect by randomly permuting the data and recalculating the test statistic multiple times to build a null distribution. It provides a way to test hypotheses without making assumptions about the underlying distribution of the data. With the help of this methodology, I want to answer the question whether the means of two distributions are statistically significantly different from each other. As null hypothesis, the hypothesis stating that there is no statistical difference, I assume that the difference in the mean values of heatwaves characteristics from the future period minus the historical period of the largest 500 heatwaves (100 x 5 members, as described in section 3.1.1) is not different from zero. Additionally, the statistic of the original data is compared to the null distribution to determine the p-value. A small p-values suggests that the observed results are unlikely to have occurred if the null hypothesis were true, leading to the rejection of the null hypothesis and suggesting that there is statistical significance in the difference being tested. On the other hand, a large p-value suggests that the observed results are plausible under the null hypothesis, and there is insufficient evidence to reject it.

4. Results and Discussion

In this chapter, I present and discuss the results derived from my analysis. It is organized as follows: In section 4.1 I analyze the percentile-based extreme frequencies and show the effect of using a bias-correction method on a grid point basis. In section 4.2 this is expanded to a global level. Section 4.3 investigates the effect of using moving thresholds to define heatwaves in a future climate. The chapter concludes by evaluating the statistical significance of using a bias-correction method for its effects on downstream impact metrics.

4.1. Percentile-based extreme frequencies

This section demonstrates that a running window for percentile-based threshold calculations introduces systematic biases. This leads to a striking underestimation of the extreme frequency, depending on the window size used (Brunner et al., 2024). By applying the bias-correction method explained in section 3.4, one cannot fully eliminate the bias, but at least reduce it substantially. Note that in the following only two of the five models are compared directly with each other, namely ACCESS-CM2 and INM-CM5-0, as these two models show the greatest differences in the simulated seasonal cycle. The results of the other models can be found in the Appendix (see A).

4.1.1. Biased extreme frequencies

The left panel of Fig. 7 shows seasonal variations in the threshold exceedances for a selected grid cell in the Central Atlantic near the West African coast for two different GCMs. From this view it is not really clear which threshold line corresponds to which individual year, essential for the exceedance to define a hot day. By using a 31-year moving threshold, there is not just one threshold line, but 31, one for each year. Therefore, I computed the monthly averaged frequencies using daily exceedances grouped by all months. The right panel of Fig. 7 shows in general an underestimation in the monthly averaged frequencies. Particularly the ACCESS-CM2 model shows monthly averages of 0% (May, July, August, October, December) leading to a mean frequency of 0.20%. This implies that in all of these months in the period 1960-1990 not a single day was exceeding the calculated extreme threshold.

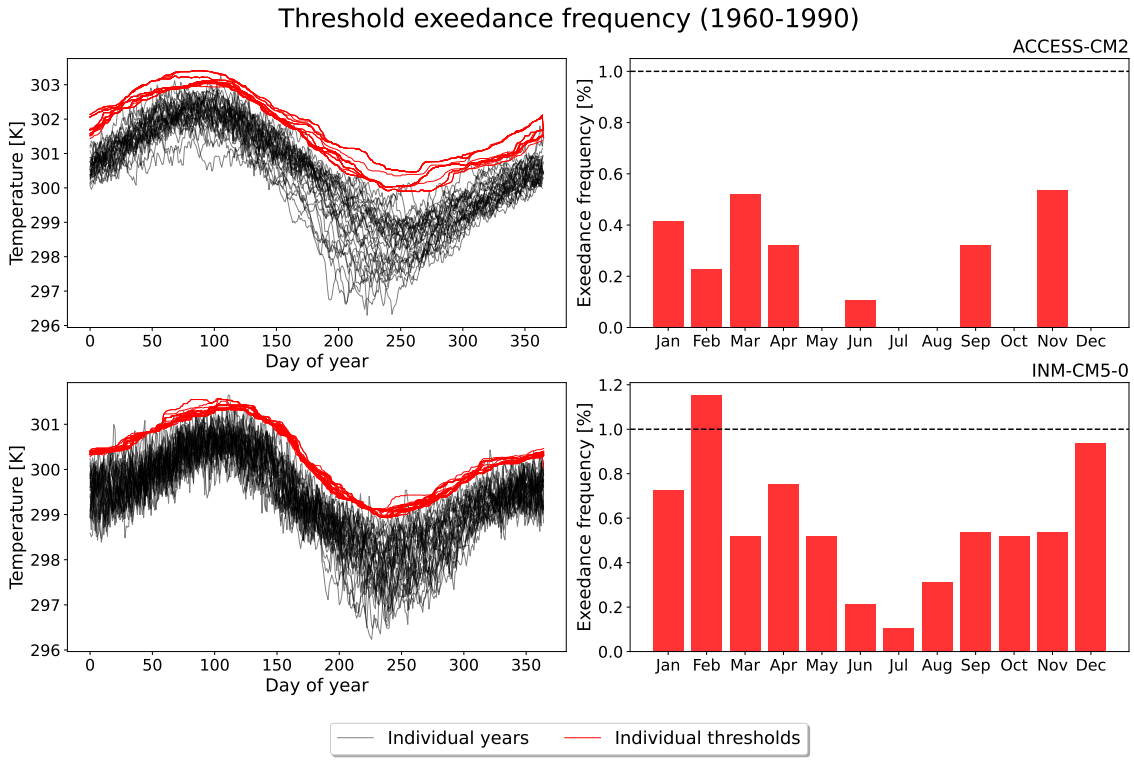

Figure 7.: Daily threshold exceedances and monthly averaged frequencies at a selected grid cell ($lat = -1.25°$, $lon = 358.75°$) from the ACCESS-CM2 (top) and INM-CM5-0 (bottom) model for the biased case. Each black line corresponds to an individual year in the period 1960-1990 while each red line represents the corresponding threshold. The mean frequency is 0.20% respectively 0.57%.

By analyzing the bottom row of Fig. 7, one can see that months with a strong seasonal temperature gradient are the drivers of the underestimation of the expected extreme frequency. In this case, it is particularly the months of June, July and August that show a significant underestimation. Against expectations, this striking underestimation of threshold exceedances may seem surprising. The 31-day window used to calculate percentile-based thresholds is symmetric and captures both seasonally colder and warmer values. This balance is expected to result in a smooth extreme threshold, exceeded for 1% of the days across the seasonal cycle and globally. Brunner et al., 2024 have highlighted that this expectation is incorrect, because the threshold is typically defined using a high percentile rather than the mean of the underlying distribution. Therefore, the threshold is dominated by the seasonally warmer days in the window. If there is a strong gradient in the seasonal cycle, it gets significantly harder for the central day to exceed the threshold. Due to the fact that the calculated threshold is actually too high, we experience a systematic underestimation of the exceedance frequency.

4.1.2. Bias-corrected extreme frequencies

The bias-correction method seems to work quite well for INM-CM5-0 while for ACCESS-CM2 it reduces the bias only slightly for the selected grid point here. The mean frequency for the ACCESS-CM2 model is only increased from 0.20% to 0.46%. As already discussed, strong variations in the seasonal cycle create a bias in the extreme frequency. A closer look at the top row of Fig. 8 shows that although the mean seasonal cycle has been removed, a residual gradient still remains within the seasonal cycle. Particularly noticeable is that in the monthly average, there are still months with an exceedance frequency of 0% (July and October). This is probably due to changes in the internal day-to-day variability (black lines are further apart) leading to a "curvy" threshold and therefore to a substantial bias and a failure to reach the expected exceedance frequency.

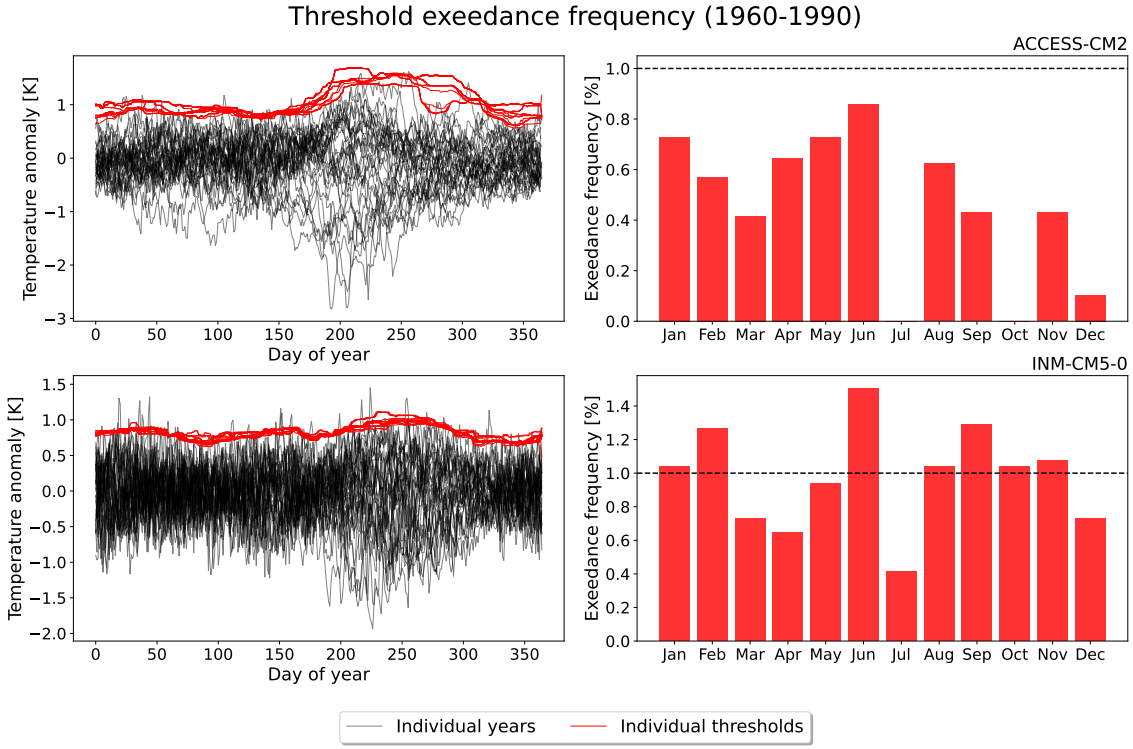

Figure 8.: Daily threshold exceedances and monthly averaged frequencies at a selected grid cell ($lat = -1.25°$, $lon = 358.75°$) from the ACCESS-CM2 (top) and INM-CM5-0 (bottom) model for the corrected case. The mean frequency is 0.46% respectively 0.98%.

Brunner et al., 2024 identified a number of errors due systematic running window bias. Especially when comparing neighboring grid cells spatial inhomogeneities can occur from differences between regions with strong and weak biases. This leads to an unequal distribution of extremes between different regions. They also found that the inhomogeneity is generally increasing with increasing percentile and window size, with differences in my case potentially exceeding 40%.

The bottom row of Fig. 7 shows the same case as in Fig. 8, highlighting the benefit of removing the mean seasonal cycle before calculating the percentile-based thresholds. For the selected grid cell in the Central Atlantic the mean frequency for the INM-CM5-0 model increased from 0.57% to 0.98% which means that there is virtually no bias anymore. For this reason, it is presumed that the bias-correction generally performs equally effectively on a global scale. To really asses the quality of the bias-correction, this assumption is examined in more detail in the following section 4.2.

As another source of error, Brunner et al., 2024 identified artificial dataset differences that arise due to differences in the spatial distribution of biases between datasets. Because the GCMs I am using misrepresent to seasonal cycle to a certain extent (e.g. compared to ERA5), respectively simulate it differently, the differences in the extreme frequency can be very pronounced (compare left and right panels of Fig. 8). However, by applying the bias-correction method one can show the convergence of the extreme frequency to the expected extreme frequency for all models on a grid point basis (see A, Fig. 26 and 27).

4.2. Relative frequency bias

This section deals with the relative frequency bias which is defined as the deviation from the expected exceedance frequency. The left column of Fig. 9 shows the full spatial distribution of the relative frequency bias averaged over all seasonal cycles of the period 1960-1990 for three different models. Locally this bias can exceed -50%, depending on the used model.

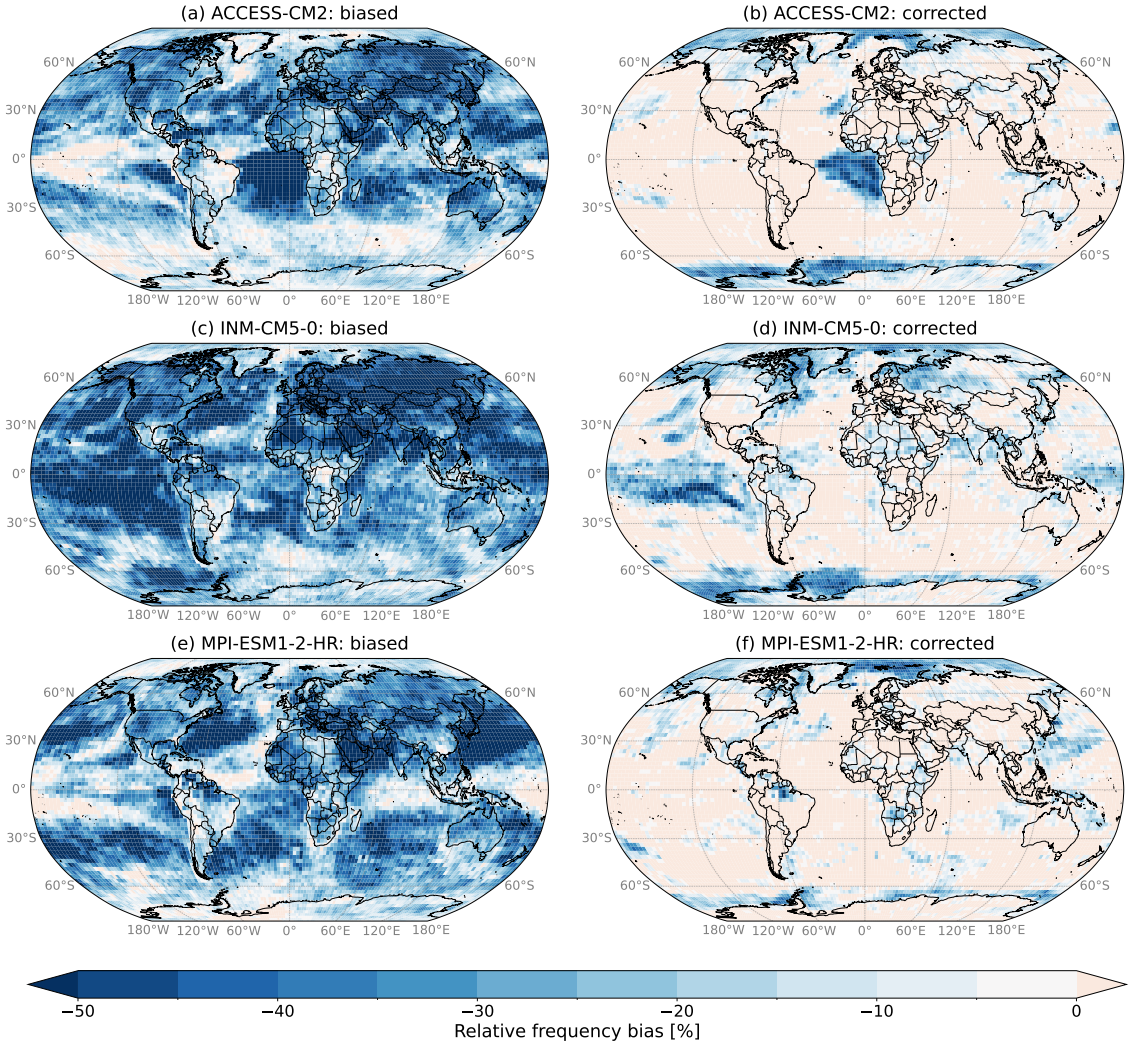

Figure 9.: Biases in the frequency of temperature extremes in ACCESS-CM2 (a, b), INM-CM5-0 (c, d) and MPI-ESM-HR (e, f). Spatial distribution of biases in the frequency of daily maximum temperatures based on exceedances of the 99th percentile using a moving threshold in the period (1960-1990). (a, c, e) without bias correction and (b, d, f) with bias correction.

In the period 1960-1990 the mean bias lies between approximately -27% and -50%, depending on the climate model considered. After removing the mean seasonal cycle, the frequency bias for the annual, global mean case is substantially reduced (see Table 3). As a consequence, the spatial inhomogeneity is also reduced, although some patterns with systematic bias still remain. In the best-case scenario, however, we would expect that the applied method is fully eliminating the bias, reducing the relative frequency bias to 0% in the global distribution. Even when considering the relative frequency bias for all selected models (see Table 3) none of them allows a complete reduction, with the BCC-CSM2-MR model having the highest bias values for both cases.

MODEL-ID	**without biascorrection [%]**	**with biascorrection [%]**
INM-CM5-0	-36.04	-4.36
MPI-ESM1-2-HR	-28.68	3.28
MPI-ESM1-2-LR	-30.59	4.50
ACCESS-CM2	-26.87	3.9
BCC-CSM2-MR	-50.03	-9.18

Table 3.: Relative frequency biases for the global mean case, averaged over the period 1960-1990.

After Brunner et al., 2024 found that in the ERA5 dataset, the 1961-1990 mean, global mean, relative frequency bias can be reduced from -10% to -0.5%, the remaining bias is likely caused by the methodology used to calculate the extreme thresholds. While they defined extreme thresholds over a fixed period of 30 years, I am using shifting periods as explained in section 3.2. This approach may lead to a discontinuity in the exceedances and is in the following referred to as the "out-of-sample problem". Since I am using running yearly windows, the data used is per definition never fully in-sample. Therefore one would never expect exactly 1% exceedance rates even when removing the mean seasonal cycle first. Specifically, being out-of-sample tends to result in extreme frequencies higher than 1%. If the correction manages to completely remove the seasonal cycle this ends up in an overestimation of the exceedance frequency, while places where some seasonal cycle remains still show a residual bias (see Fig. 9 b, d, f). The calculated global mean bias is then the average over these regions and can of course vary between models. The residual bias due to the out-of-sample problem remains uncorrected in the following, but it is acknowledged and accounted for in the interpretation of the results since it can also affect heatwave characteristics.

When contrasting the findings of section 4.1 and 4.2, particularly when comparing Fig. 9 (b) and (d), (f), one might wonder why the bias correction for the INM-CM5-0 and MPI-ESM-HR model works reasonably good in the Central Atlantic, while for the ACCESS-CM2 model, this is not the case. Initially without applying the bias-correction method, the ACCESS-CM2 model exhibits a stronger relative frequency bias compared to all other models in this region. Due to the change in the internal variability the INM-CM5-0 model simulates only weak changes in norther summer, while for ACCESS-CM2 the changes are quite strong leading to a "bend" in the threshold and consequently to a bias. Similar effects are simulated for the INM-CM5-0 model for the southeast pacific region. Together with the effect of the out-of-sample problem, the bias-correction method used may reach its limits in those cases, but only on regional scale.

What all of the analyzed models have in common, is that in the subpolar ocean on the Southern Hemisphere the relative frequency bias can still exceed 10%. In some cases, the bias even increases for some grid cells although the mean seasonal cycle has been removed. Examining a grid cell in these regions reveals that the threshold remains relatively flat, closely aligning with the path of the 0°C isotherm (see left panel of Fig 10). Since there is no pronounced gradient in the seasonal cycle in this case, there is virtually no systematic underestimation in the exceedance frequency due to the 31-day running window.

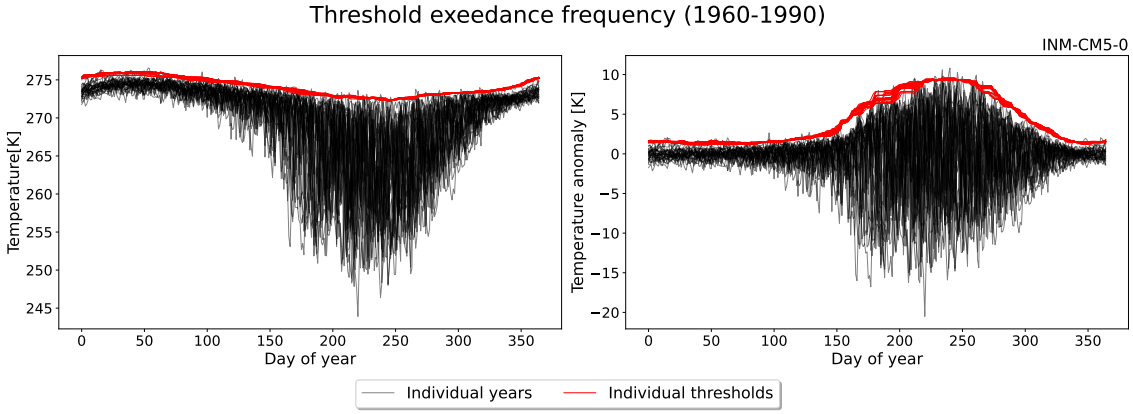

Figure 10.: Daily threshold exceedances a selected grid cell ($lat = -63.75°$, $lon = 358.75°$) from the INM-CM5-0 model with seasonal cycle not removed (left) and with seasonal cycle removed (right).

During winter in the southern hemisphere, temperatures can be much lower in comparison with the mean seasonal cycle, which leads to a pronounced amplitude in the anomalies during this period (see right panel of Fig. 10). This in turn, results in a maximum in the threshold distribution when the mean seasonal cycle is removed first, accompanied by a strong gradient leading to an increase in the relative frequency bias.

The following two illustrations show a sort of summary of the previous two sections in form of the seasonal cycle of the mean bias in the northern (see top of Fig. 11) and southern (see bottom of Fig. 11) hemisphere. On both hemispheres, a clear reduction in the bias due to the correction is detectable. The striking underestimation of threshold exceedances leads to a relative frequency bias up to -50% particularly in the transition seasons. In the summer months of the respective hemisphere, the bias is comparably small, while according to Fig. 9, the bias still present after the correction occurs in most cases over oceans.

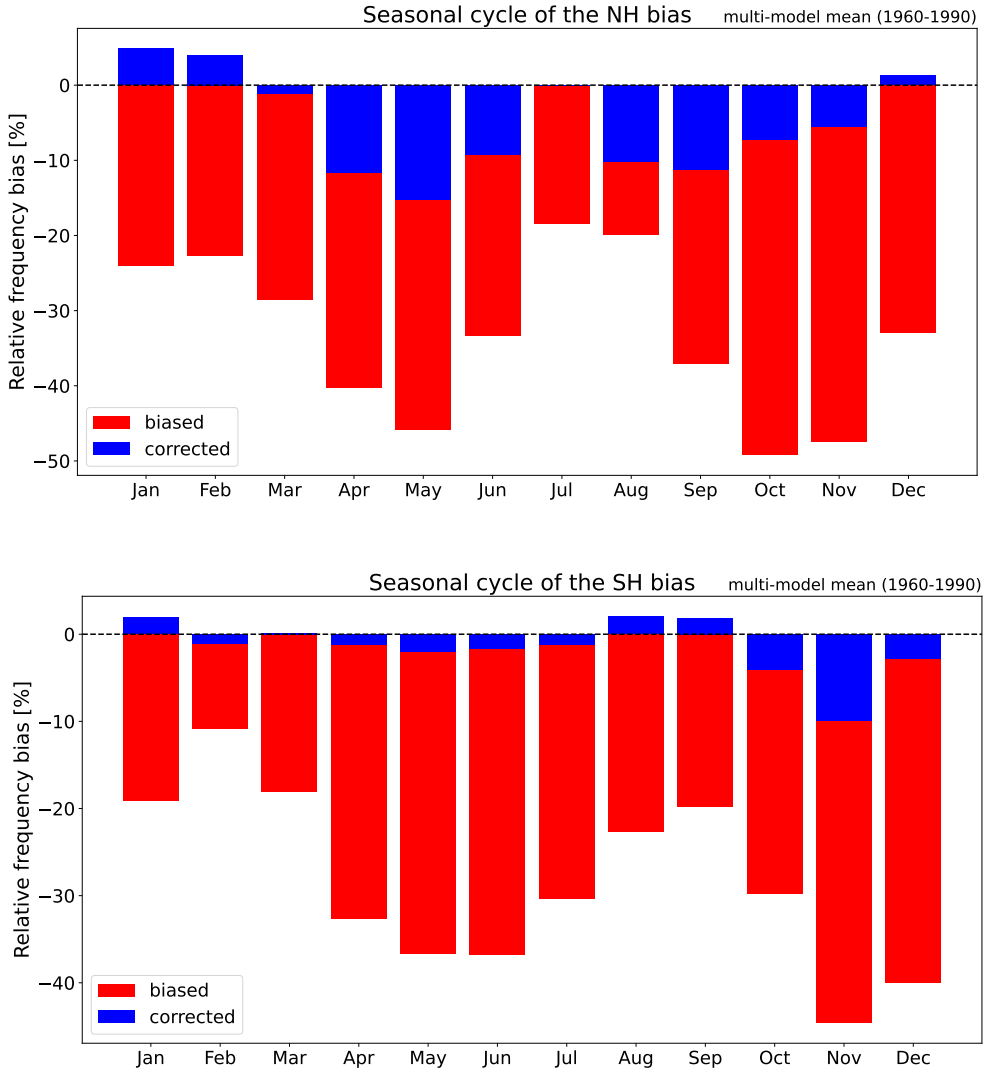

Figure 11.: Seasonal cycle of the mean bias on the northern (top) and southern (bottom) hemisphere. Monthly averages of the relative frequency bias for the multi-model mean in the period 1960-1990. The red barplots represent the biased results, while the blue barplots represent the corrected results.

4.3. Effects of the bias on heatwave properties

When examining Fig. 6, it seems like that there are more connected components when using extreme temperature thresholds with applied bias correction. This implies that there is an underestimation of connected component patterns, potentially resulting in a general underestimation of heatwave characteristics when no bias-correction is applied. This hypothesis comes from the theoretical consideration that an underestimation in the extreme frequency has an effect on derived properties. The patterns in Fig. 6 are one, but not sufficient, hint that this hypothesis is correct. To proof this assumption, I do analyze three different heatwave properties as described in section 3.6 for the biased and the corrected case. In order to proof the effect of moving thresholds, the heatwave properties of two different periods are compared to each other.

4.3.1. Heatwave Area

The cumulative heatwave area is the area affected by a heatwave over the entire duration of an event. For example, consider a heatwave that affects 5 mio. km^2 on the first day, 8 mio. km^2 on the second day and 10 mio. km^2 on the third day, then the total area would be 23 mio. km^2. The mean cumulative heatwave area of the 100 largest events with moving thresholds varies between 11.03 and 61.71 mio km^2 days depending on the model under consideration. The area for single heatwave events approximately corresponds to the country size of Algeria or 130 times the area of Austria for the lowest value and exceeds the size of Asia for the highest value.

	Past	**Future**
INM-CM5-0		
Biased	11.03	11.45
Corrected	20.86	22.44
MPI-ESM1-2-HR		
Biased	16.29	17.48
Corrected	23.44	33.33
MPI-ESM1-2-LR		
Biased	38.93	23.90
Corrected	54.44	48.01
ACCESS-CM2		
Biased	37.46	16.02
Corrected	61.71	47.18
BCC-CSM2-MR		
Biased	17.05	15.53
Corrected	34.83	40.04

Table 4.: Mean heatwave area of the 100 biggest heatwaves in mio. km^2 days.

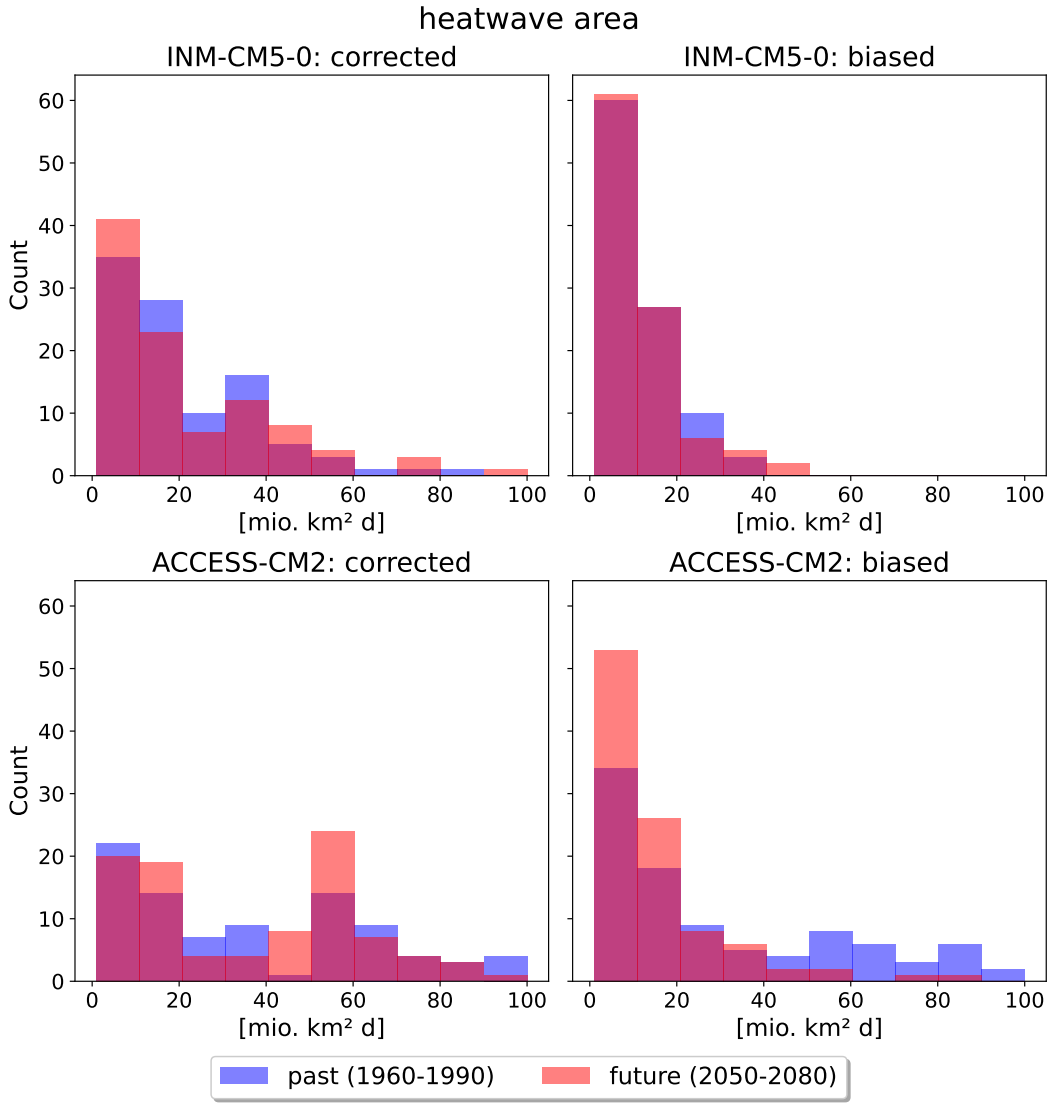

Figure 12.: Area of the 100 biggest heatwaves for INM-CM5-0 (top) and ACCESS-CM2 (bottom). Note that the histogram was cut off at 100 mio. km^2 days.

For the INM-CM5-0 model, the mean heatwave area does not show any detectable change for the two periods under consideration. To a certain extent this is the case both with and without bias-correction (see Table 4). The change with increasing global mean temperature is most pronounced for the ACCESS-CM2 model. Of particular interest here is the fact that the mean cumulative heatwave area decreases in the future period. A closer look shows that fewer large events but more small events, are simulated.

In general, the distributions of the 100 largest heatwaves shown in Fig. 12, illustrates several smaller and fewer larger heatwaves, whereby, in the biased past period ACCESS-CM2 simulates much larger heatwaves than for example INM-CM5-0. The assumption of a decrease in counts with increasing area, is supported by the findings of Zscheischler et al., 2014, who identified that spatiotemporal (three-dimensional) events exhibit power law distributions. The differences in the distributions generally indicate a relatively high intermodel variability (see A Fig. 29).

Even more important are the pronounced differences in the results for the case with and without applying the bias-correction. Every time the bias-correction is not applied, there is a significant underestimation of the heatwave area, resulting in differences of 30-60%, whereby the corrected case is taken as the reference. As suspected, this underestimation of derived heat wave properties arises from the underestimation of the expected exceedance frequency. This in turn is due to a systematic bias caused by the interaction of the seasonal cycle with the 31-day window, explained in detail in section 4.1 and 4.2.

4.3.2. Heatwave Duration

Under the settings described in chapter 3, heatwave durations between 1 and almost 60 days are simulated. (see Fig. 13). Many models show a maximum duration, typically between 50 and 60 days. The mean heatwave duration lies between 11 and 26 days (see Table 5).

	Past	**Future**
INM-CM5-0		
Biased	11.32	11.97
Corrected	16.67	17.56
MPI-ESM-1-2HR		
Biased	11.97	12.46
Corrected	13.70	17.18
MPI-ESM1-2-LR		
Biased	15.59	13.43
Corrected	19.38	19.84
ACCESS-CM2		
Biased	18.89	11.50
Corrected	25.90	20.87
BCC-CSM2-MR		
Biased	13.35	11.63
Corrected	21.12	21.73

Table 5.: Mean heatwave duration of the 100 biggest heatwaves in days.

Regarding heatwave duration, there doesn't seem to be any noticeable change for the two periods studied using the INM-CM5-0 model. Again, this remains somewhat true with and without bias-correction as the mean change in both cases is approximately 5%. In alignment with the results for the heatwave area, the ACCESS-CM2 model displays the most noticeable difference in future changes. In this case, the mean difference without applying bias-correction equals -39%, whereby a decrease in the duration of heatwaves is simulated. Whether these changes are statistically significant will be determined in section 4.4.2. No noticeable changes occur for the duration simulated by MPI-ESM-LR and BCC-CSM2-MR if the bias-correction is considered (see A Fig. 30). In general, without the usage of the bias-correction method, we detect an underestimation in the duration of heatwaves, for the same reason as for the heatwave area. In conclusion, 60% of the analyzed models show no or only minor future changes in heatwave duration.

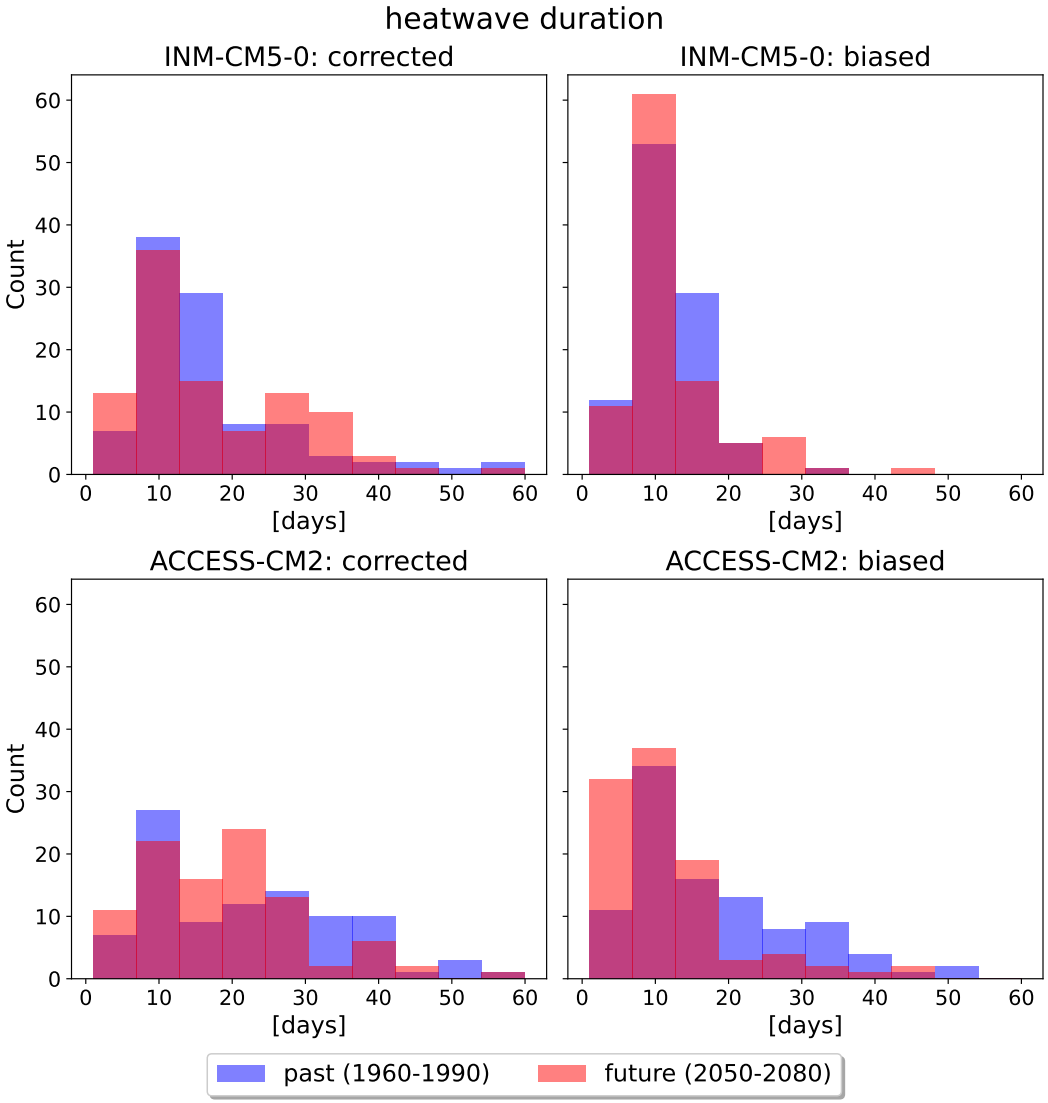

Figure 13.: Duration of the 100 biggest heatwaves for INM-CM5-0 (top) and ACCESS-CM2 (bottom). Note that the histogram was cut off at 60 days.

Vogel et al., 2020 also simulated durations between 1 and nearly 57 days, which is in this case in good agreement with my results. In order to make precise comparisons, however, one would have to look at the corresponding distribution functions of the model. Their study on the development on future heatwave for different thresholds is based on the use of CMIP5 (RCP8.5) models and the 90th percentile for threshold calculation. In addition, they only used the near-surface air temperature (tas in CMIP5) over land regions and limited the analysis to the three consecutive warmest months, whereby I am considering the full seasonal cycle and do not have to limit myself to a heatwave season. All these circumstances make a direct and exact comparison of the findings difficult, as already described in section 1.3.

In the following I want to understand to which extent heatwave duration could change under global warming assumptions. Vogel et al., 2020 related changes in heatwave duration with moving thresholds to physical drivers of heatwaves such as circulation changes of land-atmosphere feedbacks. Based on observations Horton et al., 2015 showed a trend in anticyclones, typically linked to heatwaves in the midlatitudes as described in section 1.1. While the presence of atmospheric blocking can lead to prolonged heatwaves (Pfahl, 2014), it is nevertheless difficult to observe changes in atmospheric dynamics using projections of GCMs (Woollings et al., 2018). Vogel et al., 2020 also indicated that there may be dynamical changes occurring beyond the scales resolved by present-day climate models.

4.3.3. Effect of model resolution

This section describes the effect of the native model resolution for simulating heatwave events. Therefore, I used a model that was run at two different resolutions, namely MPI-ESM1-2-HR and MPI-ESM1-2-LR, as described in section 3.1.1. Particularly striking is the fact that the largest heatwave summed over the duration simulated by MPI-ESM1-2-LR is approximately 3.7 times larger compared to the high-resolution model. Although the high-resolution model simulates much larger areas at the beginning of the event, the duration is much shorter compared to the low-resolution model (see Fig. 14).

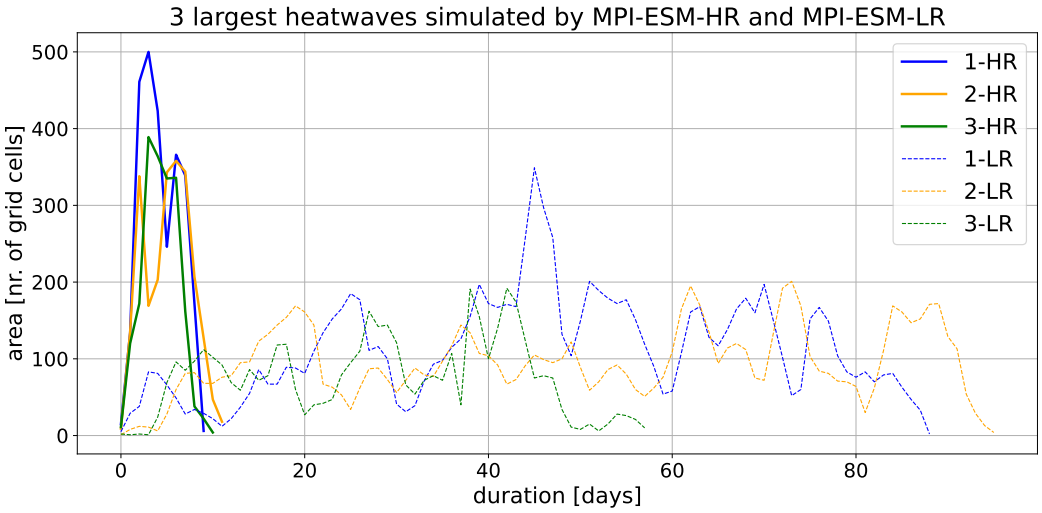

Figure 14.: Spatiotemporal evolution of the 3 largest heatwaves in the period (1960-1990) simulated by MPI-ESM-HR (thick lines) and MPI-ESM-LR (thin, dashed lines). Largest heatwaves (1-3) in terms of number of affected grid cells.

MODEL-ID	Heatwave 1	Heatwave 2	Heatwave 3
MPI-ESM1-2-HR	2662	2298	1953
MPI-ESM1-2-LR	9972	9014	4320
INM-CM5-0	2880	2105	1862

Table 6.: Three largest heatwaves simulated by MPI-ESM1-2-HR, MPI-ESM1-2-LR and INM-CM5-0 (as reference). The number of affected grid cells is summed over the entire duration of the respective heatwave.

These results are merely recorded here and will not be discussed further. The exact cause is unclear for the time being and requires further research. In any case, the reference model (INM-CM5-0) shows similar values to the high-resolution model (see Table 6), which suggests that the effect of model resolution is not negligible. Simulating heatwave events with GCMs involves several important processes such as radiative transfer processes, atmospheric dynamcis or land-surface processes, that need to be accurately represented.

4.3.4. Correlation of heatwave area and duration

A further analysis of the results from section 4.3.3 demonstrates a significant difference of heatwave area and duration between the two versions of MPI-ESM1-2 for a larger sample. Fig. 15 shows the mean heatwave area per day as function of the total duration for the 100 largest heatwaves. It is particulary striking that the correlation for the low-resolution model is always higher than for the high-resolution model for all cases considered in section 3.6.1.

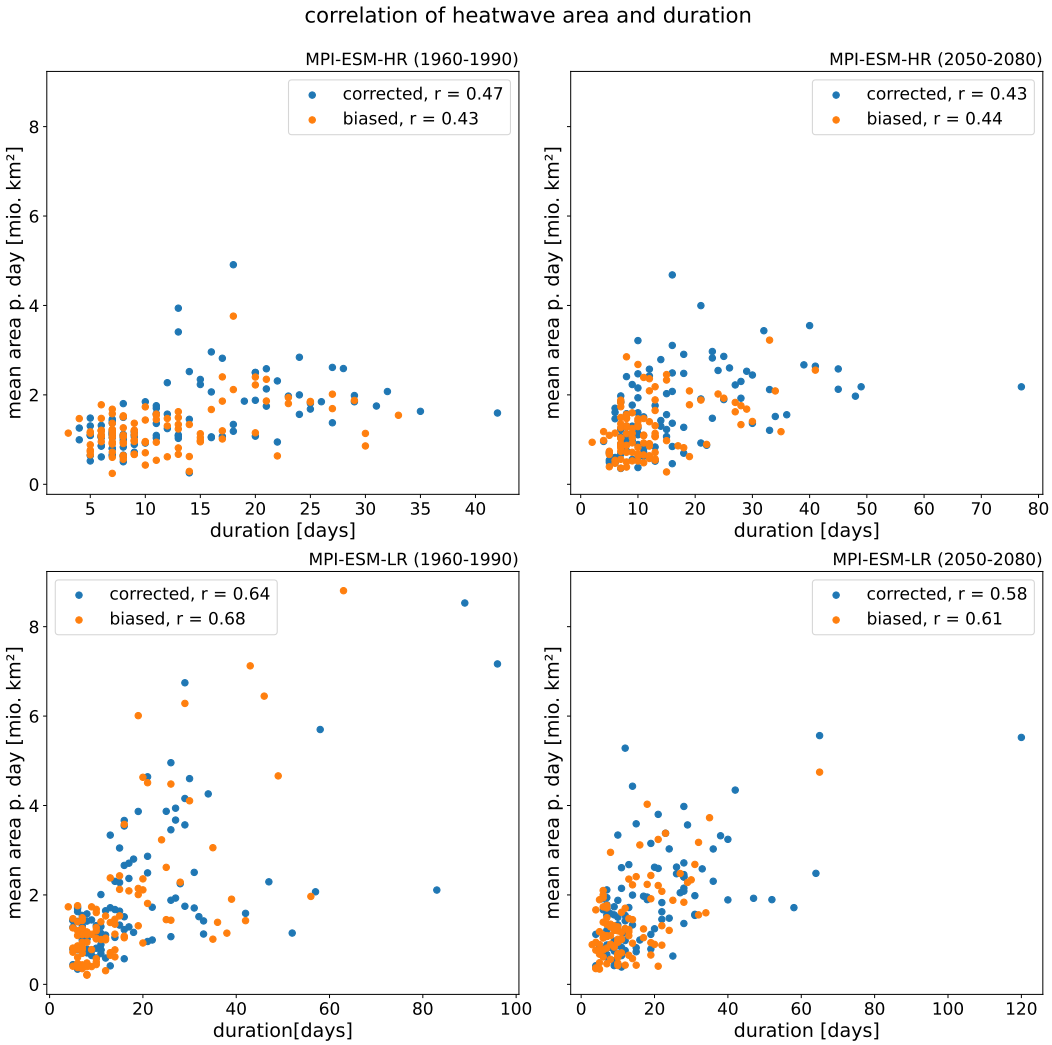

Figure 15.: Correlation of the mean heatwave area per day and duration. The x-axis represents the duration in days, while the y-axis represents the mean area per day in mio. km^2.

The correlation coefficient between the mean area per day and the duration of a heatwave ranging from 0.43 to 0.68 indicates a moderately positive relationship. This suggests that as the duration of a heatwave increases, there tends to be a corresponding increase in the

area affected. While this correlation is not extremely strong, it implies that longer-lasting heatwaves typically cover a larger geographical area. However, it's essential to consider that factors beyond duration alone may influence the extent of the area impacted by a heatwave due to the complex interplay of the physical drivers. Firstly, atmospheric circulation patterns, such as persistent high-pressure systems or jet stream ridges can trap heat over a large area for an extended period, leading to prolonged heatwaves. Secondly, land surface feedback mechanisms play a key role, as heating of the land surface during a heatwave can intensify the warming effect by releasing stored heat enhancing evaporation. Both factors collectively contribute to the simulated correlation between area and duration of heatwaves.

4.3.5. Heatwave Magnitude

The heatwave magnitude is defined as the area-weighted sum of temperature exceedances over all grid cells affected by a heatwave. The magnitude contains the area, the duration and the temperature exceedance and therefore summarizes comprehensive risks generated by heatwaves. With moving thresholds relatively large mean heatwave magnitudes between 147.85 mio. km^2 days ΔK and 355.25 mio. km^2 days ΔK are simulated.

	Past	**Future**
INM-CM5-0		
Biased	180.32	147.85
Corrected	241.46	185.44
MPI-ESM-1-2HR		
Biased	195.52	181.41
Corrected	226.82	280.28
MPI-ESM1-2-LR		
Biased	240.53	189.40
Corrected	355.25	292.75
ACCESS-CM2		
Biased	262.27	176.88
Corrected	323.50	285.37
BCC-CSM2-MR		
Biased	181.62	170.98
Corrected	225.90	175.73

Table 7.: Mean heatwave magnitude of the 100 biggest heatwaves in mio. km^2 days ΔK.

Any distribution of samples in the form of a histogram can also be represented as a probability density function. Fig. 16 shows the magnitude of the 100 biggest heatwaves for 2 different models. If those histograms are understood as distribution functions, determining the differences when using the bias-correction method becomes relatively easy. Without bias-correction, the probability density is skewed towards smaller heatwave magnitudes. However, with the application of the bias-correction, although smaller magnitudes still dominate in terms of count, the distribution shifts, leading to an increased likelihood of larger magnitudes. This leads to the inverse conclusion that we would expect more intense heatwaves when applying the bias-correction method, whereas without it, there is a tendency to underestimate the heatwave magnitude, consistent with the findings for other heatwave properties. Since the duration of heatwaves is quite similar across the selection of models when using moving thresholds (see Table 5), the differences in the heatwave magnitude probably arises from the area and temperature anomaly of the corresponding label.

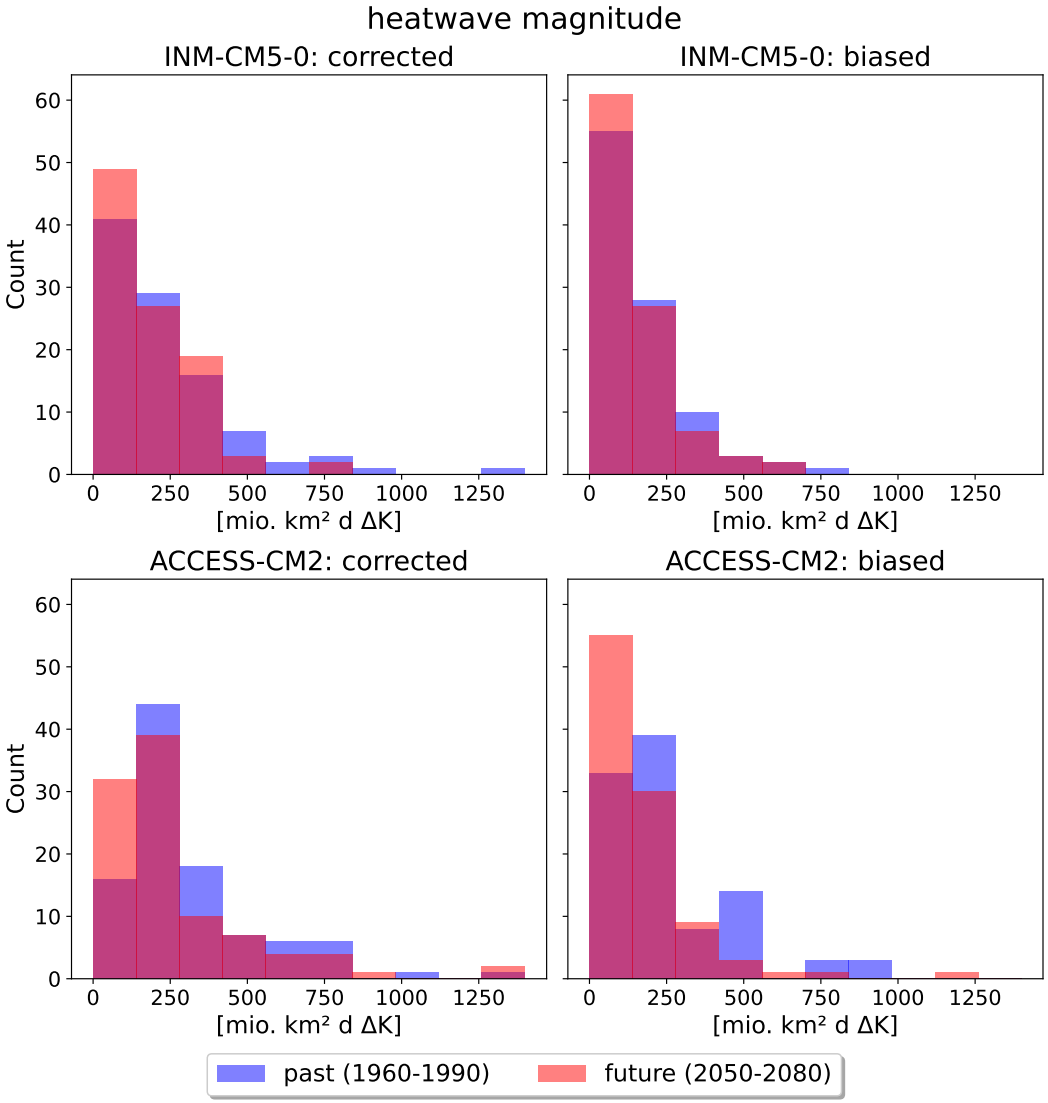

Figure 16.: Magnitude of the 100 biggest heatwaves for INM-CM5-0 (top) and ACCES-CM2 (bottom). Note that the histogram was cut off at 1400 mio. km^2 d ΔK.

All models show future changes, with these being least pronounced for MPI-ESM-HR and BCC-CSM2-MR in the biased case. On average, all other models show changes ranging between 10 and 30%. Additionally, it is noteworthy that all models except MPI-ESM-HR in the corrected case simulate a decrease in the mean heatwave magnitude. For the described scenario, an increase from 226.82 mio. km^2 d ΔK to 280.28 mio. km^2 d ΔK is modeled (see Appendix A 4.3.3). However, when considering the other properties of this scenario (HWA & HWD), one can see that both metrics show a positive trend. The change in magnitude could potentially be explained by the change in the other two metrics, although it does not necessarily imply changes in the temperature gradient.

At this point it should also be noted that I am only using two different time periods to identify future changes compared to Vogel et al., 2020 who used different warming levels. For each model, they identified the year where the warming of a 31-year running window is closest to global warming levels of +1, +1.5, +2 and +3°C relative to a preindustrial level of 1871-1890 and used this year as a center for a 21-year warming level time slice. This approach ensures comparability among warming levels of different models. However, in my case, due to model variability, the degree of warming may vary between the selected models. While they found rather constant heatwave magnitudes with moving thresholds and the distributions for higher warming levels show no substantial changes, in my case, based on qualitative analysis, there is no clear signal regarding future changes. For this reason, statistical significance tests will be conducted in the following to shed light on the matter (see section 4.4.3).

4.4. Statistical Significance

Since heatwaves are rare events, I also use an ensemble dataset of climate simulations with 5 ensemble members from the ACCESS-CM2 model to increase the sample size of heatwave events in a historic (1960-1990) and a future (2050-2080) time slice. Once again, the 100 largest events for 3 different heatwave properties are analyzed, in this case for 5 different realizations of the same model for the biased and the corrected case.

Since the results of section 4.3 did not always show clear signals with respect to future changes, a statistical significance test is now used. With the help of permutations tests, as described in section 3.6.2, I proof the assumption that there is no significant difference in the mean values of heatwave area, duration and magnitude between a future and a historical period. By defining projected heatwaves relative to a future climatology, it is hypothesized that future heatwaves do not change much.

4.4.1. Heatwave Area

Fig. 17 shows the distribution of heatwave area, for the settings described above, in form of boxplots. The range of the boxplot covers the so-called interquartile range (IQR), which is the range between the fist (Q1) and the third quartile (Q3), while the whiskers extend to show the rest of the distribution. Note that the distributions also contain some outliers, but these have been removed as they are probably artificial products of the labeling algorithm and no longer allow a physical interpretation. For this reason, the line within the boxplots also shows the median and not the arithmetic mean of the distribution, as the mean, which takes into account the value of each datapoint, is sensitive to outliers, while the median is a more robust measure, as it is less influenced by extreme data points.

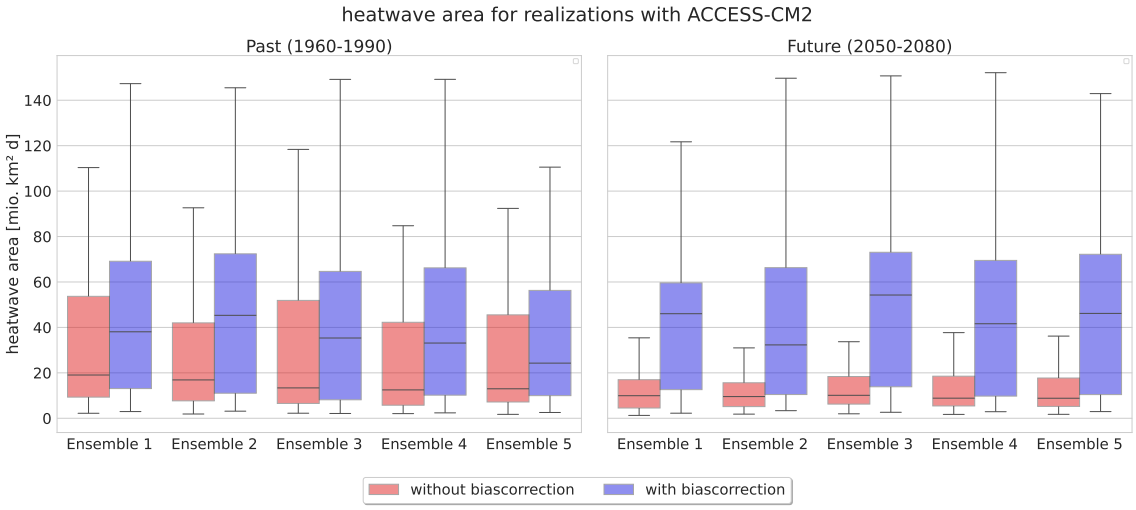

Figure 17.: Area for the 100 biggest heatwaves for each ensemble member of the ACCESS-CM2 model. The red boxplots represent the biased results, while the blue boxplots represent the corrected results. Historical period (left), future period (right).

For further analysis, the distributions of the ensemble members, shown in Fig. 17, were merged in order to have a sample size of 500 events for each of the cases considered (biased-past, corrected-past, biased-future, corrected-future). Note that the underlying distribution is based on the 100 largest events in each of the 5 ensemble members and not the overal 500 largest events. Fig. 18 illustrates the merged distributions respectively the overlap of the two considered time slices. The red and blue line shows the respective mean value of the overall distribution, whereby in the corrected scenario, it becomes apparent that these values are closer to each other.

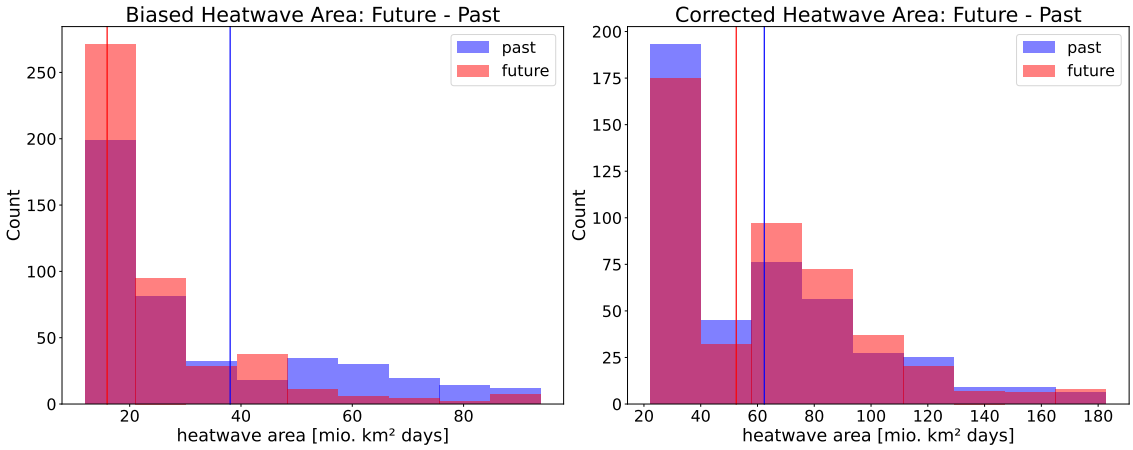

Figure 18.: Histogram of the heatwave area for all heatwaves and ensembles of the ACCESS-CM2 model. The colors indicate the time period, blue for the historical period, red for the future period. The biased case on the left and the corrected case on the right.

The used permutation test evaluates the statistical significance by randomly shuffling the data values. It then compares the test statistic from the original data to the distribution of test statistics generated from permutations. In both cases, a decreasing change is detected: Without bias-correction, a statistically significant change is simulated, whereas with bias-correction the change in the mean is not statistically significant from zero.

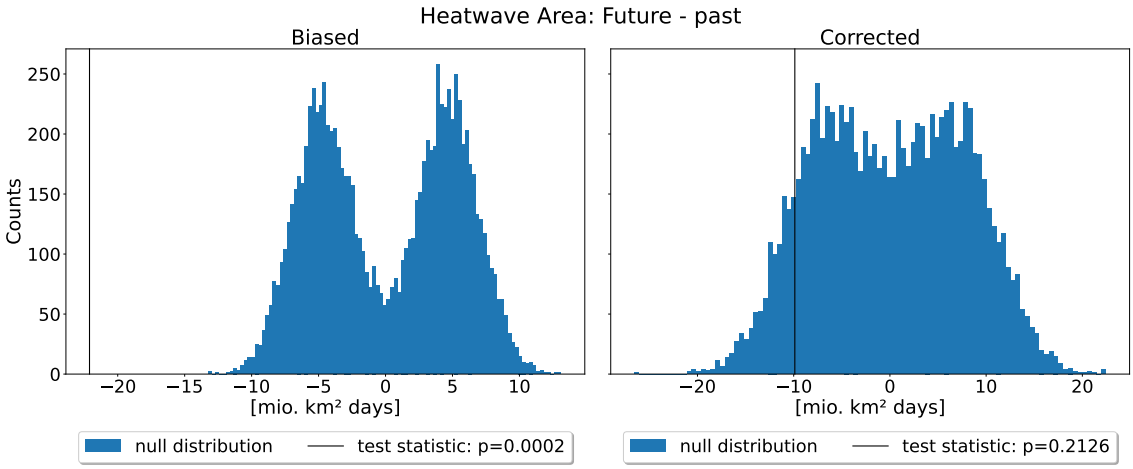

Figure 19.: Results for a permutation test for the null hypothesis that the difference in the mean heatwave area between a future and a historical period is not different from zero. Biased (left) and corrected (right).

4.4.2. Heatwave Duration

The distribution of heatwave duration, under the conditions described in section 4.4, and as illustrated in Fig. 20, highlights a striking trend: the median of the corrected cases (blue boxplots) consistently surpasses that of the biased cases (red boxplots). The underestimation of the heatwave duration in this case is therefore in good agreement with the results discussed in section 4.3.2.

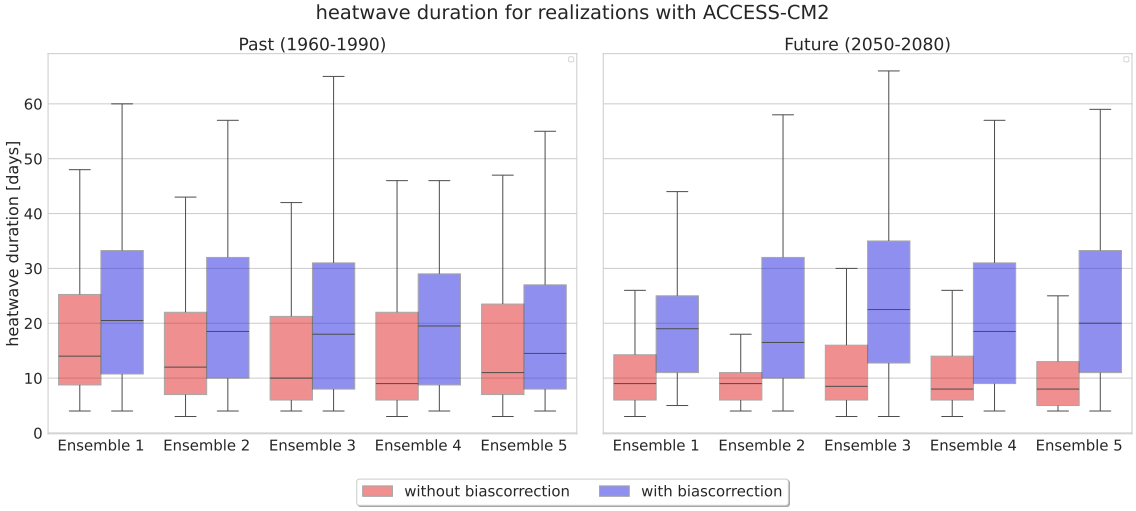

Figure 20.: Duration for the 100 biggest heatwaves for each ensemble member of the ACCESS-CM2 model.

The distributions of the individual ensemble members are again combined and the historical (blue) and the future period (red) are overlaid (see Fig. 21). Once more, it is evident that the mean values of the two periods are notably closer to each other in the corrected scenario, indicating that there are no or only minor future changes when using moving thresholds for the definition of heatwaves.

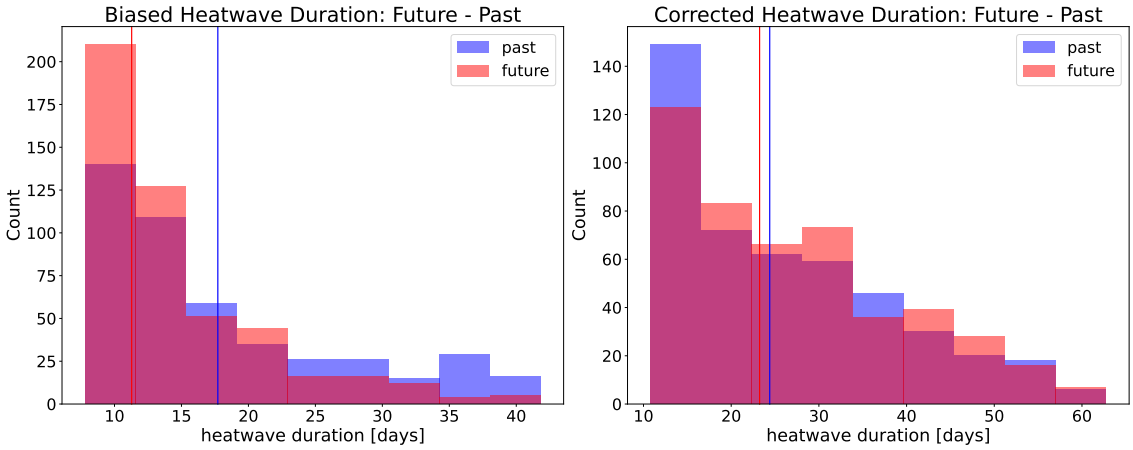

Figure 21.: Histogram of the heatwave duration for all heatwaves and ensembles of the ACCESS-CM2 model.

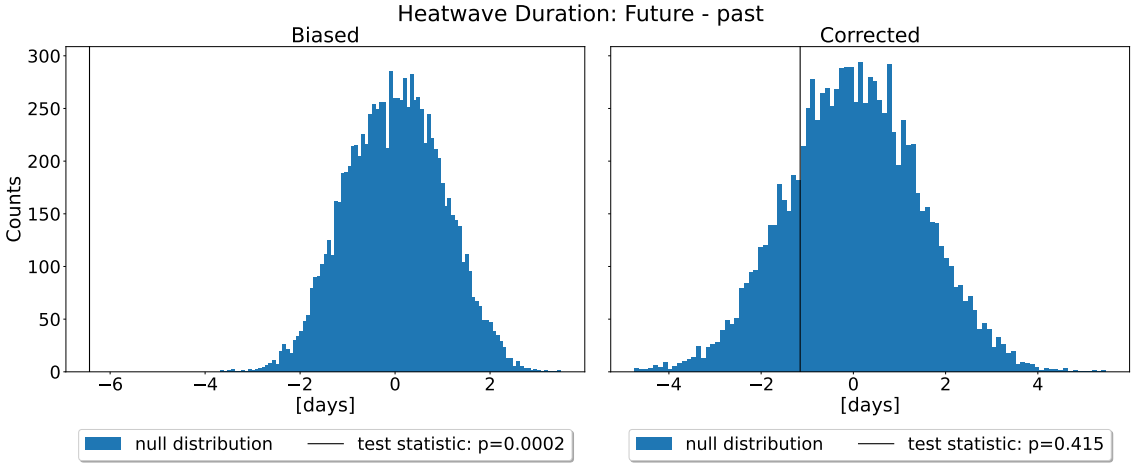

Figure 22.: Results for a permutation test for the null hypothesis that the difference in the mean heatwave duration between a future and a historical period is not different from zero.

These results are generally in line with the findings of Schielicke et al., 2022. Their study on European heatwaves in present and future climate simulations used data from model simulations based on the Community Earth System Model (CESM) Large Ensemble (CESM-LE) project with 35 ensemble members to study future changes in several European regions. Their identification of heatwaves is also based on a temperature threshold, but given by the 90th percentile of the daily maximum temperature from the climate model output within a 30-day window around the day of interest. In order to identify future changes, they define heatwaves with respect to the historic climate in the 1991-2000 time slice and with respect to future climate in the 2091-2100 time slice. In general, they observed no or only minor differences between general heatwave properties in historic

and future simulations. In the majority of regions, the observed a typical duration of less than 10 days, though under certain conditions, it can extend to a maximum of 45 days. Noticeable is that, one of their main findings is that since the method of detection is percentile-based, the total number of heatwave days is almost identical in both slices. However, this assumption in my case holds only true if an appropriate bias-correction is applied, as the results from the permutation test in Fig. 22 show. Again, in both cases, a decreasing trend is detected, whereby it is much more pronounced for the biased case as shown by the vertical lines which displays the difference in mean values of heatwave duration as calculated from the mean values shown in Fig. 21. Again, without bias-correction, a statistically significant change is simulated, whereas with bias-correction the change in the mean is not statistically significant from zero, as also indicated by the p-value.

4.4.3. Heatwave Magnitude

The magnitude of heatwaves is a good parameter to estimate the future risk of heatwaves, as it includes all basic characteristics, especially the temperature exceedance, which defines the intensity of heatwaves events. Using fixed thresholds to define heatwaves, the magnitude increases proportionally to global mean warming (Vogel et al., 2020). If they are defined with moving thresholds, as in my case, no changes would be expected.

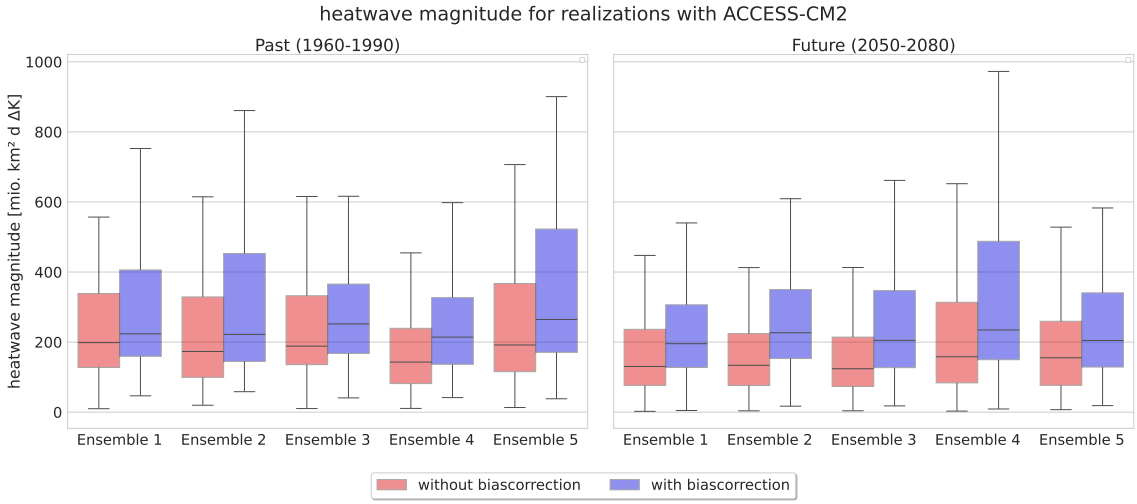

Figure 23.: Magnitude for the 100 biggest heatwaves for each ensemble member of the ACCESS-CM2 model.

The distributions of the magnitude show an underestimation in the biased case, whereby this can exceed up to 30% for selected ensemble members in the historical period (see left panel of Fig. 23). In both cases, the mean values show larger differences, whereby the p-value of the significance test for the corrected case is an order of magnitude smaller compared to area and duration, indicating little systematic future changes.

According to Vogel et al., 2020, the distributions for higher warming levels exhibit overall no significant changes. They do also note that for the difference between +3°C and +1°C warming, two models demonstrate a statistically significant change. In order to estimate the magnitude of a heatwave, Schielicke et al., 2022 used the percentile-based Heatwave Magnitude Index (HWMId), defined by Russo et al., 2014. Their comparison of historic and future distributions of the HWMid reveals that the distributions are primarily shifted towards higher HWMId, but nearly do not change shape. In certain regions, however, a broadening of the distribution towards higher HWMId, but also a narrowing away from higher HWMId was detected. At this point is should be mentioned again that no bias-correction method was used in both studies, and, as previously mentioned, the results are not directly comparable for several reasons. Upon further investigation, I have discerned variations in the magnitudes of derived heatwaves (or general heatwave characteristics) between the biased and corrected scenarios.

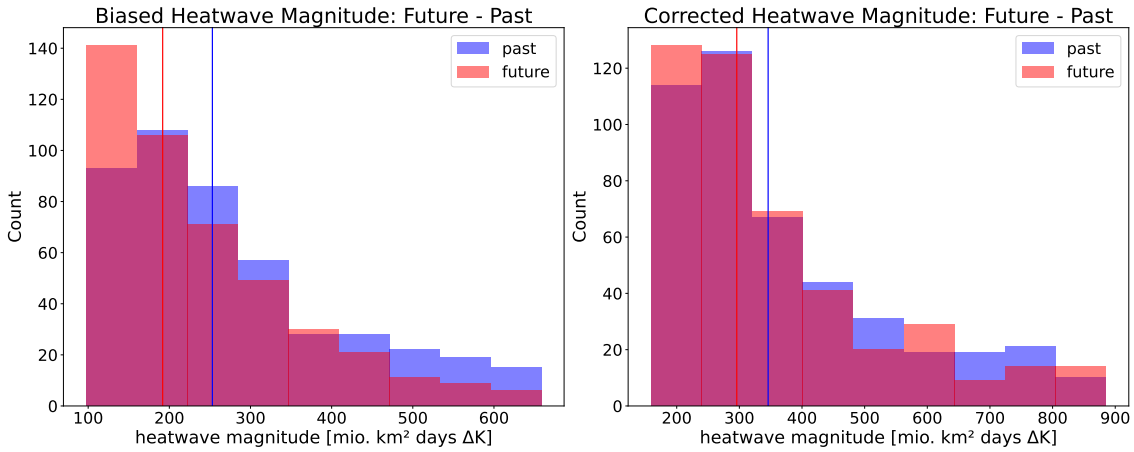

Figure 24.: Histogram of the heatwave magnitude for all heatwaves and ensembles of the ACCESS-CM2 model.

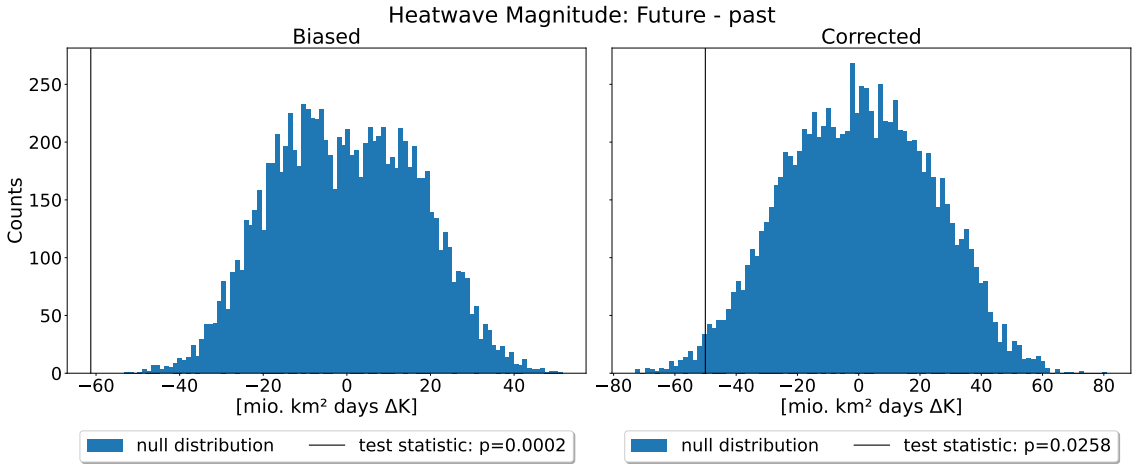

Figure 25.: Results for a permutation test for the null hypothesis that the difference in the mean heatwave magnitude between a future and a historical period is not different from zero.

Regarding physical concepts of heatwave magnitudes, Zhang et al., 2023 have identified an upper limit for extreme temperatures over land in mid-latitudes. While mechanistic drivers such as atmospheric blocking are well known for their ability to raise surface air temperature (as explained in section 1.1), the limitation of maximum surface temperature and thus the magnitude of heatwaves requires further investigation. They hypothesized that convective instability limits annual maximum surface temperature over midlatitudes, suggesting that the upper bound for surface air temperature should increase approximately 2 times as fast as 500-hPa warming. Considering this finding provides another good opportunity to estimate how the magnitude in future projections could change.

5. Conclusion

This thesis studies the changes of heatwave properties under climate change as well as a systematic bias in future heatwave diagnostics throughout the seasonal cycle. To conclude, I again explore the two research questions, formulated in chapter 2, by considering the findings of chapter 4.

1. What is the effect of a running window when using moving thresholds for defining heatwaves?

In the absence of a universal definition of a heatwave, various approaches are used to define such temperature extremes. These are generally based on the determination of relative or absolute thresholds. Thereby, the choice of method for defining thresholds is crucial in determining hot days. Many existing studies employ large running windows, typically 15 or 31 days, as in my case, when calculating relative percentile-based thresholds to increase the sample size. Despite the expectation that using the 99th percentile as a threshold would result in an overall exceedance frequency of 1%, recent research by Brunner et al., 2024 revealed that such long running windows introduce a systematic bias, leading to a significant underestimation of the expected extreme frequency, also demonstrated in this thesis using 5 selected CMIP6 models. The interaction between the running windows and the seasonal cycle has proven to be the cause of these biases in the temperature extreme frequency. By using the proposed bias correction method by Brunner et al., 2024, which includes the removal of the mean seasonal cycle, a reduction of the relative frequency bias was shown, both for a single grid cell and at global level. This bias can vary depending on which GCM is considered. Strong variations in the seasonal cycle simulated by a variety of climate models were shown to be the root source of the underestimation. As climate models simulate the seasonal cycle differently in some cases, this generally leads to an enhanced level of intermodel spread.

The method used for bias-correction was able to significantly minimize the bias, although it could not eliminate it entirely. This behavior arises due to the use of moving thresholds with shifting periods to define heatwaves relative to a future climate and leads to the so-called out-of-sample problem. The used approach uses data points outside the period actually analyzed (in-base period), which means that by construction the exceedance rate is unlikely to be exactly 1%. The underestimation of the extreme frequency, caused by the running window bias, is essential, as it can also affect derived metrics such as heatwave characteristics. This finding not only applies to my master's thesis but also extends to future studies and publications within this research field (e.g. for the recent publication by Luo et al., 2024).

2. How do different heatwave characteristics respond to global warming in CMIP6 models?

Using different definitions may lead to varying changes in heatwave properties, highlighting the strong sensitivity of simulated future heatwave properties to thresholds. To explore future changes, comparisons are made between two periods – one past and one future – employing moving thresholds to account for global warming effects. In addition, the differences caused by the running window bias were also investigated, whereby without applying an appropriate bias-correction method I generally identify an underestimation of heatwave characteristics.

While Vogel et al., 2020 detected no/minor significant changes in projected heatwaves, if they are defined relative to a future climatology, I demonstrated that simulated changes are attributed to the application or neglection of the bias-correction method. The fact that Vogel et al., 2020 only considered the 3 consecutive warmest months over land regions has the consequence, that the bias in the extreme frequency does not have such a strong effect on derived heatwave metrics.

However, in my case the application of the bias-correction method is necessary to retain the basic idea of moving thresholds. Only when using such a method there is no statistical significance in the mean change in the characteristics of heatwaves, as demonstrated in section 4.4.

If changes still occur despite the application of the bias-correction, these can of course be attributed to changes in non-linear dynamics. For example, warming in certain regions can cause the jet stream to become weaker and wavier. This causes different weather patterns to stay longer over the same place, which can possibly influence the duration of heatwaves. Furthermore, an amplified diabatic heating associated with future heatwaves, as found by Schielicke et al., 2022, could be explained by an additional drying of the land surface in future projections, potentially affecting the magnitude of heatwaves.

Summarized it can be said that my master's thesis underscores the critical role of extreme threshold selection in characterizing spatiotemporal heatwaves amid a warming planet. Through a focus on a moving threshold approach and the necessity of bias-correction techniques, the study reveals significant biases in the extreme frequency and subsequent heatwave properties. These findings underscore the importance of employing robust methodologies to ensure accurate temperature extremes analysis in the context of climate change, offering valuable insights for future climate studies and adaptation strategies.

6. Outlook

Concerning my master thesis, there exist numerous potential directions for extending the scope of my work. Therefore, I present a detailed overview of the areas that could be explored further and require additional study.

To fully understand to the development and persistance of heatwaves, it is necessary to investigate possible dynamical drivers. A study by Kautz et al., 2022 revealed that heatwaves often occur in connection with the center of a blocking system, commonly observed at the mid-troposphere around 500hPa. Since the geopotential at this pressure level is available as direct model output in CMIP6 models, it could be used to identify the connection between atmospheric blocking and heatwaves in future projections.

The bias-correction method effectively minimized the bias in the threshold exceedances but did not fully eliminate it. Primarily because moving thresholds use out-of-base data, as described in section 4.2. Because the analysis is never fully in-sample I don't need to worry about an in-sample/out-of-sample jump as described by Zhang et al., 2005. The out-out-sample problem even leads to extreme frequencies that are higher than 1%, which can even lead to a regional overestimation of the extreme frequency when applying the bias-correction.

The necessity to validate my findings across different climate models underscores the importance of utilizing multiple ensembles, especially to mitigate the impact of internal climate variability. In addition, the uncertainty due to model errors also plays an important role in representing complex small-scale atmospheric processes. However, nowadays GCMs provide a finer horizontal resolution, which makes the use of high-resolution models on a km-scale, as described in section 1.5, an insightful tool for detecting projected changes in atmospheric circulation patterns in order to better understand possible changes in heatwave properties beside mean global warming. Nevertheless, existing model errors still contribute to difficulties in accurately predicting heatwaves, highlighting the need for ongoing improvements in model development and validation to better understand and mitigate the impacts of heatwaves.

Based on the existence of model uncertainty, the results of selected CMIP6 models should be in any case evaluated against reanalysis or observational data over the historical period. This comparison allows for an assessment of the model's ability to accurately simulate heatwave characteristics in the historical period. Only when the statistics of the historical period to some extent match the reference dataset, one can make a statement about whether the model is even capable of simulating meaningful future heatwave projections.

A. Appendix

A.1. 4.1.1

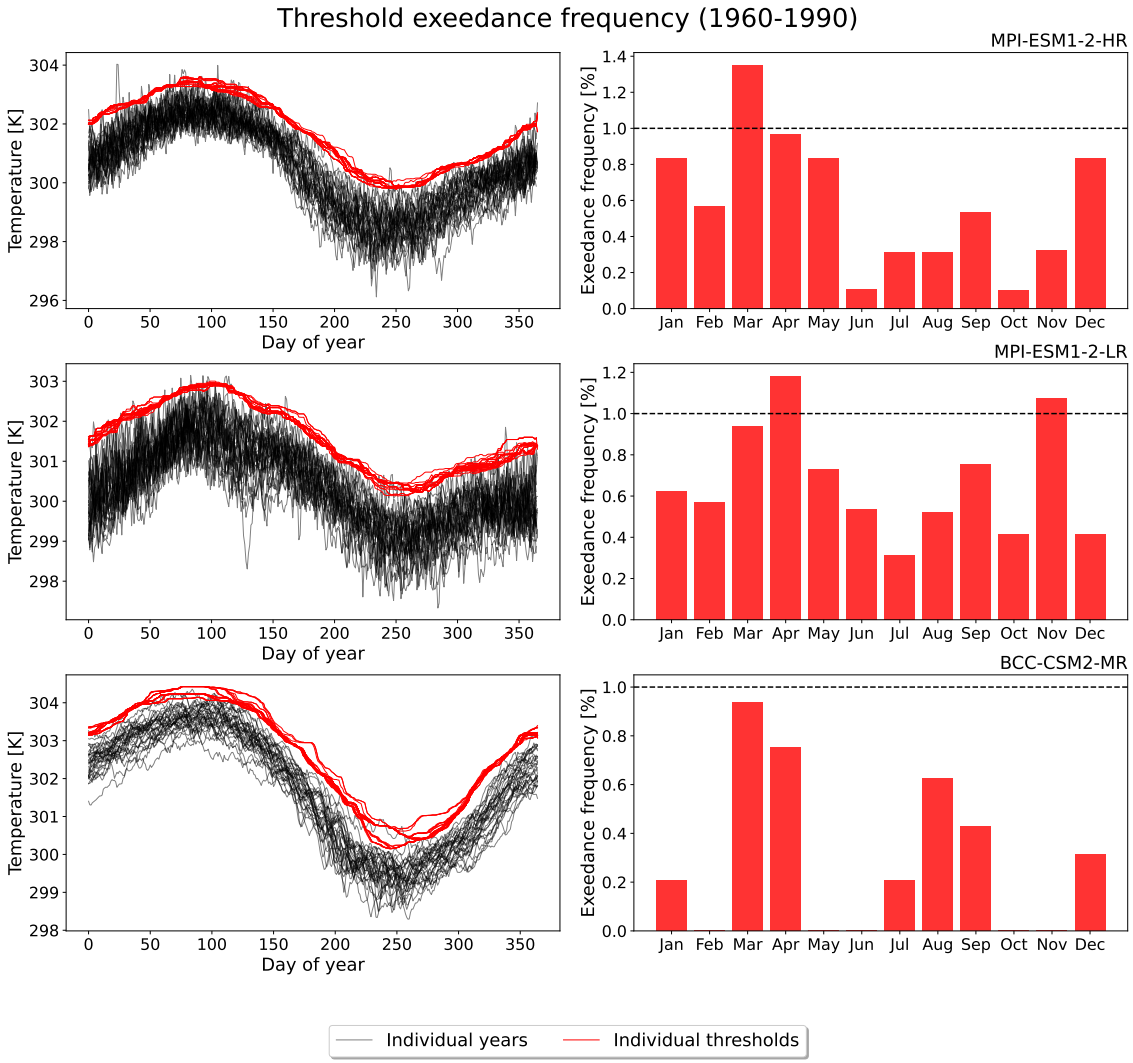

Figure 26.: Daily threshold exceedances and monthly averaged frequencies at a selected grid cell ($lat = -1.25°, lon = 358.75°$) from MPI-ESM-HR (top), MPI-ESM-LR (middle), BCC-CSM2-MR (bottom) for the biased case.

A.2. 4.1.2

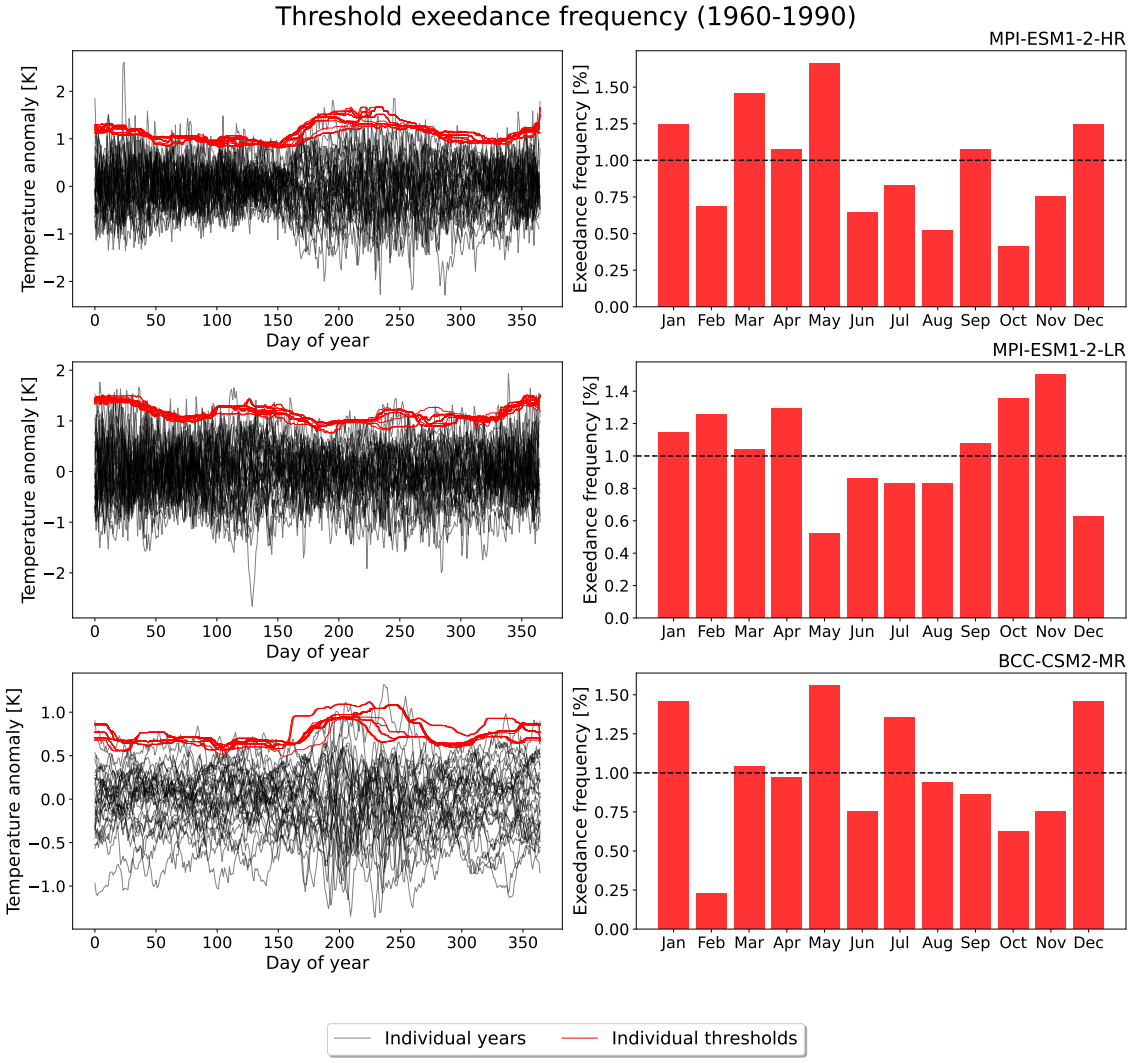

Figure 27.: Daily threshold exceedances and monthly averaged frequencies at a selected grid cell ($lat = -1.25°$, $lon = 358.75°$) from MPI-ESM-HR (top), MPI-ESM-LR (middle), BCC-CSM2-MR (bottom) for corrected case.

A.3. 4.2

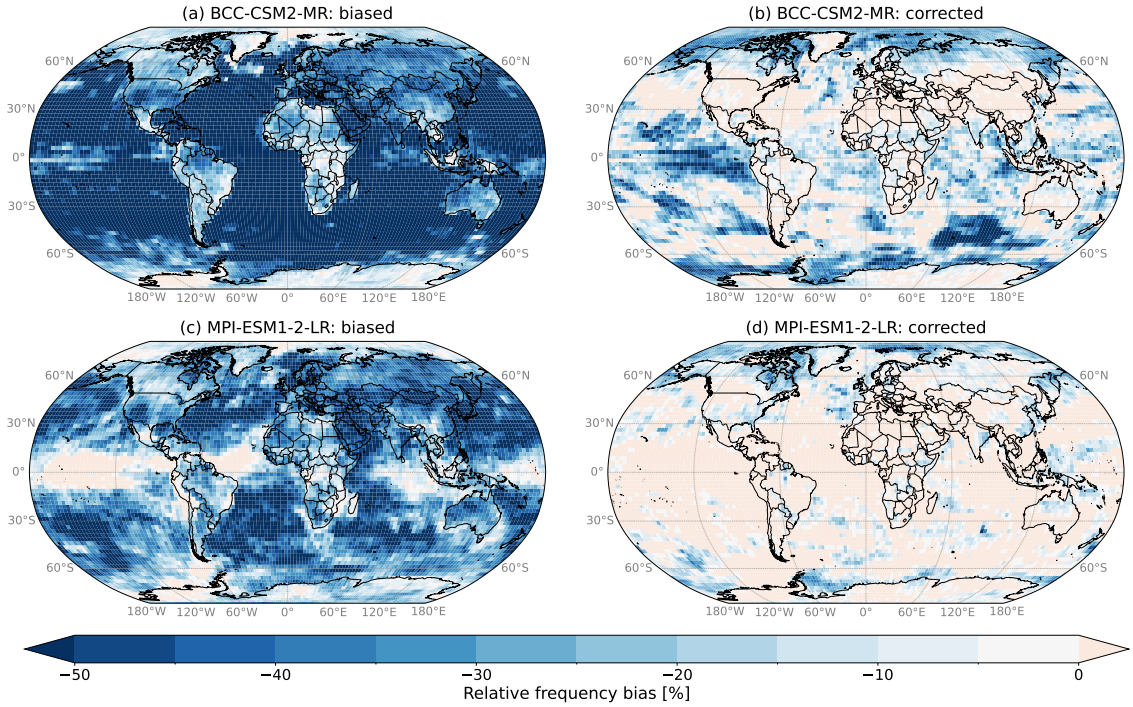

Figure 28.: Biases in the frequency of temperature extremes in BCC-CSM2-MR (a, b) and MPI-ESM-LR (c, d). Spatial distribution of biases in the frequency of daily maximum temperatures based on exceedances of the 99th percentile using a moving threshold.

A.4. 4.3.1

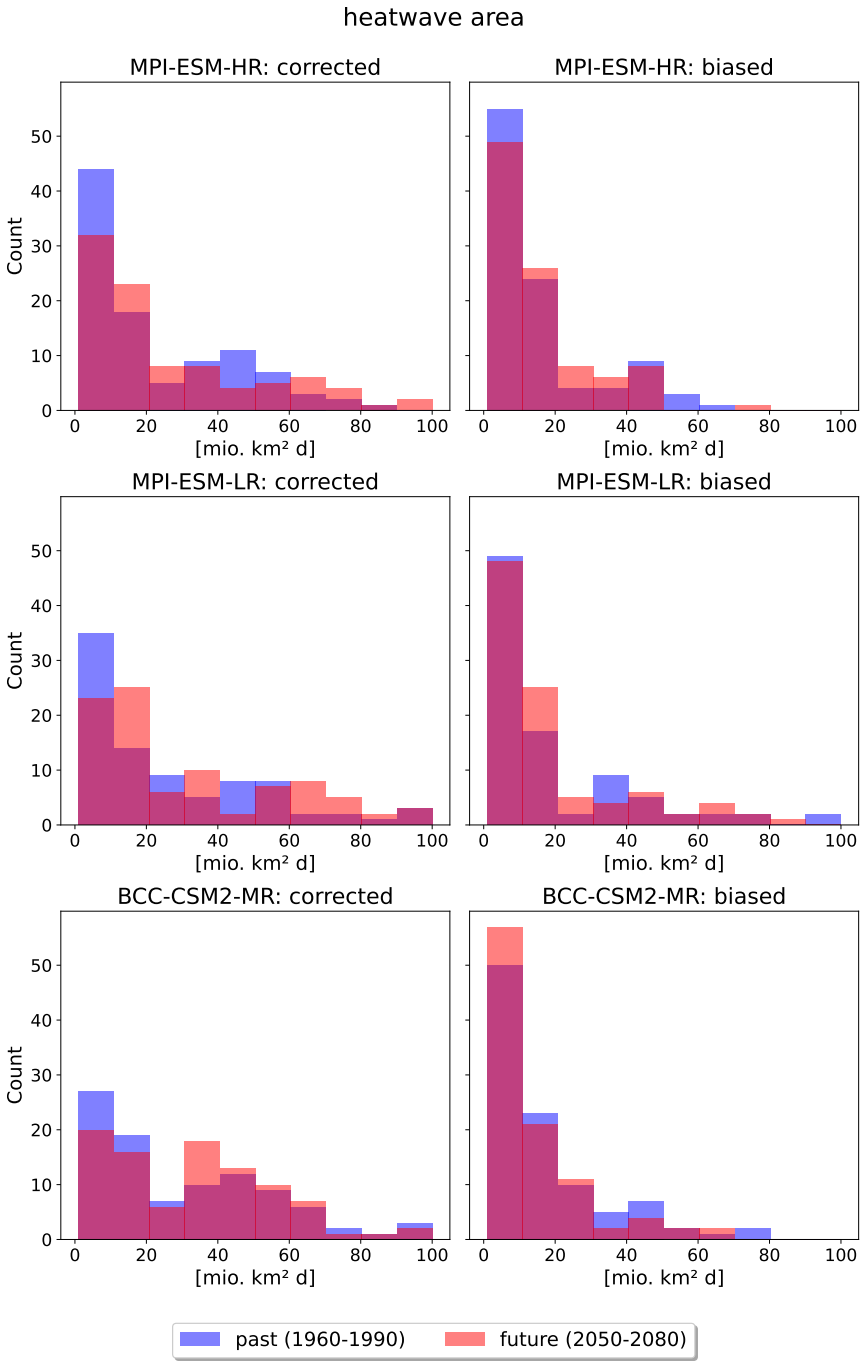

Figure 29.: Area of the 100 biggest heatwaves for MPI-ESM-HR (top), MPI-ESM-LR (middle) and BCC-CSM2-MR (bottom).

A.5. 4.3.2

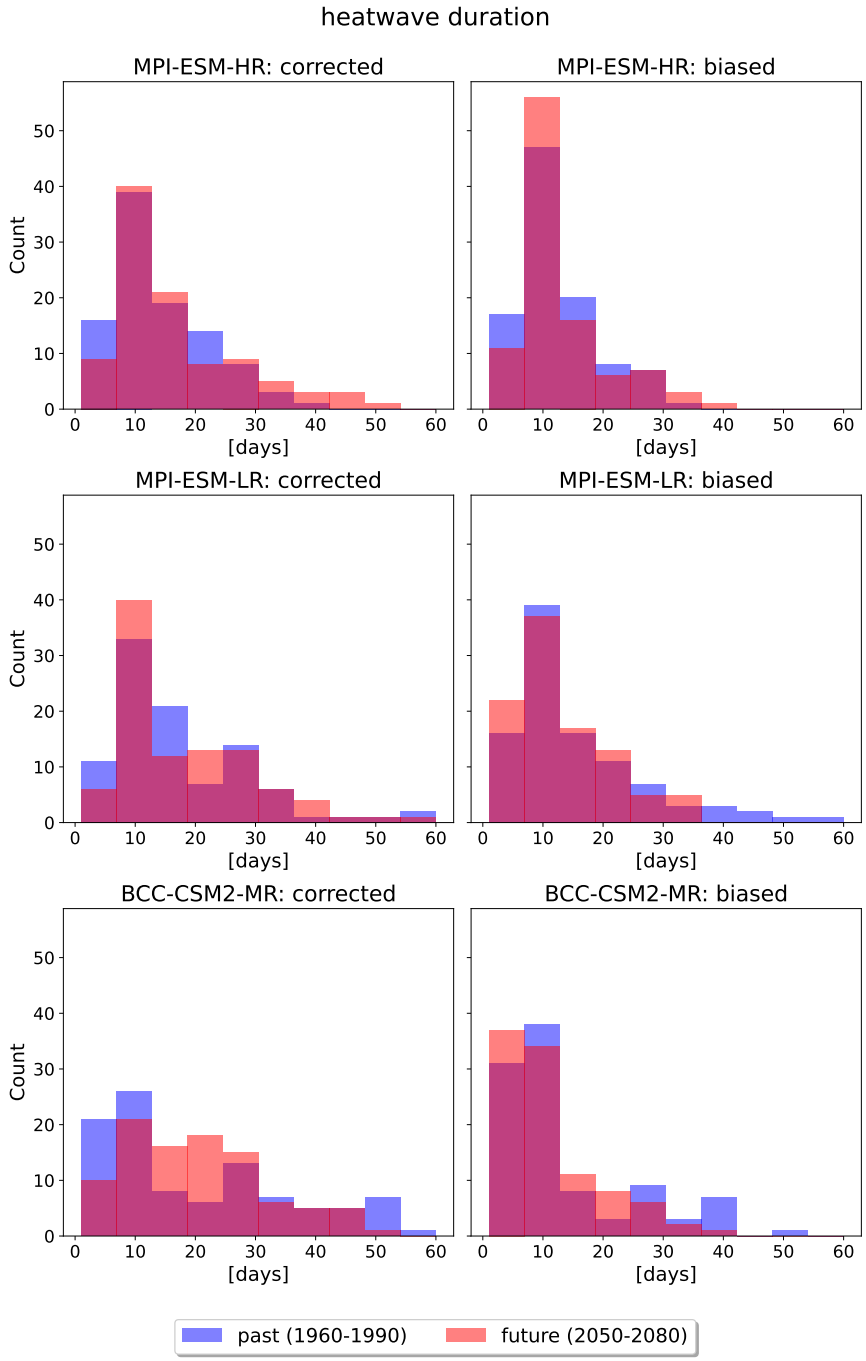

Figure 30.: Duration of the 100 biggest heatwaves for MPI-ESM-HR (top), MPI-ESM-LR (middle) and BCC-CSM2-MR (bottom).

A.6. 4.3.3

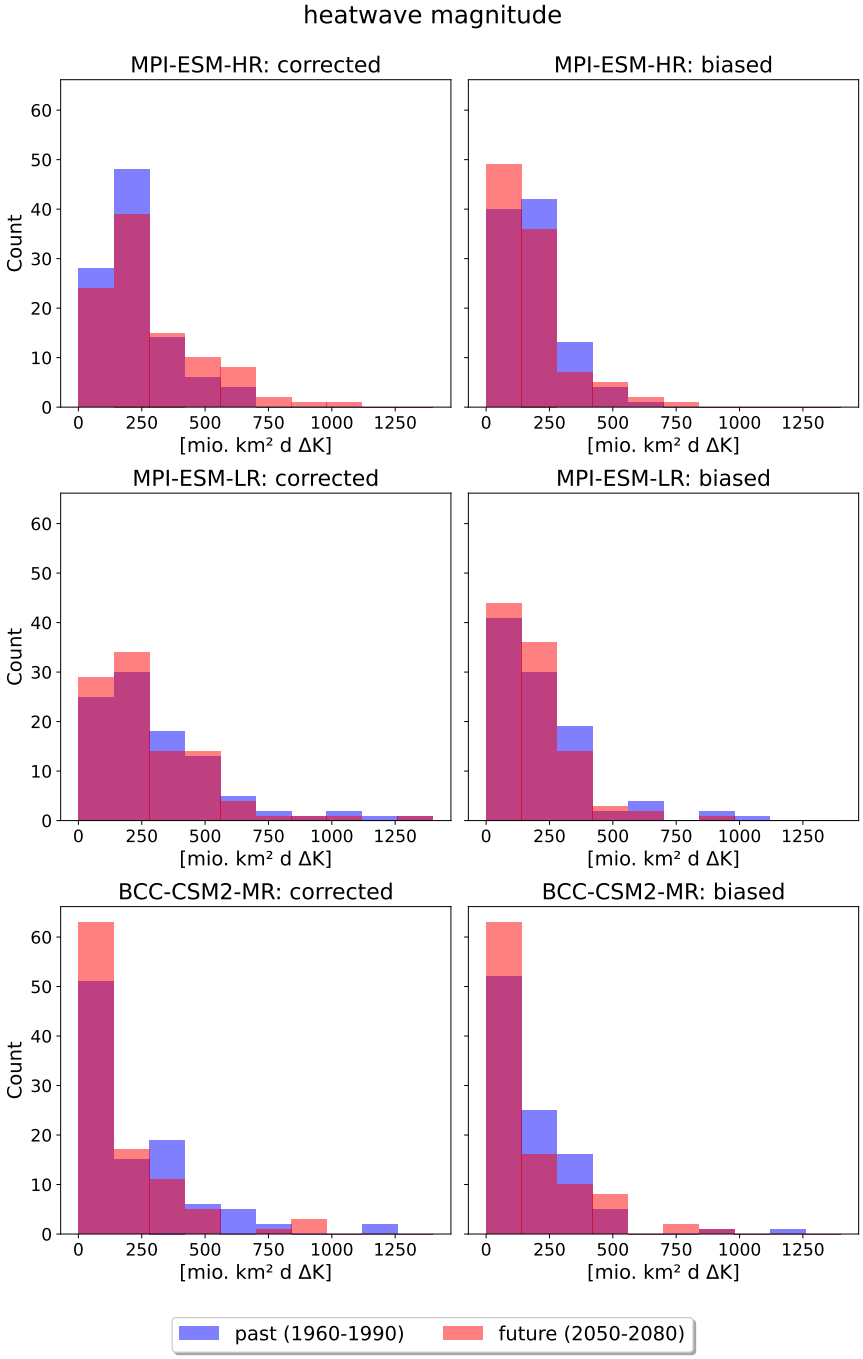

Figure 31.: Magnitude of the 100 biggest heatwaves for MPI-ESM-HR (top), MPI-ESM-LR (middle) and BCC-CSM2-MR (bottom).

Bibliography

Black, E., M. Blackburn, G. Harrison, B. Hoskins and J. Methven (2004). "Factors contributing to the summer 2003 European heatwave". In: *Weather* 59.8, pp. 217–223. DOI: https://doi.org/10.1256/wea.74.04.

Brunner, L., M. Hauser, R. Lorenz and U. Beyerle (2020). *The ETH Zurich CMIP6 next generation archive: technical documentation*. DOI: 10.5281/zenodo.3734128.

Brunner, L. and A. Voigt (Mar. 2024). "Pitfalls in diagnosing temperature extremes". In: *Nature Communications* 15.1, p. 2087. DOI: 10.1038/s41467-024-46349-x.

Collier, N., F. M. Hoffman, D. M. Lawrence, G. Keppel-Aleks, C. D. Koven, W. J. Riley, M. Mu and J. T. Randerson (2018). "The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation". In: *Journal of Advances in Modeling Earth Systems* 10.11, pp. 2731–2754. DOI: https://doi.org/10.1029/2018MS001354.

Di Luca, A., A. J. Pitman and R. de Elía (2020). "Decomposing Temperature Extremes Errors in CMIP5 and CMIP6 Models". In: *Geophysical Research Letters* 47.14. DOI: https://doi.org/10.1029/2020GL088031.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer and K. E. Taylor (2016). "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization". In: *Geoscientific Model Development* 9.5, pp. 1937–1958. DOI: 10.5194/gmd-9-1937-2016.

Folland, C., C. Anderson, D. Bader, M. Crowe, P. Jones, N. Plummer, M. Richman, D. Parker, J. Roggers and P. Scholefield (1999). "Workshop on indices and indicators for climate extremes". In: *Clim. Change 42*, pp. 31–43.

Hirsch, A. L., N. N. Ridder, S. E. Perkins-Kirkpatrick and A. Ukkola (2021). "CMIP6 MultiModel Evaluation of Present-Day Heatwave Attributes". In: *Geophysical Research Letters* 48.22. DOI: https://doi.org/10.1029/2021GL095161.

Horton, D. E., N. C. Johnson, D. Singh, D. L. Swain, B. Rajaratnam and N. S. Diffenbaugh (June 2015). "Contribution of changes in atmospheric circulation patterns to extreme temperature trends". In: *Nature* 522.7557, pp. 465–469. DOI: 10.1038/nature14550.

Kautz, L.-A., O. Martius, S. Pfahl, J. G. Pinto, A. M. Ramos, P. M. Sousa and T. Woollings (2022). "Atmospheric blocking and weather extremes over the Euro-Atlantic sector – a review". In: *Weather and Climate Dynamics* 3.1, pp. 305–336. DOI: 10.5194/wcd-3-305-2022.

Kirch, W., B. Menne and R. Bertollini (2005). "Extreme weather events and public health responses". In: DOI: https://doi.org/10.1007/3-540-28862-7.

Luo, M., S. Wu, G. N.-C. Lau, T. Pei, Z. Liu, X. Wang, G. Ning, T. O. Chan, Y. Yang and W. Zhang (2024). "Anthropogenic forcing has increased the risk of longer-traveling and slower-moving large contiguous heatwaves". In: *Science Advances* 10.13. DOI: 10.1126/sciadv.adl1598.

Lyon, B., A. G. Barnston, E. Coffel and R. M. Horton (Nov. 2019). "Projected increase in the spatial extent of contiguous US summer heat waves and associated attributes". In: *Environmental Research Letters* 14.11, p. 114029. DOI: 10.1088/1748-9326/ab4b41.

Meinshausen, M. et al. (2020). "The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500". In: *Geoscientific Model Development* 13.8, pp. 3571–3605. DOI: 10.5194/gmd-13-3571-2020.

Mukaka, M. (Sept. 2012). "Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research". In: *Malawi medical journal : the journal of Medical Association of Malawi* 24, pp. 69–71.

Palmer, T. and B. Stevens (Dec. 2019). "The scientific challenge of understanding and estimating climate change". In: *Proceedings of the National Academy of Sciences* 116, p. 201906691. DOI: 10.1073/pnas.1906691116.

Perkins, S. E., L. V. Alexander and J. R. Nairn (2012). "Increasing frequency, intensity and duration of observed global heatwaves and warm spells". In: *Geophysical Research Letters* 39.20. DOI: https://doi.org/10.1029/2012GL053361.

Perkins, S. E. (2015). "A review on the scientific understanding of heatwaves—Their measurement, driving mechanisms, and changes at the global scale". In: *Atmospheric Research* 164-165, pp. 242–267. DOI: https://doi.org/10.1016/j.atmosres.2015.05.014.

Perkins-Kirkpatrick, S. E. and S. C. Lewis (July 2020). "Increasing trends in regional heatwaves". In: *Nature Communications* 11.1, p. 3357. DOI: 10.1038/s41467-020-16970-7.

Pfahl, S. (2014). "Characterising the relationship between weather extremes in Europe and synoptic circulation features". In: *Natural Hazards and Earth System Sciences* 14.6, pp. 1461–1475. DOI: 10.5194/nhess-14-1461-2014.

Pfahl, S. and H. Wernli (2012). "Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales". In: *Geophysical Research Letters* 39.12. DOI: https://doi.org/10.1029/2012GL052261.

Rohde, R., R. Muller, R. Jacobsen, S. Perlmutter and S. Mosher (Jan. 2013). "Berkeley Earth Temperature Averaging Process". In: *Geoinformatics Geostatistics: An Overview* 01. DOI: 10.4172/2327-4581.1000103.

Rosenfeld, A. and J. L. Pfaltz (1966). "Sequential Operations in Digital Picture Processing". In: DOI: https://doi.org/10.1145/321356.321357.

Rousi, E. et al. (2023). "The extremely hot and dry 2018 summer in central and northern Europe from a multi-faceted weather and climate perspective". In: *Natural Hazards and Earth System Sciences* 23.5, pp. 1699–1718. DOI: 10.5194/nhess-23-1699-2023.

Russo, E. and D. I. V. Domeisen (2023). "Increasing Intensity of Extreme Heatwaves: The Crucial Role of Metrics". In: *Geophysical Research Letters* 50.14. DOI: https://doi.org/10.1029/2023GL103540.

Russo, S., A. Dosio, R. Graversen, J. Sillmann, H. Carrao, M. Dunbar, A. Singleton, P. Montagna, P. Barbosa and J. Vogt (Nov. 2014). "Magnitude of extreme heat waves in present climate and their projection in a warming world". In: *Journal of Geophysical Research Atmospheres* 19, pp. 12500–12512. DOI: 10.1002/2014JD022098.

Schäfer, A., B. Mühr, F. Kaiser, D. Böhnke, S. Mohr and M. Kunu (2023). "Untersuchung der globalen Hitzewelle im Jahr 2023". In: DOI: 10.5445/IR/1000161235.

Schielicke, L. and S. Pfahl (2022). "European heatwaves in present and future climate simulations: a Lagrangian analysis". In: *Weather and Climate Dynamics* 3.4, pp. 1439–1459. DOI: 10.5194/wcd-3-1439-2022.

Sillmann, J., T. Thorarinsdottir, N. Keenlyside, N. Schaller, L. V. Alexander, G. Hegerl, S. I. Seneviratne, R. Vautard, X. Zhang and F. W. Zwiers (2017). "Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities". In: *Weather and Climate Extremes* 18, pp. 65–74. DOI: https://doi.org/10.1016/j.wace.2017.10.003.

Sippel, S., J. Zscheischler, M. D. Mahecha, R. Orth, M. Reichstein, M. Vogel and S. I. Seneviratne (2017). "Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics". In: *Earth System Dynamics* 8.2, pp. 387–403. DOI: 10.5194/esd-8-387-2017.

Vautard, R. et al. (Nov. 2013). "The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project". In: *Climate Dynamics* 41. DOI: 10.1007/s00382-013-1714-z.

Vogel, M. M., J. Zscheischler, E. M. Fischer and S. I. Seneviratne (2020). "Development of Future Heatwaves for Different Hazard Thresholds". In: *Journal of Geophysical Research: Atmospheres* 125.9. DOI: https://doi.org/10.1029/2019JD032070.

Woollings, T., D. Barriopedro, J. Methven, S.-W. Son, O. Martius, B. Harvey, J. Sillmann, A. Lupo and S. Seneviratne (Sept. 2018). "Blocking and its Response to Climate Change". In: *Current Climate Change Reports* 4, pp. 1–14. DOI: 10.1007/s40641-018-0108-z.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin and F. W. Zwiers (2011). "Indices for monitoring changes in extremes based on daily temperature and precipitation data". In: *WIREs Climate Change* 2.6, pp. 851–870. DOI: https://doi.org/10.1002/wcc.147.

Zhang, X., G. Hegerl, F. Zwiers and J. Kenyon (June 2005). "Avoiding Inhomogeneity in Percentile-Based Indices of Temperature Extremes". In: *Journal of Climate - J CLIMATE* 18, pp. 1641–1651. DOI: 10.1175/JCLI3366.1.

Zhang, Y. and W. Boos (Mar. 2023). "An upper bound for extreme temperatures over midlatitude land". English (US). In: *Proceedings of the National Academy of Sciences of the United States of America* 120.12. DOI: 10.1073/pnas.2215278120.

Zscheischler, J., M. Mahecha, J. Buttlar, S. Harmeling, M. Jung, A. Rammig, J. Randerson, B. Schölkopf, S. Seneviratne, E. Tomelleri, S. Zaehle and M. Reichstein (Mar. 2014). "A few extreme events dominate global interannual variability in gross primary production". In: *Environmental Research Letters* 9, p. 035001. DOI: 10.1088/1748-9326/9/3/035001.

Zscheischler, J., S. Westra, B. J. J. M. van den Hurk, S. I. Seneviratne, P. J. Ward, A. Pitman, A. AghaKouchak, D. N. Bresch, M. Leonard, T. Wahl and X. Zhang (June 2018). "Future climate risk from compound events". In: *Nature Climate Change* 8.6, pp. 469–477. DOI: 10.1038/s41558-018-0156-3.

Zschenderlein, P., G. Fragkoulidis, A. H. Fink and V. Wirth (2018). "Large-scale Rossby wave and synoptic-scale dynamic analyses of the unusually late 2016 heatwave over Europe". In: *Weather* 73.9, pp. 275–283. DOI: https://doi.org/10.1002/wea.3278.