Significance and Robustness of Climate Change Signals for Extreme Indices over Germany in a Convection-permitting Regional Climate Model Ensemble

Master's Thesis in
Meteorology and Climate Physics
by

Annabell Weber

July 2025

INSTITUTE OF METEOROLOGY AND CLIMATE RESEARCH
KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT)

Supervisor: Prof. Dr. Joaquim Pinto
Co-supervisor: Prof. Dr. Andreas Fink

Abstract

Extreme weather events are becoming more frequent and intense due to climate change, underscoring the need for reliable climate information to support adaptation strategies. This thesis investigates how climate information on extreme events from a convection-permitting climate model ensemble (NUKLEUS) can be communicated more effectively using climate signal maps. The robustness and precision of climate change signals for ETCCDI indices, derived from daily precipitation and temperature data at $+2K$ and $+3K$ global warming levels, are evaluated. Robustness is a measure of the reliability of a climate change signal, defined as the agreement among ensemble members on the sign and significance of the individual signals. Precision refers to the degree of agreement among ensemble members regarding the magnitude of the projected climate change signal.

Climate signal maps, following Pfeifer et al. (2015), are used to visualize climate change signals and assess their robustness. Precision is assessed using the standard deviation of the signals from individual ensemble members. A threshold for the standard deviation is applied to distinguish between small and large ensemble spreads. Incorporating the ensemble spread in the climate signal maps enables a more nuanced distinction between robust and non-robust climate change signals. For non-robust signals, it indicates whether this results from high uncertainty or the absence of a climate change signal. For robust signals, it differentiates between precise and less precise signals.

Results show robust climate change signals for an increase in mean temperature and temperature-based indices, including tropical nights and the warm spell duration index. Monthly mean precipitation shows only a few robust climate change signals, with stronger and more robust signals in summer than in winter. Precipitation-based indices show an increase in the frequency of extreme precipitation events, particularly in very extreme events, and an increase in consecutive dry days. Most regions show large ensemble spreads, indicating low precision for robust and high uncertainty for non-robust signals. The number of consecutive wet days exhibits no robust climate change signals but high precision across most regions, indicating no projected climate change signal. These findings show that the additional classification improves the interpretability of the climate signal maps and supports decision-making. However, its applicability ultimately depends on the chosen threshold.

Zusammenfassung

Extremwetterereignisse wie Hitze oder Starkniederschläge werden infolge des Klimawandels immer häufiger und intensiver. Um geeignete Anpassungsmaßnahmen treffen zu können, braucht es verlässliche Klimainformationen. Diese Thesis betrachtet, wie Klimainformationen über Extremwetterereignisse mithilfe von Klimasignalkarten effektiv kommuniziert werden können. Dabei liegt der Fokus auf der Robustheit und Präzision der Klimaänderungssignale ausgewählter ETCCDI-Indizes. Die Indizes werden basierend auf täglichen Temperatur- und Niederschlagsdaten eines konvektionsauflösenden Klimamodell-Ensembles berechnet. Dabei werden zwei Szenarien mit einer globalen Erwärmung von +2K bzw. +3K betrachtet.

Klimasignalkarten nach dem Ansatz von Pfeifer et al. (2015) zeigen die Klimaänderungssignale ausgewählter Regionen und deren Robustheit, wobei die Robustheit über die Übereinstimmung der einzelnen Ensemblemitglieder in der Richtung und Signifikanz des Klimaänderungssignals definiert ist. Die Klimasignalkarten nach Pfeifer et al. (2015) werden in dieser Thesis durch ein Maß für die Präzision der Klimaänderungssignale ergänzt. Präzision beschreibt die Übereinstimmung der Ensemblemitglieder in der Stärke des Klimaänderungssignals. Dabei wird die Präzision anhand der Standardabweichung der Klimaänderungssignale innerhalb des Ensembles evaluiert. Um den Ensemble-Spread in „gering" oder „hoch" einzuordnen, wird ein Schwellenwert für die Standardabweichung festgelegt. Dies ermöglicht eine differenziertere Betrachtung von robusten und nicht robusten Klimaänderungssignalen. Nicht robuste Signale können entweder einer hohen Unsicherheit oder dem Fehlen eines Klimaänderungssignals zugeordnet werden. Robuste Klimaänderungssignale lassen sich weiter in präzise und weniger präzise unterscheiden.

Die Ergebnisse zeigen eine robuste Zunahme der mittleren Temperatur sowie der Temperatur-basierter Indizes, wie die Anzahl der Tropischen Nächte und der Warm Spell Duration Index. Für den mittleren Niederschlag zeigen sich hingegen weniger robuste Klimaänderungssignale. Die Klimaänderungssignals sind im Sommer stärker als im Winter und mehr Regionen zeigen robuste Signale im Sommer. Die Niederschlags-basierten Indizes zeigen eine robuste Zunahme der Häufigkeit von Extremniederschlagsereignissen (R10mm, R20mm und R30mm), besonders für extreme Ereignisse (R30mm), sowie eine Zunahme der aufeinanderfolgenden niederschlagsfreien Tage (CDD). Für die meisten Regionen und betrachteten Indizes ist der Ensemble-Spread hoch, was auf eine geringe Präzision für robuste und hohe Unsicherheit für nicht robuste Signale hinweist. Die Anzahl der aufeinanderfolgenden Tage mit Niederschlag (CWD) zeigt hauptsächlich nicht-robuste Klimaänderungssignale mit einer hohen Präzision, was darauf hindeutet, dass sich für diesen Index keine Klimaänderung feststellen lässt.

Die zusätzliche Klassifikation durch den Ensemble-Spread in den Klimasignalkarten ermöglicht eine leichtere Interpretation der Klimaänderungssignale und soll Entscheidungsträger bei ihrer Bewertung unterstützen. Für eine zuverlässige Klassifikation muss jedoch der Schwellenwert sinnvoll gewählt werden.

Contents

1 Introduction

Extreme weather events such as heatwaves or heavy precipitation are becoming more frequent and intense due to climate change (IPCC, 2023b). The impacts of climate change are increasingly being felt at the local level (Riach and Glaser, 2024), with 77% of 1062 municipalities surveyed in Germany stating that they were affected by extreme weather events in the last 10 years (Umweltbundesamt, Friedrich et al. (2024)). These extreme events have severe consequences for society and ecosystems, highlighting the urgent need for climate adaptation efforts, also because more and more local authorities are legally obliged to develop adaptation strategies (Riach and Glaser, 2024). To support these efforts, climate information at the local level is essential.

Extreme events are rare, and severe weather events such as heatwaves, heavy rainfall, or droughts. Changes in the frequency and intensity of extreme events are the result of thermodynamic and dynamic processes in the climate system (IPCC, 2023b). Thermodynamic processes include exchanges of heat and moisture. Due to the warming of the atmosphere and surface in response to increasing greenhouse gas concentrations, heat extremes are becoming more frequent and intense. Warmer air can also hold more moisture. According to the Clausius-Clapeyron relation, air can hold approximately 7% more moisture per degree of warming, increasing the potential for heavy precipitation events (IPCC, 2023a). Dynamic processes are associated with changes in the atmospheric and oceanic motions, such as shifts in storm tracks or large-scale circulation patterns (Suarez-Gutierrez et al., 2020).

To analyze these complex physical processes and assess future changes in extremes, climate models are essential tools. Global Climate Models (GCMs) provide a large-scale simulation of the global climate and run over long time periods. While GCMs provide valuable information on the global climate state, their spatial and temporal resolution is limited, and processes on smaller scales that require higher spatial and temporal resolution are incorporated by parametrizations. To further improve the resolution of climate data, a higher-resolution climate model that uses the output of the GCM as initial and boundary conditions is applied. These Regional Climate Models (RCMs) provide data with higher spatial and temporal resolution over a limited area and improve the representation of regional features like orography, surface properties, and small-scale atmospheric processes, such as convection (Jacob et al., 2014; Fantini et al., 2018; Prein et al., 2016). Typically, GCMs operate on spatial resolutions of 200 km, whereas RCMs run at spatial resolutions of 12-50km or finer. To further improve the model output, quantify uncertainties, and model-specific biases, climate model ensembles consisting of multiple GCMs and RCMs are used, such as EURO-CORDEX (Jacob et al., 2014) or ENSEMBLES (Hewitt, 2004).

Although climate models are valuable tools to analyze future developments of the climate state and evaluate potential impacts and risks, their coarse resolution and inherent uncertainties prevent their direct use in adaptation planning (Lorenz et al., 2017). Thus, current adaptation strategies in Germany are mainly based on past and present climate data (Lorenz et al., 2017). To address this, climate model output at higher resolution paired with a robust assessment of the uncertainties is needed.

Latest developments in climate modeling have improved the quality and resolution of climate model outputs and related uncertainties. So-called convection-permitting climate models enable climate projections on spatial resolutions of $2 - 4$km, providing more regional details, reducing model-specific biases, and improving the simulation of extreme events, especially for precipitation extremes (Prein et al., 2015; Fosser et al., 2024; Pichelli et al., 2021; Kendon et al., 2017). In the NUKLEUS project (Nutzbare Lokale Klimainformationen für Deutschland, Actionable local climate information for Germany), the first multi-model ensemble of convection-permitting regional climate simulations for Germany is provided (Sieck et al., 2025). The NUKLEUS ensemble offers the potential to provide more detailed insights into regional climate change and its impacts for Germany.

However, raw climate model output is not directly useful for adaptation planning. It needs to be translated into climate information that is accessible and relevant for users outside the scientific community (Lemos et al., 2012). Climate indices help with this by summarizing climate model output into statistics on climate extremes and potential impacts (Zhang et al., 2011; Schipper et al., 2019; Hackenbruch et al., 2017). The choice of indices depends on the application. A widely used set of indices to describe temperature and precipitation extremes is the 27 indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI, Zhang et al. (2011)).

To build more trust in the climate model output, a transparent assessment of its reliability is essential (Brasseur and Gallardo, 2016; Lorenz et al., 2017). Reliability refers to the trustworthiness of the climate model ensemble to represent the future development of the climate system accurately and consistently (Räisänen, 2007). An indicator for reliability is the robustness of the climate change signals, which indicates whether the projected climate change signal is consistent and significant among the individual ensemble members (Jacob et al., 2014; Pfeifer et al., 2015; Knutti and Sedláček, 2013; Kjellström et al., 2018; Dosio and Fischer, 2018).

For climate information to have practical applications, it must be communicated in a clear, accessible, and interpretable way. Pfeifer et al. (2015) addressed this by introducing a method to analyse and visualize the robustness of climate change signals in so-called climate signal maps. Climate signal maps aim to communicate the climate change signals for different regions alongside their robustness in an accessible and comprehensible way, also to non-scientists.

Improving the quality of climate information is essential, but to effectively support decision-making, users also need guidance in interpreting and applying this information (Brasseur and Gallardo, 2016; Hewitt et al., 2017, 2020). Current approaches, such as the climate signal maps by Pfeifer et al. (2015), classify climate signals as either robust or non-robust, aiming to guide users on which

information is reliable. However, this binary classification does not capture important nuances. For instance, a non-robust climate change signal results either from the absence of a climate change signal or from too high uncertainty (Tebaldi et al., 2011). In contrast, a robust signal can still vary in its precision. Precision refers to the degree of agreement among the ensemble members regarding the magnitude of the projected signal.

Climate signal maps visualize only the ensemble median signal and categorize its strength into three categories, without reflecting the uncertainty in this classification. If the variability among the ensemble members is large, the assigned category may not accurately represent the full range of the ensemble. This limitation is not addressed in the approach by Pfeifer et al. (2015).

In this thesis, a more granular classification of the climate change signals is proposed to enhance the interpretability of the climate signal maps and support decision-making. Building on the approach by Pfeifer et al. (2015), the climate signal maps are extended by the ensemble spread to assess the precision of the ensemble median signal. A threshold for the standard deviation is used to classify the ensemble spread as either small or large.

The classification process is illustrated in Figure 1.1, with four resulting cases indicated in the lower part. Robust signals are considered reliable, meaning that the majority of ensemble members agree on the direction and significance of the climate change signal. If the ensemble spread is small, indicating the ensemble members also agree on the magnitude of the signal, it is classified as reliable and precise. If the ensemble spread is large, the signal is still reliable but is not considered precise due to the high uncertainty in its magnitude. Non-robust signals indicate that the ensemble members disagree on the direction and significance of the climate change signals. If the ensemble spread is small for non-robust signals, it indicates that there is no or only a weak climate change signal. In contrast, non-robust signals with a large ensemble spread indicate a lack of information, as the uncertainty is too high to derive a meaningful signal.

The thesis evaluates whether these four cases can be identified based on robustness and precision of the climate change signals. To do so, the climate change signals of ETCCDI indices derived from the NUKLEUS ensemble are analyzed, aiming to provide climate information on extreme events for Germany. This concludes in two central research questions:

1. Which climate change signals of ETCCDI-indices derived from the NUKLEUS ensemble can be classified as robust?

2. Does the consideration of the ensemble spread add value to the climate signal maps?

This thesis is structured as follows: Chapter 2 gives an overview of currently available climate change information on extremes at the local level and its visualization. Chapter 3 describes the NUKLEUS climate model ensemble and methods used for the analysis of the climate change signals. The results are presented in Chapter 4 in two parts: the climate change signals of precipitation and temperature, and the ETCCDI indices. Chapter 5 discusses the results and the added value of the ensemble spread, followed by a conclusion and outlook in Chapter 6.
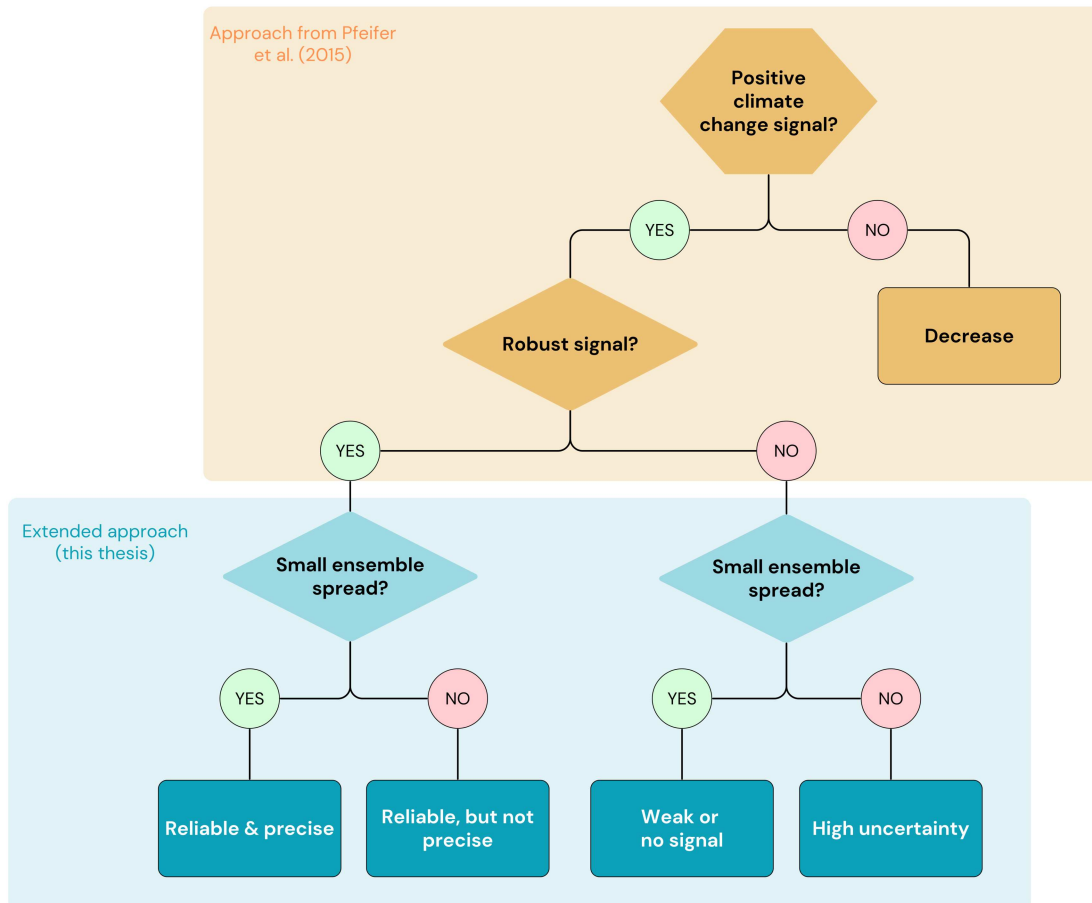
Figure 1.1: Illustration of the proposed classification approach based on the robustness and ensemble spread of a climate change signal. The orange colors underline the initial climate signal map approach introduced by Pfeifer et al. (2015). The blue colors indicate the extended approach introduced in this thesis. Here, a positive change is considered. The resulting distinct classifications of the climate change signals are illustrated in the Appendix, Figure B.1.

2 Currently available climate change information on extremes at local level

This chapter presents the current literature on assessing and visualizing climate change information on extremes at the local level. It is divided into two sections: the first describes current methods to analyse climate change signals on extremes. This includes climate models and their role in providing the basis for climate change information, what factors need to be considered providing user-relevant climate information, and methods for validating the climate information. A climate change signal refers to the projected change in a climate variable or index relative to a defined reference period (IPCC, 2023; Seaby et al., 2013). The second part describes ways to visualize climate change information.

2.1 Current methods to analyze climate change signals of extremes

Climate information can be based on observational data, reanalysis data, or climate projections from climate models. Reanalysis data refers to climate data that is generated by climate models based on past observations to provide consistent and continuous climate data of the past and present (Hersbach et al., 2020). Observational and reanalysis data give valuable insights about the past and present climate that help to better understand drivers of past climate variations and important processes within the climate system. However, their ability to predict future climate conditions is limited, especially regarding extreme events. Observational records might be too short to capture very rare extreme events, and reanalysis data can under- or over-estimate the frequency and intensity of extremes (Hu and Franzke, 2020; Donat et al., 2013). Furthermore, extreme events are expected to increase in their frequency and intensity due to climate change (IPCC, 2023b). Therefore, it is difficult to derive information about future changes in extreme events based on past and present climate records.

Climate models provide projections of the future climate under different scenarios, offering a valuable alternative to reanalysis and observational data. To validate their performance, climate models are used to simulate the historical climate, and their output is compared to reanalysis and observational data (Flato et al., 2013).

Climate models are based on physical principles and can simulate how the climate system responds to different future conditions, which makes them particularly useful for analyzing future changes in

extreme events (IPCC, 2023; Eyring et al., 2016; Sillmann et al., 2013a). The following chapter, therefore, introduces the general concept of climate models. To provide climate information on extremes at the local scale and support climate adaptation strategies, high-resolution climate data is required (Lorenz et al., 2017; Lemos et al., 2012). Convection-permitting climate models provide such high-resolution climate data. Therefore, the section further introduces regional and convection-permitting climate models.

2.1.1 Climate models

Climate models simulate the Earth's climate system, including processes in the atmosphere, biosphere, hydrosphere, and cryosphere. They are used to analyze past climate, understand crucial processes and interactions between the different components, and to project future climate. The simplest mathematical form of a climate model is an energy-balance model (EBM) describing the balance between incoming solar energy and outgoing heat (Edwards, 2011). A more advanced approach is a radiative–convective model describing vertical transfers of energy in the atmosphere (Edwards, 2011).

However, to capture the complex dynamics of the climate system, more complex models are required. General Circulation Models (GCMs) are based on the principle of numerical weather prediction and consist of two fundamental components: the dynamical core and model physics. The dynamical core solves the primitive equations, which are prognostic equations describing the conservation of mass, momentum, and energy, and describe large-scale fluid motions. These include the three-dimensional momentum equations, the ideal gas law, the continuity equation, the first law of thermodynamics, and the budget equations for water substances. The model physics refers to diagnostic equations that describe sub-grid scale processes such as radiation, convection, or turbulence. These processes cannot be resolved directly by the model and thus must be parameterized.

As the primitive equations cannot be solved analytically, they are solved numerically by discretizing the equations both in space and time. This results in a three-dimensional grid representation of the atmosphere, as illustrated in Figure 2.1. At each time step, the equations are updated for every grid cell, accounting for both vertical and horizontal exchanges between the neighboring grid cells. Here, a rectangular grid is shown. There are also other grid structures, such as the icosahedral grid (ICON, Zängl et al. (2015)).

Over time, GCMs have improved the representation of the climate system, with an increasing number of different components, including the ocean, land surface, cryosphere, and the carbon cycle. Models including all these components are called Earth System Models (ESM). A coupler connects the different components to ensure a consistent simulation of the climate system. In the following, the acronym GCM refers to Global Climate Models and describes any model from an atmosphere-only to a full coupled ESM.
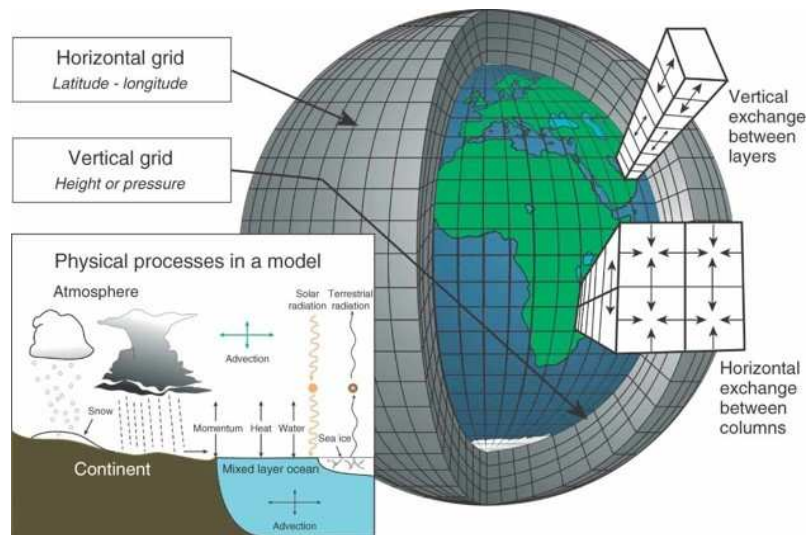
Figure 2.1: Illustration of the Cartesian grid structure in a GCM, from Edwards (2011). The horizontal and vertical grid, as well as the exchange processes between the grid cells, are shown. The box on the left depicts the different physical processes simulated by the model.

The parametrizations describing sub-grid scale processes can be formulated in many different ways and differ between different GCMs. Comparing simulations from different models is useful to analyse whether the results are consistent across models or differ. To coordinate this model intercomparison within the climate science community, the Coupled Model Intercomparison Project (CMIP) was founded by the World Climate Researcher Programme (WCRP). The goal of CMIP is to better understand past, present, and future climate changes in a multi-model context. CMIP is organized in different phases, updating the standards for climate simulations, data formats, and analysis tools. The most recent phase is CMIP6 (Eyring et al., 2016).

The CMIP provides the foundation for the climate projections assessed in the Intergovernmental Panel on Climate Change (IPCC) reports. These projections are based on standardized scenarios. Up to the fifth CMIP phase, the simulations were conducted for different Representative Concentration Pathways (RCPs, Van Vuuren et al. (2011)), which represent different greenhouse gas emission scenarios. For the sixth phase, CMIP6, the Shared Socioeconomic Pathways (SSPs, O'Neill et al. (2014)) are considered, combining different socioeconomic developments with emission scenarios, aiming to better represent a full picture of possible climate futures.

Regional climate models

GCMs provide large-scale simulations of the climate on a global scale with typical resolutions of 200 km. While GCMs provide valuable information on the global climate, their spatial and temporal resolution is limited, and processes on smaller scales that require higher spatial and temporal resolution are incorporated by parameterizations. To further improve the spatial and temporal resolution of climate data from GCMs, two main downscaling approaches are commonly used: statistical and dynamical downscaling. Statistical downscaling is based on the statistical relationships between the variables from the GCM output and regional variables, whereas dynamical

downscaling applies a higher-resolution climate model that uses the output of the GCM as initial and boundary conditions. These regional climate models (RCMs) provide data with higher spatial and temporal resolution and cover a limited area. The dynamical downscaling process is illustrated in Figure 2.2. At the resolution of the RCM, orographical features like the Alps are represented in much more detail compared to the GCM (see Figure 2.2).
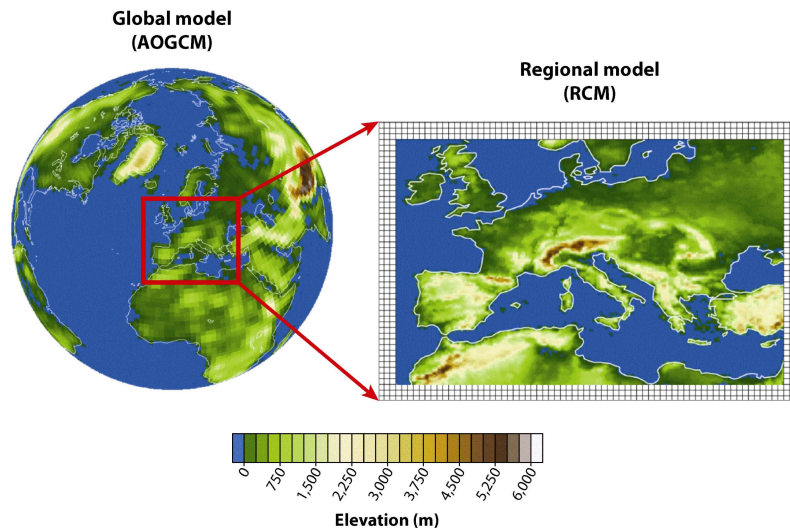
Giorgi F, Gutowski Jr WJ. 2015.
Annu. Rev. Environ. Resour. 40:467–90

Figure 2.2: Illustration of dynamical downscaling with a RCM from Giorgi and Gutowski (2015). AOGCM stands for a coupled atmosphere-ocean general circulation model. The orography in the GCM and RCM is shown. The squared area surrounding the RCM domain illustrates the lateral buffer zone.

Typically, RCMs run at spatial resolutions of $12 - 50\,$km or finer. Common domains of RCMs include, for example, regions defined by the EURO-CORDEX initiative, such as the EURO-CORDEX domain, which covers Europe and neighboring regions, and runs at resolutions of $12.5\,$km or $50\,$km (Jacob et al., 2014).

The finer resolution of RCMs allows a better representation of regional features like orography, surface properties, and small-scale atmospheric processes, such as convection and orographic precipitation (Fantini et al., 2018; Prein et al., 2016; Jacob et al., 2014). This improves the simulations of regional climate patterns and extremes, such as heatwaves and heavy precipitation, compared to the simulations of the coarser GCM (Ritzhaupt and Maraun, 2023). Although RCMs already provide more detailed simulations compared to GCMs and improve regional climate simulations, their resolution is still too coarse to fully capture crucial processes for extreme events, such as convective precipitation and wind gusts (Pichelli et al., 2021; Kendon et al., 2017).

Convection-permitting climate models

A major source of uncertainty in RCMs is the parametrization of deep convection (Prein et al., 2015). Deep convection is a major source of heavy precipitation and thunderstorms, mainly in summer, and affects vertical transport processes, influencing cloud formation and radiative effects

(Prein et al., 2015). At the scales of RCMs, typically around $12 - 50\,$km, deep convection cannot be explicitly resolved and must be parametrized. This motivated the development of convection-permitting climate models (CPMs), which operate at a spatial resolution below $4\,$km and thus, can explicitly resolve deep convection (Prein et al., 2015). CPMs are high-resolution RCMs in which the parametrization of deep convection is switched off. As such, CPMs represent an additional step in the dynamical downscaling chain, using the output from an RCM as initial and boundary conditions. This results in a two-step dynamical downscaling: first, the GCM output is downscaled by the RCM, and then, the RCM output is further downscaled by the CPM. This two-step downscaling is illustrated in Figure 2.3. The cutouts in the lower part of the figure illustrate the representation of the surface temperature for each nesting step for the city of Hamburg. This highlights the improved representation of regional features by the CPM compared to the coarser RCM and GCM, as the urban heat effect is only captured by the CPM.
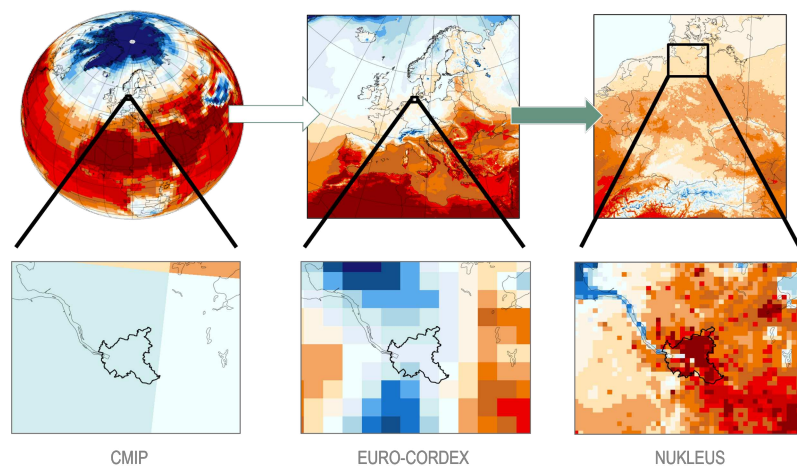
Figure 2.3: Example of a two-step dynamical downscaling from the NUKLEUS ensemble (Sieck et al., 2021). The surface temperature is shown. The cutouts in the bottom of the figure highlight the resolutions differences between the GCM (CMIP, left), RCM (EURO-CORDEX, middle), and CPM (NUKLEUS, right) for the city of Hamburg, indicated by the black contours.

The explicit resolution of deep convection along with the finer resolution of CPMs further improves the representation of precipitation patterns and reduces the bias in comparison with observations (Pichelli et al., 2021; Prein et al., 2015; Fosser et al., 2024). Especially, the diurnal cycle of precipitation in terms of the onset of convection, the intensity of precipitation, and the representation of hourly extremes is improved in CPMs compared to RCMs (Prein et al., 2015, 2017). Largest improvements of the CPMs are found for precipitation extremes and for summer precipitation, while for winter precipitation, the simulations tend to agree between the coarser RCM and CPM simulations (Kendon et al., 2017; Prein et al., 2017). The main differences between CPMs and RCMs are found for summertime rainfall intensity and duration, hourly and daily rainfall extremes in summer, and daily precipitation extremes over mountains in winter (Kendon et al., 2017). For instance, Kendon et al. (2017) compared various CPMs to the output of coarser RCMs for the UK and the Alps. The CPMs used by Kendon et al. (2017) hereby showed an increase in the summer precipitation intensity, while this increased intensity was not evident in coarser-resolution RCMs. Furthermore, Pichelli et al. (2021) found that projections based on CPM simulations compared to RCM-based simulations even modify the sign of precipitation intensity change and

heavy precipitation for several regions. This reflects the high relevance of CPM simulations, especially for representing regional precipitation patterns and precipitation extremes.

CPMs not only improve the representation of precipitation compared to RCMs. The finer resolution of CPMs provides a more accurate representation of orography and surface properties, further improving the simulations, especially in mountainous and coastal regions (Prein et al., 2015), with strong spatial heterogeneity. The improved representation of orography leads to improvements in the simulation of the 2-m temperature field (Prein et al., 2015). The finer resolution of CPM also improves the representation of urban climate and enables resolving urban features like urban heat islands (Prein et al., 2015).

Overall, these improvements in the representation of precipitation patterns, precipitation extremes, the representation of orography, and fine-scale temperature features highlight the added value of CPMs and underline their importance for providing detailed climate data on extremes at the local scale.

2.1.2 User-relevant climate information

Climate change is not only affecting the mean state of the climate system but also leads to more intense and frequent extreme events such as heatwaves, heavy precipitation, or droughts (IPCC, 2023b). These extreme events pose a growing risk for society, infrastructure, and ecosystems. Therefore, effective adaptation strategies across multiple sectors are essential for society to reduce and prevent severe consequences. For this, decision-makers need trustworthy information on which they can base their adaptation strategies and risk assessments (Lemos et al., 2012; Lorenz et al., 2017). Climate information refers to data, projections, and analyses about past, current, or future climate conditions, including both average conditions and extreme events, used to inform users on climate change and its impacts (Hewitt et al., 2020; Buontempo et al., 2022; Riach and Glaser, 2024). Several factors need to be considered when providing user-relevant climate information.

1. **Accessibility:** First of all, climate information must be presented in an accessible and understandable format, also for non-experts, and must match the user's needs (Lemos et al., 2012; Hackenbruch et al., 2017). This involves translating raw climate model output into decision-relevant indicators such as extreme indices describing changes in the occurrence or intensity of extreme events (Schipper et al., 2019; Hackenbruch et al., 2017). For example, a city developing a heat protection plan needs tailored information on the frequency and intensity of heat events.

2. **Resolution:** The resolution of the climate information must match the spatial and temporal scale at which decisions are made (Wagner et al., 2013). For example, developing a heat protection plan for a city requires climate data at a resolution that resolves the urban features like the urban heat island effect. Moreover, the information must meet the time frame of adaptation planning cycles (UNFCCC - Adaptation Committee, 2020).

3. **Reliability:** The climate information must be reliable, which includes the validation of the climate models, their output, and the analysis methods. A measure for the reliability of climate change signals is robustness, which evaluates the sign and significance of the climate change signal (Knutti and Sedláček, 2013; Pfeifer et al., 2015; Jacob et al., 2014).

4. **Communication:** As the climate information is applied by non-experts, a clear communication of the data and how it can be applied is key (Lemos et al., 2012; Brasseur and Gallardo, 2016; Hewitt et al., 2017). This includes straightforward visualization and accessible platforms.

5. **Applicability:** Ideally, the format and method to provide climate information are applicable for different variables and applications, and comparable between different regions and time frames.

In the following sections, these aspects of user-relevant climate information are examined individually, and how they can be implemented, aiming to support decision-makers with relevant climate information on extremes at the local level.

2.1.3 Extreme indices

The direct output of climate models includes physical variables like temperature or precipitation. While these variables are important to understand the physical side of climate change, they have limited application outside of the scientific context (Hackenbruch et al., 2017; Schipper et al., 2019; Lemos et al., 2012). For the user, such as stakeholders in city councils or private companies, it is difficult to apply the raw climate model output for their specific application. To provide usable information for these users, the changes in the basic variables are translated into relevant indicators of climate impacts. These climate indices translate climate model output into useful climate information on climate impacts and extreme events. Applying climate indices facilitates the interpretation of climate model outputs, as they focus on specific applications such as describing extremes, which is difficult to derive from mean changes in the physical variables themselves.

There are several different climate indices, developed for different applications and targeted end-users, aiming to improve the applicability of climate model data in practice. The requirements for the practical formulation of climate indices differ depending on the decision-maker and application. Therefore, understanding of the users' needs and close cooperation are essential to ensure the applicability of climate indices (Buontempo et al., 2020). For example, based on expert interviews, Schipper et al. (2019) defined tailored climate indices describing climate change impacts on application-oriented sectors like infrastructure, health, or human recreation. Examples include the number of salting days, which enables a better projection of future resources required for traffic safety in winter, or the number of hiking days, which is relevant for public use and the tourism sector (Schipper et al., 2019; Pinto et al., 2025).

Other indices focus on describing the more physical features of climate impacts, such as changes in maximum temperatures, the duration of dry spells, or the occurrence of heavy precipitation

events (Zhang et al., 2011). Regarding the assessment of extreme events, various indices focus on describing the frequency and intensity of extremes. The Expert Team on Climate Change Detection and Indices (ETCCDI) has developed 27 indices describing temperature and precipitation extremes. These indices offer a consistent way to describe climate extremes and enable the detection of changes in their frequency and occurrence. The ETCCDI indices focus on describing moderate extreme events with return periods of a year or less (Zhang et al., 2011). Moreover, because of the universal definition, the ETCCDI indices are comparable across different regions and studies.

The ETCCDI indices can be divided into four different types of indices. First, there are indices describing extreme characteristics, such as the hottest day of the year or the annual maximum temperature. Second, there are threshold-based indices that are defined as the number of days exceeding a specified threshold, for example, ice days or summer days. Third, duration-based indices describe the duration of certain events, such as the length of dry spells. Fourth, percentile-based indices are defined through exceeding a percentile-based threshold, such as R95pTOT, indicating the total precipitation amount per year accumulated over all days for which the daily precipitation amount exceeds the 95th percentile. A percentile indicates the value below which a given proportion of all data points from the reference period lie. Accordingly, the 95th percentile corresponds to the value below which 95% of all data points within the reference period fall. The ETCCDI-indices allow comparability with other studies and have been widely tested and validated for representing extremes (e.g. Sillmann et al., 2013a; Jacob et al., 2014; Vautard et al., 2021).

2.1.4 Uncertainties and multi-model ensembles

Understanding and quantifying the uncertainties of climate projections is essential for providing reliable climate information. The uncertainties of climate projections can be divided into three groups: scenario uncertainty, internal variability, and model uncertainty (Hawkins and Sutton, 2009).

The scenario uncertainty describes the uncertainty due to the uncertainty in the future emission of greenhouse gases (Hawkins and Sutton, 2009). Possible scenarios are described by the RCP Van Vuuren et al. (2011) or SSP (O'Neill et al., 2014) frameworks. The climate response varies depending on the considered scenario.

The internal variability refers to the chaotic nature of the climate system and can be addressed by running multiple realizations of the same model and scenario, but with slightly varying initial conditions (Hawkins and Sutton, 2009; Deser et al., 2012).

The model uncertainty is due to differences in the physical formulation and structure of climate models. The models vary in the parametrizations and numerical descriptions of the simulated processes. As there is no perfect model, multi-model ensembles are used instead of relying on a single model. These multi-model ensembles reduce the influence of individual models and model-specific biases, aiming to capture a range of possible outcomes (Ban et al., 2021). However, the model selection plays a major role in the ensemble design, and an overrepresentation of specific

models must be avoided (Seaby et al., 2013; Evans et al., 2014). This can be addressed, for example, by using a matrix-like structure combining multiple GCMs and RCMs (Vautard et al., 2021; Déqué et al., 2012). However, the equal weighting of all models within an ensemble is not ideal, as some models might perform better than others (Christensen and Kjellström, 2022). Still, multi-model ensembles enable a better and robust projection and avoid uncertainties due to the model selection. Examples of multi-model ensembles include the EURO-CORDEX ensemble (Jacob et al., 2014), ENSEMBLES (Hewitt, 2004), the CPM ensemble CORDEX-FPS (Pichelli et al., 2021; Ban et al., 2021), and the convection-permitting multi-GCM COSMO-CLM ensemble (Hundhausen et al., 2023, 2024).

2.1.5 Robustness of climate change signals

The reliability of a climate change signal can be assessed with robustness. According to the Third Assessment Report of IPCC, "a robust finding for climate change is defined as one that holds under a variety of approaches, methods, models, and assumptions, and one that is expected to be relatively unaffected by uncertainties" (Watson and UNEP, 2003). There are various different definitions and measures for robustness based on the agreement of the ensemble members or the significance of the projected changes, or both. Kjellström et al. (2018) and Kendon et al. (2008) uses the signal-to-noise ratio (SNR) to assess the robustness of climate change signals. Knutti and Sedláček (2013) introduced a robustness measure based on the signal-to-variability ratio and the ranked probability skill score used in weather prediction. Dosio and Fischer (2018) defines robustness as the interplay of significance and agreement on the sign of change.

Pfeifer et al. (2015) introduced a similar robustness definition based on two tests. The first test is based on a predefined fraction of ensemble members that need to agree on the sign of change. The second test assesses the significance of the signal for each ensemble member, and a certain fraction of members need to show significant changes to pass the test. Thus, their definition of robustness is based on the individual signals of each ensemble member and their degree of agreement in sign and significance. Assessing the significance of each ensemble member individually ensures that robustness reflects consistent significant signals across the ensemble, with significance evaluated by each ensemble member. This approach to assess the robustness of climate change signals in multi-model ensembles was further adopted in other studies (Jacob et al., 2014; Georgoulias et al., 2022). Assessing the robustness of climate change signals within ensembles improves the confidence and reliability in their projections. This supports not only decision-making for adaptation and mitigation strategies but also helps to identify regions and variables where the climate change signals are not robust and further analysis or data are required.

Limitations

If applying a robustness measure to climate data, it is also important to question the robustness of the robustness measure. The robustness measure by Pfeifer et al. (2015) assesses the sign and significance of the climate change signals in the individual ensemble members. Pfeifer et al. (2015)

tested the sensitivity of the robustness test to changing the parameters. First, they tested differences in the fraction of models in the test of agreement. Changing the degree of agreement from 66% to 90% did not cause strong changes in the number of regions with robust signals. This is not the case for significance. Here, Pfeifer et al. (2015) found that increasing the fraction to 90% of the models has a stronger influence on the number of regions with robust signals than decreasing the significance level from 0.15 to 0.1 (Pfeifer et al., 2015).

Following the approach by Pfeifer et al. (2015), the climate change signals are divided into robust and non-robust. This binary classification of climate change signals into robust or non-robust does not take into account a more nuanced differentiation. Tebaldi et al. (2011) proposed further differentiating non-robust signals into a lack of signal and a lack of information, by differentiating between agreement on the sign and significance. These aspects highlight the limitations of the robustness assessment, and ultimately, these limitations need to be communicated when applying robustness in the context of climate services. Transparently communicating what robustness does and does not capture ensures that the climate information is interpreted correctly and further enforces the reliability and trust in the climate projections.

2.1.6 Variability in climate model ensembles

Depending on the variable and region, the individual models within a climate model ensemble tend to agree or vary in the magnitude or even in the sign of the projected changes. The ensemble spread is a measure to quantify the agreement on the magnitude of the projected signal. If the ensemble spread is low, the models seem to agree, and there is high confidence in the magnitude. Instead, if the ensemble spread is high, the models disagree on the magnitude or even the sign of change, and no clear conclusion can be drawn. This shows that the ensemble spread is relevant to make assumptions about the confidence in the magnitude of a climate signal.

While the robustness definition of Pfeifer et al. (2015) indicates whether the ensemble agrees on the sign of the signal and its significance, it does not give information on how much the signal's magnitude varies between the models. As every ensemble member acts as a possible realization of the future, it is "important for decision-making to see the full range of possible consequences and associated probabilities" (IPCC Guidance Note on Consistent Treatment of Uncertainties, Watson and UNEP (2003)). Thus, it is not only important to know whether a signal is considered robust, but also if the models agree on its magnitude. For instance, if there is a medium change in tropical nights for a region, a decision-maker might want to know whether the ensemble consistently projects medium changes or if it results from a split, with half of the models showing high changes and the other half low changes.

There are several statistical measures to quantify the ensemble spread, like the standard deviation or variance. Other measures capture the distance between the maximum and minimum values, or use the interquartile range, defined as the difference between the 75th and 25th percentiles. Another measure is the SNR (e.g. Kendon et al. (2008); Kjellström et al. (2018)), which compares the spread to the signal. Knutti and Sedláček (2013) introduced a robustness metric that combines robustness

with a measure for variability. It is based on the magnitude and sign of change, natural variability, and the inter-model spread. However, this approach by Knutti and Sedláček (2013) is rather complex, and the interpretation of the results is not straightforward, particularly for non-experts. A common and well-known measure for variability is the standard deviation, which is widely used beyond the context of climate science.

2.2 Current methods to visualize climate change information on extremes

The choice of graphical representation of the results and their robustness depends on the purpose of the study and the targeted user. A common way to display the robustness of the results of a climate model ensemble is to mark robust areas with hatching in maps (e.g. Jacob et al., 2014; Kjellström et al., 2018). To support adaptation strategies, the data needs to be represented in an accessible and understandable way for decision-makers (Lemos et al., 2012; Lorenz et al., 2017). In recent years, climate services have been developed to provide user-oriented climate information tailored to the needs of stakeholders. These services aim to bridge the gap between climate science and decision-making by translating complex climate data into actionable climate information and providing the information in accessible formats for users (Hewitt et al., 2012, 2017; Buontempo et al., 2020). Examples include the Copernicus Climate Change Service (C3S) (Buontempo et al., 2022) or the Climate Service Center Germany (GERICS).

In this context, Pfeifer et al. (2015) proposed climate signal maps as a way to communicate climate change signals and their robustness in a user-friendly format. Instead of displaying the full range of projected changes at each grid point, the data is aggregated as spatial means over predefined regions and divided into 3 coarse categories given in percentage (low, medium, and high change). In a climate signal map, only one direction of change is considered, and only robust changes are displayed. Regions with non-robust signals are shown in gray, and no information about the climate signal is shown. This allows the user to easily extract information for their region, whether the signal is robust, and if so, how large the signal is. Overall, the climate signal method introduced by Pfeifer et al. (2015) provides a balance between combining information on the signal strength and robustness, while facilitating the results for easier interpretation for non-experts.

2.2.1 Limitations

This presentation of the results involves some loss of information, such as the exact magnitude of change due to the use of categories, or the consideration of only one direction of change, which does not give insights into the strength of opposite trends. The latter is less relevant in an application-oriented context, as in climate extremes, often only one direction of change is relevant. For example, an increase in the number of hot days is far more relevant than a potential decrease, as for an increase according adaptation measures need to be implemented. The choice of the limits for the levels depends on the considered variable and is not constant. Therefore, the classification

of the signals into small, medium, and large signals is up to the user and the application. As the categories are generally very broad, the choice of the limits has a significant impact on the outcome of the classification of the signals.

Moreover, the spatial aggregation of the climate data into regions has an impact on the outcome of the results. One important factor for the climate signal maps by Pfeifer et al. (2015) is that the region size must suit the spatial resolution of the underlying climate data. Due to the very different sizes of the rural districts applied by Pfeifer et al. (2015), the climate data might be too coarse for very small regions. As Pfeifer et al. (2015) suggested, for very small regions, it is advisable to consider also neighboring regions for the interpretation of the results. The different sizes of the regions also affect the comparability between the regions. For example, very small regions might cover only a few to a single grid point, whereas large regions include multiple grid points. Therefore, averaging over the spatial extent of a large and a small region might not be comparable, as extremes might be smoothed spatially for the large regions (Pfeifer et al., 2015). Moreover, for spatially heterogeneous variables or indices, including, for example, only one grid point, the spatial averaging might lead to a loss of information. Despite the disadvantages of aggregating the climate information from grid resolution to regions, this allows an accessible and practical visualization of the climate information. From the user's perspective, the primary concern is the expected signal for their region, such as "10% decrease of summer precipitation for Region X". Therefore, the spatial aggregation to regions is applied, although it leads to some loss of information.

This underlines that there is no perfect method to test or to display the robustness of climate change signals, and ultimately, the choice of method depends on the application and the user. The climate signal method of Pfeifer et al. (2015) is a good compromise between communicating not only the climate signals but also their robustness while keeping it compact and easy to interpret. In general, the applicability of climate service products relies on a close collaboration with the targeted users and across disciplines to ensure the information is relevant, understandable, and tailored to their needs (Lemos et al., 2012; Hewitt et al., 2017; Brasseur and Gallardo, 2016).

3 Data and Methods

In this chapter, an overview of the methods and data used for this thesis is provided. First, the multi-model ensemble is introduced, followed by a description of the extreme indices and their calculation. Next, the concept of climate signal maps is presented, including the definition of robustness and a description of the statistical tests applied. In addition to robustness, an additional method to include a measure for the ensemble spread within the climate signal maps is introduced.

3.1 Extreme indices in a convection-permitting climate model ensemble

3.1.1 Ensemble setup

This thesis analyzes the output from the NUKLEUS ensemble (Sieck et al., 2025), which is a convection-permitting climate model ensemble designed to provide high-resolution climate information for Germany. The ensemble is based on multiple CPMs and GCMs. A total of three CPMs are used, driven by boundary and initial conditions from three GCMs. This results in a full 3x3 matrix of CPM-GCM pairings. The CPMs used include ICON-CLM (ICON, Pham et al. (2021)), COSMO-CLM (CCLM, Rockel et al. (2008)), and REMO-NH (REMO, Jacob et al. (2012)). The boundary and initial data are provided by three GCMs, namely EC-Earth3-Veg (ECE), MPI-ESM1-2-HR (MPI), and MIROC6 (MIROC). The GCMs used are all part of the Coupled Model Intercomparison Project Phase 6 (CMIP6, Eyring et al. (2016)). For further details, see the references listed in Table 3.1.

The ensemble was generated in two dynamical downscaling nesting steps. The first nest is defined over the EURO-CORDEX EUR11 domain (Jacob et al., 2014), which covers Europe at a spatial resolution of $0.11°$ (12.5 km). The second nest corresponds to the EURO-CORDEX CEU3 domain (Sieck et al., 2025), which represents Central Europe at a resolution of $0.0275°$ (3km). Both domains are shown in Figure 3.1. For the last nesting step, the parametrization of deep convection is switched off, so that deep convection gets resolved explicitly. The update interval for the boundary conditions is 3 hours for the CPM (CEU3 domain) and 6 hours for the RCM (EUR11 domain). The CEU3 model domain (see Figure 3.1b) includes 385 x 425 grid cells with 55 vertical levels.

Table 3.1: List of GCMs applied as boundary and initial data for the CPMs within the NUKLEUS ensemble. The periods when GWL2K and GWL3K are reached relative to the historical period 1961-1990 are given for each GCM. For further details on the GCMs, see the given references.

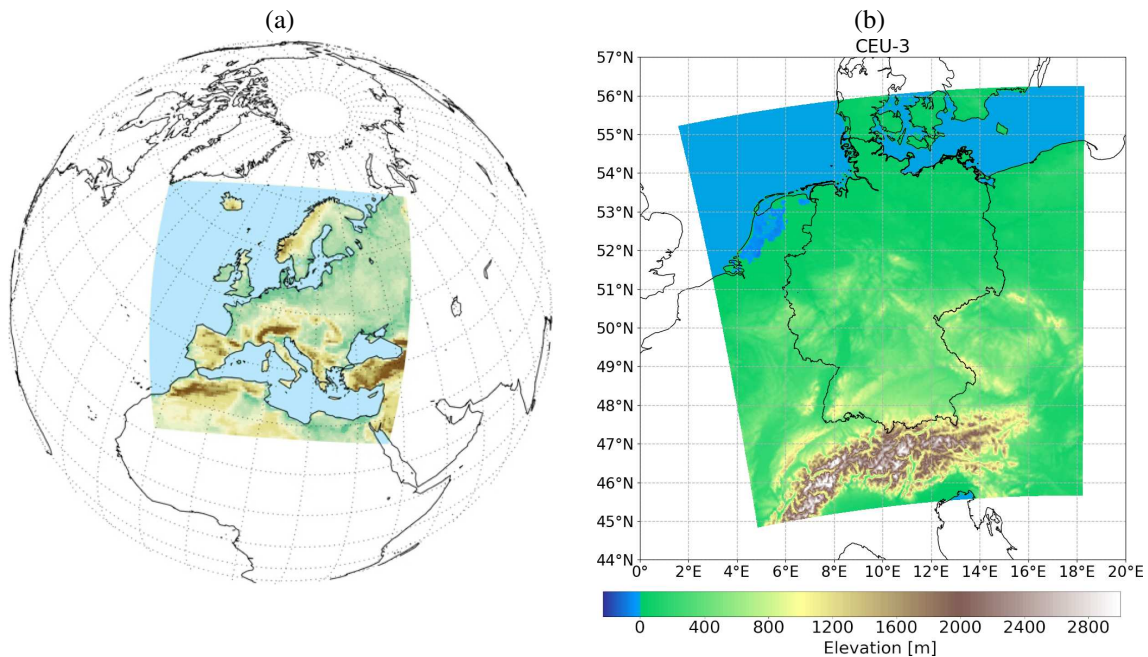| GCM	GWL2K	GWL3K	Reference
EC-Earth3-Veg	2024 - 2053	2047 - 2076	Döscher et al. (2022)
MPI-ESM1-2-HR	2037 - 2066	2066 - 2095	Mauritsen et al. (2019),Müller et al. (2018)
MIROC6	2038 - 2067	2070 - 2099	Tatebe et al. (2019)

Figure 3.1: Figure (a) shows the EUR11 domain, taken from Jacob et al. (2020). Figure (b) shows the topography of the CEU3 domain (provided by F. Ehmele).

3.1.2 Global warming levels

The fine resolution of the CPM makes running multiple CPM simulations computationally expensive. Due to these high computational costs, the simulations of the CPM ensemble were conducted for global warming levels (GWLs) instead of using a transient simulation approach. A GWL is defined by a 30-year time slice when the global mean temperature of the GCM, which forces the simulation with the CPM, reaches a mean global warming of a certain threshold relative to the historical period (Vautard et al., 2014). The period at which a GWL is reached depends on the emission scenario and the selected GCM, due to differences in climate sensitivity among the GCMs. Here, GWLs of $+2\,\mathrm{K}$ (GWL2K) and $+3\,\mathrm{K}$ (GWL3K) are evaluated relative to the historical period 1961-1990. The simulations are based on the SSP-370 emission scenario (O'Neill et al., 2014). The corresponding periods when each GCM reaches the GWLs are listed in Table 3.1. The simulations on the EUR11 domain are performed as transient simulations for the period 1950-2100.

3.1.3 ETCCDI Indices

The joint CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI) defined 27 extreme indices to describe temperature and precipitation extremes in a consistent way between different datasets, scenarios, and regions. The indices are based on daily near-surface air temperature at 2 m and precipitation data, and the output is defined annually. A detailed description of the 27 core ETCCDI indices can be found on the website of ETCCDI (`https://etccdi.pacificclimate.org/list_27_indices.shtml`). The ETCCDI indices describe moderate extreme events with one-year or less return periods.

The selection of indices for this thesis is shown in Table 3.2. All indices are selected from the 27 core indices by the ETCCDI. The indices were calculated with the ClimDexCalc2-Plugin by Freva (Free Evaluation System Framework). Freva is a scientific software framework hosted at the German Climate Computing Centre (DKRZ), supporting the analysis and post-processing of climate data (Kadow et al., 2021). It provides various evaluation tools, so-called plug-ins, to analyse and visualize climate statistics. The ClimDexCalc2-Plugin is based on the Climpact software provided by the World Meteorological Organization's Expert Team on Sector-Specific Climate Indices (ET-SCI) to calculate climate indices using the ETCCDI indices definitions (Zhang et al., 2011). The calculation of the indices for the NUKLEUS project was conducted by project partners at BTU Cottbus and is based on daily precipitation and temperature data of the individual NUKLEUS simulations. The output has annual and monthly resolution, and is available through the Freva Databrowser (Kadow et al., 2021), which is a data system implemented in Freva to support easy access to the data for scientists and the plugins themselves.

Table 3.2: List of ETCCDI indices analysed in this thesis. Detailed definitions of the ETCCDI indices can be found on the ETCCDI website (`https://etccdi.pacificclimate.org/list_27_indices.shtml`). RR is the daily precipitation amount in mm, TX stands for the daily maximum temperature, and TN for daily minimum temperatue.

Index	Full name	Definition	Unit
R10mm	Heavy precipitation days	Annual count of days with $RR \geq 10$ mm	days
R20mm	Very heavy precipitation days	Annual count of days with $RR \geq 20$ mm	days
R30mm	Extreme precipitation days	Annual count of days with $RR \geq 30$ mm	days
CDD	Maximum length of dry spell	Maximum number of consecutive days with $RR < 1$ mm	days
CWD	Maximum length of wet spell	Maximum number of consecutive days with $RR \geq 1$ mm	days
TR	Tropical nights	Annual count of days when $TN > 20\,°C$	days
ID	Ice days	Annual count of days when $TX < 0\,°C$	days
WSDI	Warm spell duration index	Annual count of days with at least 6 consecutive days when $TX > 90$th percentile of the period 1961-1990	days

Precipitation-based indices

For this study, a total of five precipitation-based indices are selected. These include R10mm, R20mm, and R30mm, which describe the number of days exceeding a certain amount of daily precipitation. They describe the annual frequency of days with heavy, very heavy, and extreme precipitation. An increase in these indices therefore, implies that more days exceed such precipitation amounts. These indices focus on the frequency of precipitation extremes. To further evaluate duration changes in precipitation extremes, CDD and CWD are selected, which describe changes in dry and wet periods. CDD describes the maximum length of dry spells as the maximum number of consecutive days with no precipitation ($RR < 1\,\mathrm{mm}$). Thus, an increase in CDD indicates longer-lasting dry periods. Similarly, CWD describes the maximum length of wet periods, through the maximum number of days with precipitation ($RR \geq 1\,\mathrm{mm}$). A change in CWD is evaluated to assess changes in wet periods.

Temperature-based indices

To describe the impacts of climate change on temperature extremes, a total of three temperature-based indices is selected. The number of tropical nights, which describes days when the minimum temperature does not fall below $20\,^\circ\mathrm{C}$. An increase in tropical nights can affect human health and recreation (Kabisch et al., 2023; He et al., 2024). To include also a measure for cold temperature extremes, the number of ice days is selected, which is an important measure for agriculture, infrastructure, and transport (Pinto et al., 2025). The warm spell duration index (WSDI) is a measure for the length of heatwaves, defined through the count of days with at least six consecutive days exceeding the 90th percentile of maximum temperature of the reference period 1961-1990. An increase in the WSDI indicates that warm spells become more frequent and longer-lasting.

3.2 Climate signal maps

3.2.1 Climate signal maps by Pfeifer et al. (2015)

The climate signal map method was introduced by Pfeifer et al. (2015). Suitable regions are selected based on the application and targeted user, in this case rural districts of Germany. For each district and ensemble member, the spatial and temporal means are computed for each period, including the historical period. The climate signal is then calculated by subtracting the mean of the historical period from the mean of the considered scenario period, for each region and ensemble member. Then the median signal for each region and scenario period is calculated across all ensemble members.

The purpose of the climate signal maps is to communicate information on climate change signals in a clear and effective way, also to non-scientific users (Pfeifer et al., 2015). Therefore, only a single direction of change is considered, either an increase or a decrease, and the signals are shown

in relative changes. The direction considered in the climate signal maps is specified in advance. Regions that show signals in the opposite direction of change are shown in white. To simplify the interpretation, the resulting signals are classified into three categories: small, medium, and large signals. The levels can be defined for each variable or index individually, depending on the application. Pfeifer et al. (2015) used green for small, orange for medium, and red for large climate change signals.

Here, the climate signal map method is applied to seasonal precipitation and temperature data as well as a selection of the ETCCDI indices, all based on the NUKLEUS ensemble. The climate signals are computed for the GWL2K and GWL3K. For temperature and temperature-based indices, absolute changes are considered instead of relative changes. The levels for the classification of small, medium, and large signals are adjusted to each variable and index individually. Here, the same colors as in Pfeifer et al. (2015) are used, except that small changes are shown in yellow instead of green, as green is often associated with positive or safe conditions. Small climate signals are distinct from the case where there is no change in the variable or index. Therefore, using the color green for small signals can lead to misinterpretation of the climate signal map.

Furthermore, Pfeifer et al. (2015) introduced a robustness test to evaluate the robustness of the climate signals. Only robust signals are shown in colors, non-robust signals in gray. The robustness test is further described in section 3.2.3.

3.2.2 Aggregation of climate change information into subregions

The goal of this thesis is to improve climate change information on extreme events at the local level to support adaptation strategies. As adaptation strategies are typically planned and implemented at the local level (Lorenz et al., 2017), here, German rural districts ("Landkreise") were selected as regional divisions in the climate signal maps. Depending on the application and user needs, alternative divisions can be considered for the climate signal maps, such as natural regions taking into account geological, hydrological, or pedological characteristics, or other administrative regions like provinces, states, Metropolitan regions, or NUTS-regions (European Commission. Statistical Office of the European Union., 2018).

There are 402 rural districts in Germany, which are referred to in the following as regions. For each region and period, the spatial mean of the considered variable or index was calculated. The spatial mean was computed taking into account all grid points falling inside a region, weighted by their proportional area within the region. The temporal mean for each region and period was then computed by aggregating the spatial means over all time steps.

3.2.3 Robustness test

To test the robustness of the climate signals, the same approach as in Pfeifer et al. (2015) is applied. Pfeifer et al. (2015) defines robustness as a combination of agreement on the sign of change and

agreement on the significance of the projected change of each simulation in the ensemble. To test for robustness, two statistical tests for each region are applied. Both tests need to be passed for a region to be classified as robust. In the following sections, these tests are described in detail.

Agreement Test

The first test considers the model agreement within the ensemble on the sign of change. A certain fraction of members within the ensemble has to agree on the sign of change, which is specified in advance. Pfeifer et al. (2015) choose 66% of the models must agree on the sign to pass the test. This threshold is adapted for this thesis. The NUKLEUS ensemble consists out of 9 ensemble member which lead to 6 out of 9 members that have to agree on the sign of change to pass the first test.

Significance test

The second test evaluates the significance of the projected climate signals for each ensemble member by applying the Mann-Whitney-Wilcoxon test (Wilcoxon, 1945; Mann and Whitney, 1947; Wilks, 2011). The test is based on a shift of ranks. First, the historical and GWL data are combined into a single dataset, and ranked across all values. After ranking the combined dataset, the data is split back into historical and GWL. Then the U-statistic is calculated based on the sum of ranks of each dataset. The U-statistic shows whether there is a significant shift of ranks based on the chosen significance level. If the ranks of the GWL dataset are shifted towards higher (or lower) ranks compared to the historical dataset, this results in a positive (or negative) climate change signal. If the ranks are equally distributed over both periods, there is no significant change. The climate signal maps consider only one direction of change. Therefore, it is only tested in one direction, so a one-sided test is applied. The Mann-Whitney-Wilcoxon test is a non-parametric statistical test, meaning it does not assume an underlying distribution of the data. This is particularly important for variables or indices, like precipitation, that tend to be skewed.

One key assumption of the Mann-Whitney-Wilcoxon test is that both datasets are independent, meaning that individual values within a dataset should not influence each other. To ensure this requirement, autocorrelation was tested using the autocorrelation coefficient. The coefficient is calculated with the Numpy Python package. The results are displayed in Table B.1 and B.2. For each variable and index, the 5th to 95th percentile range of the autocorrelation coefficients is shown, representing the 90% of the values. The mean autocorrelation across regions, models, and periods is relatively small, below 0.1. Most of the values are relatively small in the range of -0.3 to 0.3, which is still within an acceptable range of values (Pfeifer et al., 2015). Some regions exhibit higher autocorrelation for some periods and indices. Given the limited number of regions with high autocorrelation, the significance test remains valid. For the few regions with high autocorrelation, the results should be interpreted with caution.

The significance test is applied to each region, model, and GWL separately. Following the previous test, a region passes if a certain fraction of ensemble members agree on a significant change. As in Pfeifer et al. (2015), a model fraction of 66% and a significance level of 0.15 were applied. The significance level can be adjusted depending on the application. To compare the results with the findings of Pfeifer et al. (2015), the same significance level is applied here.

3.3 Integrating ensemble spread information to the climate signal maps

Climate signal maps, introduced by Pfeifer et al. (2015), communicate the climate signals along with their robustness. In this thesis, the climate signal map approach by Pfeifer et al. (2015) is extended by an additional layer of information on the ensemble spread. To achieve this, the ensemble needs to be categorized into large and small spreads, similar to the classification of the climate signals into robust and non-robust. Here, the standard deviation is used as a measure of the ensemble spread. The next step is to define a suitable threshold for the standard deviation that categorizes large and small ensemble spreads.

To find such a threshold, several factors need to be considered. First, the standard deviation depends on the unit of the variable and the magnitude of the signal. Therefore, a threshold must be defined for each variable and index individually, and the unit of the variable needs to be considered. Second, to be able to compare across different regions, the threshold must be set consistently for all regions. In simple words, the goal is to find a practical threshold definition that applies consistently to all regions and indices evaluated.

Here, two approaches to defining the threshold are proposed, depending on whether the index is precipitation-based or temperature-based. For precipitation-based indices, a threshold based on a coefficient of variation is suggested. The coefficient describes the ratio between the standard deviation and the mean. It is computed by

$$CV = \frac{\sigma}{\overline{\mu}} = 30\% \tag{3.1}$$

with the standard deviation σ of the signal among the ensemble members and the mean signal $\overline{\mu}$ of all regions with robust signals. The coefficient is set to 30 %. This results in a threshold for the standard deviation of 30 % of the mean signal of all robust signals. Only regions with robust signals are considered for the threshold to ensure that outliers are excluded, especially regions that show signals in the opposite direction. For indices that show less than 30 regions with robust signals, the mean signal of all regions is used instead to ensure a sufficient sample size and reduce the impact of individual regions' signals on the threshold.

For temperature-based indices, a different approach to setting the threshold is applied. Similar to the definition of the categories for the climate signal maps, absolute thresholds are used. The suggested thresholds are based on the limits set for the three different categories in the climate

Table 3.3: Suggested thresholds for the standard deviation of temperature-based indices indicating a small ensemble spread. These thresholds are based on the limits used in the climate signal maps.

Limits	Threshold	Indices
5 to 10	2	TR
10 to 25	5	ID
15 to 30	5	WSDI

signal maps. Depending on the interval width between the two limits, the thresholds are selected accordingly. For instance, for indices with limits of 5 and 10 days, the threshold for the standard deviation is set to 2 days, while for limits that are further apart, larger thresholds are suggested. Simply put, the larger the span between the two limits, the larger the accepted standard deviation. Table 3.3 lists the suggested thresholds based on the limits in the climate signal maps.

The thresholds for the standard deviation among the ensemble members are applied for all regions and indices. Regions with a standard deviation below the threshold are marked with hatching to indicate a small ensemble spread. Regions with a large ensemble spread are left unhatched.

4 Results

This thesis aims to investigate climate change signals of extreme indices over Germany. To evaluate the climate change signals of extreme indices and their robustness, it is also important to examine the underlying physical variables from which these indices are derived. Therefore, this chapter is divided into two parts. The first part describes the climate change signals of seasonal precipitation and temperature, focusing on winter and summer seasons. Particular emphasis is placed on the robustness of the precipitation signals and factors affecting the number of regions with robust signals. The second part presents the climate change signals of a representative selection of the ETCCDI indices, divided into precipitation- and temperature-based indices. In addition to robustness, the ensemble spread of the climate change signals is evaluated, which allows for a further distinction of robust and non-robust climate change signals (see Figure 1.1).

There is one important general remark regarding the data used to obtain the results: At the time of conducting the analyses presented here, no bias-corrected dataset based on the NUKLEUS ensemble was available. For further details on the expected effects of performing a bias correction on the results presented here, see Section 5.

4.1 Climate change signals of seasonal precipitation and temperature

This section presents the seasonal climate change signals of the physical variables, precipitation and temperature, and their robustness, focusing on the winter and summer seasons. Summer is defined as June, July, and August (JJA), and winter as December, January, and February (DJF).

4.1.1 Climate change signals of winter and summer precipitation

In the following, the climate signal map method is applied to winter and summer monthly mean precipitation data obtained from the NUKLEUS ensemble. The upper row of the figures shows the climate signal maps with the ensemble median signals relative to the historical period. For each region, the ensemble median of the spatial mean signal is shown. The signals are divided into three categories: small signals below 10 % (yellow), medium signals from 10 % to 15 % (orange), and large signals above 15 % (red). Only one direction of change is considered: an increase in monthly mean precipitation is assessed for winter, and a decrease for summer. Climate change signals in the opposite direction are shown in white. The left panel shows the climate change signals for GWL2K,

and the right panel shows the signals for GWL3K. The bottom row of the figure shows the same climate signal maps as the upper row, but with the robustness test applied to each region. The applied robustness test (see Section 3.2.3) is illustrated in the Appendix Section A.1. Regions with robust climate change signals are shown in color. Non-robust climate change signals are shown in gray.
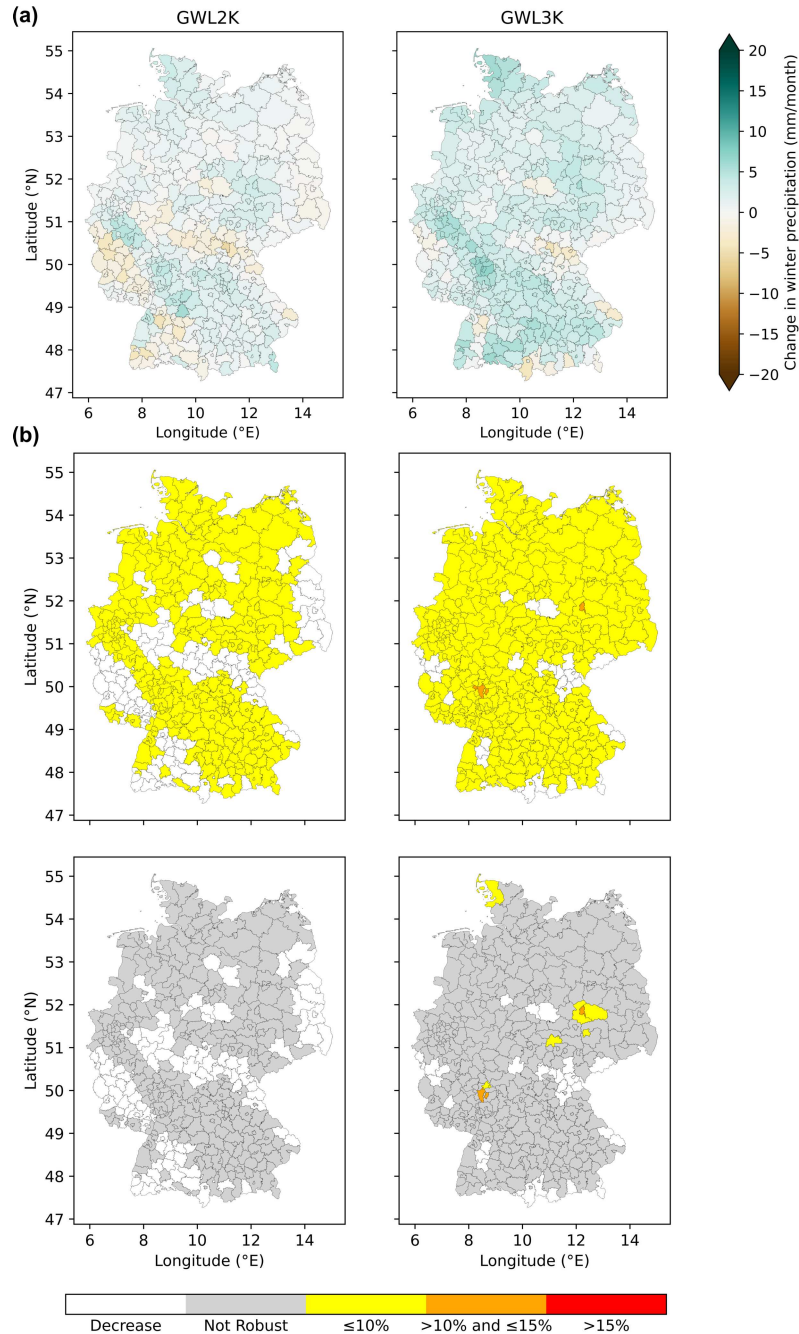

Figure 4.1: Climate change signals in winter precipitation for GWL2K (left panels) and GWL3K (right panels). (a) shows the absolute median climate change signals in mm/month. Blue colors indicate an increase in winter precipitation, while brown colors indicate a decrease. (b) shows the climate signal maps for an increase in winter precipitation. Median climate change signals are displayed divided into 3 categories, representing low (yellow), medium (orange), and high (red) relative changes. White areas indicate a decrease. The upper row shows the median climate change signals for all regions, while the bottom row shows only regions with robust climate change signals in color and non-robust in gray.

Figure 4.1a displays the median climate change signals for monthly mean winter precipitation. For GWL2K (left panel), most of the signals are relatively small. There are several regions that show a decrease in winter precipitation, particularly in the central and southwestern parts of Germany. For GWL3K (right panel), the majority of regions show an increase in winter precipitation, and the signals are generally stronger than for GWL2K. A small subset of regions shows a decrease in winter precipitation for GWL3K. The regions that show a decrease for GWL3K mostly overlap with those showing a decrease for GWL2K.

Figure 4.1b shows the climate signal maps for an increase in winter precipitation. In the upper row of the figure, the ensemble shows small increases in winter precipitation for both GWLs. For GWL2K, most regions show a small increase in winter precipitation, which aligns with the absolute values in Figure 4.1a. For GWL3K, the vast majority of the regions show a small relative change in winter precipitation, with a few regions showing medium climate change signals. Although the absolute signals in Figure 4.1a are stronger for GWL3K compared to GWL2K, these are still largely categorized as small relative climate change signals in the climate signal maps (Figure 4.1b).

In the bottom row of Figure 4.1b, the robustness test is applied to the climate signal maps from the upper row. For GWL2K, there is no region with a robust climate change signal. For GWL3K, a few regions show a robust climate change signal, including the regions with medium signal strength. There is no distinct spatial pattern with regard to which regions show a robust climate change signal for GWL3K. However, nearly all regions showing medium signals are classified as robust. Still, there are too few regions with a robust climate change signal to approve this trend.

Figure 4.2a shows the median climate change signals for monthly mean summer precipitation. For GWL2K, there is a decrease in monthly mean summer precipitation in the western part of Germany. In the eastern part of Germany, several regions show an increase in monthly mean summer precipitation for GWL2K. For GWL3K, almost all regions show a decrease in summer precipitation with stronger signals than for GWL2K. Regions in the South of Germany show weak climate change signals for GWL3K, with several regions showing an increase in summer precipitation. The strongest drying signals are located in the northwestern part of Germany for both GWLs.

Figure 4.2b shows the climate signal maps for a decrease in summer precipitation. For GWL2K, most regions show a small relative climate change signal, except for regions in the Northwest that show medium to large signals. In the eastern part of Germany, several regions show an increase in monthly mean summer precipitation for GWL2K. For GWL3K, almost all regions show a decrease in summer precipitation with small relative signals. The drying signals in the Northwest intensify, and more regions show medium to large relative climate change signals than for GWL2K.

The climate signal maps show relative climate change signals. This can be seen for the region in the South of Germany, with a very strong absolute climate change signal for GWL3K (Figure 4.2a, right panel). Although the absolute signal is very strong and stands out compared to the surrounding regions, it is still classified as a small relative signal in the climate signal maps (Figure
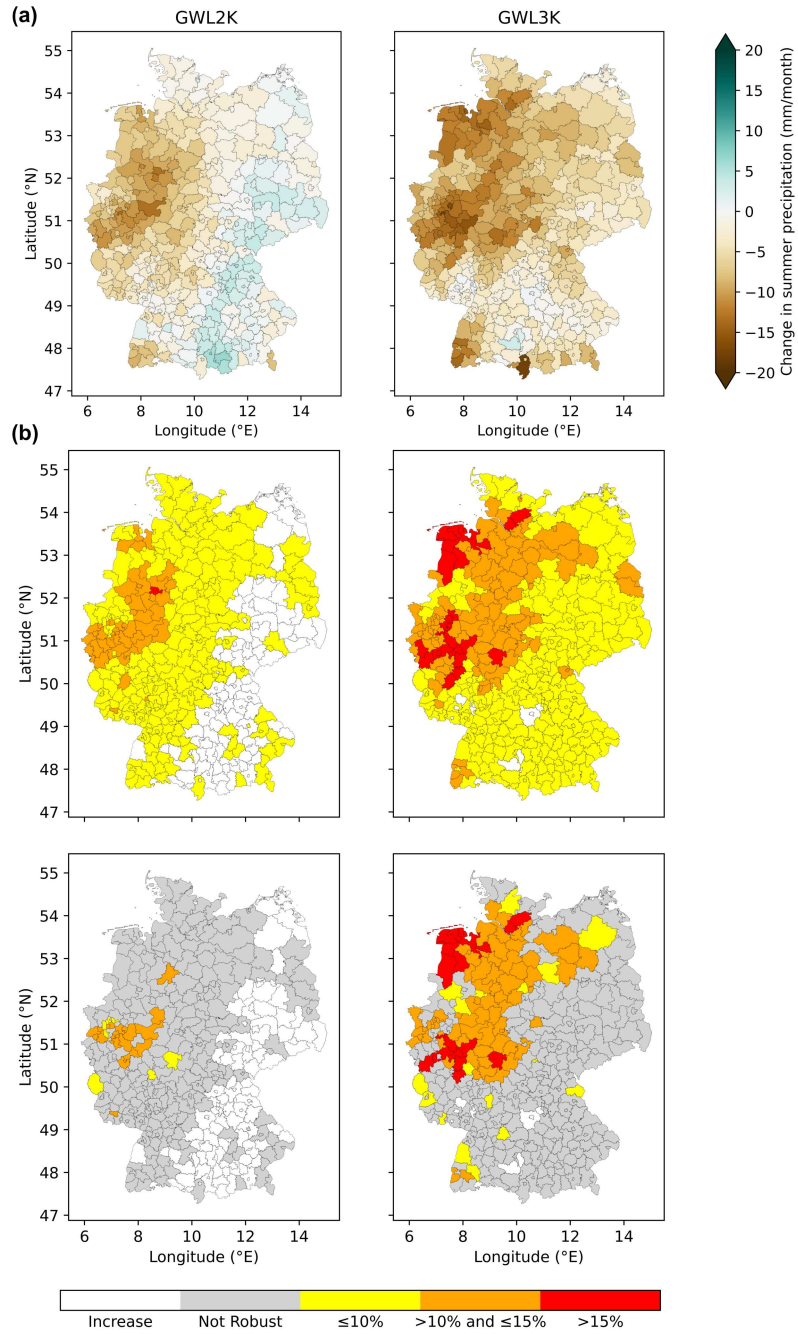
Figure 4.2: Same as Figure 4.1, but for summer precipitation. (a) shows the absolute median climate change signals in mm/month. (b) shows the climate signal maps for a decrease in summer precipitation.

4.2b, right panel). This underlines the strength of the climate signal map method to evaluate the climate change signal relative to the typical climatological conditions of an individual region, rather than regarding the absolute signals alone.

Regarding the robustness of a decrease in summer precipitation in the bottom row, GWL2K shows some regions with robust climate change signals located in the Northwest of Germany. For GWL3K, there is a higher number of regions with robust signals located in a large coherent area in the Northwest of Germany. The regions that show robust signals for GWL2K also show robust signals for GWL3K. Medium to large climate change signals tend to be more robust than small signals. This is expected, as a stronger median climate change signal indicates that individual ensemble members show stronger signals and stronger climate change signals are more likely to be statistically significant relative to the reference period.

Overall, the climate change signals for an increase in winter precipitation show small climate change signals for both GWLs and low robustness. For summer precipitation, there are diverging climate change signals for GWL2K, with a decrease in the West and an increase in the East of Germany. For GWL3K, most regions show a decrease in summer precipitation, with the largest signals in the Northwest of Germany. Both GWLs show more climate change signals classified as robust for a decrease in summer precipitation than for an increase in winter precipitation.

Why are there so few regions with robust climate change signals?

The climate signal maps for both winter and summer show a small number of regions with robust climate change signals, except for GWL3K in summer. This raises the question of why there are so few regions with robust signals. The robustness of a climate change signal is indicated by the degree of agreement of the ensemble members on the sign of the climate change signal and its significance. To pass the robustness test, 6 out of 9 models must agree on the sign and the significance of the climate change signal.

The first criterion, the agreement on the sign of the climate change signal, can be investigated based on the absolute changes in mean precipitation for each ensemble member and GWL. This is shown in Figure 4.3 for winter, and Figure 4.6 for summer. Increases in precipitation are shown in blue colors and decreases in brown colors.

The second criterion is the agreement of the ensemble on the significance of the climate change signal. The significance mask of the climate change signals for each ensemble member and GWL is shown in Figure 4.4 for an increase in winter precipitation and in Figure 4.7 for a decrease in summer precipitation. The red regions indicate regions with statistically significant climate change signals in the specified direction relative to the reference period. Note that a one-sided significance test is applied. This means that only significant climate change signals of the specified direction of change are considered significant. Significant signals in the opposite direction are still classified as non-significant.
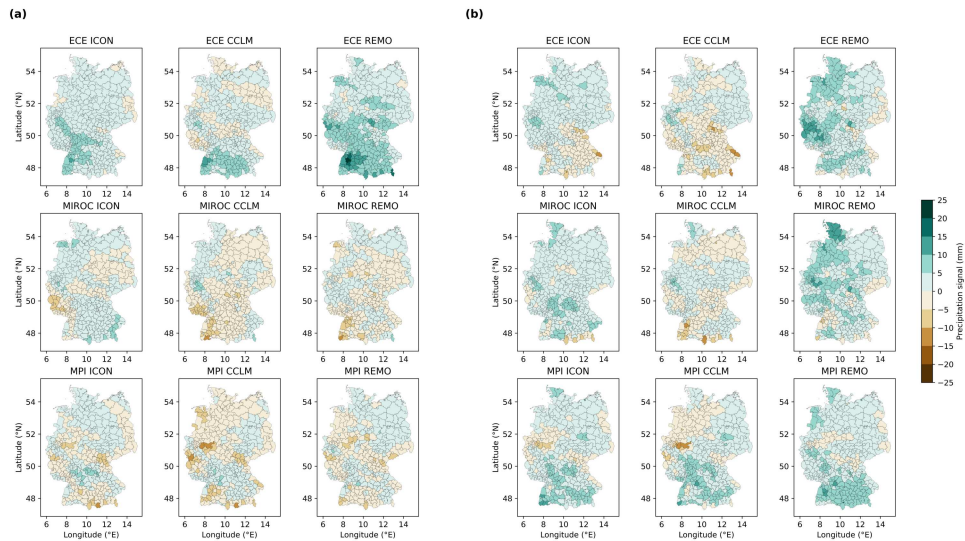
Figure 4.3: Climate change signals for monthly mean winter precipitation for each ensemble member of the NUKLEUS ensemble. The changes are shown in mm/month. Each row corresponds to a GCM, and each column corresponds to a CPM of the NUKLEUS ensemble.
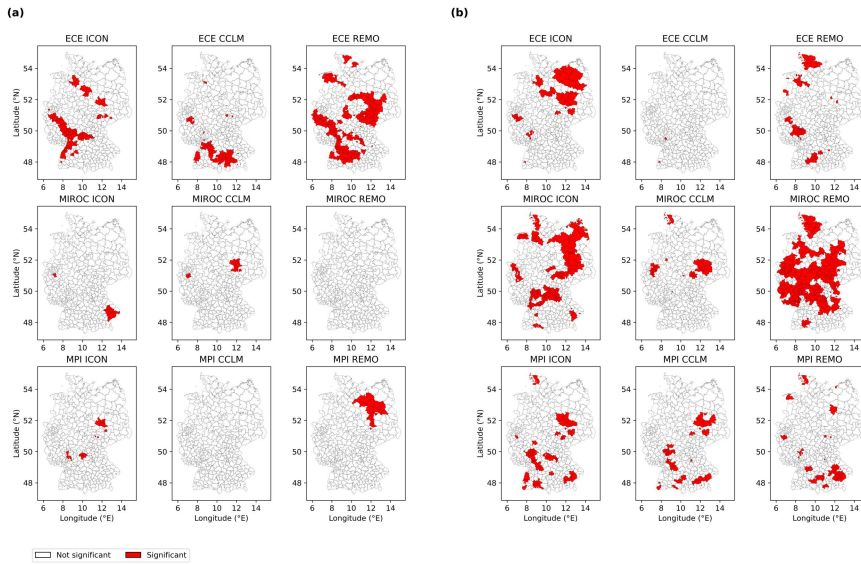

Figure 4.4: Significance mask for an increase in monthly mean winter precipitation for each ensemble member of the NUKLEUS ensemble. Red regions indicate regions with significant climate change signals. Each row corresponds to a GCM, and each column corresponds to a CPM of the NUKLEUS ensemble. A one-sided significance test is applied at a significance level of 0.15.

Regarding the climate change signals of the individual models for GWL2K in winter (Figure 4.3 a), the individual models disagree in the sign and strength of the signals among different regions. Some models indicate more regions with wetting signals. Other models indicate more regions with drying signals. There are also models that show a balanced number of regions with drying and wetting signals. For GWL3K (Figure 4.3 b), there is a higher fraction of regions with wetting signals compared to GWL2K, especially for the models driven by the GCMs MIROC and MPI (middle and lower row).

Figure 4.5 shows the spatial mean signal over the whole model domain for each model, season, and GWL. The same colors are applied as in Figures 4.3 and 4.6. For GWL2K, most of the models show a weak drying spatial mean signal in winter precipitation. For GWL3K, all models show a spatial mean wetting signal for winter precipitation. This indicates that there is a trend to more regions with wetting signals among all models and a higher agreement on an increase in winter precipitation for GWL3K.
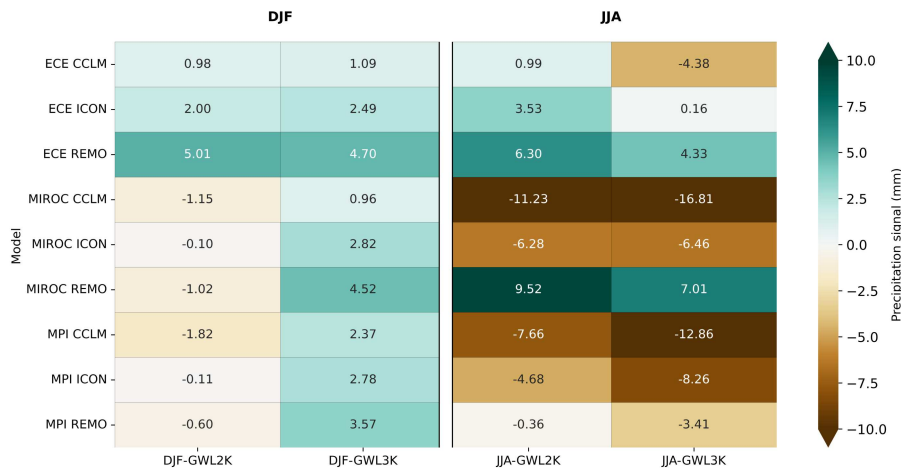

Figure 4.5: Spatial mean of the monthly mean precipitation signal over the whole model domain in winter (DJF) and summer (JJA), shown for each ensemble member of the NUKLEUS ensemble and each GWL. The climate change signals are given in mm/month.

Regarding the second criterion for robustness, the agreement on the significance of an increase in winter precipitation (Figure 4.4), only a few regions per model show significant climate change signals for GWL2K, with MPI-ICON and MIROC-REMO showing no regions with significant climate change signals. For GWL3K, there are in total more regions that show significant climate change signals. However, when individual models are considered, regions that show significant signals for GWL3K do not correspond to those regions that show significant signals for GWL2K.

For example, for GWL2K, MPI-REMO (Figure 4.4a, bottom row, right plot) shows a few regions with significant climate change signals in the Northeast of Germany, while for GWL3K (Figure 4.4b), regions with significant climate change signals are distributed across Germany. This illustrates the lack of spatial consistency in the significance of the climate change signals across the two GWLs. In winter, the sets of regions passing the significance test for GWL2K and GWL3K barely overlap, despite an increase in the number of significant regions for GWL3K. This suggests that the climate change signals remain weak compared to the median signals for the historical period, so that most signals stay close to or fall below the threshold of significance. As a result, the spatial pattern of significant changes varies between the two warming levels, even if the overall number of significant signals increases.

Consequently, in winter, the key limiting factor for the robustness of the climate change signals is the lack of significance, particularly for GWL3K. Although many regions show a consistent wetting signal among the models, only a few regions show significant climate change signals. This

results in a low number of robust climate change signals in the climate signal maps for winter (Figure 4.1).

Regarding the climate change signals for GWL2K in summer (Figure 4.6 a), the individual models show very different fractions of wetting and drying signals, similar to the signals for GWL2K in winter. There is a higher fraction of drying signals for GWL3K compared to GWL2K in summer, except for the model MIROC-REMO (middle row, right plot). However, there is a large disagreement among the models in the direction of change within individual regions.

The spatial mean signals of the model domain (Figure 4.5) show that the differences between the individual models are more pronounced in summer compared to winter and that the spatial mean signals are stronger in summer than in winter, for both GWLs. For instance, for GWL2K in summer, MIROC-CCLM, MIROC-ICON, MPI-CCLM, and MPI-ICON show a strong decrease in precipitation in the spatial mean signal, whereas MIROC-REMO, ECE-REMO, ECE-CCLM, and ECE-ICON show an increase. These strong, but opposing, spatial mean climate change signals show that the ensemble does not agree on the direction of the climate change signals in summer. The agreement on the direction of the signal is higher towards GWL3K in summer, as most models show a higher fraction of drying signals here.

The significance of the climate change signals for a decrease in summer precipitation (Figure 4.7) shows very different fractions of regions with significant climate change signals. For GWL2K (Figure 4.7a), MIROC-ICON, MIROC-CCLM, MPI-ICON, and MPI-CCLM show significant climate change signals for most of the regions, while other models show only a few to no regions with significant signals. For GWL3K (Figure 4.7b), the total number of regions with significant signals per model is higher compared to GWL2K. Especially, MIROC-CCLM, MPI-CCLM, and MPI-ICON show significant signals across nearly all regions. The higher number of significant climate change signals for GWL3K aligns with the larger number of regions with robust signals in the climate signal map for GWL3K (Figure 4.2). In contrast to winter, there are already many regions with significant climate change signals for GWL2K in summer (Figure 4.7a), and the regions with significant signals for GWL2K in summer mostly match those for GWL3K (Figure 4.7b). Along with the generally higher signal strengths in summer, this indicates that there is a consistent trend towards drying signals in summer precipitation that intensifies for higher GWLs.

However, there are still very different patterns regarding the significance of the climate change signals in summer precipitation among the models. This is due to the diverging signs of the climate change signals shown in Figure 4.6. Some models show a strong decrease in summer precipitation, while others show mostly wetting signals. As a one-sided significance test is applied, regions with a significant increase in summer precipitation are still classified as non-significant, leading to barely any significant signals for models with mostly wetting signals like MIROC-REMO.

In general, the climate change signals in summer are more significant, especially for GWL3K, as the signals intensify. This is also reflected by the higher number of regions with robust signals and stronger ensemble median signals for summer, shown in Figure 4.2. Consequently, for summer, the
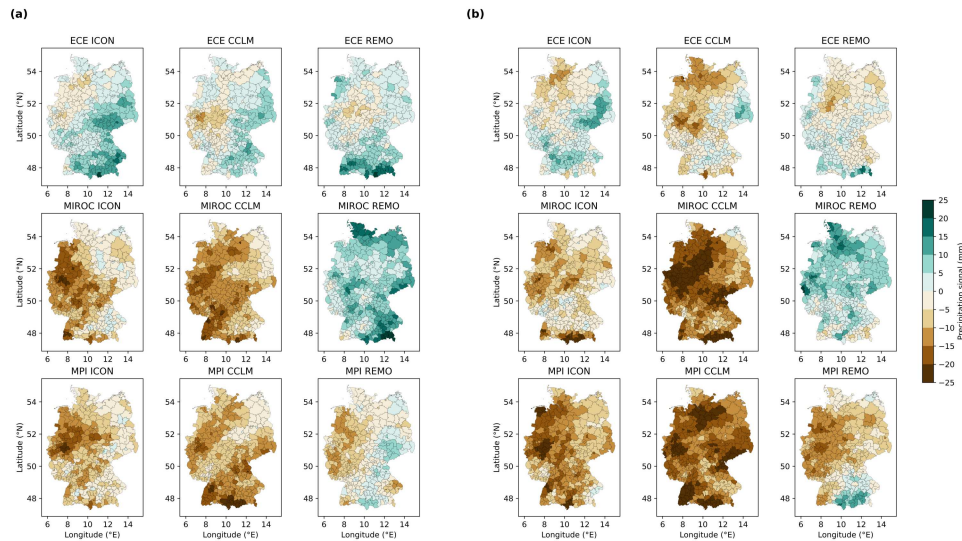
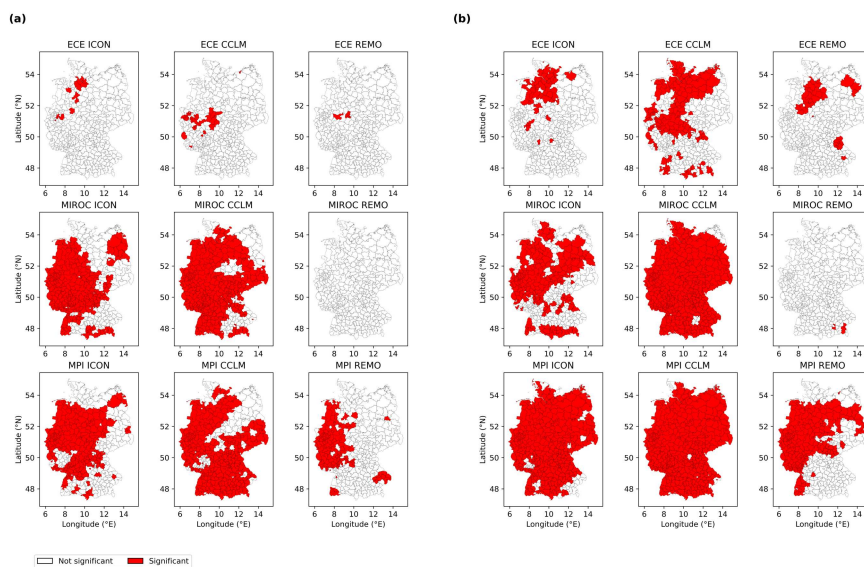Figure 4.6: Same as Figure 4.3, but for summer precipitation.

Figure 4.7: Same as Figure 4.3, but for a decrease summer precipitation.

factor limiting the number of regions with robust signals is the diverging sign of the climate change signals among the models.

Overall, both winter and summer show very different fractions of wetting or drying signals among the models, especially for GWL2K. These diverging responses of the models for GWL2K are due to high internal variability. The periods corresponding to GWL2K start in the years 2024, 2037, and 2038, depending on the driving GCM (see Table 3.1). GWL3K represents later time slices. This highlights that GWL2K represents near-future projections of the climate. Therefore, the models show diverging climate change signals for GWL2K in both seasons, as for upcoming decades, projections of precipitation signals show high uncertainties (Seaby et al., 2013; Pfeifer et al., 2015). The low agreement in the sign of the precipitation signals for near-future scenarios among different models is due to high internal variability and model uncertainty (Seaby et al., 2013; Evin et al., 2021; Hawkins and Sutton, 2009). These factors combine in a high uncertainty for near-future

precipitation projections, which is reflected in the high number of regions that show non-robust signals in the climate signal maps for GWL2K in summer and winter (Figures 4.11-4.12).

Regarding GWL3K, there is a higher fraction of wetting signals in winter and drying signals in summer. This higher agreement on the sign of the climate change signals among individual models is further underlined by the fewer white regions in the climate signal maps for GWL3K in summer and winter (Figures 4.11-4.12).

The key factor limiting the number of regions with robust signals in the climate signal map for winter is the lack of significance. Although more regions agree on an increase in winter precipitation for GWL3K, the second criterion for robustness, which is the agreement on the significance of the climate change signals, is not met. This is reflected in the climate signal map for winter precipitation (Figure 4.1), where barely any region is classified as robust. The lack of significance is due to the small signal strength in winter for both GWLs. As a result, the signals in winter precipitation mostly fall within the range of internal variability and are not statistically distinguishable from the reference period.

In contrast, the summer precipitation signals are stronger than in winter and intensify towards GWL3K, leading to more significant climate change signals. Here, the key factor limiting the number of regions with robust signals in the climate signal maps is the lack of agreement on the sign of the climate change signals among the models. Although the signals of a decrease in summer precipitation are significant for both GWLs, some models indicate an increase in summer precipitation, especially for regions in the Southeast of Germany. This is also reflected in the climate signal maps for summer (Figure 4.2) as most regions with robust signals are located in the Northwest of Germany, while for GWL2K, the regions in the Southeast indicate a median increase in summer precipitation and thus are shown in white.

A major factor driving the diverging climate responses in summer precipitation among the models is the variation in how physical processes are represented in the CPMs (Tölle et al., 2018). The CPMs differ in their parameterizations, dynamical cores, and process implementations. These differences in the model physics result in different precipitation responses. The large spread in the climate change signals of summer precipitation across different models is consistent with the findings of Coppola et al. (2021), who assessed seasonal precipitation responses using a multi-model RCM ensemble.

Overall, based on the NUKLEUS ensemble evaluated here, there is a small number of robust climate change signals for an increase in winter precipitation. This is mainly due to the lack of statistical significance of the signals, as for GWL3K, the models show high agreement on the sign of the signals of individual regions. In contrast, for summer, although the models show a high number of significant signals for both GWLs, there is less agreement among the models on the sign of the signals. Thus, the limiting factor for the robustness of the signals in summer is the agreement on the sign of the climate change signals. Although there are more regions with a robust signal in summer, the models show diverging patterns in the sign of the signals, which indicates a large uncertainty in the projected signals among the models.

Comparison to Pfeifer et al. (2015)

The climate signal map method was initially introduced by Pfeifer et al. (2015). In the following, the climate signal maps for winter and summer precipitation signals (Figures 4.11-4.12) derived from the NUKLEUS ensemble are compared to the findings by Pfeifer et al. (2015). They analyzed three different regional climate model ensembles applying the climate signal map method, which is the same applied here. Figure 4.8 shows their results for an increase in winter precipitation for an ensemble of 10 simulations based on the representative concentration pathway (RCP) 4.5 by the EURO-CORDEX initiative. Accordingly, Figure 4.9 shows their results for a decrease in summer precipitation for the same ensemble and RCP. For more details on the simulations included in the ensemble, see Pfeifer et al. (2015). Instead of GWLs, they used four 30-year time slices, starting at the years 2031, 2041, 2051, and 2061. The reference period used was 1971 to 2000. The color scheme of the climate signal maps used by Pfeifer et al. (2015) is the same as used in this work, with one exception that in this work, yellow instead of green is used to indicate small signals.

For winter (Figure 4.8), the results of Pfeifer et al. (2015) show an increase in winter precipitation for all regions, with the strongest signals in the South of Germany. Regions with robust climate change signals are mainly located in the South and Central Germany. In summer (Figure 4.9), the climate change signals for a decrease in summer precipitation are small. Many regions are showing a decrease in summer precipitation. For the first two periods, these regions are mainly located in the North of Germany. For the last two periods, they are located in the East and South. Hardly any region shows robust climate change signals among all periods.

Comparing the results of Pfeifer et al. (2015) to the results of the NUKLEUS ensemble presented in Figures 4.1 and 4.2, there are two major differences. First, the NUKLEUS ensemble shows stronger climate change signals in summer than in winter. In contrast, the findings from Pfeifer et al. (2015) show an inverse pattern with the strongest signals in winter. Second, in the NUKLEUS ensemble, the highest amount of regions with a robust climate change signal is shown for a decrease in summer precipitation, while Pfeifer et al. (2015) found most robust signals for an increase in winter precipitation.

There are several potential reasons for the differences between the results presented here and those obtained by Pfeifer et al. (2015). First, the RCM ensemble used by Pfeifer et al. (2015) has a much coarser resolution of approximately 12 km while the NUKLEUS ensemble is a CPM model ensemble with a resolution of approximately 3 km. As a consequence, deep convection is parameterized in the ensemble used by Pfeifer et al. (2015), while it is resolved explicitly in the NUKLEUS ensemble. Second, the ensembles are based on different RCMs and GCMs. Third, the ensemble used by Pfeifer et al. (2015) is based on transient simulation and considers a total of four 30-year time slices spanning from 2031 to 2090, whereas the NUKLEUS ensemble is based on two GWLs with the periods given for each model in Table 3.1. Lastly, the applied reference periods differ, with Pfeifer et al. (2015) using 1971 to 2000, while the reference period applied here is 1961 to 1990. Section 5 will discuss which of these factors contributes most to the differences between the results presented here and those obtained by Pfeifer et al. (2015).

2031 - 2060   2041 - 2070   2051 - 2080   2061 - 2090

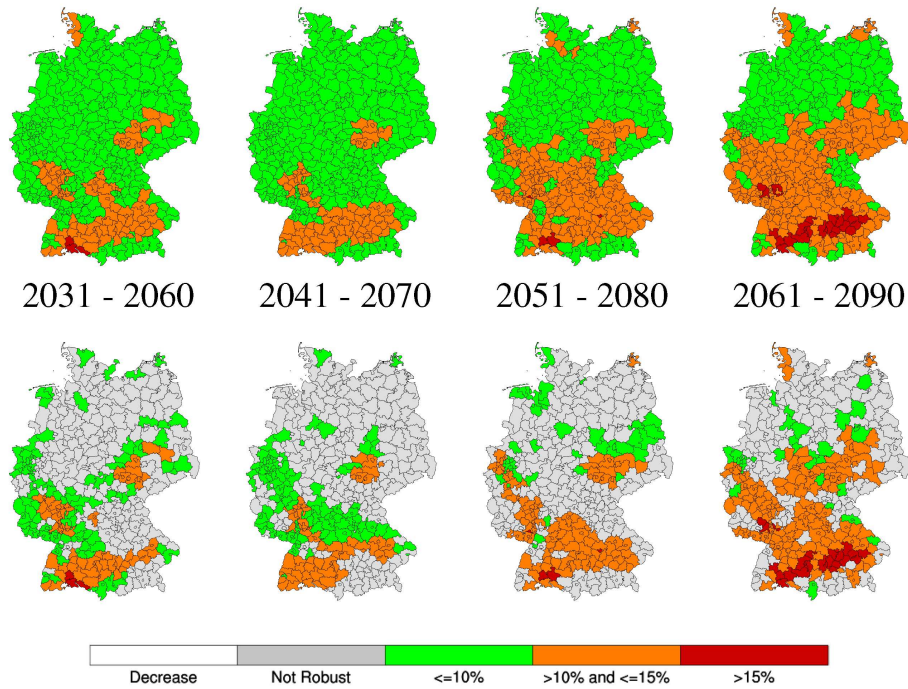| Decrease | Not Robust | <=10% | >10% and <=15% | >15% |

Figure 4.8: Climate signal maps for an increase in winter precipitation, from Pfeifer et al. (2015). The median climate change signals based on 10 RCM simulations from the EURO-CORDEX RCP4.5 simulations are shown for the periods 2031–2060, 2041–2070, 2051–2080, and 2061–2090. The reference period is 1971–2000. The upper row shows the median climate change signals in percentage. The climate change signals are displayed divided into 3 categories, representing low (green), medium (orange), and high (red) relative changes. White areas indicate a decrease. For the bottom row, the robustness test was applied, and only robust signals are shown in color. Non-robust signals are shown in gray.
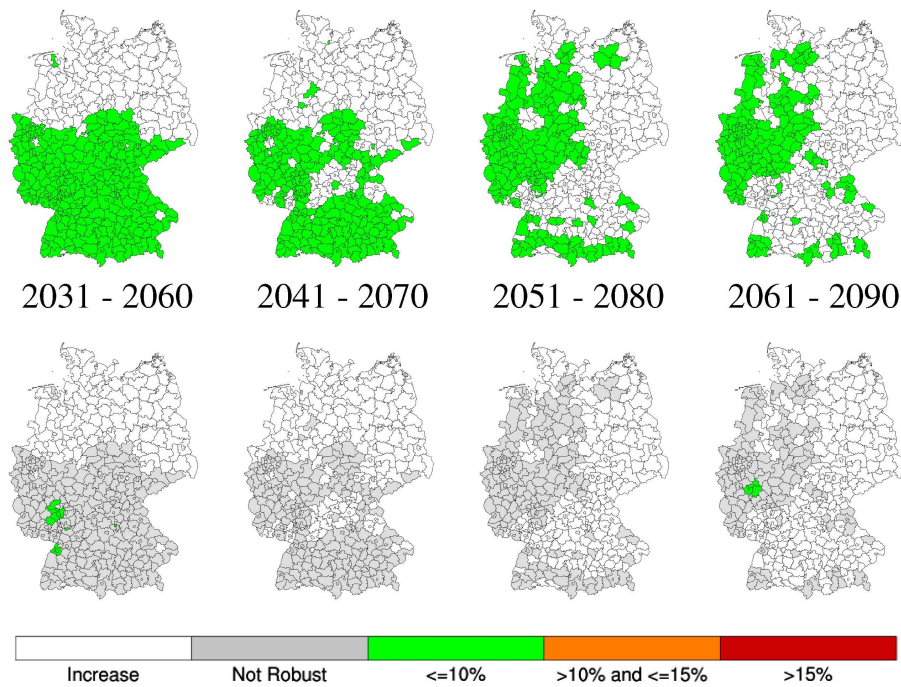

2031 - 2060   2041 - 2070   2051 - 2080   2061 - 2090

| Increase | Not Robust | <=10% | >10% and <=15% | >15% |

Figure 4.9: Same as Figure 4.8, but for a decrease in summer precipitation, from Pfeifer et al. (2015).

4.1.2 Climate change signals of winter and summer temperature

The ETCCDI indices are based on precipitation and near-surface air temperature at 2 m. The previous section presented the climate change signals for winter and summer monthly mean precipitation. The following section focuses on climate change signals in monthly mean winter and summer temperatures based on the data obtained from the NUKLEUS ensemble.

Figure 4.10 shows the ensemble median temperature signals for winter in the upper row and summer in the bottom row. In general, the climate change signals for GWL3K are more intense compared to GWL2K for both seasons, and the signals for summer are stronger than for winter. In winter, for both GWLs, the smallest signals are shown in the West of Germany, while the strongest signals are shown in the Northeast. In summer, for GWL2K, regions in Central and Eastern Germany show the strongest signals. For GWL3K, regions in the South and Northeast of Germany show distinct strong signals. However, the median absolute climate change signals show small ranges within a selected GWL and season. For instance, the summer temperature signals range for GWL2K between 1.83 °C and 2.13 °C, and for GWL3K between 3.02 °C and 3.65 °C. This highlights that the temperature differences in mean seasonal temperatures are less pronounced among the regions or seasons, and the strongest differences are found between the two GWLs.

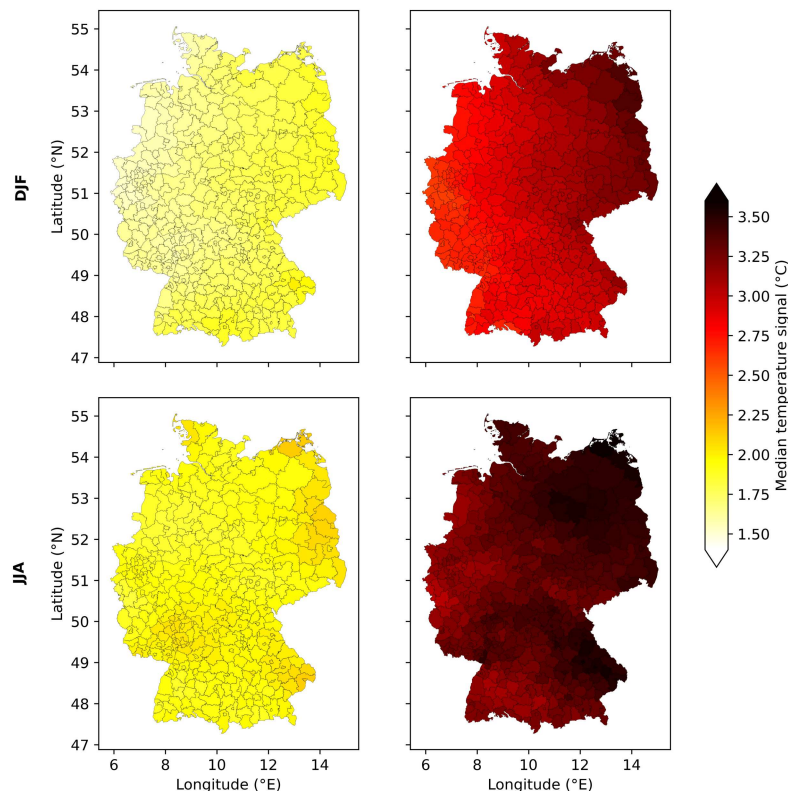Figure 4.10: Ensemble median climate change signals for winter (DJF, upper row) and summer (JJA, bottom row) temperature in °C. The left panel shows the results for GWL2K, and the right panel shows the results for GWL3K.

Figure 4.11 shows the climate signal map for winter temperature, and Figure 4.12 for summer temperature. For both seasons, an increase in monthly mean temperature is assessed. All temperature

signals are robust for both seasons and GWLs. Therefore, only one row of climate signal maps is presented, which is analogous to the bottom row of the signal maps presented in Section 4.1.1. Note that the categories for the temperature signals are expressed in absolute values given in °C.
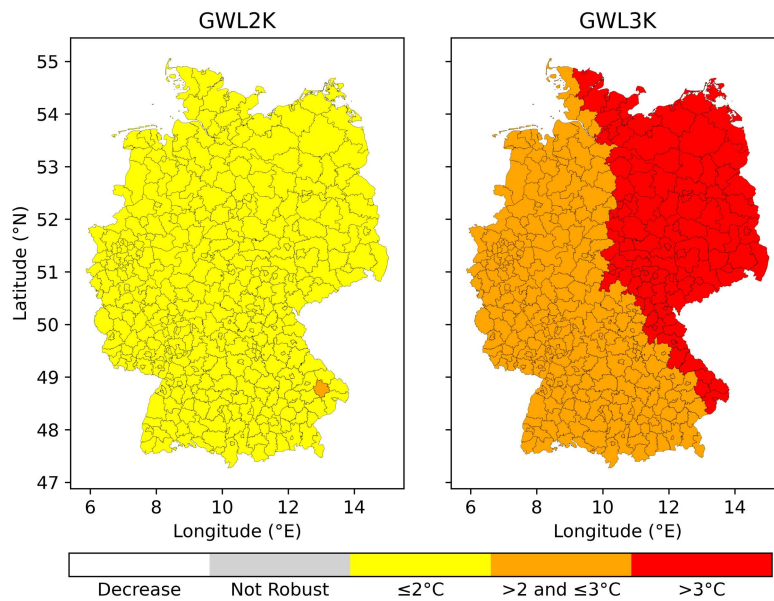
Figure 4.11: Climate signal map for an increase in monthly-mean winter temperature for GWL2K (left panels) and GWL3K (right panels). Median signals are displayed divided into 3 categories, representing low (yellow), medium (orange), and high (red) changes. All climate change signals are robust.
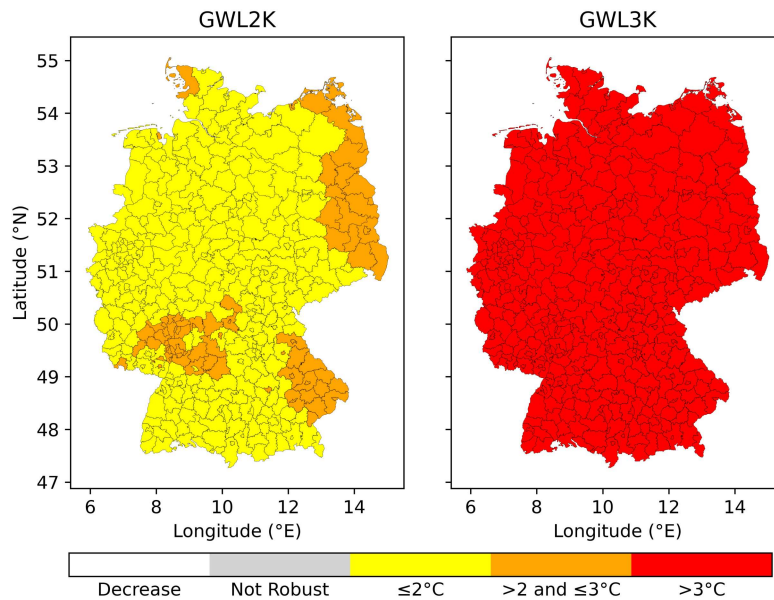
Figure 4.12: Same as Figure 4.11, but for an increase in monthly-mean summer temperature. All signals are robust.

In winter (Figure 4.11), all regions show small climate change signals for GWL2K except for one region in the Southeast of Germany, which shows a medium signal. For GWL3K, the north-eastern part of Germany shows large climate change signals above 3 °C, while the rest of the regions show medium signals. In summer (Figure 4.12), for GWL2K, most of the regions show small climate

change signals. Some regions in Central and Eastern Germany show median signals. For GWL3K, all regions show large climate change signals, with a maximum climate change signal of 3.65 °C. The temperature signals for both seasons differ strongly between the two GWLs. Within each GWL, the climate change signals show little regional variation and no pronounced regional features.

The results presented here show robust signals for both seasons and GWLs. This is consistent with other studies examining the significance and robustness of temperature changes (e.g. Jacob et al., 2014; Kjellström et al., 2018). The strong climate change signals in summer temperature for GWL3K in the Southeast of Germany might be due to the orography. As shown, for example, by Tölle et al. (2018), the orography influences the temperature response of climate simulations. However, this connection between orography and stronger temperature responses is not consistent across all regions, as other regions influenced by orography, such as the Black Forest in the Southwest of Germany, do not show enhanced signals.

Comparing the two representations of the temperature signals from Figures 4.11-4.12 and 4.10 shows that regional differences are less apparent in the climate signal maps due to the application of categories. A representation of the climate change signals without categories, like in Figure 4.10, is more suitable to show regional differences in mean temperature signals.

4.2 Climate change signals of extreme indices

This section presents the climate change signals of a selection of ETCCDI indices, illustrated with climate signal maps. The chapter is divided into temperature and precipitation-based indices. The indices are derived from daily mean precipitation and near-surface air temperature at 2 m data obtained from the NUKLEUS ensemble. In addition to the robustness of the climate change signals, the ensemble spread is evaluated. In contrast to the climate change signals for precipitation and temperature shown in Section 4.1, the climate change signals of the indices are evaluated based on annual changes rather than separated by season.

4.2.1 Climate change signals of temperature-based indices

The following section presents the climate change signals of a selection of temperature-based ETCCDI indices. The considered indices exhibit robust climate change signals for all regions and GWLs. Consequently, only a single row of climate signal maps is shown, with the robustness test applied but not visible due to the absence of non-robust signals. Similar to the climate signal maps of temperature in Section 4.1.2, the categories are given in absolute values. To evaluate the ensemble spread, a threshold is applied for the standard deviation of each region (see Section 3.3 for more details). The threshold applied for temperature-based indices is based on the limits used for the categories of the climate signal maps, given in Table 3.3. To illustrate the relevance of an ensemble spread measure, the first part of this section focuses on the climate change signals of the individual models for selected example regions.

Tropical Nights (TR)

Figure 4.13a shows the median climate change signals for the number of tropical nights per year. All regions show an increase in the number of tropical nights. For GWL2K (left panel), most of the regions show small climate change signals below 5 days, except for the northern Part of the Upper Rhine Valley in Southwest Germany, where the projected changes are stronger. For GWL3K (right panel), the median climate change signals get stronger compared to GWL2K. Strongest signals are found for the Upper Rhine Valley with an increase of more than 10 tropical nights per year. Small regions, so-called city districts ("Stadtkreise"), exhibit particularly large signals. These urban areas show a stronger increase in the number of tropical nights than their surrounding rural regions, consistent with the urban heat island effect (Kabisch et al., 2023). This suggests a stronger increase in the number of tropical nights in city districts than in their surrounding rural regions. The strong increase in tropical nights for the Upper Rhine Valley and urban areas, especially for GWL3K, is consistent with other studies (Pinto et al., 2025; van Rüth et al., 2023). The spatial pattern of an increase in the number of tropical nights aligns with the summer median temperature signals (Figure 4.10, bottom right), as expected since tropical nights are derived from temperature.

Figure 4.13b shows the climate signal map for an increase in the number of tropical nights. All climate change signals are robust. For GWL2K, small increases below 5 days are shown in most regions, except the northern part of the Upper Rhine Valley, where medium signals are shown. For GWL3K, the climate change signals intensify, with large signals in the Upper Rhine Valley, and medium signals in southern and eastern Germany. Here, the city districts also stand out with stronger climate change signals (medium to large) than their surrounding rural regions, as shown for the median signals (Figure 4.13a).

While the robustness measure from Pfeifer et al. (2015) assesses agreement on the sign and significance of the individual climate change signals within the ensemble, it does not reflect agreement on the signal's magnitude. This limitation is illustrated in Figure 4.14. Here, the increase in the number of tropical nights for GWL3K is shown for selected example regions (highlighted in Figure 4.13b, right panel). Figure 4.14 shows for each example region the individual climate change signal of each model of the NUKLEUS ensemble as well as the ensemble median signal. The different levels of the climate signal map are indicated by background colors, with the category thresholds marked by red and orange lines. Regions C, E, and G show a large median ensemble signal, Region F a medium signal, and regions A, B, and D show small median ensemble signals. Below the region labels, the standard deviation is given in days, indicating the level of agreement among ensemble members. Table 4.1 presents the corresponding values for each region and model.

All climate change signals of the individual models presented in Figure 4.14 are significant, and the ensemble median signals are robust. However, while the models agree well in some regions, others show considerable differences in the magnitude of the projected climate change signals.

Regions A and B show low variability across models, as reflected by the small standard deviations. For example, the signals for region A range from 1.3 to 5.9 days, with a median of 3.9 days, indicating good agreement on the magnitude of the signal. In contrast, regions C to G exhibit larger
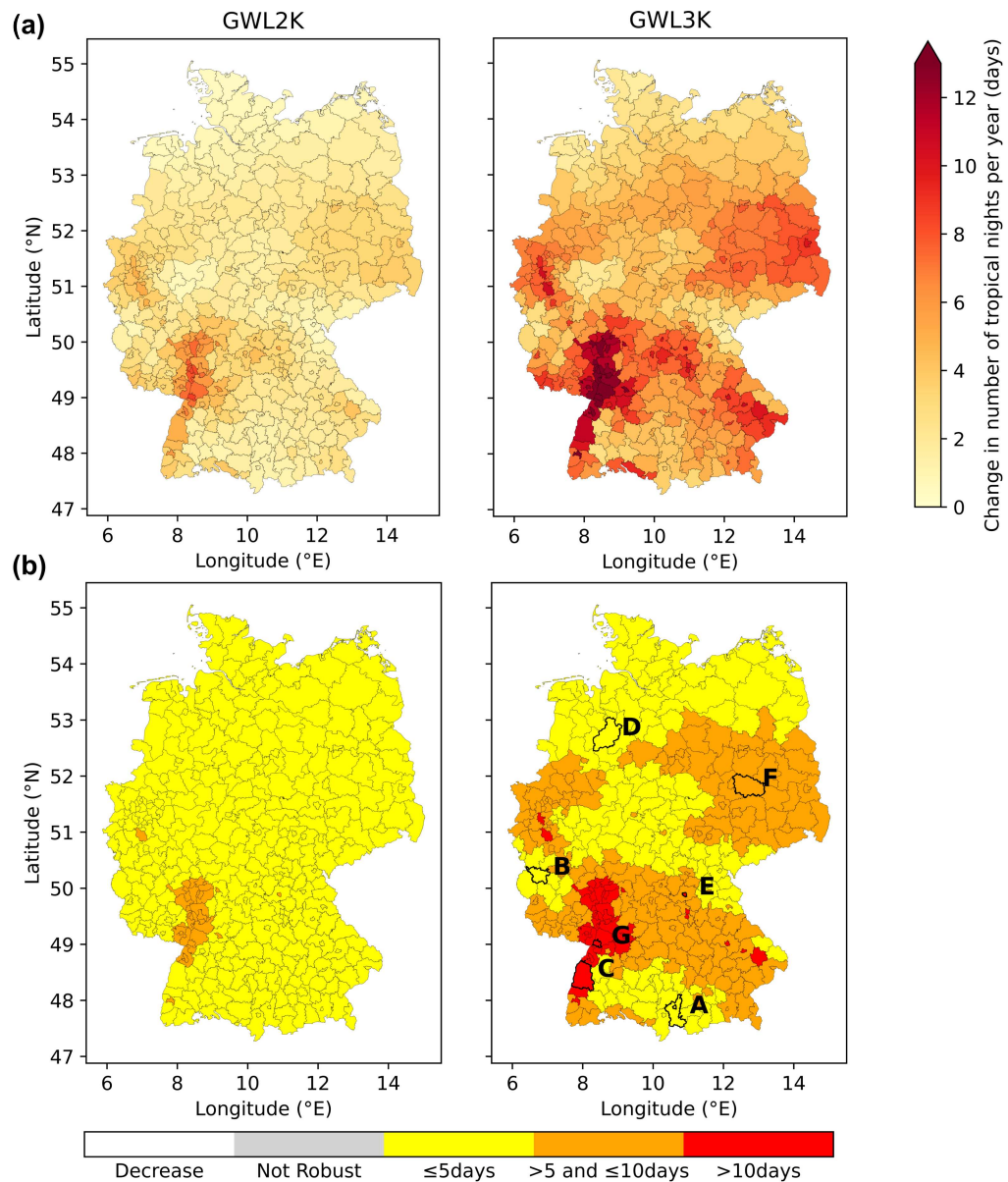
Figure 4.13: Climate change signals for a change in the number of tropical nights for GWL2K (left panels) and GWL3K (right panels). (a) shows the absolute median climate change signals in days. (b) shows the climate signal maps for an increase in the number of tropical nights. Median signals are displayed divided into 3 categories, representing low (yellow), medium (orange), and high (red) absolute changes, expressed in days. White areas indicate a decrease. All climate change signals are robust. For GWL3K, a selection of example regions is highlighted with a bold outline.
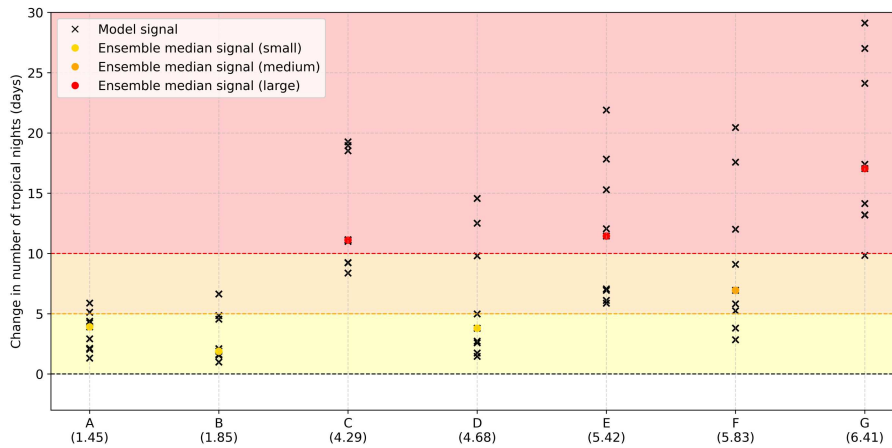
Figure 4.14: Climate change signals for an increase in the number of tropical nights for the individual models of the NUKLEUS ensemble for the selected regions (Figure 4.13b, right panel). The orange and red lines mark the limits of the corresponding levels in the climate signal map (Figure 4.13b). The standard deviation for each region is displayed in brackets below the region's letter. The ensemble spread is classified as small when the standard deviation is below the threshold of 2 days.

ensemble spreads. For example, the climate change signals in Region F range from 2.8 to 20.5 days, covering small, medium, and large categories (see also Table 4.1). Similarly, the signals for region E range from 5.9 to 21.9 days. In region G, almost all models show large signals, but still their magnitude varies strongly, ranging from 9.8 up to 29.1 days. This spread is not caused by a single outlier, but rather several models showing signals above 20 days, while others range between 10 and 20 days. Comparable variations of the climate change signals are shown for regions C, D, and E.

Interestingly, the models that show particularly high climate change signals are all driven by the GCM MIROC (see Table 4.1). The maximum signals for each region are highlighted in bold in Table 4.1. The maximum values across all regions correspond to the driving GCM MIROC, with MIROC-REMO showing the most. This aligns with the findings from Cusinato et al. (2025), who identified a warm bias in MIROC-driven simulations in the NUKLEUS-EUR11 ensemble. Since the number of tropical nights is defined through a fixed threshold of $TN > 20\,^{\circ}C$, this warm bias results in a higher number of tropical nights for the MIROC-driven simulations than for the rest of the ensemble members.

This highlights the relevance of considering the ensemble spread in addition to robustness. While all example regions show robust signals, the degree of agreement on their magnitude differs. These large differences in the signals' magnitude are not reflected in the climate signal maps following Pfeifer et al. (2015). Therefore, in the following, the ensemble spread is incorporated into the climate signal maps.

Regions where the standard deviation is below a given threshold are referred to as regions with a small ensemble spread, i.e., precise signal. In contrast, regions where the standard deviation exceeds the threshold are considered to have a large ensemble spread. Regions with a small ensemble spread are additionally marked with hatching. The threshold for the standard deviation depends on whether

the considered index is derived from precipitation or temperature data. For temperature-based indices, the threshold is determined based on the limits of the climate signal maps, given in Table 3.3.

Figure 4.15 shows the climate signal map for an increase in tropical nights with the consideration of the ensemble spread. Here, the threshold for the standard deviation is set to 2 days. For GWL2K, most of the regions show a small ensemble spread, except for the regions with a medium signal in the Upper Rhine Valley and the northeastern part of Germany. This implies general agreement among the models on the magnitude of the increase in tropical nights for GWL2K. This is also highlighted in Figure B.2. Here, the individual signals of each ensemble member for the same example regions as in Figure 4.14 are shown for GWL2K. Regions A to D exhibit a small standard deviation and are hatched. Regions E to G remain unhatched, as their standard deviation is greater than the threshold of 2 days. For GWL2K, the ensemble spread is lower than for GWL3K, which is reflected in the higher number of hatched regions.

For GWL3K, only a few regions, mainly located in the South of Germany, exhibit a small ensemble spread. This is also highlighted in Figure 4.14. For regions A and B, the standard deviation is below the given threshold, and the regions are thus shown with hatching. In contrast, the standard deviation of regions C to G exceeds the threshold and remains unhatched.

Although the ensemble spread for GWL3K is overall large, it is important to bear in mind that all signals are robust. This means that the models consistently project a significant increase in the number of tropical nights. The ensemble spread measure solely captures the agreement on the magnitude of the signal, and thus provides additional insights into the precision of the ensemble median signal displayed in the climate signal maps. This demonstrates that incorporating ensemble spread allows for a further distinction between robust climate change signals by highlighting differences in precision. This enables a more nuanced interpretation of the projected climate change signals (see Figure 1.1).

Ice days (ID)

To also consider changes in cold temperature extremes, Figure 4.16a shows the median climate change signals for the number of ice days. All regions show a decrease in the number of ice days for both GWLs. For GWL2K, the regions with the strongest climate change signals are located in the East of Germany and regions with small signals in the West. The same East-West pattern is shown for GWL3K. The climate change signals for GWL3K are stronger than for GWL2K. Strongest signals are shown for East-Central Germany, with a decrease of more than 15 days per year.

Figure 4.16b shows the climate signal maps for a decrease in the number of ice days. Here, a threshold of 5 days was applied for the standard deviation to categorize the ensemble spread into small or large. The signals for GWL2K show a West-East gradient with small signals below 10 days in the West of Germany and medium signals in the East and Central Germany. For GWL3K, this West to East gradient intensifies, with medium signals in ice days for the western part and large

Table 4.1: Climate change signals for the change in the number of tropical nights for GWL3K. The climate change signal for each simulation within the NUKLEUS ensemble are given for the regions A to G (Figure 4.13b, right panel). The unit is days. The simulations are sorted by the driving GCMs. The maximum signals for each region are highlighted in bold.

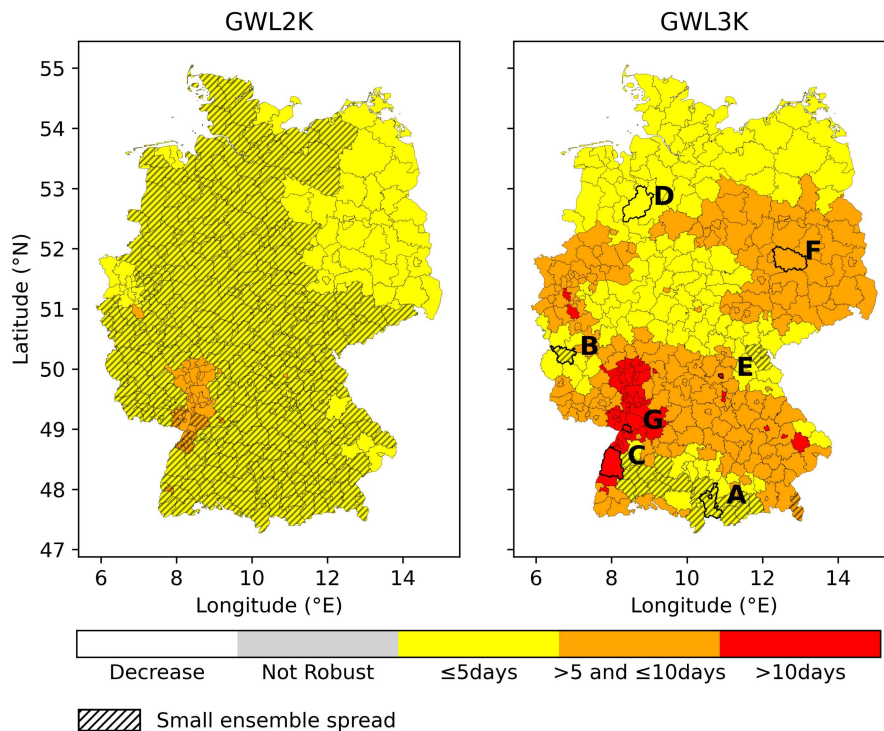| Model	A	B	C	D	E	F	G
ECE ICON	2.05	1.64	9.23	2.73	6.11	5.83	14.14
ECE CCLM	4.38	1.57	11.01	1.74	6.95	3.81	13.19
ECE REMO	1.32	2.11	11.12	4.99	12.05	9.1	17.4
MIROC ICON	3.92	4.55	**19.26**	12.51	17.83	17.58	**29.13**
MIROC CCLM	**5.89**	4.85	18.96	9.81	15.29	12.01	24.12
MIROC REMO	5.13	**6.64**	18.52	**14.57**	**21.91**	**20.45**	27.01
MPI ICON	2.92	1.55	9.25	2.59	7.06	5.24	13.21
MPI CCLM	4.25	0.99	8.38	1.46	5.88	2.84	9.84
MPI REMO	2.14	1.89	11.14	3.8	11.45	6.95	17.05

Figure 4.15: Same as Figure 4.13b, but including an evaluation of the ensemble spread for each region. The ensemble spread is classified as small when the standard deviation is below the threshold of 2 days.

signals in East and Central Germany. Some regions in the Northwestern part of Germany also show small signals for GWL3K.

Regarding the ensemble spread, regions in the western part of Germany show a small ensemble spread, as well as a few regions in Southern and Central Germany. Towards GWL3K, the number of regions with a small ensemble spread slightly decreases, but remains concentrated in the West of Germany. Overall, both GWLs show a West–East pattern in signal strength and model agreement on the magnitude of the signals. Weaker signals with a small ensemble spread in the West of Germany, and stronger signals with a large ensemble spread in the East. Despite the large ensemble spread in many regions, all climate change signals for the number of ice days are robust, indicating agreement among the models on a significant decrease in ice days. The decrease in the number of ice days found here is consistent with the increase of near-surface air temperature for the winter season (Figure 4.1). As ice days are defined by a temperature threshold, their number decreases naturally with an increase in winter temperatures. The decrease in the number of ice days for Germany aligns with previous findings (Pinto et al., 2025; van Rüth et al., 2023).

Warm Spell Duration Index (WSDI)

So far, the considered temperature-based indices were based on the frequency of temperature extremes. Next, an index that evaluates changes in the duration of temperature extremes is evaluated. The Warm Spell Duration Index, WSDI, is defined as the number of days that are part of warm spells. A warm spell is defined as 6 consecutive days that exceed the 90th percentile of daily maximum temperature. Therefore, an increase in the WSDI indicates that warm spells become more frequent and longer. Figure 4.17a shows the median climate change signals for the WSDI per year. For GWL2K, most regions show an increase below 20 days, except for one region in the North of Germany showing a higher value. For GWL3K, the climate change signals intensify with the strongest signals in northern and central Germany. This indicates that warm spells are projected to become more frequent and longer, especially for GWL3K.

Figure 4.17b shows the climate signal map for an increase in the WSDI. As for ice days, the threshold for the standard deviation is set to 5 days. The limits for the categories range from 0-15 days for small signals (yellow), 15 to 30 days for medium signals (orange), and more than 30 days for large signals (red). In GWL2K, the regions show mainly small to medium signals, with medium signals in the North and South as well as in the middle part of Germany. For GWL3K, almost all regions project a large increase in the WSDI of more than 30 days.

Regarding the ensemble spread, only some regions in the South of Germany exhibit small ensemble spreads in GWL2K. For GWL3K, no region shows small ensemble spreads. Although all signals are robust, indicating a clear increase in the WSDI, the large ensemble spread indicates the ensemble median signal shown in the climate signal map may not fully represent the range of possible magnitudes and should be interpreted with caution. Overall, the results show a robust trend towards more frequent and longer-lasting warm spells for both GWLs projected, which is consistent with previous findings (e.g. Sillmann et al., 2013b).
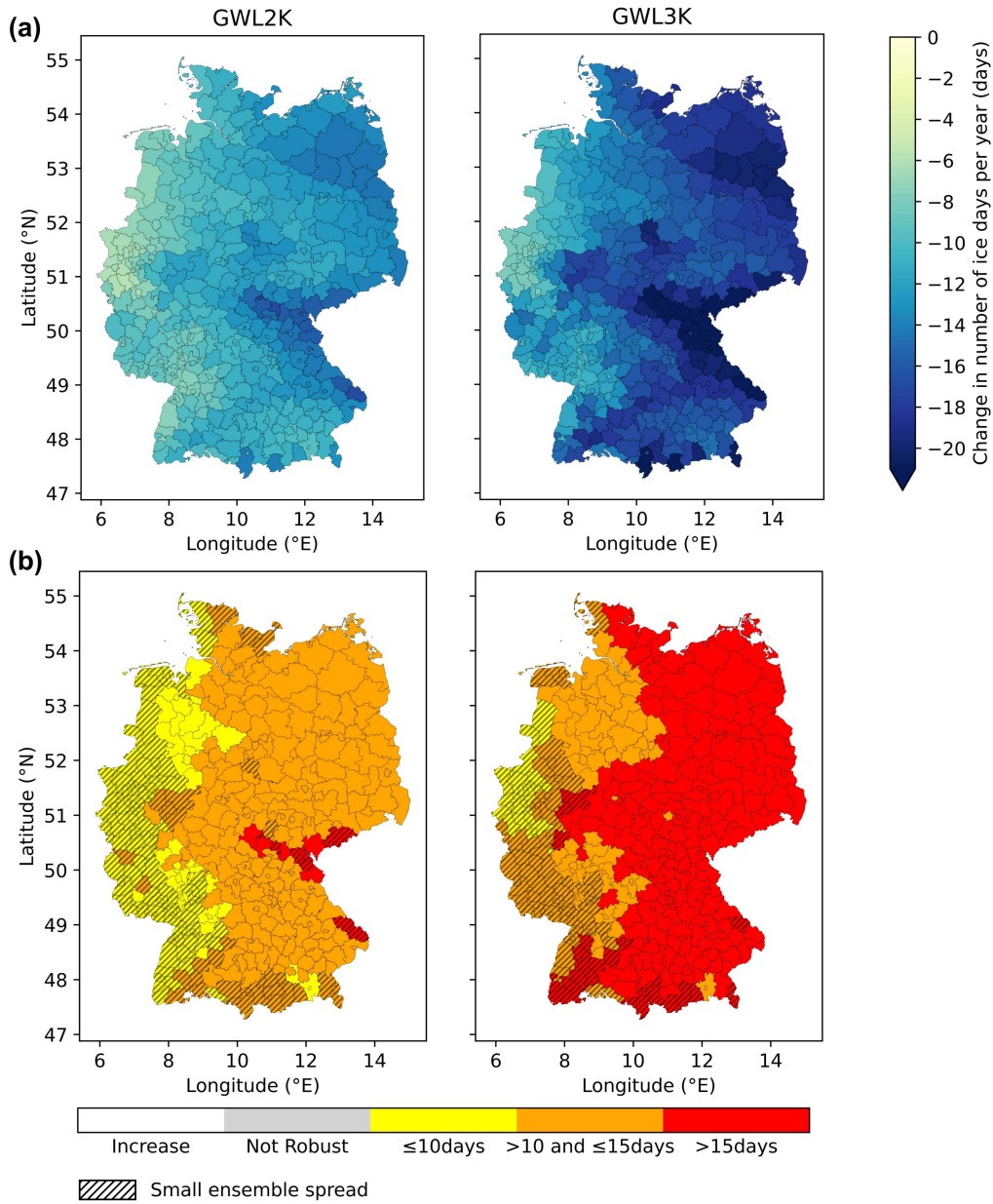
Figure 4.16: Climate change signals for a change in the number of ice days for GWL2K (left panels) and GWL3K (right panels). (a) shows the absolute median climate change signals in days. (b) shows the climate signal maps for a decrease in the number of ice days. Median climate change signals are displayed divided into 3 categories, representing low (yellow), medium (orange), and high (red) changes. All climate change signals are robust. Regions with a small ensemble spread are marked with hatching. The ensemble spread is classified as small when the standard deviation is below the threshold of 5 days.

Figure 4.17: Same as Figure 4.16, but for changes in the WSDI (warm spell duration index). All climate change signals in (b) are robust. An increase in the WSDI is considered for the climate signal maps.

4.2.2 Climate change signals of precipitation-based indices

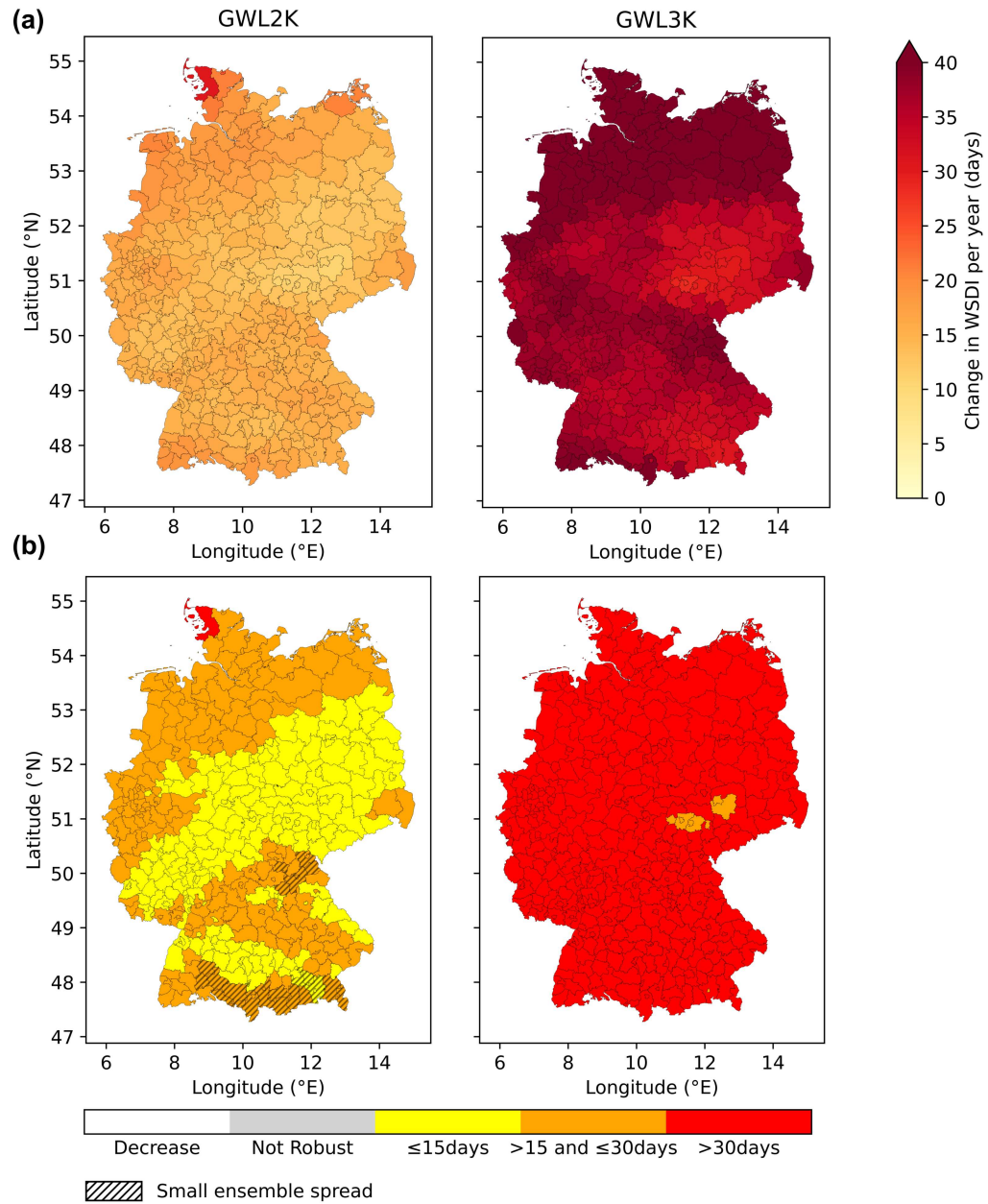In the following, climate change signals for selected ETCCDI indices based on precipitation data from the NUKLEUS ensemble are presented. For these indices, the categories in the climate signal maps are expressed in percentages. The threshold for the standard deviation is set to 30% of the mean robust signal (see Section 3.3 for details). If fewer than 30 regions show robust signals, the threshold is calculated based on the mean signal across all regions.

Days with R10mm, R20mm, and R30mm

Figure 4.18 shows the climate change signals for a change in the number of days with a) R10mm, b) R20mm, and c) R30mm. All three plots share the same colorbar. For GWL2K (left panels), the strongest climate change signals are shown for an increase in the number of days with R10mm in the East of Germany (Figure 4.18a), particularly in the South-East of Germany. The climate change signals for changes in the number of days with R20mm and R30mm (Figures 4.18b-c) are generally smaller than for R10mm, with the strongest signals located in the Southeast of Germany. For GWL3K (rights panels), the climate change signals for all three indices intensify. The strongest signals are shown for the change in R10mm. R20mm and R30mm show the strongest signals for regions in the South of Germany.

Figure 4.18 shows the climate signal maps for an increase in the number of days with d) R10mm, e) R20mm, and f) R30mm. The same categories for small (yellow, $\leq 15\%$), medium (orange, $< 15\%$ and $\geq 25\%$), and large signals (red, $> 25\%$) are applied to all three climate signal maps.

For GWL2K, the climate signal map shows mostly small signals for an increase in days with R10mm (Figure 4.18d, left panel), with some regions showing a decrease in R10mm in Western Germany. The ensemble median climate change signals for R20mm (Figure 4.18e, left panel) and R30mm (Figure 4.18f, left panel) show an increase for GWL2K across most regions, with the strongest signals in eastern Germany. The relative climate change signals for GWL2K increase from R10mm to R20mm to R30mm, with most regions showing large relative signals for R30mm.

For GWL3K, a higher fraction of regions show an increase in days with R10mm, with some medium signals in the Northeast. For R20mm, most of the regions, especially in the East of Germany, show medium to large signals for GWL3K. For R30mm, almost all regions show large climate change signals for an increase in the number of days with R30mm for GWL3K.

Although the strongest absolute signals are shown for R10mm (Figure 4.18a), the relative changes are small in the climate signal maps (Figure 4.18d). Conversely, R30mm exhibits the weakest absolute changes (Figure 4.18c), but strongest relative changes (Figure 4.18f). This indicates that the most extreme events show the strongest relative increases in frequency, as they are rare in the reference period. Similarly, while the absolute changes in R10mm are large, they represent only a small increase relative to the reference period. These results indicate an intensification of the frequency of extreme precipitation events with increasing GWLs, particularly for the more extreme

indices such as R20mm and R30mm. Moreover, the comparison of the absolute climate change signals to the relative signals underline the purpose of the climate signal maps to illustrate the climate change signals in a comprehensive way. Representing changes in relative terms makes the results more accessible than using precipitation units, as users can better visualize the potential impacts based on their knowledge of the climate conditions of their region.

The bottom row of Figure 4.18d shows that only a few regions display robust signals in the number of days with R10mm for GWL2K. In contrast, R20mm and R30mm show more regions with robust signals, again concentrated in the East (Figures 4.18e-f). For GWL3K, the number of robust signals increases for all three indices compared to GWL2K. Medium to large signals are more often classified as robust, with R30mm showing the highest number of robust signals for both GWLs. An East–West gradient in the robustness and signal strength is shown across all indices: the strongest and most robust signals occur more frequently in the East, while regions in the West show less strong and less robust signals. This pattern is consistent with earlier results, which showed that stronger climate change signals in summer precipitation are more often classified as robust (see Section 4.1.1).

Regarding the ensemble spread, only a few regions meet the threshold for a small ensemble spread, indicated by hatching. These regions are mostly found for R30mm in GWL3K, particularly in the south of Germany. Here, R20mm also shows a few regions with a small ensemble spread. Overall, the ensemble spread is high across regions and indices, indicating high variability in the magnitude of changes in extreme precipitation events.

Interestingly, for GWL2K, some regions with non-robust signals in R20mm and R30mm still show a small ensemble spread, indicated by gray colors and hatching in the bottom row of Figures 4.18e-f. Figure B.3 illustrates this further for selected example regions for R30mm and GWL2K. The considered regions are illustrated in Figure 4.18f (right panel, bottom row). Region A shows a robust medium signal and a small ensemble spread (red and hatched), while Regions B and C show non-robust signals and at the same time a small ensemble spread (gray and hatched). This shows that even if a signal is not robust, it can still be precise, indicating that no or a very weak climate change signal is projected (see Figure 1.1). In contrast, Region E shows a large ensemble spread, and the median climate change signal is not robust, indicating that the uncertainty is too high to derive a climate change signal. The signal for Region D is robust, but shows a large ensemble spread.

The values of the indices are all derived from the precipitation data obtained from the NUKLEUS ensemble. As shown in Section 4.2.2, there is a large uncertainty in the winter and summer precipitation signals of the NUKLEUS ensemble, showing diverging signals, especially for the summer season. This disagreement in the sign and magnitude of the precipitation signals among the models is reflected in the large ensemble spread of the precipitation-based indices. Therefore, only a small number of regions show small ensemble spreads. There are more robust climate change signals for the extreme indices R10mm, R20mm and R30mm presented in Figure 4.18d-f, compared to the signals for mean winter and summer precipitation. This is consistent with findings

of Pfeifer et al. (2015), where the 95th percentile of winter precipitation showed stronger signals and more robust signals than for the mean precipitation signal in winter.

Overall, the number of days exceeding extreme precipitation sums shows a robust increase compared to the reference period and intensifies towards GWL3K, particularly for the eastern part of Germany. The indices R20mm and R30mm show the strongest relative climate change signals. Regarding the climate change signals of extreme precipitation events, three main findings can be derived: first, the more extreme the index, the stronger the relative climate change signals. Second, stronger relative signals tend to be more robust. Third, the signals intensify in strength and show more robustness towards GWL3K.

The results show that the most extreme indices, such as R30mm, exhibit the strongest and most robust relative climate change signals. This pattern is consistent with previous studies, which found that rare and more extreme precipitation events tend to intensify more strongly than moderate ones under global warming (Ban et al., 2015; Hundhausen et al., 2024). The intensification of extreme precipitation can be explained by the Clausius-Clapeyron relation, which describes a scaling of 7% increase in the moisture capacity per degree warming. Moreover, for short-duration or convective events, the scaling can even exceed the Clausius-Clapeyron scaling (Lenderink and Van Meijgaard, 2008).

Another factor for the strong relative changes of R20mm and R30mm is their low occurrence in the reference period, which leads to larger relative increases in the climate change signal. In contrast, the strongest absolute changes are found for R10mm (Figure 4.18a), which is consistent with the results of Pinto et al. (2025). However, relative to the reference period, these changes are considered small (Figure 4.18d).

Maximum length of dry spells (CDD)

So far, the analysis has focused on indices that measure the frequency of precipitation extremes. Next, extreme indices that consider the duration of precipitation extremes are evaluated. Figure 4.19a shows the median climate change signals for a change in the maximum length of dry spells (CDD), indicating changes in the maximum number of consecutive dry days.

For GWL2K (left panel), the majority of regions show an increase in CDD, with the strongest climate change signals for regions in the Northwest of Germany. For GWL3K (right panel), the median climate change signals are stronger than for GWL2K, and all regions show an increase in CDD. The strongest signals are shown for regions in Central Germany, particularly in the western-central part.

Figure 4.19b shows the climate signals maps for an increase in CDD. For GWL2K, most regions show a small relative increase in CDD, except for a few regions in Southeastern Germany that show a decrease. The strongest relative signals for GWL2K are shown in Western Germany, with medium to large signal strength. For GWL3K, the largest relative climate change signals are located in
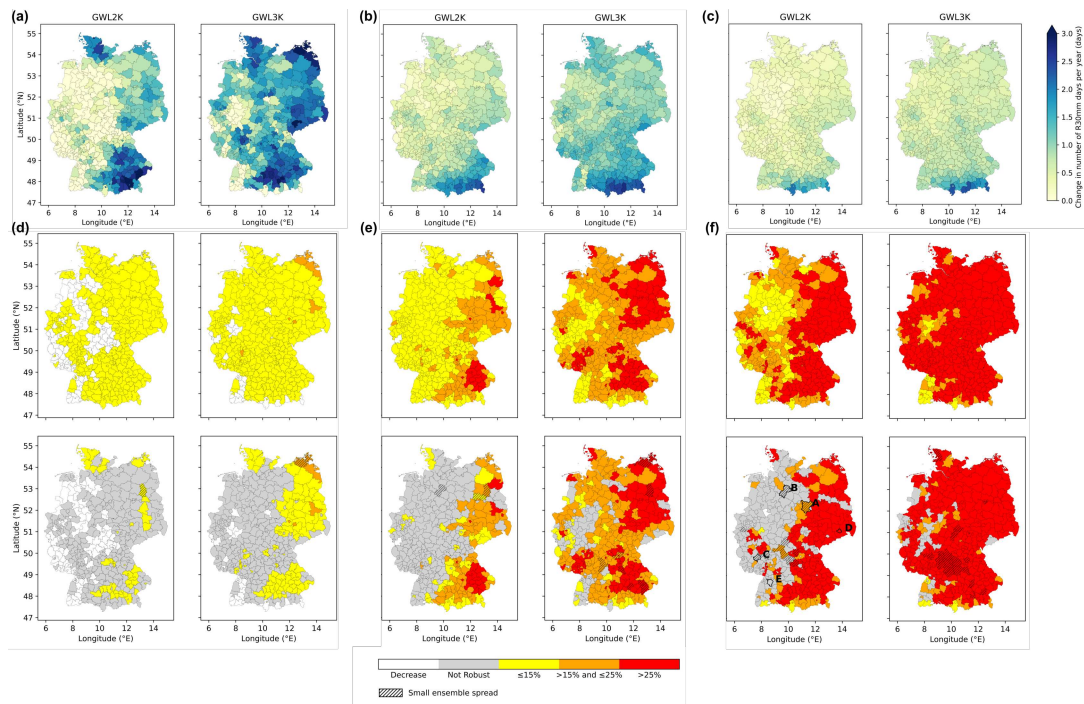
Figure 4.18: Climate change signal for a change in the number of days with (a) R10mm, (b) R20mm, and (c) R30mm for GWL2K (left panels) and GWL3K (right panels). (d-f) shows the climate signal maps for an increase in the number of days with (d) R10mm, (e) R20mm, and (f) R30mm. Median climate change signals are displayed divided into 3 categories, representing low (yellow), medium (orange), and high (red) relative changes. White areas indicate a decrease. The upper row shows the median climate change signals for all regions, while the bottom row shows only regions with robust climate change signals in color and non-robust in gray. Regions with a small ensemble spread are hatched. The ensemble spread is classified as small when the standard deviation is below the threshold based on 30% of the mean robust climate change signal. For R30mm and GWL2K (f, bottom row, left plot), selected example regions are outlined in bold.

Central Germany and the Alpine Foreland in the South of Germany, while northern and southern regions show small climate change signals.

Regarding the robustness of the signals (bottom row of Figure 4.19b), most of the regions show non-robust signals for GWL2K. There are a few robust climate change signals for regions in the West of Germany. For GWL3K, the number of regions with robust climate change signals is higher compared to GWL2K. These are mainly located in Central Germany, where the signals are also medium to large. Regarding the ensemble spread, no region exhibits a small ensemble spread for both GWLs.

In Section 4.1.1, seasonal changes in mean precipitation are evaluated. For winter (Figure 4.1b), the ensemble shows small climate change signals for an increase in mean precipitation, but the changes were not classified as robust for both GWLs. In contrast, for summer (Figure 4.2b), the ensemble projects a decrease in summer precipitation, especially for GWL3K. The regions with the strongest climate change signals for a decrease in summer precipitation for GWL3K are located in the Northwest of Germany and show robust changes. This pattern aligns with the CDD climate change signals for GWL3K (Figure 4.19b, right panel), where the strongest increases in CDD are
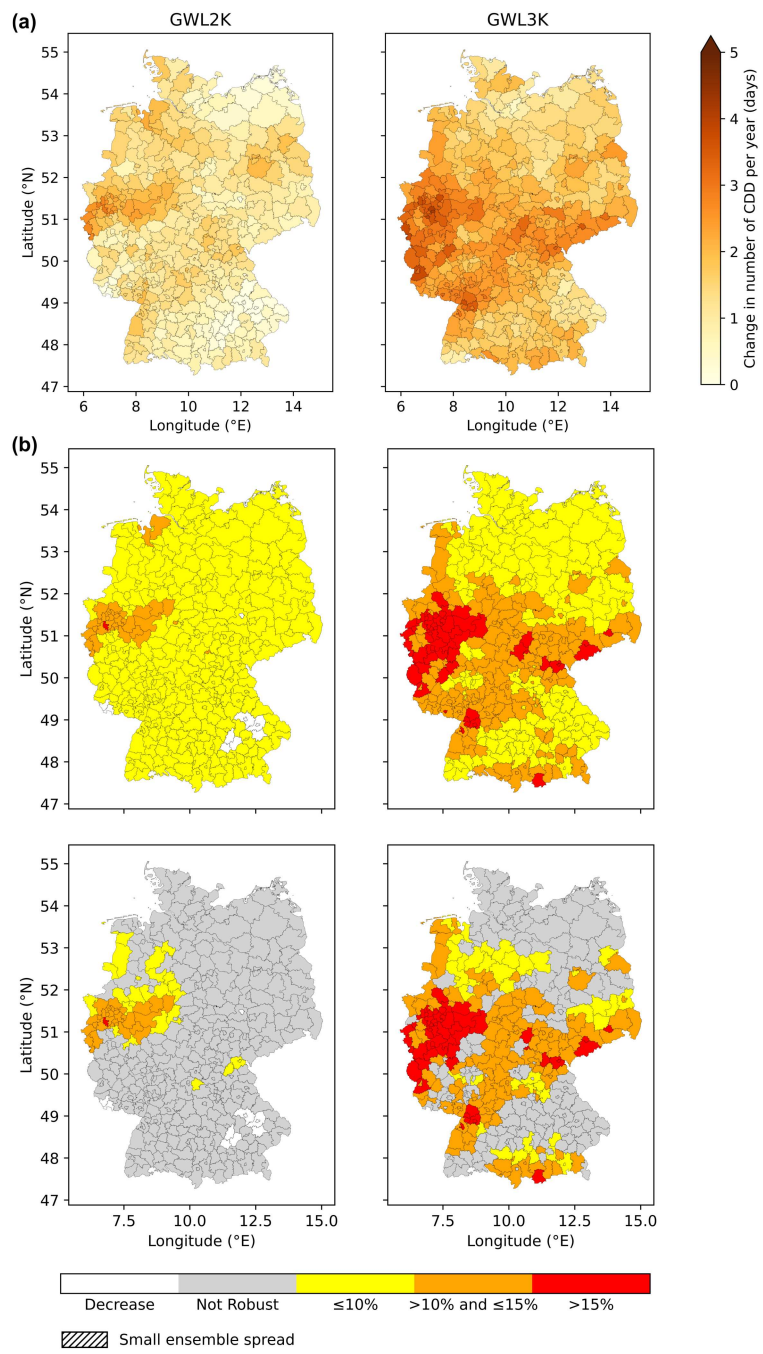
Figure 4.19: Climate change signals for a change in the maximum length of dry spells (CDD) for GWL2K (left panels) and GWL3K (right panels). (a) shows the absolute median climate change signals in days. (b) shows the climate signal maps for an increase in CDD. Color scale and hatching are consistent with Figure 4.18.

also found in Western Germany, overlapping with the regions that show the strongest decreases in mean summer precipitation for GWL3K (Figure 4.2b, right panel).

Although annual changes of CDD are considered, dry spells predominantly occur in summer (Spinoni et al., 2017). As shown in Section 4.1.1, the individual models show diverging climate change signals for summer precipitation (Figure 4.6), resulting in a large spread. This large spread is reflected in the CDD signals, as CDD is derived from the precipitation data. Consequently, there is a large ensemble spread for CDD, and no region exhibits a standard deviation below the threshold.

Overall, based on the NUKLEUS ensemble, there is an increase in CDD for both GWL2K and GWL3K. The climate change signals intensify towards GWL3K, and there is a higher number of regions with robust climate change signals compared to GWL2K. Medium to large climate change signals are generally more often classified as robust compared to small signals. This is consistent with the findings for the indices R10mm, R20mm, and R30mm. The projected increase in CDD shown here aligns with the previous studies (Coppola et al., 2021).

Maximum length of wet spells (CWD)

Figure 4.20a shows the changes in the maximum length of wet spells (CWD). The majority of regions show negative climate change signals for both GWLs, indicating a decrease in CWD. The signal strength does not considerably vary between GWL2K and GWL3K. Several regions show positive changes indicating an increase in CWD. The regions that show an increase in CWD for GWL2K are not the same regions as for GWL3K.

Figure 4.20b shows the climate signal maps for a decrease in CWD. Note that the limits applied for the categories are smaller than those used for other indices such as CDD. The threshold for large signals in CWD corresponds to the threshold for small signals in CDD. These small limits are applied to emphasize the relatively low magnitude of the relative climate change signals in CWD.

In most of the regions, a decrease in CWD is projected for both GWLs. For GWL2K, most of the regions show small signals below 5 % relative change. Some regions in Central and North of Germany show medium signals between 5-10% relative change. A few regions show an increase in CWD. For GWL3K, the number of regions with medium signals is lower than for GWL2K. The majority of regions show small signals. For GWL3K, there are also a few regions with an increase in CWD. These regions do not correspond to the regions with an increase for GWL2K.

For both GWLs, hardly any region shows a robust climate change signal for an increase in CWD. Interestingly, the number of robust signals decreases for GWL3K, from 11 regions for GWL2K to 6 regions with robust climate change signals for GWL3K. No region exhibits a small ensemble spread for both GWLs.

The absence of robust climate change signals and a decrease in the number of robust signals towards GWL3K indicates that no robust change in CWD can be derived from the NUKLEUS ensemble.
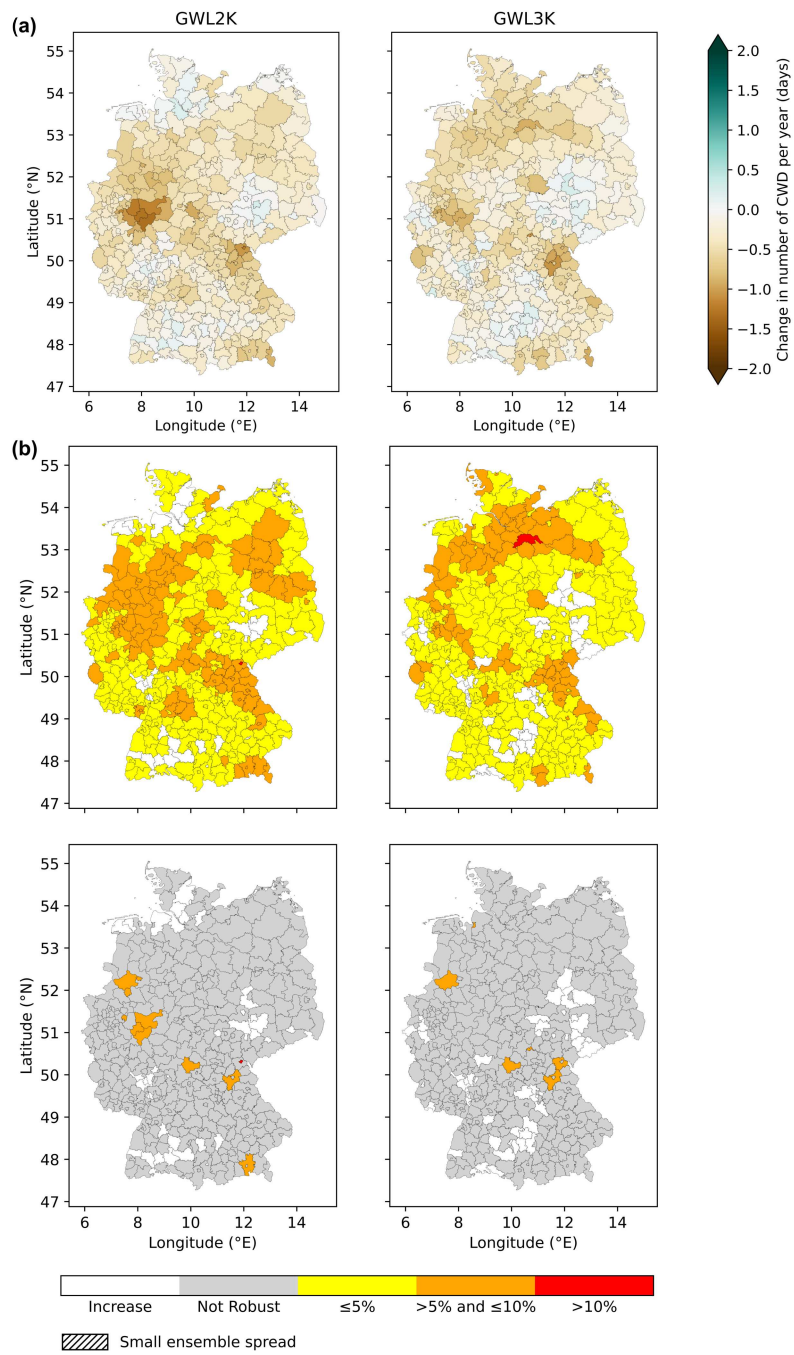
Figure 4.20: Climate change signals for a change in the maximum length of wet spells (CWD) for GWL2K (left panels) and GWL3K (right panels). (a) shows the absolute median climate change signals in days. (b) shows the climate signal maps for a decrease in CWD. Color scale and hatching are consistent with Figure 4.18.

This is further supported by the overall weak signal strengths, which even decrease for GWL3K. Although annual changes of CWD are considered, the results align with the seasonal trends in mean precipitation (see Section 4.1.1). There is no robust increase in winter precipitation for both GWLs (Figure 4.1), and there is a trend in summer precipitation towards drying signals (Figure 4.2). Together, this aligns well with the results for a small increase in CWD (Figure 4.20b), which is not robust. Therefore, no climate change signal for CWD can be derived.

Regarding the ensemble spread of both climate signal maps, no region exhibits a small ensemble spread. The climate change signals for CWD are very small. As the number of regions with robust signals is very low for both GWLs, the threshold for the standard deviation is based on the mean signal of all regions. Given that the climate change signals for CWD are very small, this results in a very small threshold for the ensemble spread of approximately 0.1 days. Consequently, the standard deviation exceeds this low threshold in all regions, and no region is classified with a small ensemble spread. This highlights a limitation of the threshold definition, which is discussed further in Chapter 5.

Figure 4.21 shows the climate signal maps for CWD, similar to Figure 4.20b, but with a different threshold for the ensemble spread. Here, the threshold is set to 0.5 days for both GWLs, resulting in many regions with a small ensemble spread.

Although most climate change signals in CWD are small and not robust, considering the ensemble spread helps distinguish regions with no or very weak signals from those where uncertainty is too high to draw reliable conclusions (see Figure 1.1). This distinction is illustrated in Figure 4.22, which shows the climate change signals of each ensemble member in the NUKLEUS ensemble for selected regions under GWL3K (see Figure 4.21, right panel, bottom row).

Regions A to C show a small ensemble spread with a standard deviation below the threshold of 0.5 days. The individual signals of the ensemble members show small values close to zero for all three regions. The good agreement on the very small magnitude among the ensemble members indicate that no climate change signal in CWD can be derived for these regions.

In contrast, for regions E and F, the individual models differ widely in their magnitude and even show opposing signs, reflecting high uncertainty. In these cases, the lack of robustness can be attributed to high uncertainty among the ensemble members instead of a lack of signal. This highlights the difference between non-robust signals with a small ensemble spread (A-C) and those with a large ensemble spread (E-F).

Region D shows a robust climate change signal with a small ensemble spread. Here, the individual models indicate a decrease in CWD, resulting in a robust and precise signal.

This demonstrates that the classification approach (see Figure 1.1) only works if the threshold is set to realistic values, ensuring a meaningful distinction between regions with small and large ensemble spread. If the threshold is set too low, however, all regions may be falsely classified as having a large ensemble spread (see Figure 4.20).
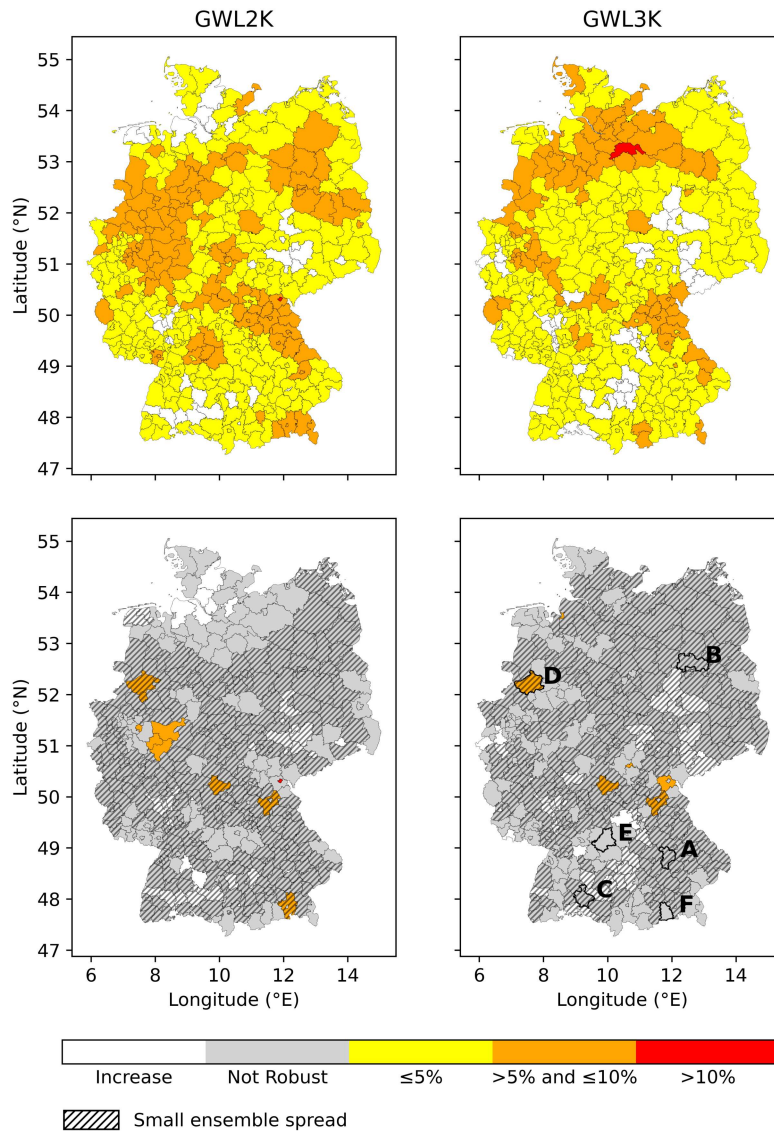
Figure 4.21: Same as Figure 4.20b, but with a different threshold for the standard deviation. The ensemble spread is classified as small when the standard deviation is below the threshold of 0.5 days.
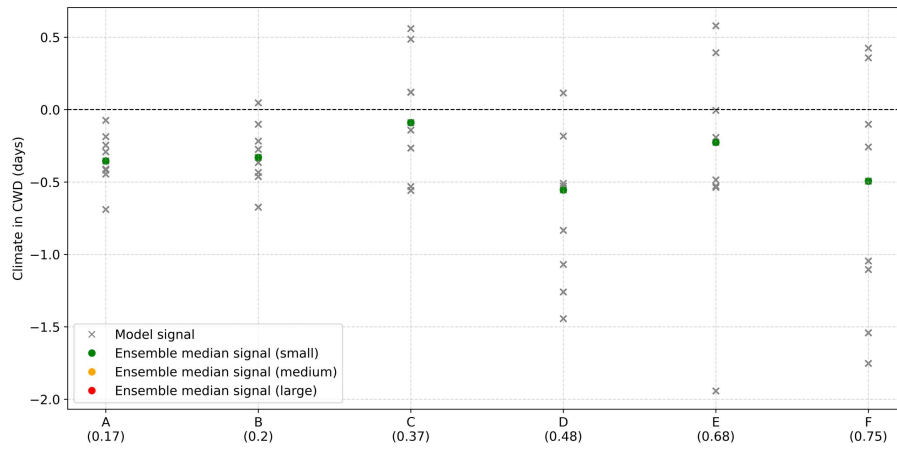
Figure 4.22: Climate change signals for a decrease in the maximum length of wet spells (CWD) for GWL3K of the individual models of the NUKLEUS ensemble for the selected regions in Figure 4.21, right panel. The individual signals of each model and each example region are displayed, along with the ensemble median signal. The standard deviation for each region is displayed in brackets. The threshold for the standard deviation applied here is 0.5 days.

5 Discussion

This thesis analyzes the robustness of climate change signals for mean summer and winter temperature and precipitation, as well as selected ETCCDI indices, based on data obtained from the NUKLEUS ensemble. The first objective is to identify which climate change signals are classified as robust. Climate signal maps are used to evaluate the robustness of the climate change signals. The second objective is to assess whether incorporating the ensemble spread into the climate signal maps adds value. For this purpose, a threshold is introduced to categorize the ensemble spread as either small or large.

5.1 Robustness of the climate change signals for mean temperature, precipitation, and the ETCCDI indices

The results show robust and spatially consistent warming signals for the mean temperature across all regions, seasons, and GWLs. In contrast, the precipitation signals are less consistent, with low robustness and diverging signals, particularly for the near-future period, GWL2K. There are more robust signals for a decrease in summer precipitation than for an increase in winter precipitation, although the signals for summer precipitation show a higher variability.

Temperature-based indices show robust climate change signals across all regions and GWLs, and an intensification for GWL3K compared to GWL2K. The precipitation extremes show generally more robust signals and consistency than for the mean precipitation signals. The results indicate more frequent and intense precipitation extremes, both for wet and dry extremes, except for CWD, where no trend is identified.

The robustness and spatial consistent trend for both an increase in mean temperature changes and temperature extremes are consistent with previous studies (Jacob et al., 2014; Kjellström et al., 2018). The higher robustness of the climate change signals of precipitation extremes, compared to mean precipitation signals, aligns with other studies showing an intensification and higher significance for precipitation extremes (Sillmann et al., 2013a; Fosser et al., 2024; Prein et al., 2013). The signals in mean summer and winter precipitation for GWL2K differ both in sign and strength, indicating substantial internal variability for near-future precipitation changes. This aligns with other studies, which also show high variability in near-future precipitation signals (Seaby et al., 2013; Pfeifer et al., 2015; Evin et al., 2021).

The high internal variability can be attributed to various effects, including the ensemble size. The NUKLEUS ensemble is based on three CPMs and three GCMs resulting in 9 ensemble members. One potential reason for the diverging climate signals and the high internal variability is that the ensemble size is not sufficient to correctly represent internal variability. As no single model can be considered "correct", the more models included in the ensemble, the more the results can be trusted. Including additional models in the ensemble might broaden the range of possible realizations of future precipitation projections, which could lead to a clearer ensemble mean signal. However, the large spread in precipitation signals does not necessarily imply a bad performance of the ensemble (Jebeile and Barberousse, 2021). Since every climate simulation based on a specific model represents a possible realization of the future climate, a wide spread implies that the ensemble covers a wide bandwidth of potential future projections. This can also be seen as an indicator of the trustworthiness of the ensemble.

5.1.1 Comparison to Pfeifer

Not only does the number of models included in the ensemble influence the outcome, but also the selection of which models are included (Evans et al., 2014; Herger et al., 2018). The relevance of the model selection is reflected in the comparison of the findings here to the results of Pfeifer et al. (2015). Although the same method as in Pfeifer et al. (2015) is applied, the robustness and strength of the climate change signals in mean precipitation presented here differ substantially from those of Pfeifer et al. (2015). The NUKLEUS ensemble shows the strongest signals and the highest number of robust signals for a decrease in summer precipitation, while Pfeifer et al. (2015) found the strongest and most of the robust signals for an increase in winter precipitation.

Several differences between the NUKLEUS ensemble and the ensemble used by Pfeifer et al. (2015) are identified in Section 4.1.1. These include differences in spatial resolution, the models included in the ensembles, the choice of time periods and GWLs, as well as the use of different scenarios, RCP4.5, and SSP370.

To address these differences, a comparison with the EUR11-NUKLEUS ensemble and the findings of Pfeifer et al. (2015) is presented in the Appendix Section A.2.1. Here, both ensembles are compared for the same model domain, at the same resolution, and for the same time periods. Note that at the time of the analysis not all simulations of the NUKLEUS-EUR11 ensemble were available. However, the NUKLEUS-EUR11 ensemble shows similar results to the NUKLEUS ensemble on the CEU3 domain and differs substantially from the results of Pfeifer et al. (2015). This implies that the model selection plays a major role in the differences between the two ensembles. The model selection is not the only difference between the two ensembles. The considered scenarios are SSP370 for the NUKLEUS ensemble and RCP4.5 in Pfeifer et al. (2015). Consequently, the differences between the two ensembles cannot solely be attributed to the selection of models within the ensembles.

Overall, the comparison of the results from Pfeifer et al. (2015) and the NUKLEUS-EUR11 ensemble shows that the model selection for a climate model ensemble is important and directly

affects the outcome of the results. This strong dependence of the model selection within ensembles is in line with previous studies showing the relevance of the ensemble constellation (Evans et al., 2014; Herger et al., 2018).

5.1.2 Effects of applying GWLs

Due to the high computational costs of CPMs, the simulations are based on GWLs instead of transient simulations. As each GCM reaches a certain GWL at a different period, the simulations represent different temporal slices (see Tab. 3.1). This can limit the comparability, as some effects such as land use changes, aerosol forcings, and natural variability like different phases of ENSO or AMOC are also dependent on the time and might be in different phases for the different time slices selected for the GWLs. As discussed before, GWL2K represents near-future periods (see Table 3.1), which limits the number of robust signals. The ECE-driven simulations, which reach GWL2K earliest, exhibit particularly few significant regions at this warming level.Pfeifer et al. (2015) showed similar findings. They compared different periods of transient simulation and showed that later periods exhibit more robust signals than near-future periods.

Furthermore, the periods for the GWLs cover 30 years. This raises concerns that long-term variability and trends may not be adequately captured within the period. Therefore, transient climate simulations would represent the statistics of extremes more adequately (Hundhausen et al., 2024). However, as the focus here is on the signals for the ETCCDI indices, which evaluate moderate extremes of return periods of approximately a year or less, 30-year time slices are suited.

5.1.3 Bias correction

Another important factor that needs to be considered is that the data evaluated in this thesis has not been bias-corrected. This is especially important for the interpretation of the climate change signals of the ETCCDI indices and for assessing climate change impacts (Teutschbein and Seibert, 2012; Chen et al., 2021).

At the time of this analysis, the output of the NUKLEUS ensemble was not bias-corrected. Bias correction is applied to climate model output to avoid systematic errors, such as consistently too high temperatures or errors in the timing or amount of precipitation (Christensen et al., 2008; Teutschbein and Seibert, 2010). To address this, the climate model output is bias-corrected based on observational data (Teutschbein and Seibert, 2010, 2012). This is especially important for climate impact studies, as uncorrected data can over- or underrepresent climate statistics like extreme events, leading to misleading results (Teutschbein and Seibert, 2012). This underlines the importance of applying a bias correction, especially for assessing climate change impacts and extremes. An additional approach to addressing systematic biases in climate model ensembles is to use full matrix structured ensembles to avoid the over- or underrepresentation of certain models (Vautard et al., 2021; Déqué et al., 2012). This applies to the NUKLEUS ensemble, which includes three CPMs and three GCMs.

During the process of bias correction, for example, with quantile delta mapping, the climate signal of a variable is kept constant and only the baseline is corrected. Thus, the seasonal precipitation and temperature signals are marginally affected by bias correction (Ugolotti et al., 2023). This is not the case for the climate signals of the extreme indices. As some indices are defined based on absolute thresholds, they are very sensitive to systematic biases like too hot temperatures. For instance, the simulations driven by the GCM MIROC show a warm bias. This warm bias is consistent with the findings of Cusinato et al. (2025), where they find a warm bias in the MIROC-driven simulations for the historical period in the NUKLEUS-EUR11 ensemble. Regarding the number of tropical nights, the MIROC-driven simulations consistently show a higher number of tropical nights (see Tab. 4.1), as more days exceed the threshold of $TN > 20\,^\circ$C than for the other simulations. This may lead to unrealistic results and, when evaluating an ensemble of different models driven by various GCMs, to a large ensemble spread, as shown, for example, in the signals for tropical nights for GWL3K (Figure 4.15).

Therefore, uncorrected data must be analyzed with caution. The missing bias correction is one key factor explaining why the models exhibit such large ensemble spreads for the climate signal in the extreme indices, particularly for temperature-based indices, due to the warm bias of the MIROC-driven simulations. However, since the bias-corrected data were not available at the time of this analysis, uncorrected data were used to assess and demonstrate the proposed method of extending the climate signal maps by the ensemble spread.

5.1.4 Limitations of the climate signal maps

The analysis of the climate change signals is conducted using climate signal maps introduced by Pfeifer et al. (2015). The climate signal maps categorized the signals into three levels of signal strengths. The choice of the limits for the levels depends on the considered variable or index and is not fixed. Therefore, the classification of the signals into small, medium, and large signals is up to the user and the application. The limits applied here are therefore only one possible realization. As the categories are generally very broad, the choice of the limits has a significant impact on the outcome of the classification of the signals. This is shown in Figure 4.20, where very small limits were applied compared to the limits for example of CDD (Figure 4.19). If the same categorization were applied to CWD, all signals would be classified as small. However, setting lower limits for the categories aimed to show how small the signals of CWD are. In addition, the use of the categories leads to a loss of information on regional differences of smaller scales than the chosen levels.

Climate signal maps consider only one direction of change and do not give any information on the strength or robustness of the opposite direction signals. Therefore, the method is only suitable for variables or indices that show a single direction of change, or in the case that the user is only interested in one direction of change. For instance, regarding summer precipitation signals in GWL2K (Figure 4.1), the results show approximately half of the regions indicating an increase in summer precipitation. As a decrease was considered, no information on the strength of the regions with positive changes in summer precipitation is displayed.

Pfeifer et al. (2015) applied the climate signal map method only to precipitation data with categories based on relative changes. Here, the climate signal maps were also applied to seasonal temperature and temperature-based indices. As the values of the temperature-based indices, such as tropical nights, are close to zero for the reference period, expressing the climate signals in relative terms leads to unrealistic high values. Therefore, the climate change signals for temperature and temperature-based indices are shown in absolute values. The evaluation of the temperature signals using climate signal maps showed that the climate signal maps are less suited for temperature signals compared to precipitation signals. As the signal strengths differ strongly between both GWLs, the broad categories lead to nearly no regional features being displayed. Thus, for climate change signals of mean winter and summer temperature, the climate signal maps are not suited, and the ensemble median signals should be displayed for the individual regions without categories.

The results for mean precipitation and the precipitation-based indices show that representing the signals in relative terms enables a clearer visualization of climate change signals for users. Precipitation units are less intuitive for users outside the academic context. Representing the signals in relative terms allow them to directly relate the changes to their knowledge of the regional climatic conditions of their region. This is shown, for example, for the climate change signals of R10mm, R20mm and R30mm (Figures 4.18a-f). Although the strongest absolute changes are shown for R10mm, R30mm shows stronger relative signals. This means that compared to the reference period, the frequency of very extreme events (R20mm and R30mm) increases stronger than for moderate extreme events (R10mm). Therefore, representing the climate change signals of mean precipitation and precipitation-based indices in relative terms reveals important additional information to the users than when solely considering absolute values. It allows to assess the projected signals in the climatic context of the respective region and takes into account the previous state of the considered variable or index in a reference period.

The data analyzed in this thesis is based on the NUKLEUS ensemble, which is a convection-permitting climate model ensemble. Although, the high resolution of the CPMs represent a well-suited resolution for the rural district of Germany, the resolution might be still too coarse for very small regions. This is reflected in the results for the autocorrelation. Most of the regions exhibit values for autocorrelation within the acceptable range for the significance testing, however, there are some outliers with very high autocorrelation (see Table B.1). As Pfeifer et al. (2015) suggested, for very small regions, neighboring regions should be considered for the interpretation of the results.

In this thesis, the climate signal maps are used to evaluate climate change signals of extremes. Two aspects need to be considered when averaging extremes over regions: the spatial heterogenity of extremes and the comparability of large and small regions. Some extremes, such as heavy precipitation events, occur at small spatial scales and spatially heterogeneously. This can lead to a loss of information when averaged spatially (Pfeifer et al., 2015). Moreover, the different sizes of the regions also affect the comparability between the regions (Pfeifer et al., 2015). For example, very small regions cover only a few to a single grid point, whereas large regions include multiple grid points. Therefore, averaging over the spatial extent of a large and a small region might not be comparable, as extremes might be smoothed spatially for the large regions (Pfeifer et al., 2015).

5.2 Added value of ensemble spread

For the analysis of the ETCCDI indices, the ensemble spread is incorporated in the climate signal maps by Pfeifer et al. (2015) using a threshold for the standard deviation. The threshold divides the ensemble spread for each region into either small or large. The climate change signals of the considered indices show a large ensemble spread for most of the regions.

The large ensemble spread for the temperature-based indices is primarily due to the warm bias in the temperature data of the MIROC-driven simulations. This warm bias leads to high ensemble spreads in the indices, as this bias is carried forward in the derivation of the indices. Since the indices presented here are defined by a temperature threshold, a warm bias in the temperature data artificially increases the number of days that exceed this threshold.

The large ensemble spread for the precipitation-based indices is reflected in the diverging precipitation signals for summer and winter (Figs. 4.3 and 4.6). As discussed before, the climate change signals in seasonal mean precipitation not only show diverging signs among the individual models but also do not agree on the significance of the signals (Section 4.1.1).

Although a large ensemble spread is not necessarily a drawback, the question is how to communicate it (Jebeile and Barberousse, 2021). The initial climate signal map approach by Pfeifer et al. (2015) categorized the climate change signals into three classes: robust, non-robust, and opposite direction of change. The incorporation of the ensemble spread into the climate signal maps adds new information on the precision of the illustrated signal magnitude in the climate signal maps. This allows a more nuanced interpretation of robust and non-robust signals. Four categories of climate change signals can be distinguished: robust and precise, robust but not precise, weak or no signal, and high uncertainty (see Figures 1.1 and B.1). With this approach, robust signals are further evaluated based on the precision of the signals, and non-robust signals can be distinguished into a lack of signal and a lack of information. This distinction is important, as highlighted by Tebaldi et al. (2011), who show that it matters whether a signal is not robust because there is no climate change, or because the disagreement among the models of an ensemble is too large to identify a climate change signal.

This classification is shown for several indices. The relevance of further distinction of robust signals is shown, for example, in the climate change signals for tropical nights (Figure 4.15) or days with R20mm or R30mm (Figure 4.18). Here, the results show with both robust and precise signals (colored and hatched), and robust but not precise signals (colored and unhatched). For the latter, the ensemble median does not represent the magnitude of the change precisely, and thus the classification into small, medium, and large signals within the climate signal maps is not reliable. This is, for example, shown for region G in Figure 4.1. Although the ensemble median is classified as a medium signal, the values range from small, medium, to large signals.

The more nuanced distinction for non-robust signals is reflected, for example, in the climate change signals for CWD (Figure 4.21). Most of the regions show non-robust climate change signals. Including ensemble spread helps to separate regions where no climate change signal is projected

from regions where the ensemble shows high uncertainty in the climate change signal. This adds valuable information for the interpretation of the non-robust signals, which is particularly relevant for determining adaptation measures. The information on whether a region shows no climate change signal or whether the uncertainty is too high affects the next course of action. In the case that there is no climate change signal, no adaptation measure is needed. In the case where the uncertainty is too high, the next course of action depends on the application and potential consequences. Here, the decision-maker can decide on whether further studies should be conducted to obtain more clarity.

Similarly, if a region shows a robust climate change signal, the precision of the magnitude of the signal is important for the decision. In the case that the climate change signal is robust and precise, appropriate adaptation measures can be applied. If the signal is robust but not precise, the decision-maker knows that there is a reliable climate change signal for their region, but the magnitude of the signal is uncertain. In this case, the range of signals or the maximum value should be considered in the decision process. Depending on the potential consequences of the considered index or variable, further studies can potentially provide more precise information on the climate change signal. This clearly shows that the suggested incorporation of the ensemble spread into the climate signal maps adds valuable information for its user on how to interpret the results and how to derive the next course of action.

5.2.1 Limitations of the extended climate signal map approach

However, there are some limitations to the presented method here. The ensemble spread is evaluated based on a threshold for the standard deviation. Although the standard deviation is a good choice as a measure due to its robustness against single outliers and as it is a commonly known statistical measure, it has the same unit as the considered variable or index. This complicates finding a universal threshold definition that applies to all indices and enables a good comparison.

For precipitation-based indices, a threshold based on 30 % the mean robust signal is set. As a result, the threshold increases as the signal increases, which makes it suitable to compare indices that show different magnitudes and enables a universal threshold definition and comparison between different indices. However, the dependency on the signal strength leads to differences among the GWLs. Furthermore, it is not a given that robust signals show a better ensemble spread than non-robust signals. Therefore, deriving the threshold from only robust signals does not guarantee a better choice. However, as robust signals agree on the same sign of change and exhibit significant climate change signals, which exclude outliers of opposite or very small signals from setting the threshold.

Moreover, if the climate change signals are very small, the resulting threshold is also very small. This is reflected in the climate change signals for CWD (Figure 4.20). As the climate change signals are very small for CWD, the resulting threshold is very small, and no region exhibits a standard deviation below the threshold.

This clearly demonstrates the limitations of a universal threshold definition. Instead of using a universal threshold definition that applies to all indices, it is more effective to choose the threshold

individually for each variable or index. This ensures that the threshold is set to realistic values. To further enhance the usefulness of climate signal maps, the threshold for the ensemble spread should ideally be determined in collaboration with the targeted users and tailored to the specific application.

Since the initial universal threshold applied to the climate signal maps of CWD did not perform as intended (Figure 4.20), a new threshold was tested for a small ensemble spread illustrated in Figure 4.21. The adjusted threshold allows for the classification of non-robust signals with either a small or a large ensemble spread. This shows that if the ensemble spread threshold is defined individually, incorporating the ensemble spread adds valuable information to the climate signal maps.

For temperature-based indices, the thresholds are chosen based on the levels of the categories of the climate signal maps. These thresholds serve as a suggestion. Ultimately, the choice of the threshold depends on the user and application and should be chosen accordingly.

To mark regions with a small ensemble spread, hatching was suggested and applied here. This visualization might not be ideal due to several factors. At first, it is difficult to recognize whether small regions show a small spread or not, as hatching is difficult to identify for small regions. Second, other studies use hatching to mark significant or robust signals (Jacob et al., 2014; Tebaldi et al., 2011). This could lead to a misinterpretation of the hatching. Third, in the results presented here, only a small number of regions exhibit climate change signals with a small ensemble spread. For example, no regions with a small ensemble spread are identified for the climate signals in CDD (Figure 4.19). The absence of regions with hatching can falsely lead to the conclusion that no additional criterion, except for robustness, is applied. Moreover, if no signals with a small ensemble spread are identified, no hatched regions are displayed, which can make the distinction between large and small spread unclear to the user. However, adding a new layer of information on the ensemble spread through hatching is still a well-suited choice, if clearly communicated what the hatching indicates.

5.2.2 Applicability of the extended climate signal maps

Apart from the methodological aspects, the question of whether the ensemble spread adds value to the climate signal map is closely connected to its applicability for users. This includes two key aspects: first, the selection of the indices and their applicability for the user, and second, whether the extended climate signal maps are useful in practice.

This thesis analyze the robustness of climate change signals of extremes using the ETCCDI indices. The ETCCDI indices are suitable to assess changes in extremes like heatwaves or heavy precipitation events, and offer a good comparison to other studies. However, the overall goal is to find an effective way of communicating climate information to users. Therefore, more application-oriented indices are recommended to improve the applicability for users outside of the scientific context. Hackenbruch et al. (2017) showed that climate information tailored to the specific needs of the decision-makers and the targeted application improves the communication of

climate information. This can be achieved by tailored climate indices that are directly relevant for the targeted application and support decision-making.

Furthermore, the bias correction is expected to reduce the ensemble spread in the climate change signals of the indices. Therefore, the analysis should be repeated based on a bias-corrected dataset to evaluate the added value of the ensemble spread.

The value of the extended climate signal maps ultimately is dependent on their applicability in practice. Several difficulties in providing climate information effectively to users have been identified, including the format in which the information is provided (Brasseur and Gallardo, 2016). Moreover, regarding climate information products, there is a gap between the focus on improving the data and focusing on supporting the decision (Findlater et al., 2021). Incorporating the ensemble spread into the climate signal maps aims to close this gap by providing additional support in the interpretation of the climate change signals. This is achieved by distinguishing between four cases of reliability and precision instead of solely differentiating between robust and non-robust signals as in the approach by Pfeifer et al. (2015). However, to truly assess the added value of the extended climate signal map approach presented here, it is essential to collaborate directly with users and include their needs in setting the ensemble spread threshold and limits for the climate signal maps (Findlater et al., 2021; Brasseur and Gallardo, 2016; Buontempo et al., 2020). The presented approach should therefore be seen as a suggestion to support useful climate information.

6 Conclusions and Outlook

Extreme events are becoming more frequent and intense due to climate change (IPCC, 2023b). Consequently, there is a growing need for climate adaptation strategies. While climate models provide important insights into the future development of extreme events under climate change, their coarse resolution and inherent uncertainties often hinder a direct application in adaptation strategies (Hackenbruch et al., 2017; Lorenz et al., 2017). One way to address this is to use high-resolution climate data from convection-permitting climate models and assess the robustness of the climate change signals.

The analysis is based on the NUKLEUS ensemble (Sieck et al., 2021), which is a convection-permitting climate model ensemble. Climate change signals are evaluated for two global warming levels of +2K and +3K relative to the reference period of 1961-1990. The analysis includes changes in winter and summer precipitation and temperature, along with eight ETCCDI indices representing extreme events: TR, ID, WSDI, R10mm, R20mm, R30mm, CDD, and CWD.

For each variable and index, the climate change signals and their robustness are evaluated using climate signal maps introduced by Pfeifer et al. (2015). For the indices, the climate signal maps are extended by a measure for the ensemble spread, aiming to provide additional information on the precision of the climate change signals.

Which climate change signals of ETCCDI-indices derived from the NUKLEUS ensemble can be classified as robust?

The results show a weak increase in mean winter precipitation, and only a few robust climate change signals for both +2K and +3K global warming levels can be identified. For summer, a decrease in mean precipitation is shown, with the strongest and robust signals in the Northwest of Germany that intensify for the +3K global warming level. There is a generally low number of robust climate change signals in the +2K global warming level due to large internal variability in the near-future precipitation response compared to the magnitude of the signals. The absence of robust climate change signals in winter is attributed to a lack of significance of the individual simulations, whereas the limiting factor in summer is a lack of agreement on the sign of change.

All regions show robust climate change signals for winter and summer temperatures, and both global warming levels of +2K and +3K. The difference between the signals in winter and summer is less pronounced than the difference between the two global warming levels. The climate change signals of the temperature-based indices also show robustness for all regions and both global warming levels. There is an increase in the number of tropical nights, with the strongest signals

in the Upper Rhine Valley and urban areas. The number of ice days decreases for all regions and shows a West-to-East gradient in the signal strength, with the strongest climate change signals in the East of Germany. These findings are consistent with Pinto et al. (2025), who also showed an increase in the number of tropical nights, particularly for the Upper Rhine Valley, as well as a pronounced decrease in ice days with a stronger signal in eastern Germany and for the +3K global warming level. The warm spell duration index increases under both warming scenarios, indicating an intensification of the length and intensity of warm spells.

The indices for extreme precipitation show more robust climate change signals than the mean precipitation signals. Regarding the frequency of heavy precipitation events (R10mm, R20mm, and R30mm), the most extreme index (R30mm) exhibits the strongest relative climate change signals. The signals intensify, and there are more robust climate change signals for the +3K compared to the +2K global warming level. The intensification of the very extreme precipitation indices (R20mm, R30mm) compared to the moderate ones (R10mm) is consistent with previous findings (Ban et al., 2015; Hundhausen et al., 2024; Pinto et al., 2025). The maximum length of dry spells (CDD) increases for both global warming scenarios, with the strongest and robust signals located in Central and Western Germany. There is no robust trend in the maximum length of wet spells (CWD) under the two global warming levels +2K and +3K. For CWD, the climate change signals are very weak, and there are only very few regions with a robust climate change signal.

Does the consideration of the ensemble spread add value to the climate signal maps?

In most of the regions, the climate change signals of the ETCCDI indices exhibit a large ensemble spread. The large ensemble spread of the precipitation-based indices is attributed to the diverging signals of the mean precipitation among the ensemble members. For the temperature-based indices, the large ensemble spread is attributed to a warm bias in the simulations driven by the initial and boundary conditions of the global climate model MIROC, which deviates from the other global climate models. This warm bias is also found in the NUKLEUS-EUR11 simulations (Cusinato et al., 2025).

The results show that the assessment of the ensemble spread alongside the robustness adds valuable information on the precision of a climate change signal. Four distinct classifications of climate change signals are determined based on the robustness and precision (see classification process, Figure 1.1). Robustness indicates that the climate change signal is reliable and can be further distinguished between precise and not precise. This provides valuable information for a decision-maker, whether the magnitude of the projected climate change signal is certain or if a range of possible values needs to be considered. The relevance of this additional distinction becomes evident in the case of tropical nights, where all regions show robust climate change signals, but the large ensemble spread in most regions highlights considerable uncertainty in the magnitude of the signals, particularly for global warming level +3K. This increased variability, especially for the global warming level +3K is consistent with other findings (Pinto et al., 2025). Non-robust signals can be further divided into the case where no climate change signal can be derived, or the case where the uncertainty is too high to derive any information. By introducing a more differentiated classification of robust and non-robust signals, this classification approach is proposed to support decision-makers

in the interpretation of the climate information and in determining the next course of action. This added value becomes evident in the case of consecutive wet days. Although most of the climate change signals are not robust, the additional distinction based on the ensemble spread shows that many regions exhibit no or a very weak climate change signal.

This thesis applied uncorrected climate data obtained from the NUKLEUS ensemble due to the lack of bias-corrected data at the time of the analysis. Future work with bias-corrected data is expected to reduce the ensemble spread of the climate change signals, particularly for temperature-based indices, which are affected by the warm bias in the MIROC-driven simulations. Additionally, the comparison to the results of Pfeifer et al. (2015) underlines the strong influence of the model selection on the results, emphasizing the important role of the ensemble design. This is in line with previous studies showing the importance of the model selection on the results (Evans et al., 2014; Herger et al., 2018). To fully assess the effect of the model selection, the comparison to Pfeifer et al. (2015) should be repeated with the full NUKLEUS-EUR11 ensemble. Furthermore, the differences between the NUKLEUS-EUR11 and NUKLEUS-CEU3 ensembles can be evaluated to quantify the influence of using global warming levels instead of transient simulations, as well as the effects of the higher resolution and using convection-permitting models.

The proposed classification framework can be applied to all kinds of climate variables and indices. However, the results show that the visualization using climate signal maps is particularly useful for precipitation signals and precipitation-based indices. For a meaningful classification, setting suited thresholds for the ensemble spread and for the limits of the climate signal maps is essential and depends on the application and considered variable. To ensure the applicability in practice, the thresholds should be tailored to the application in close collaboration with the targeted users. Additionally, to improve the applicability of the results for adaptation planning, more user-oriented indices tailored to the user and application can be employed (e.g. Schipper et al., 2019; Pinto et al., 2025). To fully assess the added value of the extended climate signal map approach, collaboration with the targeted users is crucial (Findlater et al., 2021; Brasseur and Gallardo, 2016; Buontempo et al., 2020). The proposed framework should therefore be seen as a suggestion to provide useful climate information, with the next step being direct engagement with decision-makers and stakeholders.

Abbreviations

GCM	Global Climate Model
IPCC	Intergovernmental Panel on Climate Change
RCM	Regional Climate Model
ETCCDI	Expert Team on Climate Change Detection and Indices
CPM	Convection-permitting Climate Model
NUKLEUS	Nutzbare Lokale Klimainformationen für Deutschland, Actionable local climate information for Germany
CMIP	Coupled Model Intercomparison Project
RCP	Representative Concentration Pathways
SSP	Shared Socioeconomic Pathway
ESM	Earth System Model
SNR	Signal-to-noise Ratio
GWL	Global Warming Level
GWL2K	Global Warming Level +2K
GWL3K	Global Warming Level +3K
WSDI	Warm spell duration index
ID	Ice days
TR	Tropical nights
R10mm	Annual count of days with daily precipitation amount $\geq$ 10 mm
R20mm	Annual count of days with daily precipitation amount $\geq$ 20 mm
R30mm	Annual count of days with daily precipitation amount $\geq$ 30 mm

CDD   Maximum length of dry spells

CWD   Maximum length of wet spells

CDD   Maximum length of dry spells

Bibliography

Ban, N., J. Schmidli, and C. Schär, 2015: Heavy precipitation in a changing climate: Does short-term summer precipitation increase faster? *Geophysical Research Letters*, **42 (4)**, 1165–1172, https://doi.org/10.1002/2014GL062588, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2014GL062588.

Ban, N., and Coauthors, 2021: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Clim Dyn*, **57 (1-2)**, 275–302, https://doi.org/10.1007/s00382-021-05708-w, URL https://link.springer.com/10.1007/s00382-021-05708-w.

Brasseur, G. P., and L. Gallardo, 2016: Climate services: Lessons learned and future prospects. *Earth's Future*, **4 (3)**, 79–89, https://doi.org/10.1002/2015EF000338, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015EF000338.

Buontempo, C., and Coauthors, 2020: Fostering the development of climate services through Copernicus Climate Change Service (C3S) for agriculture applications. *Weather and Climate Extremes*, **27**, 100 226, https://doi.org/10.1016/j.wace.2019.100226, URL https://linkinghub.elsevier.com/retrieve/pii/S2212094719300994.

Buontempo, C., and Coauthors, 2022: The Copernicus Climate Change Service: Climate Science in Action. *Bulletin of the American Meteorological Society*, **103 (12)**, E2669–E2687, https://doi.org/10.1175/BAMS-D-21-0315.1, URL https://journals.ametsoc.org/view/journals/bams/103/12/BAMS-D-21-0315.1.xml.

Chen, J., R. Arsenault, F. P. Brissette, and S. Zhang, 2021: Climate Change Impact Studies: Should We Bias Correct Climate Model Outputs or Post-Process Impact Model Outputs? *Water Resources Research*, **57 (5)**, https://doi.org/10.1029/2020wr028638, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020WR028638, publisher: American Geophysical Union (AGU).

Christensen, J. H., F. Boberg, O. B. Christensen, and P. Lucas-Picher, 2008: On the need for bias correction of regional climate change projections of temperature and precipitation. *Geophysical Research Letters*, **35 (20)**, https://doi.org/10.1029/2008gl035694, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2008GL035694, publisher: American Geophysical Union (AGU).

Christensen, O. B., and E. Kjellström, 2022: Filling the matrix: an ANOVA-based method to emulate regional climate model simulations for equally-weighted properties of ensembles of opportunity. *Clim Dyn*, **58 (9-10)**, 2371–2385, https://doi.org/10.1007/s00382-021-06010-5, URL https://link.springer.com/10.1007/s00382-021-06010-5.

Coppola, E., and Coauthors, 2021: Assessment of the European Climate Projections as Simulated by the Large EURO-CORDEX Regional and Global Climate Model Ensemble. *JGR Atmospheres*, **126 (4)**, e2019JD032 356, https://doi.org/10.1029/2019JD032356, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019JD032356.

Cusinato, E., C. Braun, H. Feldmann, B. Geyer, P. Ludwig, K. Trachte, and J. G. Pinto, 2025: From GCMs to RCMs: Consistency Assessment of Surface Water and Energy Fluxes. URL https://meetingorganizer.copernicus.org/EGU25/EGU25-10693.html, https://doi.org/10.5194/egusphere-egu25-10693.

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: the role of internal variability. *Clim Dyn*, **38 (3-4)**, 527–546, https://doi.org/10.1007/s00382-010-0977-x, URL http://link.springer.com/10.1007/s00382-010-0977-x, publisher: Springer Science and Business Media LLC.

Donat, M., L. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013: Global Land-Based Datasets for Monitoring Climatic Extremes. *Bulletin of the American Meteorological Society*, **94 (7)**, 997–1006, https://doi.org/10.1175/bams-d-12-00109.1, URL https://journals.ametsoc.org/doi/10.1175/BAMS-D-12-00109.1, publisher: American Meteorological Society.

Dosio, A., and E. M. Fischer, 2018: Will Half a Degree Make a Difference? Robust Projections of Indices of Mean and Extreme Climate in Europe Under 1.5°C, 2°C, and 3°C Global Warming. *Geophysical Research Letters*, **45 (2)**, 935–944, https://doi.org/10.1002/2017GL076222, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2017GL076222.

Déqué, M., S. Somot, E. Sanchez-Gomez, C. M. Goodess, D. Jacob, G. Lenderink, and O. B. Christensen, 2012: The spread amongst ENSEMBLES regional scenarios: regional climate models, driving general circulation models and interannual variability. *Clim Dyn*, **38 (5-6)**, 951–964, https://doi.org/10.1007/s00382-011-1053-x, URL http://link.springer.com/10.1007/s00382-011-1053-x.

Döscher, R., and Coauthors, 2022: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geosci. Model Dev.*, **15 (7)**, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, URL https://gmd.copernicus.org/articles/15/2973/2022/.

Edwards, P. N., 2011: History of climate modeling. *WIREs Climate Change*, **2 (1)**, 128–139, https://doi.org/10.1002/wcc.95, URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.95, publisher: Wiley.

European Commission. Statistical Office of the European Union., 2018: *Regions in the European Union: nomenclature of territorial units for statistics (NUTS) : 2024 edition.* Publications Office,

LU, URL https://data.europa.eu/doi/10.2785/714519.

Evans, J. P., F. Ji, C. Lee, P. Smith, D. Argüeso, and L. Fita, 2014: Design of a regional climate modelling projection ensemble experiment – NARCliM. *Geosci. Model Dev.*, **7 (2)**, 621–629, https://doi.org/10.5194/gmd-7-621-2014, URL https://gmd.copernicus.org/articles/7/621/2014/.

Evin, G., S. Somot, and B. Hingray, 2021: Balanced estimate and uncertainty assessment of European climatechange using the large EURO-CORDEX regional climate modelensemble. URL https://esd.copernicus.org/preprints/esd-2021-8/esd-2021-8.pdf, https://doi.org/10.5194/esd-2021-8.

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. *Geosci. Model Dev.*, **9 (5)**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, URL https://gmd.copernicus.org/articles/9/1937/2016/.

Fantini, A., and Coauthors, 2018: Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. *Clim Dyn*, **51 (3)**, 877–900, https://doi.org/10.1007/s00382-016-3453-4, URL http://link.springer.com/10.1007/s00382-016-3453-4.

Findlater, K., S. Webber, M. Kandlikar, and S. Donner, 2021: Climate services promise better decisions but mainly focus on better data. *Nat. Clim. Chang.*, **11 (9)**, 731–737, https://doi.org/10.1038/s41558-021-01125-3, URL https://www.nature.com/articles/s41558-021-01125-3.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 741–866.

Fosser, G., and Coauthors, 2024: Convection-permitting climate models offer more certain extreme rainfall projections. *npj Clim Atmos Sci*, **7 (1)**, 51, https://doi.org/10.1038/s41612-024-00600-w, URL https://www.nature.com/articles/s41612-024-00600-w.

Friedrich, T., and Coauthors, 2024: Teilbericht: Kommunalbefragung, Klimaanpassung 2023. Tech. rep., Umweltbundesamt.

Georgoulias, A. K., D. Akritidis, A. Kalisoras, J. Kapsomenakis, D. Melas, C. S. Zerefos, and P. Zanis, 2022: Climate change projections for Greece in the 21st century from high-resolution EURO-CORDEX RCM simulations. *Atmospheric Research*, **271**, 106 049, https://doi.org/10.1016/j.atmosres.2022.106049, URL https://linkinghub.elsevier.com/retrieve/pii/S0169809522000357.

Giorgi, F., and W. J. Gutowski, 2015: Regional Dynamical Downscaling and the CORDEX Initiative. *Annu. Rev. Environ. Resour.*, **40 (1)**, 467–490, https://doi.org/10.1146/annurev-environ-102014-021217, URL https://www.annualreviews.org/doi/10.1146/

annurev-environ-102014-021217, publisher: Annual Reviews.

Hackenbruch, J., T. Kunz-Plapp, S. Müller, and J. Schipper, 2017: Tailoring Climate Parameters to Information Needs for Local Adaptation to Climate Change. *Climate*, **5 (2)**, 25, https://doi.org/10.3390/cli5020025, URL http://www.mdpi.com/2225-1154/5/2/25.

Hawkins, E., and R. Sutton, 2009: The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bull. Amer. Meteor. Soc.*, **90 (8)**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1, URL https://journals.ametsoc.org/doi/10.1175/2009BAMS2607.1.

He, C., and Coauthors, 2024: Nocturnal heat exposure and stroke risk. *European Heart Journal*, **45 (24)**, 2158–2166, https://doi.org/10.1093/eurheartj/ehae277, URL https://academic.oup.com/eurheartj/article/45/24/2158/7676519.

Herger, N., G. Abramowitz, R. Knutti, O. Angélil, K. Lehmann, and B. M. Sanderson, 2018: Selecting a climate model subset to optimise key ensemble properties. *Earth Syst. Dynam.*, **9 (1)**, 135–151, https://doi.org/10.5194/esd-9-135-2018, URL https://esd.copernicus.org/articles/9/135/2018/.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart J Royal Meteoro Soc*, **146 (730)**, 1999–2049, https://doi.org/10.1002/qj.3803, URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.3803, publisher: Wiley.

Hewitt, C., S. Mason, and D. Walland, 2012: The Global Framework for Climate Services. *Nature Clim Change*, **2 (12)**, 831–832, https://doi.org/10.1038/nclimate1745, URL https://www.nature.com/articles/nclimate1745.

Hewitt, C. D., 2004: Ensembles-based predictions of climate changes and their impacts. *EoS Transactions*, **85 (52)**, 566–566, https://doi.org/10.1029/2004EO520005, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2004EO520005.

Hewitt, C. D., N. Golding, P. Zhang, T. Dunbar, P. E. Bett, J. Camp, T. D. Mitchell, and E. Pope, 2020: The Process and Benefits of Developing Prototype Climate Services—Examples in China. *J Meteorol Res*, **34 (5)**, 893–903, https://doi.org/10.1007/s13351-020-0042-6, URL https://link.springer.com/10.1007/s13351-020-0042-6.

Hewitt, C. D., R. C. Stone, and A. B. Tait, 2017: Improving the use of climate information in decision-making. *Nature Clim Change*, **7 (9)**, 614–616, https://doi.org/10.1038/nclimate3378, URL https://www.nature.com/articles/nclimate3378.

Hu, G., and C. L. E. Franzke, 2020: Evaluation of Daily Precipitation Extremes in Reanalysis and Gridded Observation-Based Data Sets Over Germany. *Geophysical Research Letters*, **47 (18)**, https://doi.org/10.1029/2020gl089624, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL089624, publisher: American Geophysical Union (AGU).

Hundhausen, M., H. Feldmann, R. Kohlhepp, and J. G. Pinto, 2024: Climate change signals of extreme precipitation return levels for Germany in a transient convection-permitting simulation

ensemble. *Intl Journal of Climatology*, **44 (5)**, 1454–1471, https://doi.org/10.1002/joc.8393, URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.8393.

Hundhausen, M., H. Feldmann, N. Laube, and J. G. Pinto, 2023: Future heat extremes and impacts in a convection-permitting climate ensemble over Germany. *Nat. Hazards Earth Syst. Sci.*, **23 (8)**, 2873–2893, https://doi.org/10.5194/nhess-23-2873-2023, URL https://nhess.copernicus.org/articles/23/2873/2023/.

IPCC, 2023: *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* 1st ed., Cambridge University Press, https://doi.org/10.1017/9781009157896, URL https://www.cambridge.org/core/product/identifier/9781009157896/type/book.

IPCC, 2023a: Water Cycle Changes. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, 1055–1210, https://doi.org/10.1017/9781009157896.010, URL https://www.cambridge.org/core/product/1E7F2B90411645E36048670B826F131A.

IPCC, 2023b: Weather and Climate Extreme Events in a Changing Climate. *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 1513–1766.

Jacob, D., and Coauthors, 2012: Assessing the Transferability of the Regional Climate Model REMO to Different COordinated Regional Climate Downscaling EXperiment (CORDEX) Regions. *Atmosphere*, **3 (1)**, 181–199, https://doi.org/10.3390/atmos3010181, URL https://www.mdpi.com/2073-4433/3/1/181.

Jacob, D., and Coauthors, 2014: EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg Environ Change*, **14 (2)**, 563–578, https://doi.org/10.1007/s10113-013-0499-2, URL http://link.springer.com/10.1007/s10113-013-0499-2.

Jacob, D., and Coauthors, 2020: Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community. *Reg Environ Change*, **20 (2)**, 51, https://doi.org/10.1007/s10113-020-01606-9, URL https://link.springer.com/10.1007/s10113-020-01606-9.

Jebeile, J., and A. Barberousse, 2021: Model spread and progress in climate modelling. *Euro Jnl Phil Sci*, **11 (3)**, 66, https://doi.org/10.1007/s13194-021-00387-0, URL https://link.springer.com/10.1007/s13194-021-00387-0.

Kabisch, N., F. Remahne, C. Ilsemann, and L. Fricke, 2023: The urban heat island under extreme heat conditions: a case study of Hannover, Germany. *Sci Rep*, **13 (1)**, 23 017, https://doi.org/10.1038/s41598-023-49058-5, URL https://www.nature.com/articles/s41598-023-49058-5.

Kadow, C., and Coauthors, 2021: Introduction to Freva – A Free Evaluation System Framework for Earth System Modeling. *JORS*, **9 (1)**, 13, https://doi.org/10.5334/jors.253, URL https://

openresearchsoftware.metajnl.com/article/10.5334/jors.253/.

Kendon, E. J., D. P. Rowell, R. G. Jones, and E. Buonomo, 2008: Robustness of Future Changes in Local Precipitation Extremes. *Journal of Climate*, **21 (17)**, 4280–4297, https://doi.org/10.1175/2008JCLI2082.1, URL http://journals.ametsoc.org/doi/10.1175/2008JCLI2082.1.

Kendon, E. J., and Coauthors, 2017: Do Convection-Permitting Regional Climate Models Improve Projections of Future Precipitation Change? *Bulletin of the American Meteorological Society*, **98 (1)**, 79–93, https://doi.org/10.1175/BAMS-D-15-0004.1, URL https://journals.ametsoc.org/doi/10.1175/BAMS-D-15-0004.1.

Kjellström, E., and Coauthors, 2018: European climate change at global mean temperature increases of 1.5 and 2 $^{\circ}$C above pre-industrial conditions as simulated by the EURO-CORDEX regional climate models. *Earth Syst. Dynam.*, **9 (2)**, 459–478, https://doi.org/10.5194/esd-9-459-2018, URL https://esd.copernicus.org/articles/9/459/2018/.

Knutti, R., and J. Sedláček, 2013: Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Clim Change*, **3 (4)**, 369–373, https://doi.org/10.1038/nclimate1716, URL https://www.nature.com/articles/nclimate1716.

Lemos, M. C., C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nature Clim Change*, **2 (11)**, 789–794, https://doi.org/10.1038/nclimate1614, URL https://www.nature.com/articles/nclimate1614.

Lenderink, G., and E. Van Meijgaard, 2008: Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geosci*, **1 (8)**, 511–514, https://doi.org/10.1038/ngeo262, URL https://www.nature.com/articles/ngeo262.

Lorenz, S., S. Dessai, P. M. Forster, and J. Paavola, 2017: Adaptation planning and the use of climate change projections in local government in England and Germany. *Reg Environ Change*, **17 (2)**, 425–435, https://doi.org/10.1007/s10113-016-1030-3, URL http://link.springer.com/10.1007/s10113-016-1030-3.

Mann, H. B., and D. R. Whitney, 1947: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, **18 (1)**, 50–60, URL http://www.jstor.org/stable/2236101, publisher: Institute of Mathematical Statistics.

Mauritsen, T., and Coauthors, 2019: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2. *J Adv Model Earth Syst*, **11 (4)**, 998–1038, https://doi.org/10.1029/2018MS001400, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018MS001400.

Müller, W. A., and Coauthors, 2018: A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *J Adv Model Earth Syst*, **10 (7)**, 1383–1413, https://doi.org/10.1029/2017MS001217, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2017MS001217.

O'Neill, B. C., E. Kriegler, K. Riahi, K. L. Ebi, S. Hallegatte, T. R. Carter, R. Mathur, and D. P. Van Vuuren, 2014: A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, **122 (3)**, 387–400, https://doi.org/10.1007/s10584-013-0905-2, URL http://link.springer.com/10.1007/s10584-013-0905-2.

Pfeifer, S., and Coauthors, 2015: Robustness of Ensemble Climate Projections Analyzed with Climate Signal Maps: Seasonal and Extreme Precipitation for Germany. *Atmosphere*, **6 (5)**, 677–698, https://doi.org/10.3390/atmos6050677, URL http://www.mdpi.com/2073-4433/6/5/677.

Pham, T. V., and Coauthors, 2021: ICON in Climate Limited-area Mode (ICON release version 2.6.1): a new regional climate model. *Geosci. Model Dev.*, **14 (2)**, 985–1005, https://doi.org/10.5194/gmd-14-985-2021, URL https://gmd.copernicus.org/articles/14/985/2021/.

Pichelli, E., and Coauthors, 2021: The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation. *Clim Dyn*, **56 (11-12)**, 3581–3602, https://doi.org/10.1007/s00382-021-05657-4, URL https://link.springer.com/10.1007/s00382-021-05657-4.

Pinto, J., M. Hundhausen, A. Weber, R. Kohlhepp, C. Mihalyfi-Dean, J. W. Schipper, and H. Feldmann, 2025: User-relevant climate indices and associated uncertainties from transient convection-permitting climate model projections. *International Journal of Climatology (under review)*.

Prein, A. F., A. Gobiet, M. Suklitsch, H. Truhetz, N. K. Awan, K. Keuler, and G. Georgievski, 2013: Added value of convection permitting seasonal simulations. *Clim Dyn*, **41 (9-10)**, 2655–2677, https://doi.org/10.1007/s00382-013-1744-6, URL http://link.springer.com/10.1007/s00382-013-1744-6.

Prein, A. F., R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland, 2017: The future intensification of hourly precipitation extremes. *Nature Clim Change*, **7 (1)**, 48–52, https://doi.org/10.1038/nclimate3168, URL https://www.nature.com/articles/nclimate3168.

Prein, A. F., and Coauthors, 2015: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, **53 (2)**, 323–361, https://doi.org/10.1002/2014RG000475, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2014RG000475.

Prein, A. F., and Coauthors, 2016: Precipitation in the EURO-CORDEX 0.11◦ and 0.44◦ simulations: high resolution, high benefits? *Clim Dyn*, **46 (1-2)**, 383–412, https://doi.org/10.1007/s00382-015-2589-y, URL http://link.springer.com/10.1007/s00382-015-2589-y.

Riach, N., and R. Glaser, 2024: Local climate services. Can municipal climate profiles help improve climate literacy? *Climate Services*, **34**, 100 449, https://doi.org/10.1016/j.cliser.2024.100449, URL https://linkinghub.elsevier.com/retrieve/pii/S2405880724000049.

Ritzhaupt, N., and D. Maraun, 2023: Consistency of Seasonal Mean and Extreme Precipitation Projections Over Europe Across a Range of Climate Model Ensembles. *JGR Atmospheres*, **128 (1)**,

e2022JD037 845, https://doi.org/10.1029/2022JD037845, URL https://agupubs.onlinelibrary. wiley.com/doi/10.1029/2022JD037845.

Rockel, B., A. Will, and A. Hense, 2008: The Regional Climate Model COSMO-CLM (CCLM). *metz*, **17 (4)**, 347–348, https://doi.org/10.1127/0941-2948/2008/0309, URL http://www.schweizerbart.de/papers/metz/detail/17/56726/The_Regional_Climate_Model_ COSMO_CLM_CCLM?af=crossref.

Räisänen, J., 2007: How reliable are climate models? *Tellus A: Dynamic Meteorology and Oceanography*, **59 (1)**, 2, https://doi.org/10.1111/j.1600-0870.2006.00211.x, URL https://a. tellusjournals.se/article/10.1111/j.1600-0870.2006.00211.x/, publisher: Stockholm University Press.

Schipper, J. W., J. Hackenbruch, H. S. Lentink, and K. Sedlmeier, 2019: Integrating Adaptation Expertise into Regional Climate Data Analyses through Tailored Climate Parameters. *metz*, **28 (1)**, 41–57, https://doi.org/10.1127/metz/2019/0878, URL http://www.schweizerbart.de/papers/metz/ detail/28/90522/Integrating_Adaptation_Expertise_into_Regional_Cli?af=crossref.

Seaby, L., J. Refsgaard, T. Sonnenborg, S. Stisen, J. Christensen, and K. Jensen, 2013: Assessment of robustness and significance of climate change signals for an ensemble of distribution-based scaled climate projections. *Journal of Hydrology*, **486**, 479–493, https://doi.org/10.1016/j.jhydrol. 2013.02.015, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169413001273.

Sieck, K., B. Tiedje, H. Feldmann, and J. Pinto, 2021: NUKLEUS - User-relevant and applicable kilometer-scale climate information for Germany. URL https://meetingorganizer.copernicus.org/ EGU21/EGU21-12137.html, https://doi.org/10.5194/egusphere-egu21-12137.

Sieck, K., and Coauthors, 2025: NUKLEUS - A First Kilometer Scale Multi-model Climate Ensemble for Germany: Evaluation. *In preparation.*

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *JGR Atmospheres*, **118 (4)**, 1716–1733, https://doi.org/10.1002/jgrd.50203, URL https://agupubs. onlinelibrary.wiley.com/doi/10.1002/jgrd.50203.

Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *JGR Atmospheres*, **118 (6)**, 2473–2493, https://doi.org/10.1002/jgrd.50188, URL https://agupubs.onlinelibrary.wiley. com/doi/10.1002/jgrd.50188.

Spinoni, J., G. Naumann, and J. V. Vogt, 2017: Pan-European seasonal trends and recent changes of drought frequency and severity. *Global and Planetary Change*, **148**, 113–130, https://doi.org/10.1016/j.gloplacha.2016.11.013, URL https://linkinghub.elsevier.com/retrieve/ pii/S0921818116301801, publisher: Elsevier BV.

Suarez-Gutierrez, L., W. A. Müller, C. Li, and J. Marotzke, 2020: Dynamical and thermodynamical drivers of variability in European summer heat extremes. *Clim Dyn*, **54 (9-10)**, 4351–4366, https://doi.org/10.1007/s00382-020-05233-2, URL https://link.springer.com/10.1007/s00382-020-05233-2, publisher: Springer Science and Business Media LLC.

Tatebe, H., and Coauthors, 2019: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.*, **12 (7)**, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019, URL https://gmd.copernicus.org/articles/12/2727/2019/.

Tebaldi, C., J. M. Arblaster, and R. Knutti, 2011: Mapping model agreement on future climate projections: MAPPING MODEL AGREEMENT. *Geophys. Res. Lett.*, **38 (23)**, n/a–n/a, https://doi.org/10.1029/2011GL049863, URL http://doi.wiley.com/10.1029/2011GL049863.

Teutschbein, C., and J. Seibert, 2010: Regional Climate Models for Hydrological Impact Studies at the Catchment Scale: A Review of Recent Modeling Strategies. *Geography Compass*, **4 (7)**, 834–860, https://doi.org/10.1111/j.1749-8198.2010.00357.x, URL https://compass.onlinelibrary.wiley.com/doi/10.1111/j.1749-8198.2010.00357.x, publisher: Wiley.

Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, **456-457**, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, URL https://linkinghub.elsevier.com/retrieve/pii/S0022169412004556, publisher: Elsevier BV.

Tölle, M. H., L. Schefczyk, and O. Gutjahr, 2018: Scale dependency of regional climate modeling of current and future climate extremes in Germany. *Theor Appl Climatol*, **134 (3-4)**, 829–848, https://doi.org/10.1007/s00704-017-2303-6, URL http://link.springer.com/10.1007/s00704-017-2303-6.

Ugolotti, A., T. Anders, B. Lanssens, T. Hickler, L. François, and M. H. Tölle, 2023: Impact of bias correction on climate change signals over central Europe and the Iberian Peninsula. *Front. Environ. Sci.*, **11**, 1116 429, https://doi.org/10.3389/fenvs.2023.1116429, URL https://www.frontiersin.org/articles/10.3389/fenvs.2023.1116429/full.

UNFCCC - Adaptation Committee, 2020: Data for adaptation at different spatial and temporal scales. Technical paper, UNFCCC. URL https://unfccc.int/documents/267555, technical paper by the UNFCCC Adaptation Committee.

van Rüth, P., K. Schönthaler, S. von Andrian-Werburg, M. Wolf, and M. Gabriel, 2023: Monitoringbericht 2023. Tech. rep., Umweltbundesamt.

Van Vuuren, D. P., and Coauthors, 2011: The representative concentration pathways: an overview. *Climatic Change*, **109 (1-2)**, 5–31, https://doi.org/10.1007/s10584-011-0148-z, URL http://link.springer.com/10.1007/s10584-011-0148-z.

Vautard, R., and Coauthors, 2014: The European climate under a 2 °C global warming. *Environ. Res. Lett.*, **9 (3)**, 034 006, https://doi.org/10.1088/1748-9326/9/3/034006, URL https://iopscience.iop.org/article/10.1088/1748-9326/9/3/034006.

Vautard, R., and Coauthors, 2021: Evaluation of the Large EURO-CORDEX Regional Climate Model Ensemble. *JGR Atmospheres*, **126 (17)**, e2019JD032 344, https://doi.org/10.1029/2019JD032344, URL https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019JD032344.

Wagner, S., P. Berg, G. Schädler, and H. Kunstmann, 2013: High resolution regional climate model simulations for Germany: Part II—projected climate changes. *Clim Dyn*, **40 (1-2)**, 415–427, https://doi.org/10.1007/s00382-012-1510-1, URL http://link.springer.com/10.1007/s00382-012-1510-1, publisher: Springer Science and Business Media LLC.

Watson, R. T., and UNEP, Eds., 2003: *Climate change 2001: synthesis report*. United Nations, New York.

Wilcoxon, F., 1945: Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, **1 (6)**, 80, https://doi.org/10.2307/3001968, URL https://www.jstor.org/stable/10.2307/3001968?origin=crossref.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*. 3rd ed., No. v. 100, International geophysics series, Academic Press, Oxford Waltham, MA.

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. *WIREs Climate Change*, **2 (6)**, 851–870, https://doi.org/10.1002/wcc.147, URL https://wires.onlinelibrary.wiley.com/doi/10.1002/wcc.147.

Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON(ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quart J Royal Meteoro Soc*, **141 (687)**, 563–579, https://doi.org/10.1002/qj.2378, URL https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.2378, publisher: Wiley.

A Appendix

A.1 Illustration of the robustness test

To illustrate the robustness definition used for the climate signal maps, two example regions are selected, of which one is robust and the other one non-robust. Figure A.1 shows the Empirical Cumulative Distribution Functions (ECDF) and the Kernel Density Estimates (KDE) for both example regions and each ensemble member within the NUKLEUS ensemble. The monthly-mean summer precipitation for the historical period and GWL3K is evaluated. Both the ECDF and KDE are computed in Python using the Seaborn package. The ECDF is illustrated as solid lines, and the KDE as shaded curves. The historical period is shown in gray and GWL3K in orange. A decrease in monthly mean summer precipitation is evaluated. Models that show a significant decrease are enclosed by the dashed lines. Figure A.1a shows the region Ludwigslust-Parchim, where the median climate change signal is classified as robust, with 6 out of 9 models showing a significant signal. Figure A.1b displays the region Wittenberg, where the signal is non-robust. Here, only 4 out of 9 models show significant signals.

A significant decrease in summer precipitation under GWL3K is indicated when the ECDF shifts to the left of the historical ECDF. This shift can be seen, for example, in Figure A.1a for the models MIROC-CCLM and MPI-CCLM (middle panel). If both ECDFs largely overlap or show minimal differences, the change is not considered significant. This is, for example, visible for the ECE-driven models in Figure A.1b (upper row), where both ECDFs overlap. Likewise, a right shift of the ECDF for GWL3K compared to the historical one indicates an increase in summer precipitation. An increase is also classified as non-significant, since only decreases are assessed. This is illustrated, for example, for MIROC-REMO in Figure A.1a (right panel, middle row), where a rightward shift is shown.

The KDE illustrates whether the distribution of the value changes from the historical distribution (gray) to the GWL3K distribution (orange). Significant signals show distributions that shift their position or change their shape, while non-significant models show no pronounced differences in the KDE. This is shown, for example, for the simulations MPI-ICON and MIROC-CCLM in Figure A.1a or for MPI-ICON and MPI-CCLM in Figure A.1b.

It is important to note that a one-sided test is being applied. Thus, if the distribution for the GWL is shifted to higher values compared to the historical distribution, it is still classified as non-significant. This is the case, for example, for MIROC-REMO in Figure A.1a (right panel, middle row), where the GWL3K KDE is shifted rightwards. Although there is a change in the distribution compared

to the historical, the change is not in the considered direction of change and thus, classified as non-significant.

The robustness definition is divided into two criteria. A fraction of 66 % of the models have to agree first on the sign of change, and second, on the significance of this change. Both of these criteria are combined within the one-sided significance test. Therefore, a region is classified as robust when at least 66 % of the models, which correspond to 6 out of 9 models, show a significant change. This is true for the region illustrated in Figure A.1a. Regions that show less than 6 significant models, as in Figure A.1b, are classified as non-robust. This underlines that the robustness definition applied in the following is based on the significance of each model's signal individually, instead of evaluating the ensemble as a whole. This ensures that the significance of the signal of each model is evaluated independently from the signals of other models. As a result, the robustness measure does not rely on the ensemble median or other ensemble-based statistics, and model-specific biases are prevented from influencing the robustness assessment.
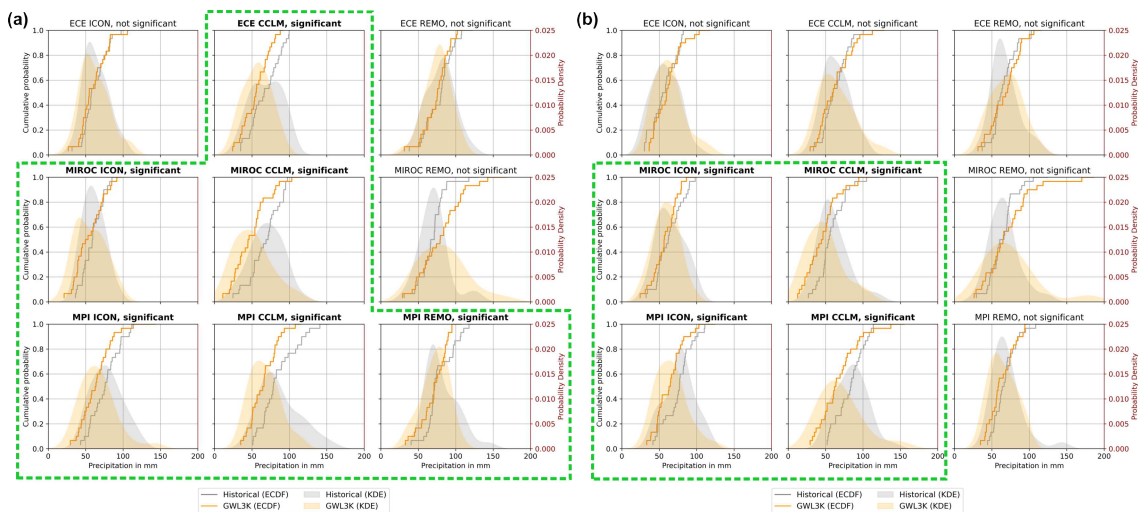

Figure A.1: Empirical Cumulative Distribution Functions (ECDF) and the Kernel Density Estimates (KDE) of a region with (a) robust climate change signal (Ludwigslust-Parchim) and (b) non-robust signal (Wittenberg) for decrease in summer precipitation for GWL3K. The green dashed lines enclose models that show a significant decrease compared to the reference period. For a signal to be considered robust, 6 out of 9 models must show significant changes.

A.2 The NUKLEUS-EUR11 ensemble

A.2.1 Comparison of the NUKLEUS-EUR11 ensemble with Pfeifer et al. (2015)

In the following chapter, the results of Pfeifer et al. (2015) (Figures 4.8-4.9) are compared to the climate signal maps based on the NUKLEUS-EUR11 ensemble. The NUKLEUS-EUR11 ensemble has the same spatial resolution and grid as the simulations that Pfeifer et al. (2015) used. For this comparison, the NUKLEUS-EUR11 is divided into the same four time slices as in Pfeifer et al. (2015), and the same historical reference period is applied. Figure A.2 shows the climate signal
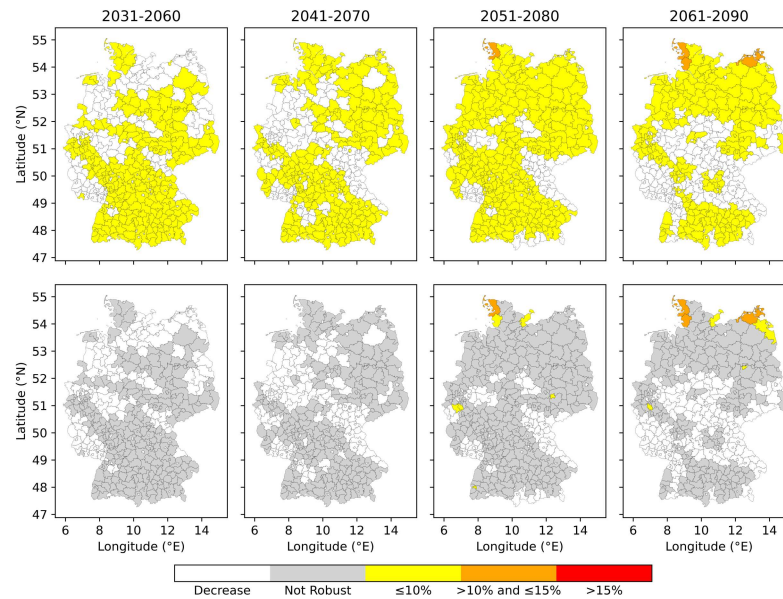
Figure A.2: Increase of winter precipitation for the periods 2031–2060, 2041–2070, 2051–2080, and 2061–2090. The reference period is 1971–2000. The upper panel shows the median climate change signals in %. For the bottom panel the robustness test was applied and only regions with robust signalss are shown in colors, non-robust in gray.

maps for an increase in winter precipitation based on the NUKLEUS-EUR11 ensemble. Figure A.3 shows the climate signal maps for a decrease in summer precipitation.

Due to the unavailability of some NUKLEUS-EUR11 simulations at the time of this analysis, only 5 out of 9 ensemble members are included. The threshold for robustness is accordingly adjusted to 3 out of 5 models, which corresponds to 60% of the models that have to agree on the sign and significance of the signals to be classified as robust. The available models include ECE-ICON, ECE-CCLM, MIROC-ICON, MIROC-CCLM, and MPI-ICON. It should be noted that the CPM REMO is not represented in this available subset of the NUKLEUS EUR11 ensemble.

Comparing the NUKLEUS-EUR11 simulations to the findings of Pfeifer et al. (2015), the same key differences can be observed as in the previous comparison of the convection-permitting NUKLEUS ensemble to Pfeifer et al. (2015) (see Section 4.1.1). For an increase in winter precipitation, the NUKLEUS-EUR11 ensemble shows only small signals (Figure A.2). Furthermore, the NUKLEUS-EUR11 ensemble indicates regions with a decrease in winter precipitation in all four time slices. The number of these regions that project a decrease in winter precipitation even increases for the latest time slice. Considering the robustness of climate change signals based on the NUKLEUS-EUR11 ensemble, only the last two time slices show regions with robust climate change signals. For the period 2051-2080, it is 7 out of 402 regions with robust signals, and for the period 2061-2090, it is 8 out of 402 regions with robust signals. In contrast, the results of Pfeifer et al. (2015) show a higher number of regions with robust climate change signals for an increase in winter precipitation compared to NUKLEUS-EUR11, and the signals get stronger in later time slices (Figure 4.8). For instance, several regions in the period 2061-2090 show large climate change signals (red regions) in the South of Germany that are also robust.
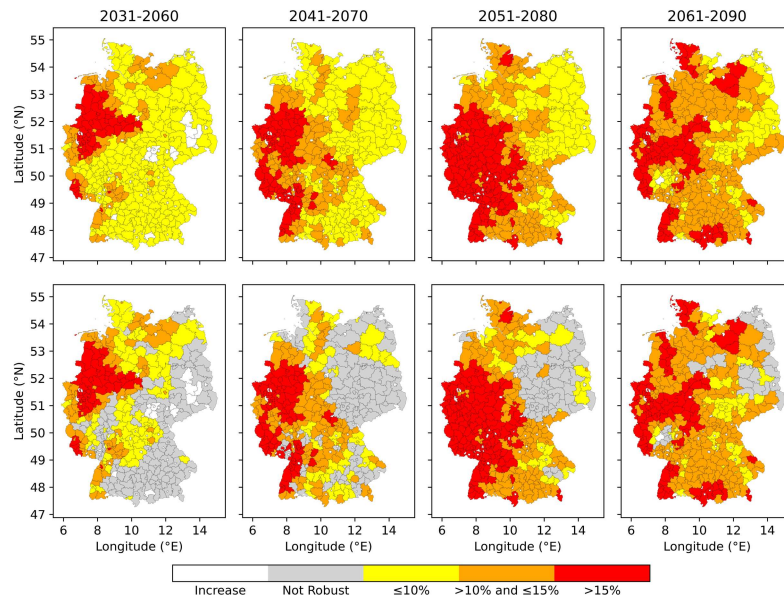
Figure A.3: Same as Figure A.2, but for a decrease in summer precipitation

A different pattern is observed in summer precipitation (Figure A.3). The NUKLEUS-EUR11 ensemble shows strong climate change signals for a decrease in summer precipitation, especially for the western part of Germany. Regarding the robustness, most regions with robust signals are shown in the Southwest, and the number of regions with robust climate change signals increases with time. Stronger signals tend to be more robust. In the first time slice, some regions in Northeastern Germany show an increase in summer precipitation. Most areas with non-robust climate change signals are also located in this part of Germany. By contrast, the findings of Pfeifer et al. (2015) (Figure 4.9) show only small climate change signals for a decrease in summer precipitation, and many regions exhibit an increase instead throughout all periods. This leads to hardly any regions with robust signals identified along all periods in the findings by Pfeifer et al. (2015).

This comparison shows that there are very contrasting results between the NUKLEUS-EUR11 ensemble and the findings of Pfeifer et al. (2015), similar to the differences seen in the comparison of the convection-permitting NUKLEUS ensemble and Pfeifer et al. (2015). These differences can largely be attributed to the model selection within each ensemble. The models analyzed by Pfeifer et al. (2015) differ from the ones included in the NUKLEUS ensemble. The selection of the models included in an ensemble can have substantial impacts on the results, and therefore, models have to be chosen wisely (Seaby et al., 2013; Evans et al., 2014; Herger et al., 2018). Although both ensembles use the same grid and time slices, differences also arise from the use of different scenarios. The NUKLEUS-EUR11 ensemble is based on the SSP-370, while the results of Pfeifer et al. (2015) are based on RCP4.5. Furthermore, only a subset of the NUKLEUS-EUR11 ensemble was available at the time of this analysis, which does not fully represent the complete NUKLEUS ensemble.

A.2.2 Comparison of the NUKLEUS-EUR11 ensemble with the convection-permitting NUKLEUS ensemble

The NUKLEUS ensemble analyzed in this thesis is a convection-permitting climate model ensemble and runs at a resolution of 3 km. To evaluate differences due to the finer resolution and explicit resolution of deep convection, the findings are compared to the results based on the NUKLEUS-EUR11 ensemble. In general, both ensembles show similar trends in winter and summer precipitation. This is also expected, as regarding the two nesting steps, the NUKLEUS-EUR11 ensemble is the boundary and initial data that drives the convection-permitting NUKLEUS ensemble.

For the increase in winter precipitation, both ensembles agree on small changes for both GWLs and all time slices. Regarding the robustness of the climate change signals, both ensembles show very few, if any, robust signals. The EUR11 ensemble (Figure A.2) shows many regions that project a decrease in winter precipitation for all four periods, while the convection-permitting NUKLEUS ensemble shows regions with a decrease in winter precipitation only for GWL2K (Figure 4.1b).

For the decrease in summer precipitation, based on the NUKLEUS-EUR11 ensemble, almost all regions agree on a decrease in summer precipitation across all periods (Figure A.3). In contrast, for GWL2K, the convection-permitting ensemble exhibits a considerable number of regions with an increase in summer precipitation and shows a strong disagreement on the sign of the signals (Figure 4.2b). Both ensembles agree on an intensification of the summer precipitation signals and an increasing number of robust signals over time. Considering the regional patterns in the convection-permitting NUKLEUS ensemble, the most robust signals are identified in the Northwestern part of Germany. For the NUKLEUS-EUR11 ensemble, the signals for the first time slice are also robust in the Northwestern part and expand towards later time slices southwards. The NUKLEUS-EUR11 ensemble shows a noticeably higher number of regions with robust climate change signals in summer, already for the early periods, compared to the convection-permitting NUKLEUS ensemble. Moreover, the EUR11 ensemble exhibits a stronger agreement on the direction of change within all periods considered, whereas the convection-permitting NUKLEUS ensemble also shows regions with an increase in summer precipitation, particularly for GWL2K.

Overall, the results from the NUKLEUS EUR11 ensemble are similar to the results based on the convection-permitting NUKLEUS ensemble. For winter precipitation signals, no regions with robust climate change signals can be identified in either ensemble, which leads to the conclusion that there is no clear trend in increasing winter precipitation based on the NUKLEUS ensemble. In summer, the ensembles generally agree, but the NUKLEUS-EUR11 ensemble exhibits more regions with robust climate change signals compared to the convection-permitting NUKLEUS ensemble. This is in line with the findings of Kendon et al. (2017), that RCMs are more likely to provide robust signals in mean seasonal precipitation. The NUKLEUS-EUR11 ensemble shows almost no regions with an increase in summer precipitation, whereas the convection-permitting NUKLEUS ensemble shows an increase in summer precipitation for many regions. The contrasting signs of the climate change signals in the convection-permitting NUKLEUS ensemble (see Figures 4.3 and 4.6) can be explained by the finer resolution and explicit resolution of deep convection. Deep convection is a major source of precipitation in summer, thus CPMs show most improvements in the projection

of summer precipitation (e.g. Prein et al., 2013; Ban et al., 2021; Fosser et al., 2024). For example, Kendon et al. (2017) showed that the major improvements of CPMs compared to RCMs are found in the representation of summer precipitation, as the coarser RCM did not capture increasing signals of summer precipitation. This is in line with the results presented here, as for summer precipitation, only the convection-permitting NUKLEUS ensemble projects an increase in summer precipitation. Although CPMs generally improve the representation of summer precipitation, this leads not to more agreement among the ensemble members of the convection-permitting NUKLEUS ensemble. The models project contrasting signs of change, leading to more uncertainty and fewer regions with robust signals. In contrast, other studies find that CPMs reduce the uncertainty due to the explicit representation of deep convection (Fosser et al., 2024). Note that the differences in the results between the two ensembles are not solely due to the effects of the downscaling using the CPMs. For the EUR11 ensemble, not all simulations were available, and it is based on transient simulations, whereas in the CPM ensemble, GWLs are considered. Moreover, different reference periods are applied.

B Supplementary Material

Table B.1: Autocorrelation coefficients for each variable and scenario. The lower and upper limits indicate the 80% range across all models and regions.

Variable/Index	Scenario	Lower	Upper
pr, DJF	historical	-0.38	0.19
	GWL2K	-0.3	0.22
	GWL3K	-0.28	0.35
pr, JJA	historical	-0.34	0.33
	GWL2K	-0.5	0.29
	GWL3K	-0.37	0.25
tas, DJF	historical	-0.46	0.39
	GWL2K	-0.18	0.37
	GWL3K	-0.32	0.14
tas, JJA	historical	-0.42	0.23
	GWL2K	-0.17	0.36
	GWL3K	-0.44	0.27

Table B.2: Autocorrelation coefficients for each index and scenario. The lower and upper limits indicate the 80% range across all models and regions.

| Index	Scenario	Lower	Upper
TR	historical	-0.26	0.24
	GWL2K	-0.14	0.4
	GWL3K	-0.25	0.28
ID	historical	-0.45	0.36
	GWL2K	-0.33	0.34
	GWL3K	-0.09	0.43
WSDI	historical	-0.29	0.35
	GWL2K	-0.28	0.38
	GWL3K	-0.23	0.46
R10mm	historical	-0.42	0.23
	GWL2K	-0.31	0.3
	GWL3K	-0.36	0.21
R20mm	historical	-0.35	0.24
	GWL2K	-0.35	0.25
	GWL3K	-0.38	0.22
R30mm	historical	-0.32	0.23
	GWL2K	-0.32	0.24
	GWL3K	-0.38	0.23
CDD	historical	-0.37	0.23
	GWL2K	-0.32	0.36
	GWL3K	-0.4	0.15
CWD	historical	-0.29	0.27
	GWL2K	-0.3	0.21
	GWL3K	-0.37	0.24
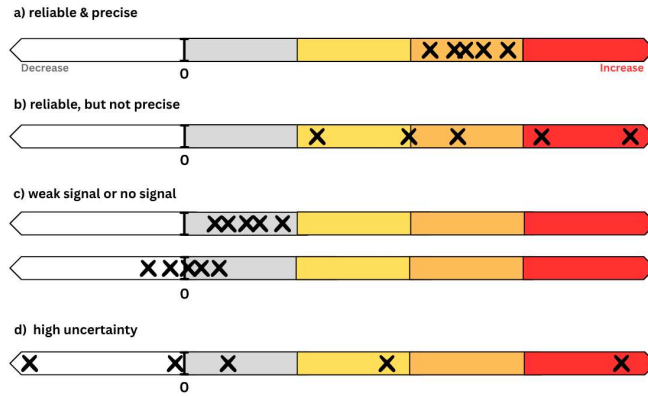
Figure B.1: Schematic representation of the four resulting classifications of climate change signals following the classification approach in Figure 1.1. Each cross represents the signal of an individual model. The colors indicate the different categories of the climate signal map: white indicates a decrease, gray indicates a non-significant increase, and yellow, orange, and red indicate different signal strengths.
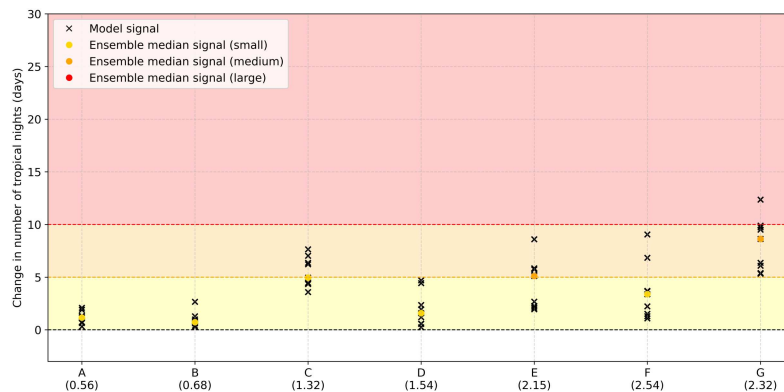

Figure B.2: Climate change signals for an increase in the number of tropical nights for GWL2K and for the individual models of the NUKLEUS ensemble for the selected regions (Figure 4.13b, right panel). The orange and red lines mark the limits of the corresponding levels in the climate signal map (Figure 4.13b). The standard deviation for each region is displayed in brackets below the region's letter. The ensemble spread is classified as small when the standard deviation is below the threshold of 2 days.
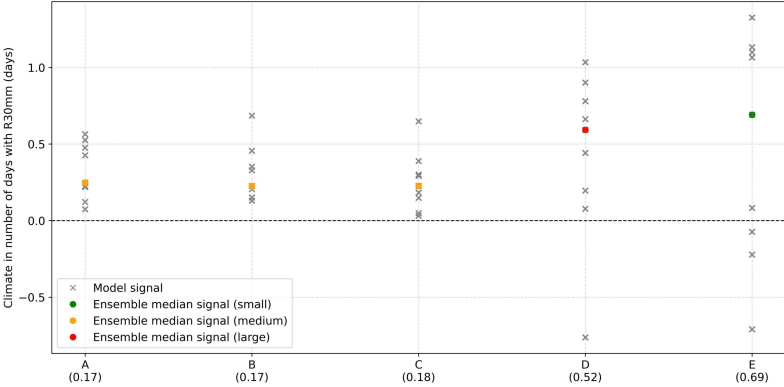
Figure B.3: Same as Figure B.2, but for an increase in the number of days with R30mm for GWL2K. The threshold applied here is 0.18 days, which corresponds to 30% of the mean robust signals.

Acknowledgments

First of all, I would like to thank the entire working group for the warm welcome. I truly enjoyed my time with you and the many shared lunches!

Special thanks to Christoph for the excellent supervision and guidance, for always being available to answer any questions, and for supporting me throughout this journey with patience and constant encouragement, especially during the final phase.

Many thanks to Joaquim for his support and guidance throughout this year, not only for this thesis but also on my way toward pursuing a PhD. Thank you for your support, valuable feedback, and for writing countless reference letters!

I also want to thank Andreas Fink for his valuable feedback on the specialization phase presentation and for his role as co-supervisor.

Thanks to Marie for her helpful advice and for being patient and flexible with my Hiwi.

Thanks to Serkan for all the laughs in the student room and for constantly motivating each other along the way.

Also, a big thank you to Linus, my family, and friends for their endless support and encouragement.

Looking back, I have learned a lot during this year and I'm grateful for the experiences, and especially for the many fun moments shared with all of you.

Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, den 16.07.2025

(Annabell Weber)