Single model large ensembles as a tool to quantify the role of natural variability in climate projections and for robust change detection of droughts

Dissertation zur Erlangung des Doktorgrades an der Fakultät für Geowissenschaften (Department für Geographie) der Ludwig-Maximilians-Universität München

Vorgelegt von Fabian von Trentini

München, 03.05.2023

“Prognosen sind schwierig – vor allem wenn sie die Zukunft betreffen“

Nils Bohr

Main supervisor: *Prof. Dr. Ralf Ludwig*
Department of Geography
Ludwig-Maximilians-Universität München

Second supervisor: *Prof. Dr. Erich Fischer*
Department of Environmental Systems Science
Eidgenössische Technische Hochschule Zürich

Date of disputation: *12.01.2024*

Summary

Climate change projections try to look into the future and are therefore uncertain by definition, since the climate system inherits a chaotic component. This component can be described in many ways and is usually referred to as natural or internal variability. Natural variability is an important part of the equation, yet often underestimated. People and decision makers are usually not able to act appropriately based on projections with large uncertainties. While changes in the mean state are usually at the focus of climate change impact assessments, the variability has often not been recognized as an important feature. However, it gets more and more attention in the scientific community and the broader public that variability changes might even be a more crucial challenge for mankind than mean climate changes. Year-to-year variability called interannual variability (IAV) is key for many impact-related assessments. For example, farmers need to adapt their practices if less “normal” and more extreme years appear to happen in the future.

This thesis tries to shed light on parts of the scientific discussion concerning natural variability from a relatively new perspective: Single Model Initial-conditions Large Ensembles (SMILEs). SMILEs use the exact same model setup and just change the initial conditions for many runs (members). This can result in an ensemble of e.g., 50 members of one model that just differ in their specific manifestation of weather and climate. However, all members share the same general path, defined by the ensemble mean (forced response). SMILEs serve as a tool to better understand and quantify changes in the variability of the climate system (e.g., IAV) but due to the large data amount for relatively short periods, they can also be used for the robust assessment of extreme events like droughts.

Three aspects of natural variability in SMILEs form the structure of this thesis and the respective three publications:

A) Interannual variability (IAV) in the SMILEs
B) Natural variability as part of the overall uncertainty of climate projections
C) SMILEs as a statistically sound data base for extreme event analysis

A) Interannual variability (IAV) is usually quantified by removing the general trend from a time series and looking at the standard deviation of the residuals. This can also be done for every member of a SMILE individually. As the intention of removing the trend is to get rid of the forced response of the climate model at consideration, it is better to remove the ensemble mean (EM) from each member in SMILEs. Another way to quantify year-to-year variability in SMILEs is the inter-member variability (IMV) which represents the standard deviation of all members from the EM for each year. It has no physical meaning like IAV but is a precise estimation of IAV with the advantage of a drastic extension of data points for statistical analysis. The first paper showed that the three regional SMILEs are capable of representing the IAV of observed climatology (E-OBS data set) for several indicators and regions, with very few exceptions. In the future, the variability of summer temperature and precipitation, heatwaves and dry periods will increase, while winter temperature and precipitation variability will decrease. The time of emergence when the variability significantly changes in the future, depends on the methodology to quantify IAV. While the single usage of IAV (for each member) or the single use of IMV often do not reveal significant changes before the end of the 21st century, a combined approach, taking into account both IAV and IMV, reveals much earlier time of emergence signals. The three models generally agree on change signals but vary in strength and timing of these changes.

B) The uncertainty of climate projections can stem from three sources: scenario uncertainty, model response uncertainty and internal variability. A classical way to look at climate projections is to run a multi-model ensemble (for one or more scenarios) and interpret the spread of the model runs (so far often only one member per model) as model response uncertainty. The initiative for European coordinated regional climate projections is called EURO-CORDEX. A multi-model ensemble of 22 EURO-CORDEX members was analyzed with respect to the variability in change signals. A baseline group was built by the CRCM5-SMILE that represents isolated internal variability. By comparing the standard deviations of these two ensembles, it was possible to show that the internal variability can play a major role in climate projection uncertainty. Results show that the single model spread is usually smaller than the multi-model spread for temperature. The contribution of internal variability is generally higher for precipitation. There is a significant decrease of the contribution of internal variability over time for both variables. However, even in the far future (2070-2099) for most regions and seasons 25–75 % of the overall variability can be explained by internal variability. This means that in the interpretation of multi-model ensembles, their variability can not directly be assigned to model differences, especially not in the case of single member representatives for each model.

C) Finally, the CRCM5-LE was used to detect changes in frequency of the 30-year return level (RL30) of low precipitation months and seasons. At the end of the 21st century, former RL30 months in summer and fall will occur with much higher frequency in large parts of Europe (up to each year in Spain). These dramatic changes will pose huge challenges to water management practices, especially in moderate climatic regions that did not suffer water scarcity so far. Consecutive dry summers will also drastically increase their frequency for 2 consecutive summers in the Mediterranean region. For 3 or more consecutive dry summers, the frequency changes decline but are still relevant in France and the Iberian Peninsula. The 2018 summer drought in Germany was the second most severe at this point in history. With a simple quantile transfer approach, the future probability for such an event can be estimated with the CRCM5-LE. It will rise from 1.4 % to 20 % at the end of the 21st century, meaning every fifth year will be a summer drought year comparable to 2018. The former drought analysis was purely based on precipitation. Looking at the propagation of the RL30 monthly changes from the CRCM5-LE's precipitation into the RL30 changes of river discharge from a hydrological model, reveals the high non-linearity of the two variables. The river discharge follows the higher frequencies of RL30 in the summer, while they also show massive frequency gains in winter that are not apparent in the precipitation data.

The analysis from this thesis shows the immense possibilities that SMILEs offer to explore the internal variability of the climate system and extreme events. The huge data base SMILEs provide gives scientists the tool at hand to quantify these phenomena and their future changes more robustly than ever. This thesis tried to look at three different aspects of the climate system's variability but can only catch a glimpse of all the possibilities that lie within SMILEs.

Zusammenfassung

Klimaprojektionen versuchen in die Zukunft zu blicken und sind damit aufgrund der chaotischen Komponente des Klimasystems automatisch mit Unsicherheiten behaftet. Diese Komponente kann vielfältig beschrieben werden und wird häufig als natürliche oder interne Variabilität bezeichnet. Natürliche Variabilität ist ein wichtiger Bestandteil des Systems, wird jedoch häufig unterschätzt. Die Bevölkerung und auch Entscheidungsträger sind gewöhnlich nicht in der Lage basierend auf Projektionen mit Unsicherheiten angemessen zu handeln. Während mittlere Änderungen gewöhnlich im Fokus bei Beurteilungen von Klimawandel-auswirkungen stehen, wurde die Variabilität oft nicht als wichtiger Bestandteil der Diskussion gesehen. In der wissenschaftlichen Gemeinschaft und der breiten Öffentlichkeit erlangt es jedoch immer größere Aufmerksamkeit, dass Veränderungen der Variabilität unter Umständen sogar die größere Herausforderung für die Menschheit sein könnte als mittlere Klimaänderungen. Die Variabilität von Jahr zu Jahr – interannuelle Variabilität (IAV) – ist für die Beurteilung vieler Klimawandelauswirkungen essentiell. So müssen Landwirte beispielsweise ihre Bewirtschaftungsformen anpassen, sollten immer weniger „normale“ und mehr extreme Jahre auftreten.

Diese Arbeit versucht Teile der wissenschaftlichen Diskussion um natürliche Variabilität aus einer relativ neuen Perspektive zu beleuchten: Sogenannte Single Model Initial-conditions Large Ensembles (SMILEs). SMILEs nutzen das exakt gleiche Modelsetup mehrfach und ändern dabei nur die die Initialbedingungen für eine Vielzahl von Läufen (member). Das führt zu einem Ensemble bei dem sich zum Beispiel 50 member eines Modells nur durch ihre spezifische Manifestation von Wetter und Klima unterscheiden, jedoch alle dem generell gleichen Pfad folgen, der durch den Ensemblemittelwert definiert ist (forced response). SMILEs dienen als Werkzeug um die Variabilität des Klimasystems besser zu verstehen und zu quantifizieren (z.B. IAV), und um eine robustere Einordnung von Extremereignissen wie Dürren vornehmen zu können.

Drei Aspekte von natürlicher Variabilität in SMILEs bilden die Struktur dieser Arbeit und der entsprechenden drei Publikationen:

A) Interannuelle Variabilität (IAV) in SMILEs
B) Natürliche Variabilität als Teil der Gesamtunsicherheit in Klimaprojektionen
C) SMILEs als robuste statistische Datengrundlage für Extremwertanalysen

A) Interannuelle Variabilität wird gewöhnlich quantifiziert, indem man den Langzeittrend einer Zeitserie entfernt und die Standardabweichung der Residuen betrachtet. Das kann auch für jeden einzelnen der member eines SMILEs vollzogen werden. Da die Intention in der Entfernung der grundsätzlichen Reaktion des entsprechenden Modells auf Veränderungen der Atmosphöre (forced response) liegt, ist es bei SMILEs jedoch sinnvoller die Residuen jedes members zum Ensemblemittelwert (EM) zu berechnen. Ein weiterer Weg um die Variabilität von Jahr zu Jahr in SMILEs zu beschreiben ist die inter-member Variabilität (IMV), welche die Standardabweichung aller member zum EM repräsentiert. Sie hat keine physikalische Bedeutung wie IAV, bietet jedoch eine sehr präzise Annäherung an IAV und erweitert erheblich die Datenpunkte für statistische Analysen. Die erste Publikation zeigt, dass die drei regionalen SMILEs in der Lage sind die beobachtete IAV für fast alle Kombinationen aus Indikatoren und Regionen widerzugeben. In der Zukunft wird sich die Variabilität von Sommertemperatur und -niederschlag, sowie für Hitzewellen und Dürreperioden erhöhen, während die Variabilität von Wintertemperatur und -niederschlag geringer ausfallen wird. Der Zeitpunkt wann sich die

Variabilitäten signifikant unterscheiden werden (time of emergence) hängt dabei stark von der gewählten Methode zur Quantifizierung der IAV ab. Während die alleinige Verwendung der IAV jedes members bzw. die alleinige Verwendung von IMV kaum signifikante Änderungen der Variabilität bis zum Ende des 21. Jahrhunderts anzeigen, führt ein kombinierter Ansatz, der sowohl IAV als auch IMV zusammen verwendet, zu deutlich früher detektierbaren Änderungen. Die drei SMILEs sind sich grundsätzlich in den Änderungssignalen einig, variieren jedoch in Stärke und Timing dieser Änderungen.

B) Die Unsicherheiten in Klimaprojektionen können aus drei Quellen stammen: Szenariounsicherheit, Modellunsicherheit und natürliche Variabilität. Eine klassische Herangehensweise an Unsicherheiten in Klimaprojektionen beinhaltet meist das Aufstellen eines Multimodell-Ensembles (für ein oder mehr Szenarios) und der Interpretation der Spannbreite des Ensembles (oftmals nur ein member pro Modell) als Modellunsicherheit. EURO-CORDEX ist der Name der europäischen Initiative für koordinierte regionale Klimaprojektionen. Ein Multimodell Ensemble aus 22 EURO-CORDEX Läufen wurde auf die Variabilität der Änderungssignale hin untersucht. Als Kontrollgruppe wurde das CRCM5-SMILE ausgewählt, das die isolierte interne Variabilität repräsentiert. Durch den Vergleich der Standardabweichungen der beiden Ensembles kann gezeigt werden, dass die interne Variabilität einen wichtigen Anteil an der Unsicherheit von Klimaprojektionen einnehmen kann. Die Ergebnisse zeigen, dass die Streuung des SMILEs bei Temperatur gewöhnlich kleiner als der des Multimodell Ensembles ist. Der Anteil interner Variabilität ist bei Niederschlag grundsätzlich höher. Für beide Variablen zeigt sich über die Zeit ein signifikanter Rückgang des Beitrags interner Variabilität. Trotzdem kann für die meisten Regionen und Saisons selbst in der fernen Zukunft (2070-2099) noch 25-75 % der Gesamtunsicherheit mit interner Variabilität erklärt werden. Das bedeutet für die Interpretation von Multimodell Ensembles, dass die dort vorgefundene Variabilität nicht so einfach direkt den unterschiedlichen Modellen zugeordnet werden kann, insbesondere wenn nur ein member pro Modell in das Ensemble eingeht.

C) Schlussendlich wurde das CRCM5-LE genutzt, um Veränderungen in der Häufigkeit von Trockenmonaten und -saisons mit einer Wiederkehrdauer von heute 30 Jahren (RL30) zu bestimmen. Am Ende des 21. Jahrhunderts treten im Sommer und Herbst heutige RL30 Monate in ganz Europa deutlich häufiger auf (bis zu jedes Jahr in Spanien). Diese dramatischen Änderungen werden die Wasserwirtschaft vor gewaltige Herausforderungen stellen, insbesondere in klimatisch gemäßigten Regionen die bisher nur wenig oder gar nicht unter Wassermangel gelitten haben. Aufeinanderfolgende Trockensommer werden sich im Mittelmeerraum für 2 aufeinanderfolgende Jahre deutlich erhöhen. Für 3 oder mehr Jahre gehen die Häufigkeitsänderungen deutlich zurück, bleiben jedoch in Frankreich und der iberischen Halbinsel durchaus relevant. Der Trockensommer von 2018 in Deutschland war zu diesem Zeitpunkt der zweittrockenste Sommer seit Beginn der Aufzeichnungen. Mittels eines einfachen Quantiltransfer-Ansatzes kann die zukünftige Wahrscheinlichkeit eines solchen Events mit Hilfe des CRCM5-LE abgeschätzt werden. Die Wahrscheinlichkeit wird sich demnach von 1.4 % auf 20 % bis zum Ende des 21. Jahrhunderts erhöhen. Somit wird jeder fünfte Sommer eine vergleichbare Sommerdürre wie 2018 aufweisen. Die bisherige Dürreanalyse basierte auf Niederschlagsdaten von Klimamodellen. Betrachtet man die Übertragung der monatlichen RL30 Signale für Niederschlag aus dem CRCM5-LE in entsprechende Werte für den Abfluss aus einem hydrologischen Modell, zeigt sich eine hohe Nichtlinearität der beiden Variablen. Der Abfluss wird im Sommer ebenfalls deutlich häufiger Niedrigwasser aufweisen und folgt damit dem Niederschlagssignal. Zusätzlich, und im Gegensatz zu den Niederschlagssignalen des RL30, ist auch im Winter deutlich häufiger mit Niedrigwasser zu rechnen.

Die Analysen in dieser Arbeit zeigen das immense Potential von SMILEs auf, die natürliche Variabilität des Klimasystems und Extremereignisse zu erforschen. Die enorme Datenbasis die SMILEs zur Verfügung stellen, ermöglicht es der Wissenschaft in bisher ungekanntem Ausmaß und deutlich robuster diese Phänomene und ihre zukünftigen Änderungen zu quantifizieren. Diese Arbeit versucht drei Aspekte der Variabilität des Klimasystems zu beleuchten, kann jedoch nur eine Vorahnung auf all die Möglichkeiten geben, die in der Analyse von SMILEs liegen.

Acknowledgements

Although this work was mainly performed by me, I had guidance and help from many people to create this thesis. First, I want to thank my supervisor Ralf Ludwig, who accompanied me through my whole academic career, as a student and then as a PhD candidate. He provided me excellent guidance, always supported me in my research topics and was patient enough to wait until this moment when I finally submit. The scientific exchange has always been fruitful, and I really appreciate him as a good boss and nice colleague, also on a personal level.

The specific work on my papers was also greatly supported by absolute experts in their fields. I want to thank Martin Leduc, Emma Aalbers and Erich Fischer for their inputs and ideas and proof-reading efforts. The work with you has really been inspiring and significantly improved the scientific output of my work as a young researcher. Erich was also willing to accept the role as second supervisor for this thesis – I really appreciate this.

In my daily life at the university, the working group provided a very nice environment and I want to thank everybody for the nice and inspiring exchange during all the years. Namely my closest colleagues Raul Wood and Florian Willkofer – you accompanied me for many years and worked with me on different projects. The exchange with you was also extremely important to sharpen ideas and helped me in so many cases to overcome technical issues. We also gave each other confidence in all the critical times of project and publication work!

I also want to thank all partners of the ClimEx project from the Landesamt für Umwelt Bayern, Ouranos, ETS, INRS and LRZ for this great project teamwork and the exciting research we conducted together over many years.

Finally, I want to thank my wife and parents for always supporting and pushing me towards the finish line. Especially in my time after working at the university you gave me the freedom and confidence to finalize my thesis.

Table of contents

Table of figures

List of Abbreviations

Abbreviation	**Explanation**
CanESM2	Canadian Earth System Model
CCLM	Regional Climate Model of COSMO community
CESM	Community Earth System Model
ClimEx	Project “Climate change and hydrological extreme events”
CMIP	Coupled Model Intercomparison Project
CORDEX	Coordinated Regional Downscaling Experiment
CRCM5	Canadian Regional Climate Model
EC-EARTH	European Community Earth System Model
EM	Ensemble Mean
ENSO	El Nino Southern Oscillation
ESD	Earth System Dynamics (journal)
ESM	Earth System Model
EURO-CORDEX	European branch of CORDEX
GCM	Global Climate Model
GHG	Greenhouse Gases
GRL	Geophysical Research Letters (journal)
IAV	Interannual Variability
IMV	Intermember Variability
IPCC	Intergovernmental Panel on Climate Change
LE	Large Ensemble
NAO	North Atlantic Oscillation
NV	Natural Variability
pr-DJF	Winter precipitation sum (Dec, Jan, Feb)
pr-DP-MAX	Maximum length of dry periods per year
pr-JJA	Summer precipitation sum (Jun, Jul, Aug)
RACMO	Dutch Regional Climate Model
RCM	Regional Climate Model
RCP	Representative Concentration Pathways
SMILE	Single-model Inital-conditions Large Ensemble
tas-DJF	Mean winter temperature indicator (Dec, Jan, Feb)
tas-HW-Nr	Number of heatwaves per year
tas-JJA	Mean summer temperature indicator (Jun, Jul, Aug)

1 Structure of the thesis

The thesis is organized as a cumulative dissertation, consisting of this framework and three publications. Two of these have been published in peer-reviewed journals and one is submitted to a peer-reviewed journal.

The common thread of the thesis explores two great advantages of SMILEs: quantification of the role of natural variability in climate projections and robust change detection in extreme events (droughts).

The framework first gives a thematic introduction to better understand the scope and background of the publications (chapters 2 to 4). This includes information on climate projections, their uncertainties and the concept of Single Model Initial conditions Large Ensembles (SMILE) – the data foundation of this thesis. This is followed by the presentation of the main topics (publications) and associated research questions that the publications address (chapter 5). Finally, the answers to the research questions and the drawn conclusions are presented together with the respective publications:

A) Interannual variability in the SMILEs (chapter 6)
B) Natural variability as part of the overall uncertainty of climate projections (chapter 7)
C) SMILEs as statistically sound data base for extreme event analysis (chapter 8)

2 Climate projections and impacts

2.1 Climate projections in general and politics

Climate change is the one of the most pressing issues for mankind on all levels of urgency. While the most severe impacts are projected to happen when the global mean temperature has increased by several degrees Celsius (far future, probably at the end of the 21st century), the impacts of the global temperature rise since pre-industrial conditions (1.0-1.5 °C) can already be felt now and will impact the near and mid-term future. Great effort has been made in the recent decades to reveal the fact that anthropogenic greenhouse gas emissions are responsible for the rapid climate change we experience, leading to a common agreement not only within the scientific community, but also in politics that the change is man-made and needs to be mitigated. Since it is already too late to prevent a certain part of climate change and its impacts to happen, adaptation strategies need to be developed to inevitable and potential future changes in climatic conditions. This requires good knowledge about the future conditions. Climate projections from models, which try to describe the earth system as good as possible, give an outlook on how the future might look like. They rely on socioeconomic pathways that inherit certain emissions that lead to different greenhouse gas (GHG) concentrations in 2100 and beyond. These GHG concentrations can be associated to a certain value of radiative forcing. Radiative forcing is the additional energy that is held back in the atmosphere by the GHGs and is expressed in W/m^2. In the framework of the Intergovernmental Panel on Climate Change (IPCC), a selection of Representative Concentration Pathways (RCP) was developed to define scenarios to drive climate models in a coordinated manner. These RCPs are named after the radiative forcing in 2100 caused by GHG concentrations: RCP8.5 relates to a radiative forcing of 8.5 W/m^2 and is usually considered a worst-case scenario with almost no reductions in GHG emissions and global economic growth. It results in major temperature rises, especially in the second half of the 21st century and usually also causes the most extreme changes in weather patterns, extreme events, and related impacts. Given the fact that many of the recent years were under the warmest ever recorded, the RCP8.5 scenario seems more plausible than 10 years ago when the discussion about the speed of climate changes evolved from a ca. 10 years lasting period of relatively stable global temperatures. This “global warming hiatus” raised a lot of attention from both science and politics, as negotiations about the international collaboration on measures to reduce GHG emissions were ongoing during that time. From a scientific point of view it was relatively clear that this can have many reasons and that a 10-15 year long period can just be an expression of natural variability (Medhaug et al. 2017). The discussion quickly receded as the coming years were under the warmest ever recorded. It however shows how much influence natural variability (NV) can have and how much the public opinion changes with the real-life experiences people make in the most recent years at a certain time. The last years were characterized by major floods and major heat and drought extremes alternating almost every year, with hardly any “normal” years anymore. As a result, the awareness of climate change not only as a long-term phenomenon of mean state changes in temperature, but also as a driver of more extreme years/events and higher variability, became apparent in broad parts of the society.

2.2 Impacts of natural variability on climate projection uncertainty

Decision makers need to know which uncertainties are included in the projections by typical multi-model ensembles. They need to understand where they come from and how much of the uncertainty can be reduced by advances in science. Decision makers also need to better understand that NV is an inherent part of the climate system, which is an irreducible part of the uncertainty when looking into the future. Even if we had a model that captures all relevant processes perfectly, we could not make a deterministic forecast of the climatic conditions for years and decades. The chaotic nature of the earth system will always leave space for a range of possible future conditions. Of course, main drivers of the path the system will take are the socioeconomic scenarios (and therefore emission scenarios), but around this general path will always remain a corridor of possible manifestations of the path’s boundary conditions. The quantification of the role of NV is thus crucial to understand the level of irreducible uncertainty.

2.3 Impacts of more and longer drought situations

Higher variability of the climate system on short time scales (from year to year) leads to more dry years or drought situations in general. The worst case are several dry years/seasons in a row that can have devastating effects on natural and anthropogenic systems. Agriculture, forestry, energy production, drinking water supply, shipping navigation, tourism, human health, and ecosystems all suffer under a shift towards more dry conditions in various ways. Many sectors therefore learn more and more how crucial it is to plan for a more diverse climate with a sequence of very dry and wet years in a row and less “mean” years. Planning of increasing interannual variability will be a major challenge for these sectors already now and in the near future. Adaptations in a technical and economical way are crucial to establish sustainable systems. Farmers need to adapt their selection of crops to be more resilient to extreme conditions or develop new sorts of crops that do so. The forestry sector needs to change the dominant monocultural woods into a more resilient mixture of species to prevent huge losses due to diseases and wildfires. Energy producers and inland navigation need to find ways to deal with longer periods of extreme low flows in rivers. The management of water resources in reservoirs and groundwater is one of the most important tasks since they often provide large parts of the drinking water supply. While arid regions usually already have management practices for droughts in place, currently moderate climatic regions will need to change their routines noticeably when droughts become more dominant. This also includes the increasing need for irrigation of agricultural land. While tourism may profit from drier conditions in some parts of the world, larger parts of the sector will need to reduce their water footprint considerably. With water becoming a sparse resource, conflicts between users are inevitable and increase the risk of conflicts. Finally, it is unclear how ecosystems can adapt or "move" as a reaction to the dramatic pace of changes that are much faster than typical climate changes in the past.

3 Uncertainties in climate projections

Looking into the future is complex and difficult by definition. However, the general idea to use models that resemble the earth system is the only possible way to enable mankind to plan for future conditions on Earth. The models cannot be perfect since the system they try to mimic is infinitely complex. To understand it, an enormous number of measurements (stations, drones, satellites, balloons, ships and submarines) would be required to better understand the system as it is now. Weather forecast is usually reliable for only one to three days, not more. And the reason is not so much the inaccuracy of weather models, but that information on the starting conditions is insufficient to feed the models properly. Another restriction, even if the knowledge of the earth system was enhanced to a degree where we fully trust the models, is the computational power needed to let these complex models run. We already use the most powerful supercomputers in the world, but it still requires a lot of personal and technical effort, time, energy, and money to produce, store and analyze these data sets.

The models have, however, shown that they can reproduce historical climatic conditions well in many cases (Flato et al. 2014; Hausfather et al. 2020). In general, the models are robust enough to trust them in terms of trends for the future. If many models predict the same behavior, the prediction is usually considered "robust". The assumption behind that is that all models are independent in their structure and implementation of processes, although many share the same physical principles as a basis. Some parts of the climate system are only parameterized and not fully solved by physical equations. This is where the most differences between models lie. The idea of model democracy is more and more challenged, since a number of models originate from the same branches. In multi-model ensembles, it might also be a good idea to give higher weight to models that performed better during hindcast simulations of the past than those who performed worse (Eyring et al. 2019). One model – one vote is still common, however.

Hawkins und Sutton (2009) and Hawkins und Sutton (2011) introduced a framework for the analysis of uncertainty in climate projections and divided it into three different components:

- **Scenario uncertainty**, since the future emissions of greenhouse gases are unknown
- **Model response uncertainty**, since different models simulate differently when driven by the same emission scenario
- **Natural variability**, i.e. natural fluctuations of the climate system that are independent of radiative forcing

Uncertainty of future conditions generally increases from a global view down to regional and local considerations. It is also higher for more complex climate phenomena and higher for extreme values than for mean states. For example, the global mean surface temperature uncertainty is lower than the uncertainty for heavy precipitation events in a certain country or region.

4 Climate model ensembles and their meaning in the light of NV

4.1 The GCM-RCM coupling

“Climate models” is the general term used in the public debate to describe a number of numeric models with different characteristics. To understand the work performed in the publications it is necessary to get an overview of the different model types and their interactions. First of all, climate models are better described as Earth System Models (ESM) nowadays, since they do not only include the atmosphere but also other components and their interactions with the atmosphere like the cryosphere (snow, ice), land cover, vegetation and the oceans, which are of huge importance to climatic behavior. These models try to mimic the behavior of this complex earth system to the best extent possible and are continuously improved (Bonan und Doney 2018). Since a global system needs to be simulated, the models cover the whole earth and are therefore also called Global Climate Models or General Circulation Models (GCM). The current framework for the execution of GCM simulations, called Coupled Model Intercomparison Project Phase 6 (CMIP6) includes more than 100 models from 50 institutions (Copernicus 2021).

Their spatial resolution typically ranges from 200-600 km, 10 to 20 vertical atmosphere and up to 30 ocean layers (IPCC 2018). This rather coarse resolution does of course not match the heterogeneity of the real world and is mostly caused by the huge amount of processing power that increases quadratically with spatial resolution improvements. A common technique for the production of better resolution is called dynamical downscaling and uses the GCM outputs of the atmospheric and ocean states at each timestep as input (e.g. 6 hours) for another model that is “nested” over a particular region of the earth (e.g. Europe). This secondary model is called Regional Climate Model (RCM) and has a much higher resolution of typically 10-50 km and is therefore representing the earth system and its processes like mountains, shorelines, clouds, precipitation and radiation much more accurate. The driving GCM hands over the boundary conditions at the edges of the area that is processed by the RCM, called domain, as well as ocean states (e.g. temperature) inside and outside of the domain. The RCM takes these boundary conditions and uses them to drive its own model algorithms. The transition zone at the boundaries of the domain are usually excluded from analysis because the influence of the GCM is still strong near the edges and the performance of the RCM really evolves after a certain distance from the edges (ca. 100-200 km). Depending on how much freedom should be attributed to the RCM, there is a technique to maintain the influence of the GCM in the RCM, called spectral nudging. It prevents large and mostly unrealistic departures between the GCM driving fields and the RCM fields at the GCM spatial scales. It is therefore meant to prevent the RCM to develop its own dynamic that cannot be brought in line with the surrounding global conditions.

The main advantage of dynamical downscaling with an RCM is the higher resolution that can better simulate important processes in the atmosphere and the interaction with land and ocean surfaces. A good overview of the history of RCMs and their added value was published by Giorgi (2019).

4.2 RCMs over Europe: EURO-CORDEX

The global climate modeling community learned quickly that an integrated approach for their modeling activities benefits the possibilities for scientific analysis, but also for communication to the public and to decision makers. The CMIP framework coordinates model runs on the GCM level. The regional climate modeling community is organized in the so called CORDEX initiative (Coordinated Regional Downscaling Experiments) that is separated into continental divisions, one of them being EURO-CORDEX for the European domain. The spatial resolution (0.11° corresponding to about 12 km) and the domain size and position is fixed, which makes data processing easy. A number of GCM-RCM combinations have been realized, mostly without systematic approaches. This ensemble of opportunity thus has flaws when it comes to systematic evaluations. However, it gives a good first order estimate of the variability of regional climate projections over Europe. As more and more simulations became available, the matrix of GCM/RCM combinations was filling up in the last years, but still leaves some empty spaces.

An overview of the first EURO-CORDEX simulations is given in Jacob et al. (2014), showing the general direction of future climatic conditions in Europe in a detailed manner. Many following studies completed the picture from various angles often considering extreme high precipitation, drought and heat wave statistics (e.g., Prein et al. 2016; Rajczak und Schär 2017; Christensen et al. 2019; Vautard et al. 2014). A set of two studies evaluated the current status of the GCM-RCM matrix in terms of model performance against historical observations (Vautard et al. 2021) and assessing the projections of this most recent EURO-CORDEX ensemble (Coppola et al. 2021).

4.3 SMILE concept

The concept of building ensembles of different models (and GCM-RCM combinations for CORDEX) to get an overview of the spread of past and future behavior of climate projections is a major pillar of climate change research and thus got and still gets a lot of scientific attention. These multi-model ensembles can give a corridor of possible future conditions. This is typically done by giving ensemble means and information about the minimum/maximum values of single models or by quantile approaches to show the range of results. This is important information for decision makers as these corridors can serve as a baseline for adaptation measures. The multi-model ensembles mostly express model response uncertainty since they give information on how much different models deviate in the future projections if a subsample of model runs with the same external forcing (RCP scenario) is chosen.

A different concept of a climate model ensemble is to take one model and create a larger number of runs by this model alone. One possibility is to change the setup of the model in terms of parametrization changes to see how these changes influence the results of the model. This is a sensitivity analysis and exposes the behavior of a certain model in different setups (Hourdin et al. 2017). The model community can learn about their models and their limitations by such an experiment. Another concept of multiple runs of a single model is to leave the model in the exact same setup and just alter the initial conditions in a very tiny way. This is called a Single Model Initial conditions Large Ensemble (SMILE). At first glance, it might not be obvious why to perform such an experiment with the exact same model many times. But the effect of small changes in the initial conditions can have massive effects on the different runs (called members of the ensemble) on short and medium, but also longer time spans of several decades. To popular science, this phenomenon is known as the “butterfly effect”, where even a small change, initiated by the swing of a butterfly to stay in the metaphor, can cause global changes. And indeed, the chaotic nature of the climate system makes it possible that two members that deviate just by extremely small differences in the initial conditions (e.g., manipulating one variable field by factors of 10^{-13}) can provide totally different sequences of weather after several days already. As all members are forced by the same RCP scenario, they generally follow a similar path into the future, which can be best expressed as the ensemble mean. However, single member behavior can deviate from this ensemble mean quite distinctly. The spread of the ensemble is thus an expression of the isolated natural variability of the chaotic climate system only, without influences of scenario and model-response uncertainty. This “clean” approach makes a systematic analysis and quantification of natural variability possible in a way that was not possible before SMILEs.

An increasing number of GCM SMILEs is available and as there is a vital community, efforts are undertaken to coordinate the data archiving and future experiment designs (https://www.cesm.ucar.edu/projects/community-projects/MMLEA/, and the associated github: https://large-ensemble.github.io/). A publication by Lehner et al. (2020) gives a good overview of the available SMILEs and tries to evaluate the findings by Hawkins und Sutton (2009) with the new tools that SMILEs provide. While the natural variability on global scale can already be detected by the original approach of Hawkins und Sutton (2009), the importance of natural variability on a regional scale is often underestimated by up to 50 % when SMILEs are not considered (Lehner et al. 2020).

Only few of these GCM ensembles have been downscaled with RCMs, and at the time of the publications only three were available. They are presented in the following, with the CRCM5-LE being used in all three publications, while the other two were only used in one publication.

4.4 Concept of Interannual and intermember variability

SMILEs provide a range of equally possible manifestations of the same boundary conditions. A short explanation with conceptual figures is given here to better understand the relation between different dimensions of a SMILE data set and how they interact. The mean of all members is called ensemble mean (EM) and represents the model’s response to a certain forcing like an RCP scenario. The interannual variability (IAV) is now just calculated as the residuals from one member’s time series against the EM. It directly represents the year-to-year variability of the model, but only within that specific member. The dimension of IAV is thus “time”. The residuals can also be extracted against the member’s own trend (linear or higher order trends), resulting in different results. However, the EM is a better estimate of the forced response and thus gives a better estimation of IAV. This is of course only possible for SMILEs. A different assessment of natural variability is also only possible with SMILEs: the intermember variability (IMV) does not depict the variability of one member along the time dimension but the variability of time series of all members for each time step (years). The dimension of IMV is thus “ensemble”. The IMV has no physical meaning like IAV but gives a very good estimate of the SMILE’s internal variability which is directly connected to IAV as well. This is shown in one of the papers: von Trentini et al. (2020).

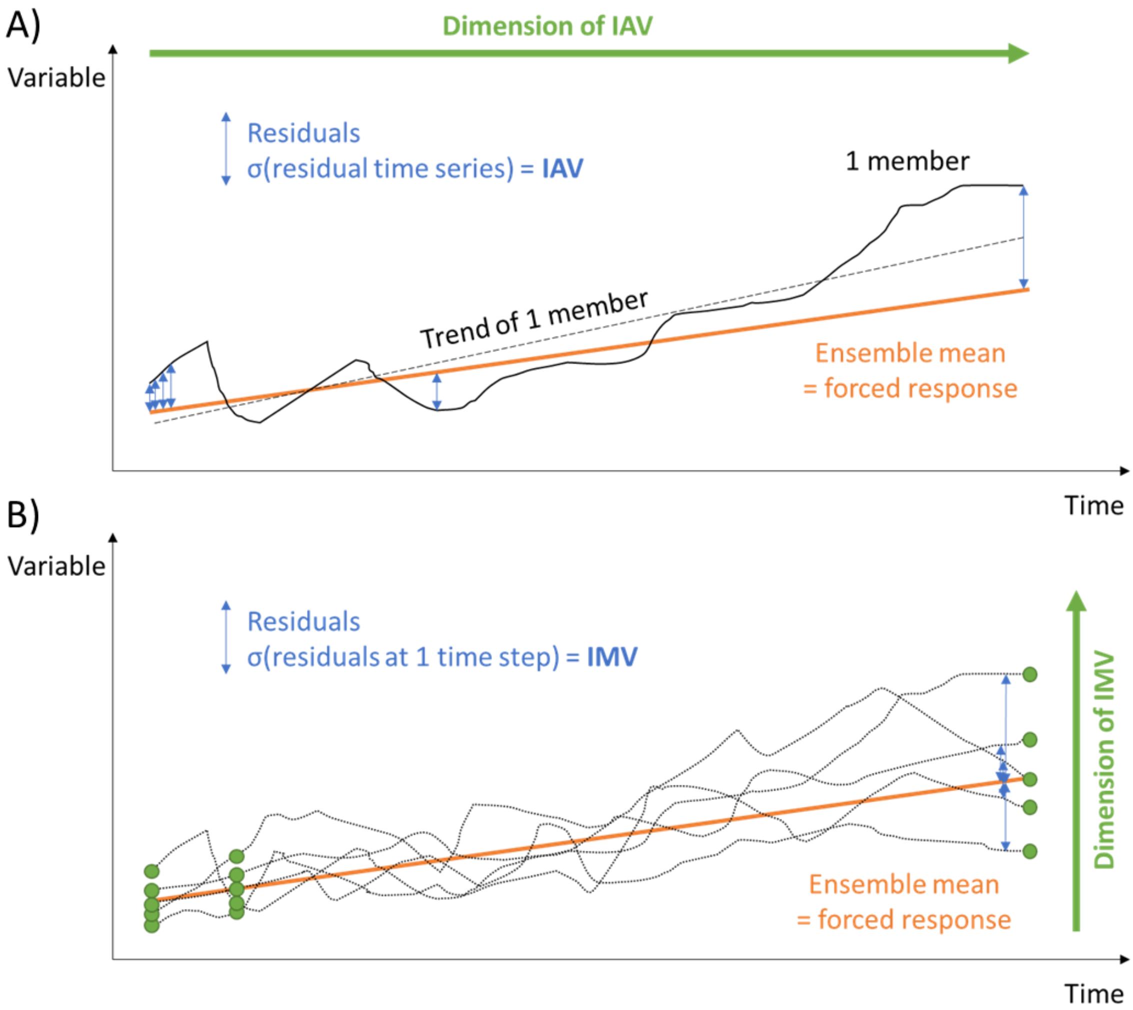


Figure 1: Conceptual depiction of IAV (A), IMV (B) and the EM in a SMILE

4.5 The CRCM5-LE

The CRCM5-LE dataset is a 50-member ensemble of the Canadian earth system model CanESM2 (Fyfe et al. 2017), downscaled by the Canadian regional model CRCM5 (Version 3.3.3.1, Šeparović et al. 2013; Martynov et al. 2013) for two domains: Northeastern North America (not part of this study) and Europe. The 50 members originate from five families of simulations, each starting at different 50-year intervals of a preindustrial run with a stationary climate and run from 1850-1950. They are then separated into ten members each by small atmospheric perturbations and run from 1950 to 2099. This 50 member CanESM2-LE was then dynamically downscaled within the ClimEx project with CRCM5 over a domain covering Europe (EU11d2) with a horizontal grid-size mesh of 0.11 degrees on a rotated latitude-longitude grid, corresponding to a 12-km resolution, using 5-minute time steps, which fits the common EURO-CORDEX grid specifications. Further information on the settings of the whole experiment, as well as a detailed description and analysis of the dataset (also for the American domain) can be found in Leduc et al. (2019). The stored variables and the terms of use can be found in the respective documentation on the project homepage (www.climex-project.org). The CRCM5-LE 12-km grid equals the one used in EURO-CORDEX simulations, although the CRCM5-LE domain is slightly smaller, still covering most of Europe (Figure 2). Many publications have made use of the CRCM5-LE already. An overview can be found on the ClimEx project homepage (www.climex-project.org).

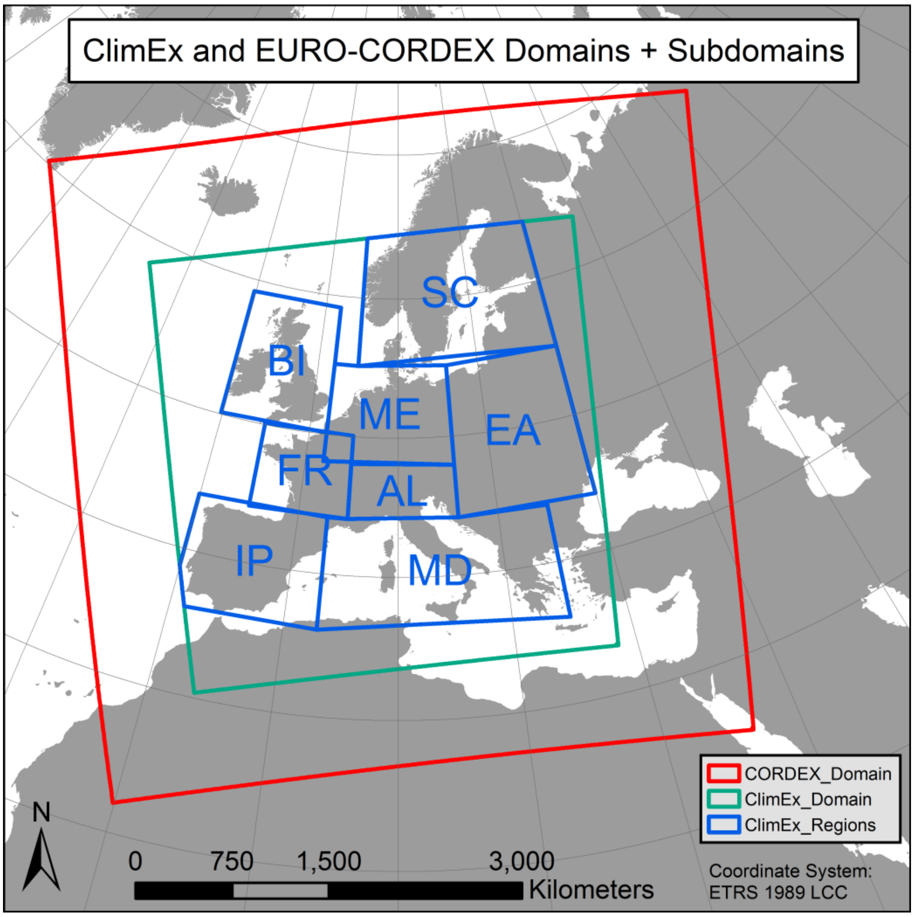


Figure 2: The EURO-CORDEX domain (red), the CRCM5-LE domain (green) and the subdomains used in the first two publications

4.6 Two additional RCM SMILEs for Europe

Additional to the CRCM5-LE, two similar SMILEs have been used in one publication and are briefly introduced here. The first ensemble consists of 21 members of the global model CESM, downscaled by the regional model CCLM over a slightly smaller domain as the CRCM5-LE domain (Fischer et al. 2013; Addor und Fischer 2015; Brönnimann et al. 2018). In the second ensemble 16 members of the global model EC-EARTH were downscaled with the regional model RACMO for a smaller domain over Western Europe (Aalbers et al. 2018). Differences in variability of the three RCM-SMILEs may stem from the differences in the resolution of both GCMs and RCMs, the different domain sizes, the different models, differences in aerosol forcing in the RCM simulations (constant in CCLM and CRCM, transient in RACMO) and in the application of an ocean slab model in the EC-EARTH-RACMO ensemble. The initialization is carried out differently in the three driving GCM-SMILEs: CanESM2 builds on a hybrid approach, where five members of different ocean conditions starting in 1850 were divided into ten members each by atmospheric perturbations in 1950 (see Leduc et al. 2019 for details). The CESM members stem from small atmospheric perturbations of the order of 10^{-13} on January 1st 1950 (Fischer et al. 2013). EC-EARTH uses the first 16 days in the year 1850 of an initial run to start the 16 members (Aalbers et al. 2018).

4.6.1 The CRCM5 and CCLM models as representatives of a strong drying mechanism

The CRCM-LE seems to be quite sensitive to land-atmosphere feedback mechanisms that accelerate dry and hot conditions during summer, especially in southern Europe. The extreme conditions on the Iberian Peninsula are a typical manifestation of this behavior, where hot and dry meteorological periods dry out the soils, which leads to a negative feedback loop, creating more hot and dry air near the surface, accelerating the further drying of the soil. Other regional models like CCLM also seem to have this feedback loop included as the CCLM shows much drier and hotter conditions in the arid parts of Europe compared to other RCMs, driven by the same GCM in the CORDEX ensemble. When looking at the past years in real-world measurements, not climate model data, this feedback mechanism seems rather plausible, and these “extreme” RCM simulations might be a realistic description of Europe’s future climate conditions in summer.

5 Setting natural variability and extreme events into context of climate projections

Since the importance of natural variability and extreme events from the impact side and the basics of climate simulations on various scales have been discussed now, the two parts need to be integrated. And this is the core of my PhD thesis. The huge amount of data provided by SMILEs give us a new source of information to answer questions on two major topics:

- the relevance of natural variability in the context of climate change (simulations)
- the probabilities of extreme events in an unprecedented manner

NV of the climate system can be defined in many ways and there is no clear definition that serves all aspects and fields of research. The term generally covers all deviations in climate variables on all possible timescales from paleoclimatology, where cycles of hot and cold eras can take millions of years easily, up to ice and warm periods in the last ice ages with eras of several thousands of years. When going to the most recent 100-200 years, many oscillations have been identified with frequencies of several decades (e.g., ENSO, NAO), which in the context here will be called long-term natural variability. Short term NV is the variability that covers shorter time periods. Interannual variability (IAV) expresses the variability of climate as captured instinctively by humans over their life and has effects on the planning activities of all kinds of sectors. The IAV can most simply be quantified as the standard deviation of a group of consecutive years, e.g., a 30-year period.

Under all the possible ways to explore this data, I chose to look at three different topics and corresponding research questions. They are listed in Table 1 and summaries of the main conclusions for these questions are given in the following chapters 6 to 8, accompanied by an overview page and the full text manuscript for each paper separately.

The *Earth System Dynamics (ESD)* paper looked at three RCM-SMILEs and how their IAV can be described and evaluated both for historical and future time periods. The *Climate Dynamics* paper compared the CRCM5-LE against the “common” EURO-CORDEX ensemble and quantifies the role of natural variability in a multi-model ensemble. Both papers investigate the opportunities that SMILEs provide for the analysis of natural variability in climate projections. The third paper then moves to the other above mentioned bullet point: SMILEs as valuable tools for extreme events analysis by investigating drought probability changes.

Table 1: Research topics, corresponding research questions and the associated publications where they are addressed.

Topics and research questions	**Journal article**
Interannual variability (IAV) in the SMILEs • **How well do the SMILEs represent the observed IAV?** • **How does the IAV develop in the future in SMILEs?** • **What methods are suitable to robustly detect changes in IAV within SMILEs?**	Published in ESD
Natural variability as part of the overall uncertainty of climate projections • **How large is the spread of signals from the CORDEX ensemble with special consideration of the GCM-RCM combinations?** • **How much of the uncertainty in a multi-model ensemble like CORDEX can be explained by NV?** • **What does that mean for the interpretation of multi-model ensembles?**	Published in Climate Dynamics
SMILEs as statistically sound data base for extreme event analysis • **How will dry period frequencies change over Europe in the future?** • **How does that relate to the 2018 summer drought in Germany?** • **How does this meteorological signal evolve into a hydrological model response at two gauges in Germany?**	Submitted to Geophysical Research Letters

6 Interannual variability in the SMILEs

The variability of three regional SMILEs is assessed against the variability of an observational data set (E-OBS). After a successful assessment, the future development of the interannual variability (IAV) is analyzed by different approaches to detect the time of emergence in IAV changes.

6.1 How well do the SMILEs represent the observed interannual variability?

We find that the large ensembles analyzed generally represent observed IAV correctly, but care needs to be taken during the analysis for specific regions and indicators. Cases of many individual members showing both higher and lower variability compared to observational IAV can be found for all ensembles for specific indicators and regions. However, the single observed realization of historical climate makes it difficult to evaluate systematic errors of the ensembles, as the E-OBS distribution is not necessarily representative of the perfectly sampled IAV. A large study on the global mean surface temperature in GCM-SMILEs with the same methodology also concludes that most models reproduce the internal variability well, but regional differences can be large (Suarez-Gutierrez et al. 2021).

6.2 How does the interannual variability develop in the future?

The results found for changes in IAV are generally in line with existing literature on Europe. We likewise find increasing variability for the summer temperature and precipitation (Fischer und Schär 2009; Fischer und Schär 2010; Vidale et al. 2007; Yettella et al. 2018; Suarez-Gutierrez et al. 2018) and decreasing variability for winter temperature and precipitation (Bengtsson und Hodges 2019; Holmes et al. 2016). The summer extreme indicators for the number of heatwaves and maximum dry period length also show increased variability in two of the three models, in conjunction with increases in their mean states.

6.3 What methods are suitable to robustly detect changes in interannual variability?

Three different methods were tested to detect if – and when – changes in IAV become significant.

The first method applied the classical IAV definition to each member individually over moving time periods. Counting the members with significant changes (both positive and negative) showed that for most indicators the number of members stays well below 50 % for the whole 21st century (except maximum length of dry periods in CRCM5 and CCLM).

The second approach makes use of the assumption that IMV is a good approximation of IAV and calculates the IMV for each year. The resulting time series also show significant changes in variability for only few cases (mostly for CRCM5).

The third approach combines the two approaches to define variability (IAV and IMV), and pools together all years of a 30-year period from all members, resulting in a much larger sample size per period. With this approach significant changes in variability can be detected much more robust. Significant changes in variability thus already start to emerge early in the 21st century for some indicators. Differences occur between an absolute and relative treatment of changes in precipitation-based indicators.

TITLE
Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe

AUTHORS
Fabian von Trentini, Emma F. Aalbers, Erich M. Fischer, Ralf Ludwig

JOURNAL
Earth System Dynamics 11, 1013-1031

PUBLISHING DATE
18.11.2020

DOI
https://doi.org/10.5194/esd-11-1013-2020

JOURNAL IMPACT FACTOR (WEB OF SCIENCE)
5.54

AUTHOR CONTRIBUTIONS
FvT designed the concept of the study, performed the analysis, and created all figures. EEA and EMF provided the RACMO and CCLM data and helped improve the concept and analysis. FvT led the paper writing with input from all authors.

CITATIONS (CROSSREF)
9

PLAIN LANGUAGE SUMMARY
We compare the inter-annual variability of three single-model initial-condition large ensembles (SMILEs), downscaled with three regional climate models over Europe for seasonal temperature and precipitation, the number of heatwaves, and maximum length of dry periods. They all show good consistency with observational data. The magnitude of variability and the future development are similar in many cases. In general, variability increases for summer indicators and decreases for winter indicators.

Earth Syst. Dynam., 11, 1013–1031, 2020
https://doi.org/10.5194/esd-11-1013-2020

Earth System Dynamics
Open Access

Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe

Fabian von Trentini[1], **Emma E. Aalbers**[2,3], **Erich M. Fischer**[4], **and Ralf Ludwig**[1]

[1]Department of Geography, Ludwig-Maximilians-Universität, Munich, 80333, Germany
[2]Royal Netherlands Meteorological Institute (KNMI), P.O. Box 201, 3730 AE De Bilt, the Netherlands
[3]Institute for Environmental Studies (IVM), Vrije Universiteit, Amsterdam, 1081 HV, the Netherlands
[4]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, 8092, Switzerland

Correspondence: Fabian von Trentini (fabian.trentini@lmu.de)

Received: 8 November 2019 – Discussion started: 20 November 2019
Revised: 24 August 2020 – Accepted: 17 September 2020 – Published: 18 November 2020

Abstract. For sectors like agriculture, hydrology and ecology, increasing interannual variability (IAV) can have larger impacts than changes in the mean state, whereas decreasing IAV in winter implies that the coldest seasons warm more than the mean. IAV is difficult to reliably quantify in single realizations of climate (observations and single-model realizations) as they are too short, and represent a combination of external forcing and IAV. Single-model initial-condition large ensembles (SMILEs) are powerful tools to overcome this problem, as they provide many realizations of past and future climate and thus a larger sample size to robustly evaluate and quantify changes in IAV. We use three SMILE-based regional climate models (CanESM-CRCM, ECEARTH-RACMO and CESM-CCLM) to investigate downscaled changes in IAV of summer and winter temperature and precipitation, the number of heat waves, and the maximum length of dry periods over Europe. An evaluation against the observational data set E-OBS reveals that all models reproduce observational IAV reasonably well, although both under- and overestimation of observational IAV occur in all models in a few cases. We further demonstrate that SMILEs are essential to robustly quantify changes in IAV since some individual realizations show significant IAV changes, whereas others do not. Thus, a large sample size, i.e., information from all members of SMILEs, is needed to robustly quantify the significance of IAV changes. Projected IAV changes in temperature over Europe are in line with existing literature: increasing variability in summer and stable to decreasing variability in winter. Here, we further show that summer and winter precipitation, as well as the two summer extreme indicators mostly also show these seasonal changes.

1 Introduction

In addition to the changes in mean climatological states, the variability of the climate system is an important feature of climate change. This variability of the climate system is subject to various drivers. Variability can be caused by natural forcings on different timescales, such as changes in solar radiation or volcanic eruptions. Variability of single components of the climate system can also be caused by the redistribution of heat and momentum between and within different components (e.g., ocean and atmosphere) of the coupled climate system, referred to as unforced internal variability. Next to these variations, anthropogenic changes in greenhouse gas concentrations contribute to a changing climate. Climate variability can be sampled on different timescales from hours and days up to multi-decadal variations.

For impact analysis of climate change, the future development of interannual variability (IAV) is of utmost importance in addition to changes in the mean climate state. Particularly increases in the IAV can be crucial for many impact sectors, as this makes it much harder for stakeholders to plan from year to year. In this study, daily data are used to calculate six indicators (summer and winter mean surface air temperature (tas) and accumulated seasonal precipitation (pr)) and two indicators for climatological extremes with high societal impact: the number of heat waves per year (tas-HW-Nr) and the maximum length of dry periods per year (pr-DP-MAX); see Table 1 for definitions. Heat waves can cause an increase in health problems and even fatalities among the population, as well as damage to infrastructure (e.g., highways) and ecological problems, as seen during the most recent heat waves in Europe (e.g., 2003, 2018, 2019). Long dry periods can have major impacts on ecology, forestry, agriculture, drinking water supply, power plant cooling outages, transport on rivers and many more. All these sectors should implement adaptation strategies to face changing climatic conditions, including IAV.

Early studies with regional climate models from PRUDENCE showed a distinct increase in IAV for the 21st century in summer temperatures (Fischer and Schär, 2009, 2010; Vidale et al., 2007) as well as decreasing winter temperature variability (Vidale et al., 2007). Later work with ENSEMBLES models revealed a less pronounced increase in summer temperature variability (Fischer et al., 2012). Analysis of SMILEs also showed future increases in variability of European summer temperatures with increasing global warming (Suarez-Gutierrez et al., 2018; Yettella et al., 2018). Holmes et al. (2016) and Tamarin-Brodsky et al. (2020) also find increasing temperature variability in summer and decreasing variability in winter for the future. European winter temperature variability has already today decreased since the pre-industrial era in another large climate model ensemble (Bengtsson and Hodges, 2019).

For large areas of the globe, including Europe, an increase in precipitation variability from daily to multi-decadal timescales is expected due to higher temperatures (Pendergrass et al., 2017). However, Ferguson et al. (2018) find significant changes in the IAV of monthly precipitation only in a small fraction of CMIP5 models for a western European domain until the end of the 21st century. Earlier analysis with regional climate models revealed future increases in summer and decreases in winter for IAV of precipitation over similar domains as used in this study (Giorgi et al., 2004).

Uncertainty of future climate projections can stem from at least three sources (Hawkins and Sutton, 2009): emission scenario, model response to a selected forcing and internal variability of the climate system. Internal variability is often referred to as "irreducible uncertainty" at timescales beyond seasons to decades. While scenario and model response uncertainty have been referred to in many climate simulation experiments (CMIP and CORDEX), the internal variability component had received less attention for many years. In recent years, a new tool for the assessment of internal variability has become quite popular: single-model initial-condition large ensembles (SMILEs), where the same model is forced with the same emission scenario several times – with the runs (members) just differing in their initial conditions. This setup is able to isolate the internal variability component from the scenario and model response uncertainty for the respective model. Based on SMILEs, it has been shown that the contribution of internal variability to the total uncertainty of multimodel ensembles (CMIP, CORDEX) can be large, especially for mid-term projections and precipitation (Kumar and Ganguly, 2018; von Trentini et al., 2019) at the regional level.

The terms large ensemble (LE) and SMILE usually describe the same thing, but we prefer SMILE as it incorporates the type of large ensemble which is built upon different initial conditions. Up to now, a number of large ensembles have been produced. Deser et al. (2020) give the latest overview of the different SMILEs available. However, most studies only use one SMILE for their analysis and the rare comparisons are usually just between two ensembles: similar patterns of internal variability of temperature and precipitation trends for the middle of the 21st century were found for a CCSM3 and an ECHAM5 ensemble over North America by Deser et al. (2014). Martel et al. (2018) showed a consensus of the IAV of annual mean and extreme precipitation in a CanESM2 large ensemble (which is also used for boundary conditions of the CRCM5 in this study; see Sect. 2) and CESM-LE with two global observational data sets.

All these simulations were performed with global climate models (GCMs), and only a few were dynamically downscaled with regional climate models (RCMs). Here, we compare three dynamically downscaled large ensembles, all forced by the Representative Concentration Pathway (RCP) 8.5, for Europe. It is the first time that regional large ensembles are compared with respect to forced changes and their internal variability. The added value of RCM simulations is well documented for EURO-CORDEX (Giorgi et al., 2009; Torma et al., 2015; Sørland et al., 2018; Giorgi, 2019). Downscaled climate data are also a necessity for impact modeling at regional to local scales (e.g., for hydrology, agriculture, biodiversity research) due to their more accurate representation of topography, complex coastlines and the heterogeneity of land surface properties.

Terminology in the context of climate variability is not always clear in the literature, as the terms natural variability, internal variability, IAV and inter-member variability (IMV) are often used synonymously or mixed up. Here, the term internal variability describes the variability at timescales from seconds up to multiple decades caused by unforced internal effects of a model or the real world due to the chaotic nature of the climate system only, without incorporating naturally forced variability due to volcanic eruptions and solar forcing. Anthropogenic changes in greenhouse gas and aerosol concentrations as well as natural solar and volcanic forcing can cause changes in the mean climate state that are superimposed by internal variability. Moreover, higher greenhouse gas concentrations can cause changes in the internal variability itself in the future as well – adding another component to climate change effects.

Table 1. Indicators and their definitions.

Indicator	Used variable	Definition
tas-JJA	tas	Summer mean temperature (June–August)
tas-DJF	tas	Winter mean temperature (December–February)
tas-HW-Nr	tas	Number of heat waves per year; a heat wave is defined as a minimum of 3 d above the 95th percentile of daily mean temperature of the reference period; no filtering on summer months is applied at any stage; however, by definition the heat waves will occur during summer in the reference period. They might, however, extend to spring and fall under RCP8.5.
pr-JJA	pr	Summer precipitation sum (June–August)
pr-DJF	pr	Winter precipitation sum (December–February)
pr-DP-MAX	pr	Maximum length of a dry period per year; dry periods are a minimum of 11 consecutive days with every day showing less than 1 mm of precipitation; no filtering on summer months is applied at any stage; the periods can thus also occur in winter (but this is rather unlikely in Europe).

In this study, IAV is calculated as the standard deviation of anomalies of each member from the ensemble mean (EM), which represents an estimate of the forced response of the respective model. SMILEs have the advantage that the EM is a much better estimate of the forced response than a trend fitted to single members. After removing the forced response, the residual IAV equals the total unforced internal variability – including low-frequency variations. We will show that IAV can be well estimated by the IMV of a SMILE in many cases, as both metrics sample the unforced internal variability of a SMILE, just on different dimensions: IAV is sampled on the time dimension of a single member, while IMV is sampled on the member dimension for each year (see Sect. 3.4). Both IAV and IMV are terms used to describe the more general term "internal variability" throughout this paper. Also, note that the EM of each SMILE can be regarded as the change signal with the highest probability, but which specific member would become realized depends on internal variability.

The usage of three RCM-SMILEs has some advantages compared to multi-model ensembles consisting of single realizations that enable us to go beyond the current literature on IAV changes. First, we can better evaluate the models against observations as (a) the forced response of the model is better estimated by the EM and (b) we thus reduce the problem of having only one realization of climate to the observational data side of the evaluation. Second, we can more reliably quantify changes in the IAV and rule out that potential changes only occur as a result of sampling uncertainty. Additionally, we can better demonstrate when changes are significantly different from historical conditions. In recent literature, often no significance test of detected changes in interannual variability is performed (e.g., Bengtsson and Hodges, 2019). Many studies just provide information about the robustness of change (e.g., by stippling in maps), measured by the accordance in the sign of change of (usually) 67 % of the models of multi-model ensembles (e.g., Holmes et al., 2016). This does, however, not allow information about the significance compared to a reference climate. Third, SMILEs allow a better separation of models as they are not only described by one member each. Additionally, we combine these general SMILE advantages with the higher resolution of RCMs.

The remaining paper is structured as follows: First, the model ensembles and the observational data set E-OBS are briefly presented. Then, the change in mean temperature and precipitation together with the inter-member spread of projected changes is analyzed for each ensemble, as the mean changes are important baseline information for variability changes. Next, the IAV of the three regional large ensembles is evaluated against E-OBS to assess the abilities of the models to represent observed IAV for the selected indicators. Finally, IAV in historical climate and future changes in IAV are compared between SMILEs. This includes a discussion on different methods to estimate IAV and detect significant changes in IAV. In the main text, most results will only be presented for mid-Europe (ME), with references to the other regions and their figures in the Supplement.

2 Data

The climate model ensembles each consist of a GCM single-model initial-condition large ensemble, which has been dynamically downscaled over Europe with a single regional climate model: a 50-member CanESM2-CRCM5 ensemble (Kirchmeier-Young et al., 2017; Leduc et al., 2019), a 21-member CESM-CCLM ensemble (Fischer et al., 2013; Addor and Fischer, 2015; Brönnimann et al., 2018) and a

Table 2. Specifications of the three ensembles used in this study.

	CRCM	CCLM	RACMO
Scenario	RCP8.5	RCP8.5	RCP8.5
GCM	CanESM2	CESM 1.0.4	EC-EARTH 2.3
GCM resolution	2.8°	2.0°	1.0°
RCM	CRCM5	CCLM4-18-7	RACMO22E
RCM resolution	0.11°	0.44°	0.11°
No. of members	50	21	16

16-member EC-EARTH-RACMO ensemble (Aalbers et al., 2018), all forced with the RCP8.5 scenario, resolved on different spatial resolutions (Table 2). Hereafter we indicate the GCM-RCM combinations with the RCM names only (CRCM, RACMO and CCLM). This setup with a shared scenario but different models enables us to analyze differences in internal variability between the three ensembles. Differences in variability may stem from the differences in the resolution of both GCMs and RCMs, the different domain sizes, the different models, and differences in aerosol forcing in the RCM simulations (constant in CCLM and CRCM, transient in RACMO) and in the application of an ocean slab model in the EC-EARTH-RACMO ensemble. RACMO also uses slightly different grid specifications. The domain size of CCLM equals the EURO-CORDEX domain, while CRCM uses a slightly smaller domain, and RACMO only captures central and northeastern Europe (Fig. 1). The initialization is carried out differently in the three driving GCM-SMILEs: CanESM2 builds on a hybrid approach, where five members with different ocean conditions starting in 1850 were divided into 10 members each using atmospheric perturbations during the initialization in 1950 (see Leduc et al., 2019 for details). The CESM members stem from small atmospheric perturbations of the order of 10^{-13} on 1 January 1950 (Fischer et al., 2013). EC-EARTH uses the first 16 d in the year 1850 of an initial run to start the 16 members (Aalbers et al., 2018). These climate model data sets will be compared but will also be compared to observations: the gridded observational data set E-OBS has daily precipitation and temperature available for Europe (version v12.0, spatial resolution of 0.22° on a rotated pole grid). We use the E-OBS data set because of its availability for Europe and it has similar spatial resolution to the regional climate models under consideration. We accept the known weaknesses of the data set (mostly caused by inhomogeneities in the sparse station network; E-OBS is also known for rather low precipitation fields; see Hofstra et al., 2009) and assume that it is nonetheless suitable for the purpose of this study.

3 Methods and results

3.1 Spatial aggregation

All indicators (Table 1) are calculated on a grid basis for each ensemble. For comparison, the indicators are spatially aggregated to four regions in Europe, for which all three RCM domains overlap (Fig. 1): the British Isles (BI), France (FR), mid-Europe (ME) and the Alps (AL). These regions are well known from other European climate model studies (Lenderink, 2010; Lorenz and Jacob, 2010; Kotlarski et al., 2014; von Trentini et al., 2019) and were introduced by Christensen and Christensen (2007). The procedure of calculating the indicators on the grid level and spatially aggregating them afterwards has the advantage that no regridding of data is needed. However, the different spatial resolutions of the models alone can potentially lead to higher variability in the 0.11° data (CRCM and RACMO), compared to the 0.22° (E-OBS) and 0.44° (CCLM) data. This is especially the case for spatially heterogeneous variables and indicators (Giorgi, 2002; Kendon et al., 2008). The indicators in this study, however, have relatively low spatial heterogeneity (seasonal temperature and precipitation, heat waves, and dry periods are rather large-scale phenomena), where the range of spatial resolutions of the data used here (between 0.11 and 0.44°) is not expected to be significantly sensitive. The effect of regridding before the calculation of indicators is shown by a short experimental analysis, where 1 year of five members of the 0.11° CRCM data is regridded to 0.44° (simply averaging 4×4 grid cells each), before the indicators are calculated. The results show that the effect of regridding on the IMV is indeed minor for the indicators considered (Fig. S1). The approach of direct regional aggregation of the indicators calculated on the grid level is therefore applied for the further analysis of this study.

3.2 Ensemble spread of projected mean climate change

Before analysis of IAV, simple scatterplots of the changes in the mean climatological states of each member for temperature and precipitation for summer and winter between 1980–2009 and 2070–2099 are shown for ME; see Figs. 2 and 3. They give a first impression on the spread of projected changes between the members of SMILEs and on the differences in the mean changes between models. We test the similarity of means between all models with a two-sample t test ($\alpha = 0.05$) and the similarity of spreads with a Brown–Forsythe test (BF test with $\alpha = 0.05$) on equal variances. The BF test does not show significant differences in variance for both temperature and precipitation in winter and summer for all model combinations. The spread of signals between members of one SMILE can be solely attributed to the internal variability of the respective model.

In summer in ME, all models show decreasing precipitation; between −3 % and −16 % for RACMO and −14 % to

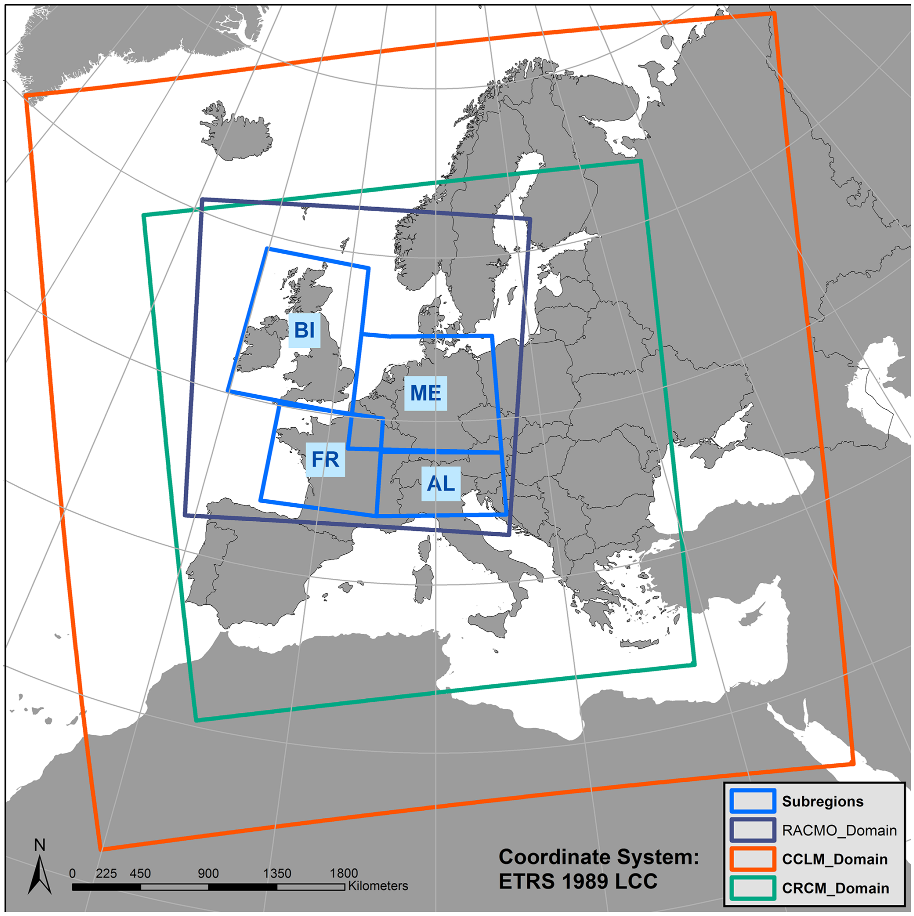


Figure 1. Domains of the three RCMs and the boundaries of the four analysis regions; BI: the British Isles; FR: France; ME: mid-Europe; AL: Alps; the CCLM domain matches the EURO-CORDEX domain.

−35 % for CRCM and CCLM (Fig. 2). Increases in summer temperature between 3 and 5 °C are projected by RACMO and CCLM, while CRCM shows much higher changes between 5 °C and more than 6 °C. Thus, RACMO and CCLM show similar changes in temperature (although statistically different in their means), while CCLM and CRCM show similar changes in precipitation. The spread of changes for both temperature and precipitation of RACMO and CRCM are similar, both in terms of standard deviation and total range, while CCLM shows a higher standard deviation and total range (Table 3). Similar results as discussed here for mid-Europe (mean changes and spread) are found for France and the Alps (not shown), with the largest decrease in summer precipitation over France and the strongest warming over the Alps. The British Isles region shows less pronounced changes for both temperature and precipitation (although they are consistent in sign). CRCM shows a closer similarity of precipitation decreases to RACMO in BI rather than to CCLM, as is the case for the other three regions ME, FR and AL.

In winter, all models project increasing precipitation (1 %–32 %) and temperature increases between 1.4 and 5 °C by the end of the 21st century (Fig. 3). RACMO and CRCM show a similar standard deviation and range of temperature and precipitation changes again, together with similar mean changes as well (significant for temperature but not for precipitation). CCLM shows distinctly smaller changes in combination with a smaller spread of changes (Table 3). Similar results also appear for FR, AL and BI, although some members of CCLM and CRCM also project a slight decrease in precipitation in these regions.

3.3 Evaluation against E-OBS

For the evaluation of the models' IAV against E-OBS, we apply an approach proposed by Suarez-Gutierrez et al. (2018) and Maher et al. (2019). For the observations and for each model and member separately, the anomalies relative to the reference period 1961–1990 are calculated for the years 1957–2099 and 1957–2015 in E-OBS, respectively (Fig. 4). Model mean state biases of the indicators, which can be quite large (see Fig. S2) are thereby removed. For each year,

Table 3. Standard deviation and total range for changes in Figs. 2 and 3.

	tas (°C)			pr (%)		
	CRCM	RACMO	CCLM	CRCM	RACMO	CCLM
Summer SD	0.27	0.28	0.39	4.2	3.8	5.1
Summer range	1.16	0.96	1.59	16.2	13.2	21.1
Winter SD	0.37	0.38	0.30	5.5	5.3	4.2
Winter range	1.87	1.47	1.33	26.1	21.2	12.9

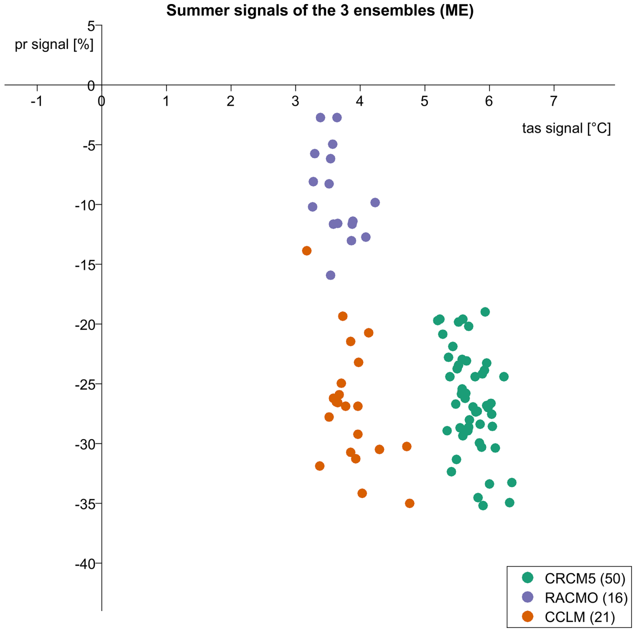


Figure 2. Change in mean summer temperature and precipitation for every member of the three ensembles in mid-Europe (2070–2099 against 1980–2009). Changes are relative to each members' value in 1980–2009 for precipitation, while temperature changes are absolute.

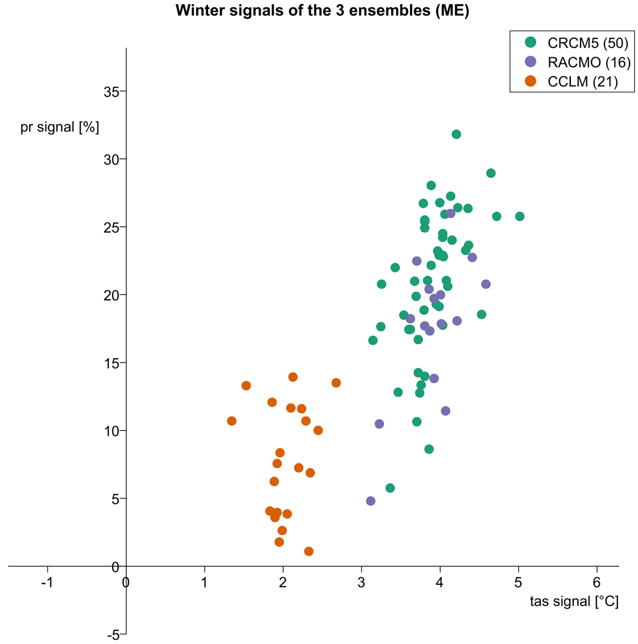


Figure 3. Change in winter temperature and precipitation for every member of the three ensembles in mid-Europe (2070–2099 against 1980–2009). Changes are relative to each members' value in 1980–2009 for precipitation, while temperature changes are absolute.

we then plot the ensemble median, minimum and maximum member, the area between the 12.5th and 87.5th percentile, within which 75 % of the members are situated, and the E-OBS data. For a perfect model, the E-OBS data are expected to occur normally distributed within the range spanned by the ensemble and are concentrated in the inner 75 %, several years situated in between the minimum and maximum of members, but also outside this range from time to time. If the E-OBS data are concentrated too much inside the total range or even the 75 % area, the variability of the ensemble overestimates the observational variability. By contrast, if too many E-OBS data points exceed the ensemble spread, SMILE underestimates observational variability. To quantify this further, the probability density function of the anomalies in the period 1957–2015 is plotted for each member and E-OBS separately. The functions are estimated probability densities based on a normal kernel function, similar to an approach by Lehner et al. (2018).

The forced response (ensemble median) increases for all indicators analyzed, except for summer precipitation, which decreases in all models, and there is no clear change in pr-DP-MAX in RACMO. Note that this approach does not only compare the IAV of the models and E-OBS, but also the forced response in the historical period. Differences in the distributions can thus also arise from a false representation of the forced response in a model, compared to the trend in E-OBS. On the other hand, if the modeled and observed distributions largely coincide, both the forced response and the IAV are well represented by a model. All three models generally seem to reproduce the forced response in the historical part quite well, as the models are consistent with the trends of the E-OBS points (e.g., increase in tas-JJA). Only for summer precipitation (pr-JJA) do all models show a decrease in

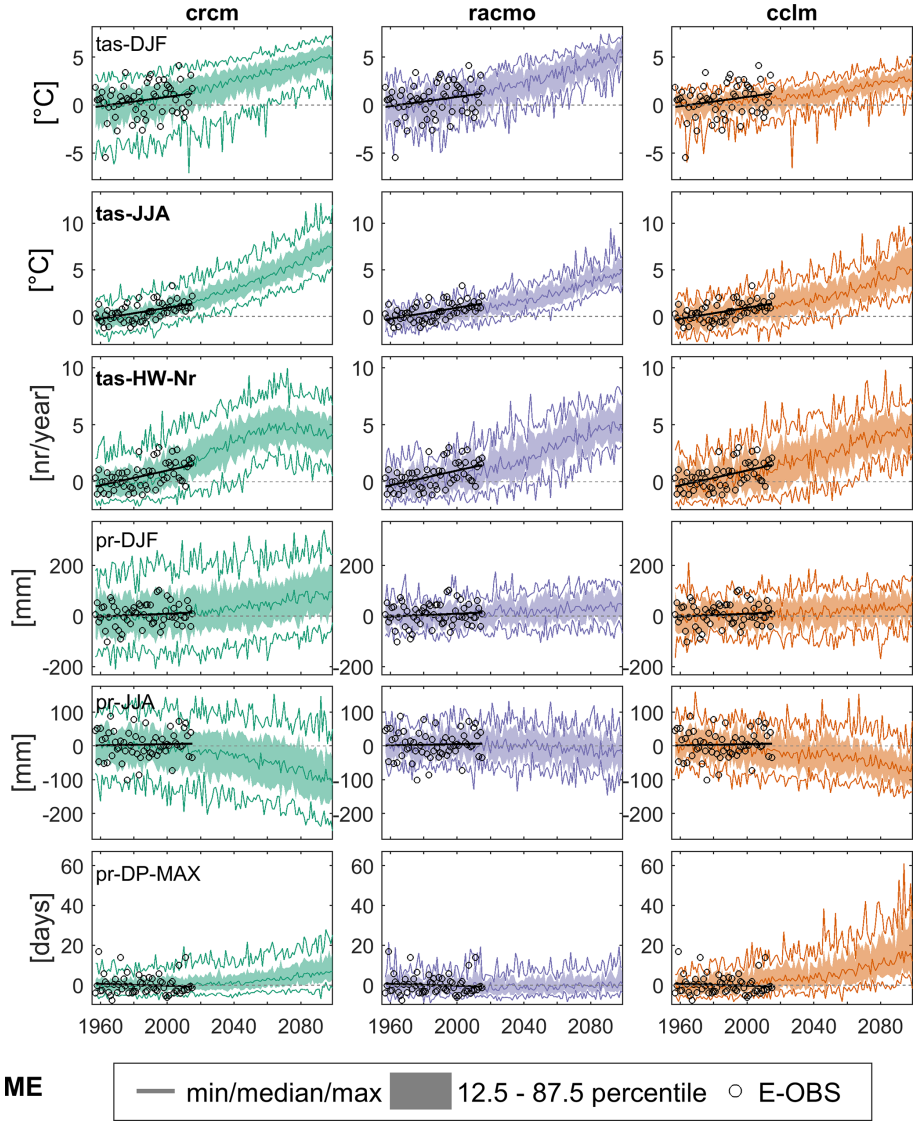


Figure 4. Anomalies from 1961 to 1990 of the six indicators in mid-Europe (ME) for E-OBS (circles 1957–2015) and the three ensembles (1957–2099), represented by the median, minimum and maximum (solid lines) of the ensemble and an area from the 12.5th and 87.5th percentile, spanning the range of the inner 75 % of the members (shading). Black lines show the linear trend for the E-OBS points. The indicator names are in bold when the trend is significant using a Mann–Kendall test ($\alpha = 0.05$).

the forced response, whereas E-OBS shows no significant negative trend. However, in all ensembles, not all members show decreasing trends. The observations may thus still be consistent with the simulated forced response.

The comparison of E-OBS and the three SMILEs during the historical period from 1957 from 2015 in ME shows largely good representations of IAV in the ensembles, as seen by well distributed E-OBS points within the 75 % range (12.5 %–87.5 % quantile) and minimum and maximum range of the ensembles (Fig. 4). However, too a strong clustering of the E-OBS points in the 75 % area occurs for winter precipitation in CRCM (97 % fall inside) and number of heat waves in CCLM (90 %), meaning the simulated IAV is too high. On the other hand, too many outliers beyond the minimum and maximum members appear in winter temperature in CCLM (22 % outside of total range), winter precipitation in RACMO (17 %) and maximum duration of dry periods in CRCM (10 %), i.e., for these models and indicators the simulated IAV is too low. To demonstrate this further we calculate probability density functions of the annual anomalies for each member and E-OBS (Fig. 5). Note that probability density functions could also be somewhat inflated by the underlying mean trend, but we expect this effect to be small because trends in the observational period are small and largely consistent between models and observations. To evaluate the ability of SMILEs to represent observational IAV, we test whether the E-OBS distribution looks like a possible member of the respective ensemble. The observations are not expected to fall near the ensemble median but rather should be ideally indistinguishable from a random additional member of the ensemble, since E-OBS only represents one possible realization of historical climate. Significant differences can be seen for the examples already mentioned: the distribution of CRCM in winter precipitation is much broader than the E-OBS distribution, whereas the winter temperature distribution for CCLM is concentrated too much in the middle compared to E-OBS.

Similar results as in ME can be found in the other three regions as well (Figs. S3–S5), with only a few cases where the E-OBS distributions show a distinctly different shape than all members of the ensembles (Fig. 6 for AL and Figs. S6 and S7 for BI and FR, respectively), especially for the maximum duration of dry periods of CCLM in France. This is not too surprising, as the maximum duration of dry periods is an extremely sensitive indicator because of its potentially extreme differences in magnitude between (model) years/members (one wet day can make a huge difference). The other two SMILEs are able to represent the E-OBS variability for this indicator in France though. Other remarkable features are the underestimation of variability in RACMO for all six indicators in the British Isles region (Fig. S6), as well as the relatively good performance of the models for the Alps (Fig. 6), which is probably the most difficult region for a model to represent correctly due to the strong spatial heterogeneity. However, the Alps show some "outlier-members" with distinctly different distributions in the ensembles (e.g., winter precipitation in CRCM), which cannot be found in the other regions – at least not this pronounced. These outliers demonstrate how large the influence of internal variability between members can be in a single realization of climate, as these outlier members just deviate from all other members by their initial conditions. Estimating the IAV of a model thus needs a large number of members, as even with 49 members that give a uniform range of distributions (pr-DJF in CRCM5 in the Alps, Fig. 6), one single additional member can change the picture and add more information on the range of IAV for the respective model. The evaluation of E-OBS gets rather difficult in these cases, as the methodology is based on the assumption that the E-OBS distribution should somehow "fit" to the uniform range of distributions of the model. If the E-OBS data showed such an outlier behavior, it means that the one realization of climate variability as seen by the E-OBS data might still be part of a SMILE's range of possible variability manifestations. However, from a probability perspective, the conclusion of similar variabilities becomes rather unlikely. It just makes it harder to prove that E-OBS has a different distribution than all members of a SMILE.

3.4 Projected changes in internal variability and the connection between IAV and IMV

The temporal development of the internal variability is important information along with the underlying forced response (change in the EM) for a better understanding of changing climatic conditions. We discuss three possible ways to describe changes in the internal variability on annual timescales within a SMILE. All three methods are based on the application of a BF test on equal variances. While IAV and IMV are expressed as standard deviations (SDs), the BF test analyses differences in the variance, which is just the square of SD. In the cases of IAV (methods 1 and 3), moving time periods of 30-year length, shifted by 1 year each (1961–1990, 1962–1991, ..., 2070–2099) are used. For the second method, IMV is sampled over the dimension of the ensemble size per year. Thereby we test whether the internal variability changes significantly over time. Differences in the methods arise from the different data samples used for the testing.

The first method is based on the methodology that one would choose for single members and observations. By looking at the IAV for different periods within one member, changes in IAV can be detected. Usually the forced response is taken out of the data by fitting a polynomial to the data and only using the residuals. However, the estimate of the forced response of a model based on only one member may deviate from the true forced response (Lehner et al., 2020). Therefore, we choose the EM as an estimate of the forced response and use the residuals from each member with respect to the EM for the BF test. The BF test results is Boolean information for each member and each moving period on whether the variance has significantly changed with respect to a reference period (here: 1961–1990) or not. This information can be used to show the percentage of members with a significant change (separated for positive and negative changes) in each period. The advantage of SMILEs within this approach is the better estimate of the forced response and a more robust detection of changes, as they are built on multiple members. One member alone could be an outlier in its representation of (changes in) IAV just as it could be for the trend. The method is sensitive to the chosen reference period of course, as the variance of this period determines the baseline variance. Since we use moving periods, the results do not change significantly when using different periods starting in the 1960s.

For the second method, we make use of the assumption that the IMV for a given year is a good approximation for the IAV in a period around that year (e.g., ± 15 years to get a sample size of the typical 30 years for climate analysis). The sampling of variability is thus not based on consecutive time series within each member but on a compound of annual data for 1 year from all members of a SMILE. The IMV is also based on residuals from the EM as for IAV. Under the assumption of small influence of low-frequency variability, IMV should be a good estimate of total unforced IAV, as

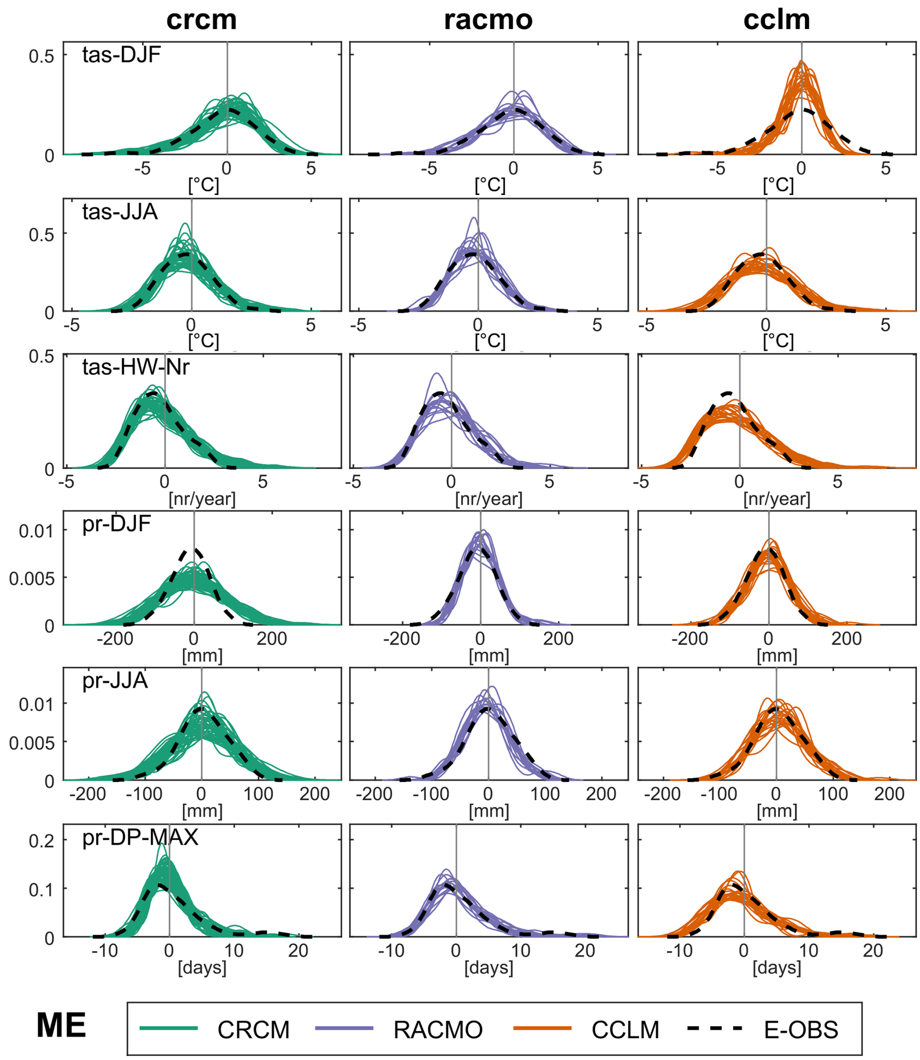


Figure 5. Probability density functions of the annual anomalies during the period 1957–2015 in E-OBS and each ensemble member for all six indicators in mid-Europe (ME).

both sample the annual variability during a similar state of climate for a given time horizon. This concept is particularly relevant in the presence of non-linear forcing. For instance, the response to a volcanic eruption cannot be separated easily from unforced IAV. In addition, the anthropogenic forcing since 1950 has not been linear in time. Using IMV is an elegant way to get around this challenge. Some recent publications support the concept of using IMV as an approximation of IAV, although the two have different background meanings: while IAV has a physical meaning and represents the variability of a consecutive sequence of weather phenomena, IMV is a measure of variability without a direct physical meaning (Nikiéma et al., 2018).

In Leduc et al. (2019) the authors state that "In the case of a climate system under transient forcing, the use of [IMV equation] to assess temporal variability using the inter-member spread involves weaker assumptions than calculating the residual temporal variability from detrended time series." (Leduc et al., 2019, p. 681), based on the study by Nikiéma et al. (2018). A recent publication by Wang et al. (2019) even concludes that the IMV of winter sea level pressure over Eurasia in a SMILE is driven by the same mechanisms as observed IAV via an EOF analysis. Another example is the analysis of seasonal mean and heavy precipitation in Europe where long-term variations are small compared to the IAV in the RACMO ensemble (Aalbers et al., 2018). A comparison of IAV and IMV in each ensemble is carried out by comparing the means and standard deviations of these two variability metrics – calculated over different dimensions of the ensemble data. The IAV is calculated for each member during a 30-year reference period (1980–2009) and three future periods. The mean and standard deviation of these 50, 21 and 16 values is calculated for IAV. The IMV is calculated for each of the 30 years of the respective period between the 50,

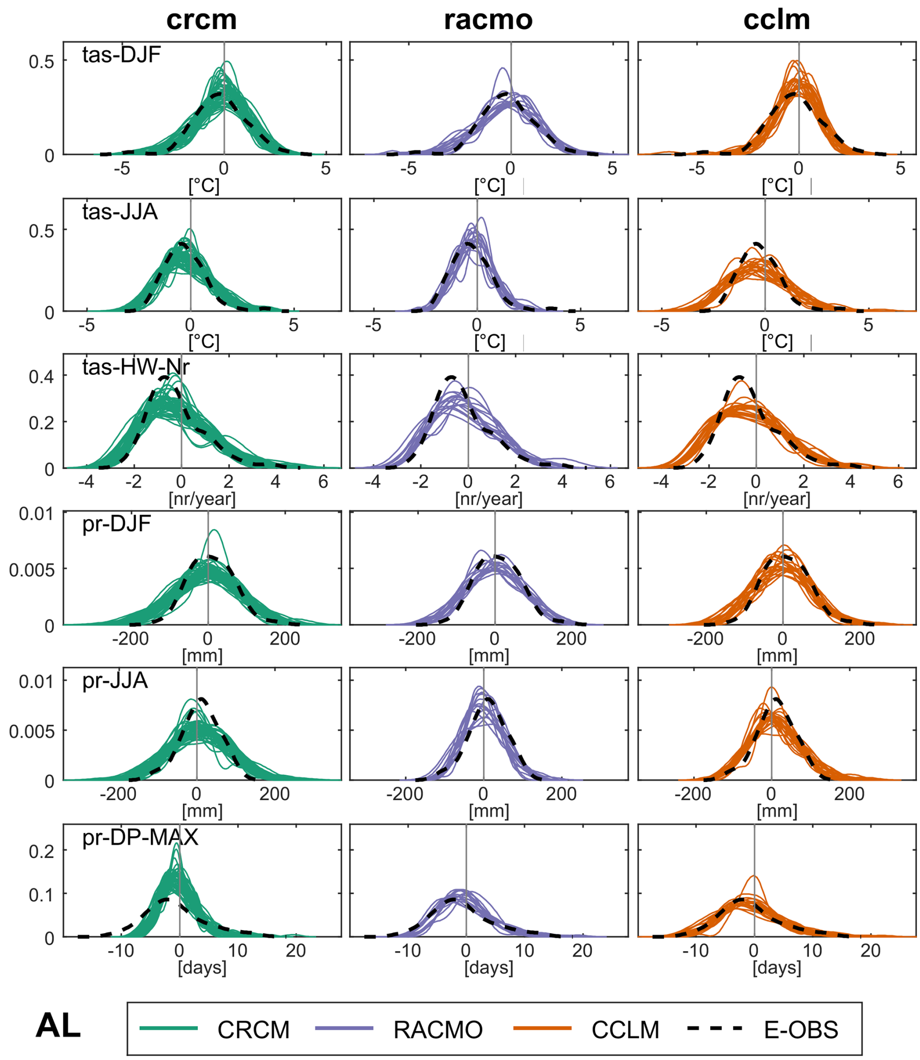


Figure 6. Probability density functions of the annual anomalies for all six indicators in the Alps (AL). For details, see Fig. 5.

21 and 16 members, leading to a mean and standard deviation, calculated from these 30 values. The mean and standard deviation of IAV and IMV are indeed very similar for all indicators, periods and regions (exemplarily shown for winter temperature in Fig. 7). Especially the similarity in future changes suggests a similar response to external forcing for the two variability metrics IAV and IMV. The IMV has the advantage that it is insensitive to inflation effects of the variability due to an existing trend and forced effects like cooling after volcanic eruptions for example. However, although IAV and IMV seem to be similar in many cases (see also literature above), they can potentially also differ under special circumstances in the external forcing like volcanic eruptions. Note that according to our results IMV is always slightly larger than IAV. This may be caused by two factors. First, detrending the time series is more likely to remove than to add some of the variability, and it affects only IAV. Second, also without detrending the data, in the presence of low-frequency variability, IAV is likely smaller than IMV, which has no auto-correlation in the underlying data. For the variables considered here, differences are small though, implying that the low-frequency variability is indeed small compared to the high-frequency variability.

Given the similarity of IAV and IMV, the third approach pools together the annual anomalies from the EM from all members for a given 30-year period (30 times the ensemble size, e.g., 30 × 21 values for CCLM). It is therefore a mixture of IAV and IMV, enabling a more robust BF test result for changes in variance by a larger sample size.

While the interpretation of temperature-based indicators is always based on absolute anomalies from the EM, it can be useful to look at both absolute and relative anomalies from the EM for precipitation-based indicators (in contrast to the previous evaluation against E-OBS, where they were only absolute anomalies). Relative anomalies thus give information on how much the standard deviation changes with respect

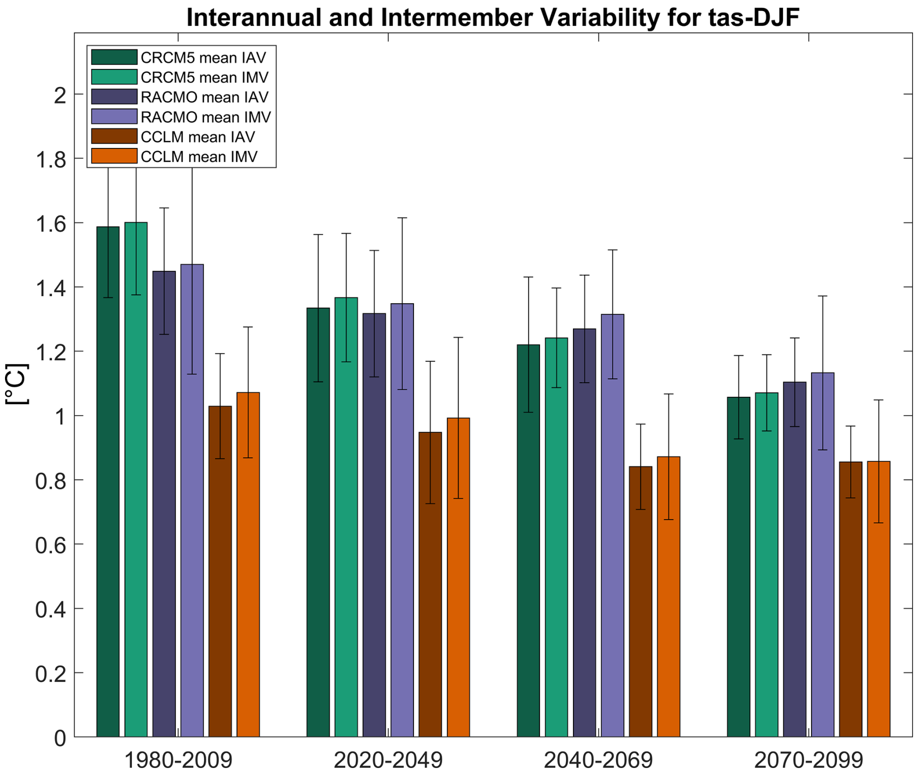


Figure 7. IAV and IMV of winter temperature in the three ensembles for the reference period (1980–2009) and three future periods. Bars: mean over the variability of each member (IAV) or year (IMV). Error bars: ± standard deviation (members or years); IMV: 16/21/50 members; IAV: 30 years of the respective period.

to changes in the EM. For example, a stable IMV in absolute terms will result in a decrease in the relative IMV when the EM increases. Increasing relative IMV, together with an increasing EM on the other hand means that the internal variability is increasing even more than the mean.

The percentage of members with significant changes in IAV as a function of time is shown in Fig. 8 for all indicators, for mid-Europe. Significantly decreasing IAV for winter temperature and increases for summer temperature and heat waves are found, but only for a minority (<50 %) of the members for all models, even at the end of the 21st century. While all three models point to the same direction of change, percentages differ substantially. For winter and summer precipitation, an even smaller percentage of members shows significant changes in IAV, and there is no clear direction of change in any model. Only CRCM for pr-JJA shows an increasing number of members with significant positive changes throughout the second half of the 21st century. For dry periods, RACMO has a very small number of members showing significant changes in both directions, while CRCM and CCLM show marked increases in the number of members with significant positive changes in IAV throughout the 21st century. For the last period 2070–2099, even all members of CCLM show significant increases.

The temporal evolution of IMV (relative to the EM for precipitation-based indicators, Fig. 9) generally supports the direction of changes as seen by the method using the percentage of members with significant changes in IAV. However, when testing for significant changes in the variance between members, hardly any of the changes are significant. CRCM shows significant changes in the majority of years for tas-DJF from 2040 on, for tas-JJA from 2080 on and for pr-JJA from 2060 on. As the IMV is calculated for each year, the plot shows the noise in the IMV per year, which can be large.

Figure 10 shows the change in variability determined from the pooled annual anomalies from the EM for moving 30-year periods from all members. This means, all 30 anomalies from all members are pooled together before calculating the standard deviation (i. e. pooled IAV) and tested for significant changes in the variance with a Brown–Forsythe test. Given the much larger sample size per 30-year period, in contrast to the two former methods, we can now see significant changes in many combinations of indicator and model (Fig. 10). As expected from the previous two methods, internal variability decreases for winter temperature and increases for summer temperature and the number of heat waves. In contrast to the former methods, however, significant changes can be detected earlier. In these cases, the internal variability has already changed significantly in the historical simulations

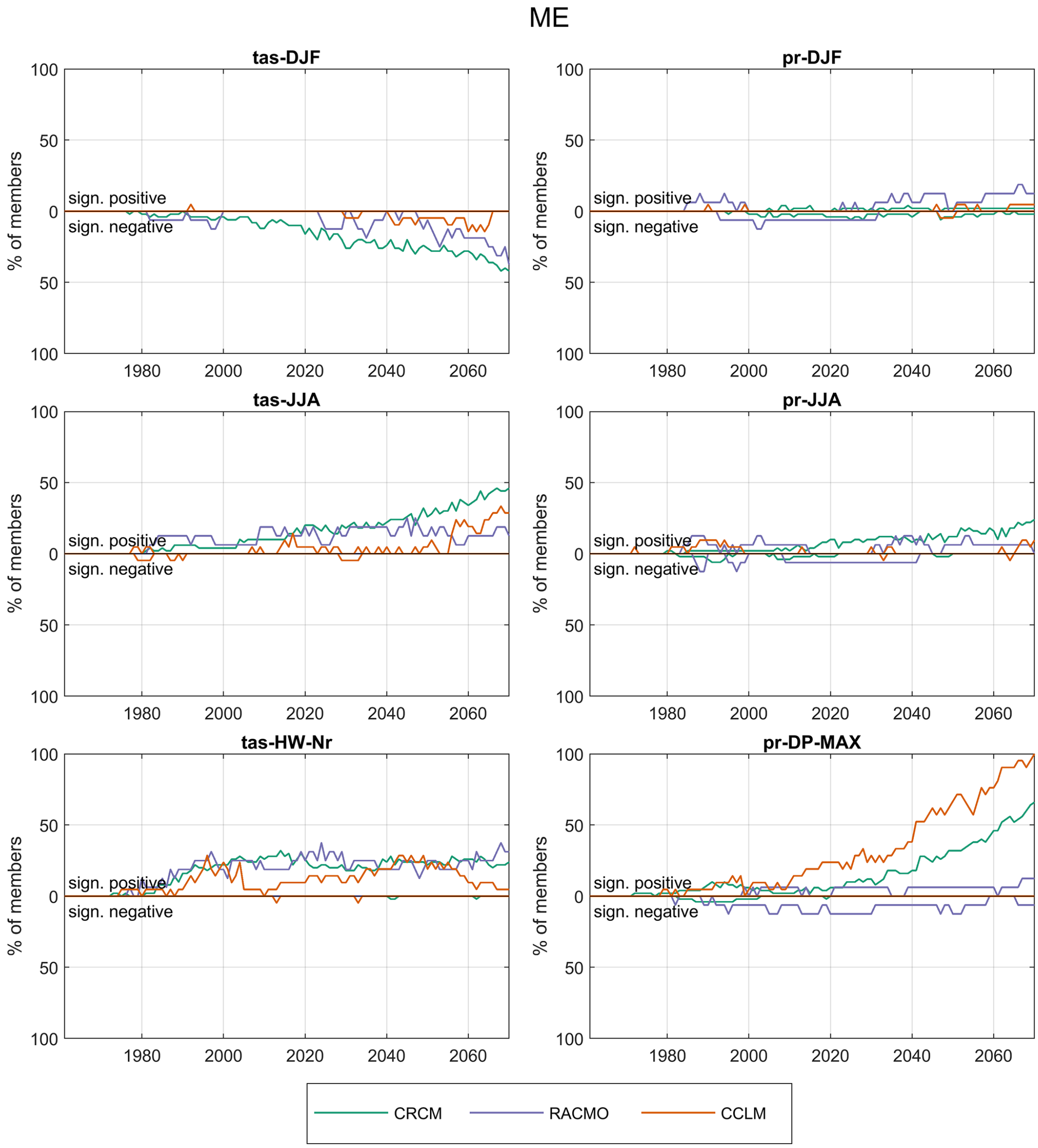


Figure 8. Percentage of members with significantly different variance (Brown–Forsythe test with $\alpha = 0.05$) with respect to the reference period 1961–1990 in mid-Europe. The analysis is based on residuals after removing the EM from each member. The years on the x axis denote to the starting year of moving 30-year periods.

of SMILEs or it changes in the present/near future around 2020. The internal variability in the number of heat waves increases until about 2010–2030, reaches a plateau for about 30–40 years and then decreases again. This behavior can be explained by the forced response of the indicator, which shows strong increases until around 2060, when the number of heat waves stabilizes around 6 (and even decreases afterwards in CRCM, Fig. S2), because the heat waves become so long that their number per year cannot increase anymore. This is especially true for CRCM, where the mean duration of heat waves at the end of the 21st century is much longer than for CCLM and RACMO and about 16 d (not shown), leading to a rough estimate of $6 \times 16 = 96$ heat wave days per year, equal to about 3 months. Since heat waves are defined by the 95th percentile of temperature in the reference period (thus describing extreme conditions), the former extreme heat becomes a regular condition during the summer months at the end of the 21st century in CRCM. For the pooled IAV, both absolute and relative changes in IAV are shown for precipitation-based indicators to demonstrate the effect of the two different approaches. For pr-DJF, CRCM does not show any change in absolute IAV, while this stable behavior in combination with the increase in pr-DJF in the EM leads to a decreasing relative IAV, which is significant

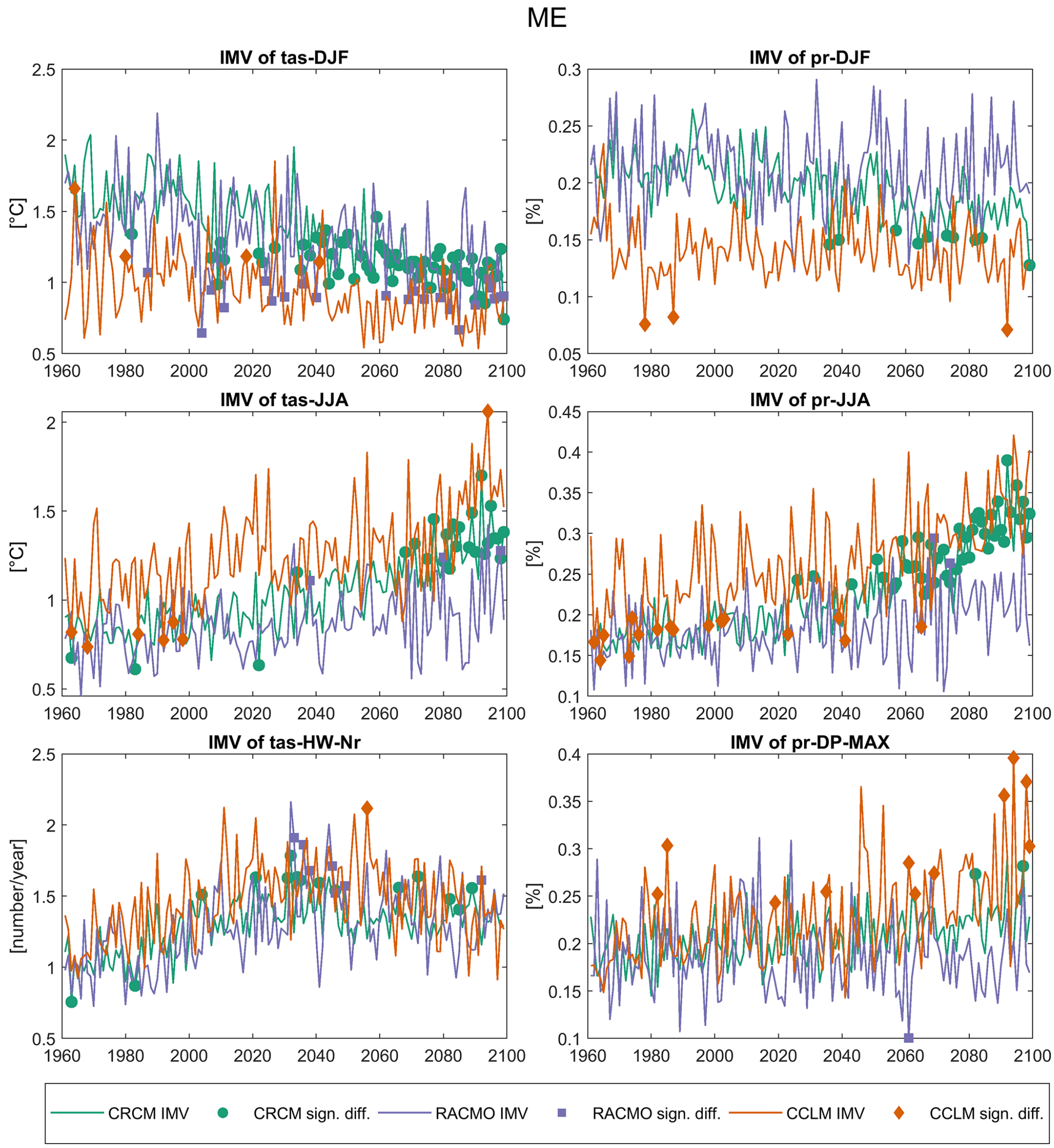


Figure 9. IMV per year sampled on the dimension of the respective ensemble size (50, 21 and 16) for mid-Europe. The analysis is based on residuals after removing the EM from each member. The markers highlight years with a significantly different variance than the reference year 1961. Precipitation-based indicators are shown with their relative anomalies from the ensemble mean (percentage).

from the early 21st century onwards. CCLM and RACMO show increasing variability in absolute terms, but changes are significant for RACMO only, from ∼ 2060 onwards. For both CCLM and RACMO there is no clear change in IAV relative to the change in EM for pr-DJF. Note that while RACMO shows the lowest absolute IAV it shows the highest relative IAV. This originates from the lower EM for winter precipitation in RACMO compared to CRCM and CCLM, which both have quite distinct wet biases (Fig. S2). For summer precipitation, absolute IAV increases according to all models, while EM decreases. Changes in absolute IAV are largest and significant for CRCM and RACMO from ∼ 2000, respectively ∼ 2045 onwards. For CCLM, changes are not significant. Owing to the decreasing EM, increases in relative IAV are significant for all models and significant changes occur earlier in time (∼ 1970 for CCLM, ∼ 1990 for CRCM and ∼ 2040 for RACMO). The changes in EM and IAV in both summer and winter have also been detected by Pendergrass et al. (2017) for CMIP5 and CESM-LE precipitation data in extratropical regions. IAV of pr-DP-MAX does not change according to RACMO, while CRCM and CCLM show distinct increases that go hand in hand with increases in the EM that is also much stronger in these two models than in RACMO (Fig. 4). The changes are significant for relative IAV later in time than for absolute IAV.

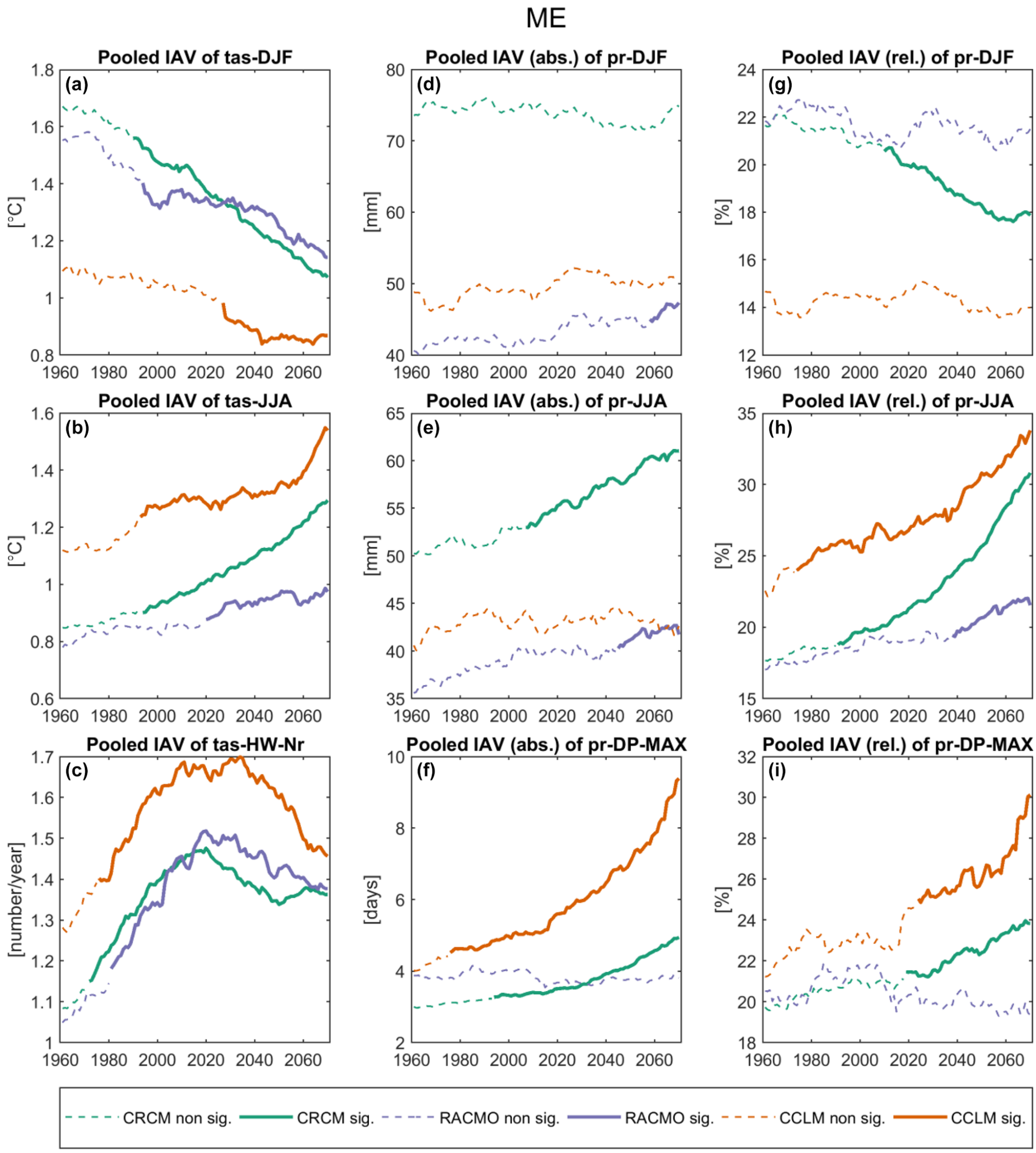


Figure 10. "Pooled IAV" for mid-Europe. The analysis is based on residuals, pooled together from all members, after removing the EM from each member. Temperature-based indicators are shown in absolute terms **(a–c)**. Precipitation-based indicators are shown both in absolute terms **(d–f)** and relative to the ensemble mean **(g–i)**. The change from dashed to solid lines marks the point in time when all following periods show significant changes in variance (BF Test with $\alpha = 0.0.5$).

The abovementioned results are mostly valid for the other regions as well. Differences in the magnitudes of variability and its changes are briefly discussed in the following (see Figs. S8–S10): ME shows higher winter temperature variability than the other three regions, especially than BI. Lower levels of variability compared to the other regions occur over the British Isles for winter temperature and winter precipitation (relative to EM). The Alps show a smaller variability for the number of heat waves than the other regions. The variability of pr-DP-MAX for all three ensembles is similar to ME in AL, while BI and FR hardly show any significant changes. If all regions are considered, RACMO generally has the highest internal variability in winter and the lowest variability in summer for temperature and precipitation (relative to EM), while CCLM has the highest internal variability for summer temperature and precipitation (relative to EM) as well as for heat waves and dry periods (both absolute and relative to EM). Significant changes generally occur similar to ME for winter and summer temperature and precipitation (both absolute and relative to EM). Changes in the number of heat waves are not significant in CCLM in all three regions and in RACMO in AL.

The first method, testing the percentage of members with significant changes in IAV, gives a good overview of the behavior of the members in general. It can, however, only inform about the direction of change in IAV. Additional information on the magnitude of the changes in IAV is needed to get the whole picture. The second method using IMV is in general agreement with the first method when looking at the direction of change. It incorporates the magnitude of internal variability and can show significant changes in the IMV in the same figure. The variations in IMV from year to year are relatively large. Therefore, the BF test results also largely depend on the choice of a reference year to test all other years against. Changes from one year to the next can give very different results. Unstable significance testing is the result. Method 1 is less sensitive to the reference given the overlapping periods. Considering the sample sizes in this study, the best results can be obtained with method 3, where sensitivity can be tested with a much larger sample size. For method 1 it is 30 values per period, for method 2 it is 16/21/50 values per year, and for method 3 it is the product of both 30 years and the ensemble size of the respective ensemble. However, if the sample size was larger for the first two methods, they would probably also result in the detection of significant changes, e.g., when using more members.

To see how the ensemble size impacts the results for the pooled IAV, we reduce the largest ensemble (CRCM with 50 members) to the size of the other ensembles (16 and 21) and repeat the analysis for ME. Even after reducing the ensemble size to only 16 members, the changes at the end of the 21st century in CRCM are still significant for indicators that already showed significant changes for all 50 members. However, the detection of significant changes is only possible at a later time horizon (Fig. S11) for changes in tas-DJF (around 40 years later), relative changes in pr-DJF (20), absolute changes in pr-JJA (20) and relative changes in pr-DP-MAX (15). Tests with a number of ensemble sizes suggest that around 10 members are sufficient to detect the significance of changes and around 20 are sufficient to detect the timing of these significant changes additionally.

4 Discussion

The number of SMILEs available for the quantification of internal variability in this study is still relatively small – we only used three GCM-RCM combinations (to the knowledge of the authors these three ensembles are the only regionally downscaled SMILEs over Europe). More simulations with RCM-SMILEs could help to make results even more robust – especially for winter precipitation and dry periods, where the three ensembles do not agree on the change in variability.

The effect of regional aggregation after the calculation of indicators on the grid level and the potential effects of the original resolution of different data sets on the internal variability seem to be minor for the selected indicators, as seen in the experimental analysis conducted on a subset of the data (Fig. S1). This estimate of sensitivity to differing spatial resolutions might be conservative, however. Nevertheless, the methodology seems to be suitable for the selected indicators of this study.

Methods based on anomalies from the EM are chosen in order to compare results despite different biases in historical and future mean climate states. It can, however, not be ruled out that differences in the variability may originate from mean state biases of the models. CRCM for example shows much higher precipitation sums than the other two ensembles, leading to higher variability in absolute terms. The normalization with the ensemble mean covers these differences in absolute amounts. In the end it largely depends on the definition of variability: is one interested in absolute deviations (mm) or in the fluctuations in relative terms (%)? Results are sensitive to a relative versus an absolute definition or vice versa. The relative approach has the advantage that it allows for a fair comparison of models with different mean precipitation amounts. This is also why a recent publication by Giorgi et al. (2019) gave preference to the relative definition, for example.

The scatterplots of projected changes for seasonal temperature and precipitation (Figs. 2 and 3) show both agreement and dissent, but usually at least two of the three models show similar ranges for one variable. Better agreement might be possible when comparing the data sets not for a fixed period but for periods with the same global warming level in each driving GCM.

We find that the large ensembles analyzed here generally represent observed IAV correctly, but care needs to be taken during the analysis for specific regions and indicators. Cases of many individual members showing both higher and lower variability compared to observational IAV can be found for all ensembles for specific indicators and regions. However, the single observed realization of historical climate makes it difficult to evaluate systematic errors of the ensembles, as the E-OBS distribution is not necessarily representative of the perfectly sampled IAV. It would be interesting to compare large ensembles against an observational large ensemble as proposed by McKinnon and Deser (2018) to better see systematic deficiencies of large ensembles compared to observations.

The results found for changes in IAV are generally in line with existing literature on Europe. We likewise find increasing variability for the summer indicators tas-JJA and pr-JJA (Fischer and Schär, 2009, 2010; Vidale et al., 2007; Yettella et al., 2018; Suarez-Gutierrez et al., 2018) and decreasing variability for the winter indicators tas-DJF and pr-DJF (Bengtsson and Hodges, 2019; Holmes et al., 2016). The summer extreme indicators tas-HW-Nr and pr-DP-MAX also show increased variability in two of the three models, in conjunction with increases in their mean states. Several mechanisms contribute to the changes in all indicators. For changes in the summer temperature IAV, land–atmosphere coupling is becoming more important in central or northern Europe in the future because the transitional zone between dry and wet climates moves northwards from the Mediterranean region, leading to enhanced alternation of dry and wet summer soil moisture (Seneviratne et al., 2006; Fischer et al., 2011). Moreover, stronger warming over land than over the oceans causes the land–ocean temperature gradient in summer to increase. This results in increased variability in thermal advection, which is suggested to play a role in the increase in temperature variability in Europe as well (Holmes et al., 2016). Analysis of observations shows that in the Mediterranean more than half of summer temperature variability can be explained by large-scale atmospheric circulations and sea surface temperatures (Xoplaki et al., 2003). The decrease in winter temperature IAV is suggested to be influenced by changing circulation patterns (Vautard and Yiou, 2009), and a decrease in variability of advected heat due to the decrease in the winter land–ocean temperature gradient (Holmes et al., 2016) and arctic amplification and sea ice loss (Screen, 2014; Sun et al., 2015; Tamarin-Brodsky et al., 2020), even under unchanged circulation variability (Holmes et al., 2016; Tamarin-Brodsky et al., 2020).

The increase in summer precipitation variability that would not be expected under decreasing mean summer precipitation might be caused by a reduction in the number of wet days (>1 mm) that exists in all three ensembles (not shown), as discussed by Räisänen (2002). A reduction in wet days implies an increase in variability since the seasonal precipitation sum becomes more dependent on individual precipitation events.

Land–atmosphere feedback mechanisms are not yet fully understood, and there are still improvements needed in their implementation in earth system models and regional climate models (Vogel et al., 2018). Uncertainties in the future regional development of heat waves and dry periods are thus rather large (Miralles et al., 2019). Nevertheless, increasing frequency, intensity and variability in the number of heat waves as projected by SMILEs using RCP8.5 in this study seem plausible, although the magnitudes can be uncertain. The strong increase in the maximum length of dry periods in two of the models is not necessarily what could be expected. While the length of severe dry periods increases in the future for southern Europe, central and northern Europe do not show any change in the EURO-CORDEX data for dry period length (Jacob et al., 2014). Analysis of precipitation changes shows that both CRCM and CCLM (the RCM itself, not SMILE) are on the dry end of projections for summer precipitation (von Trentini et al., 2019). This might be related to sensitive implementations of land surface modules in these two RCMs. RACMO does not show an increase in the maximum length of dry periods.

Although beyond the scope of this paper, which only analyzed the manifestations of internal model variability in surface variables (tas, pr and associated indicators), there is a need for a better understanding of the mechanisms leading to the model-inherent characteristics of internal variability and why differences between the models appear (e.g., circulation patterns, ocean characteristics).

Using SMILEs for studying changes in IAV allows for much more robust statements on the direction, magnitude and emergence of changes, when using a certain model and RCP scenario. Analyzing the individual members, significant changes in IAV are found in less than half of the members for almost all indicators and ensembles, even at the end of the 21st century (Fig. 8). However, pooling the data of all ensemble members, all three ensembles show significant changes in internal variability of most indicators and often from early in the 21st century onwards (Fig. 10).

Thompson et al. (2015) showed that a statistical model based on a historical period could be as good as a SMILE for predicting future variability of seasonal temperature and precipitation trends up to 2060. The detection of significant increases in internal variability in summer and winter temperature much earlier than 2060 (Fig. 10) challenges this assumption of stable variability. For precipitation (especially absolute changes), however, the changes are often not significant before 2060, confirming the results of Thompson et al. (2015).

5 Conclusions

There is an increasing interest on the part of the scientific community to use single-model initial-condition large ensembles in a wide variety of applications, ranging from deeper levels of understanding of natural climate variability to impact assessments in different fields. The rich data basis which these ensembles provide for the analysis of internal variability is very valuable and enables new insights into this critical part of the climate system. Future changes, in particular, can be better set into context. The effects of dynamical downscaling of GCM large ensembles with regional climate models are not yet sufficiently explored. Further research is needed in this direction to see whether and by how much the internal variability is altered in the RCM simulations of a respective GCM large ensemble. However, downscaling is an important step to make climate simulation information attractive for local adaptation research and impact modellers. The results from this study can be helpful for these research communities to better understand and quantify the role of IAV in the climate system. In particular, increases in variability as seen for summer temperature, relative summer precipitation, heat waves and dry periods in most regions and models can be a huge burden for sectors like agriculture, ecology and hydrology.

The evaluation and comparison of the three RCM-SMILEs in this study gives a first overview of the agreement of SMILEs with observations and among each other. The generally high agreement with observations suggest that the internal variabilities of the RCM-SMILEs at the regional scale are useful approximations of the observed IAV of the climate system in Europe. The direction of changes in internal variability is also mostly the same between the ensembles, suggesting a relatively robust signal. While the "summer indicators" mostly show increasing variability in the future, winter temperature and precipitation show decreasing variability or no change. The change in variability is potentially impact-relevant as it suggests that the most extreme summers and winters may warm more strongly than the corresponding mean.

Despite an increasing number of studies that compare SMILEs (Deser et al., 2014; Martel et al., 2018; Rondeau-Genesse and Braun, 2019; Deser et al., 2020; Lehner et al., 2020), one limitation of many publications using SMILEs is the use of only one model and thereby one estimate of internal variability, leaving it unclear how representative the results are. Although the respective SMILE is usually evaluated against observations in these studies, the uncertainty in future changes in IAV cannot be quantified in the same way as in this study. A further challenge is also the fact that low-frequency variability at decadal and multi-decadal timescales remains uncertain and cannot be rigorously evaluated against observations due to the relatively short observational record and the difficulty of separating forced changes from unforced internal variability in observations.

Overall our results underline the great potential of SMILEs in quantifying the changes in IAV and when they become significant, also at the regional scale.

Data availability. The CRCM5 data are publicly available at https://climex-data.srv.lrz.de/Public/CanESM2_driven_50_members/ (Ouranos, 2020). RACMO data are available upon request directed at erik.van.meijgaard@knmi.nl. CCLM output is available upon request directed at erich.fischer@env.ethz.ch.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/esd-11-1013-2020-supplement.

Author contributions. FvT designed the concept of the study, performed the analysis and created all figures. EEA and EMF provided the RACMO and CCLM data and helped improve the concept and analysis. FvT led the paper writing with input from all authors.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "Large Ensemble Climate Model Simulations: Exploring Natural Variability, Change Signals and Impacts". It is not associated with a conference.

Acknowledgements. We acknowledge Urs Beyerle, Daniel Lüthi and Sven Kotlarski for their support in setting up the multi-member CCLM experiment. We thank Erik van Meijgaard and Camiel Severijns from KNMI for producing the 16-member KNMI RACMO-EC-EARTH and EC-EARTH ensembles. The CRCM5 was developed by the ESCER centre of Université du Québec à Montréal (UQAM; http://www.escer.uqam.ca, last access: 12 November 2020) in collaboration with Environment and Climate Change Canada. We acknowledge Environment and Climate Change Canada's Canadian Centre for Climate Modelling and Analysis for executing and making available the CanESM2 large ensemble simulations used in this study and the Canadian Sea Ice and Snow Evolution Network for proposing the simulations. The production of ClimEx was funded within the ClimEx project by the Bavarian State Ministry for the Environment and Consumer Protection. Computations with the CRCM5 for the ClimEx project were made on the SuperMUC supercomputer at the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. The operation of this supercomputer is funded via the Gauss Centre for Supercomputing (GCS) by the German Federal Ministry of Education and Research and the Bavarian State Ministry of Education, Science and the Arts.

Financial support. This research has been supported by the European Commission, H2020 Research Infrastructures (IMPREX (grant no. GA 641811) and EUCP (grant no. GA 776613)).

Review statement. This paper was edited by Sebastian Milinski and reviewed by three anonymous referees.

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., and van den Hurk, B. J. J. M.: Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability?, Clim. Dynam., 50, 4745–4766, https://doi.org/10.1007/s00382-017-3901-9, 2018.

Addor, N. and Fischer, E. M.: The influence of natural variability and interpolation errors on bias characterization in RCM simulations, J. Geophys. Res. Atmos., 120, 10180–10195, https://doi.org/10.1002/2014JD022824, 2015.

Bengtsson, L. and Hodges, K. I.: Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability?, Clim. Dynam., 52, 3553–3573, https://doi.org/10.1007/s00382-018-4343-8, 2019.

Brönnimann, S., Rajczak, J., Fischer, E. M., Raible, C. C., Rohrer, M., and Schär, C.: Changing seasonality of moderate and extreme precipitation events in the Alps, Nat. Hazards Earth Syst. Sci., 18, 2047–2056, https://doi.org/10.5194/nhess-18-2047-2018, 2018.

Christensen, J. H. and Christensen, O. B.: A summary of the PRUDENCE model projections of changes in European climate by the end of this century, Climatic Change, 81, 7–30, https://doi.org/10.1007/s10584-006-9210-7, 2007.

Deser, C., Phillips, A. S., Alexander, M. A., and Smoliak, B. V.: Projecting North American Climate over the Next 50 Years: Uncertainty due to Internal Variability, J. Climate, 27, 2271–2296, https://doi.org/10.1175/JCLI-D-13-00451.1, 2014.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M.: Insights from Earth system model initial-condition large ensembles and future prospects, Nat. Clim. Chang., 10, 277–286, https://doi.org/10.1038/s41558-020-0731-2, 2020.

Ferguson, C. R., Pan, M., and Oki, T.: The Effect of Global Warming on Future Water Availability: CMIP5 Synthesis, Water Resour. Res., 54, 7791–7819, https://doi.org/10.1029/2018WR022792, 2018.

Fischer, E. M. and Schär, C.: Future changes in daily summer temperature variability: driving processes and role for temperature extremes, Clim. Dynam., 33, 917–935, https://doi.org/10.1007/s00382-008-0473-8, 2009.

Fischer, E. M. and Schär, C.: Consistent geographical patterns of changes in high-impact European heatwaves, Nat. Geosci., 3, 398–403, https://doi.org/10.1038/ngeo866, 2010.

Fischer, E. M., Lawrence, D. M., and Sanderson, B. M.: Quantifying uncertainties in projections of extremes – a perturbed land surface parameter experiment, Clim. Dynam., 37, 1381–1398, https://doi.org/10.1007/s00382-010-0915-y, 2011.

Fischer, E. M., Rajczak, J., and Schär, C.: Changes in European summer temperature variability revisited, Geophys. Res. Lett., 39, L19702, https://doi.org/10.1029/2012GL052730, 2012.

Fischer, E. M., Beyerle, U., and Knutti, R.: Robust spatially aggregated projections of climate extremes, Nat. Clim. Change, 3, 1033–1038, https://doi.org/10.1038/nclimate2051, 2013.

Giorgi, F.: Dependence of the surface climate interannual variability on spatial scale, Geophys. Res. Lett., 29, 16-1–16-4, https://doi.org/10.1029/2002GL016175, 2002.

Giorgi, F.: Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next?, J. Geophys. Res.-Atmos., 124, 5696–5723, https://doi.org/10.1029/2018JD030094, 2019.

Giorgi, F., Bi, X., and Pal, J.: Mean, interannual variability and trends in a regional climate change experiment over Europe. II: climate change scenarios (2071–2100), Clim. Dynam., 23, 839–858, https://doi.org/10.1007/s00382-004-0467-0, 2004.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin, 58, 175–183, 2009.

Giorgi, F., Raffaele, F., and Coppola, E.: The response of precipitation characteristics to global warming from climate projections, Earth Syst. Dynam., 10, 73–89, https://doi.org/10.5194/esd-10-73-2019, 2019.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, B. Am. Meteorol. Soc., 90, 1095–1107, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hofstra, N., Haylock, M., New, M., and Jones, P. D.: Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, J. Geophys. Rese.-Atmos., 114, D21101, https://doi.org/10.1029/2009JD011799, 2009.

Holmes, C. R., Woollings, T., Hawkins, E., and de Vries, H.: Robust Future Changes in Temperature Variability under Greenhouse Gas Forcing and the Relationship with Thermal Advection, J. Climate, 29, 2221–2236, https://doi.org/10.1175/JCLI-D-14-00735.1, 2016.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P.: EURO-CORDEX: New high-resolution climate change projections for European impact research, Reg. Environ. Change, 14, 563–578, https://doi.org/10.1007/s10113-013-0499-2, 2014.

Kendon, E. J., Rowell, D. P., Jones, R. G., and Buonomo, E.: Robustness of Future Changes in Local Precipitation Extremes, J. Climate, 21, 4280–4297, https://doi.org/10.1175/2008JCLI2082.1, 2008.

Kirchmeier-Young, M. C., Zwiers, F. W., and Gillett, N. P.: Attribution of Extreme Events in Arctic Sea Ice Extent, J. Climate, 30, 553–571, https://doi.org/10.1175/JCLI-D-16-0412.1, 2017.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297–1333, https://doi.org/10.5194/gmd-7-1297-2014, 2014.

Kumar, D. and Ganguly, A. R.: Intercomparison of model response and internal variability across climate model ensembles, Clim. Dynam., 51, 207–219, https://doi.org/10.1007/s00382-017-3914-4, 2018.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguére, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: ClimEx project: a 50-member ensemble of climate change projections at 12-km resolution over Europe and northeastern North America with the Canadian Regional Climate Model (CRCM5), J. Appl. Meteorol. Clim., https://doi.org/10.1175/JAMC-D-18-0021.1, 2019.

Lehner, F., Deser, C., and Sanderson, B. M.: Future risk of record-breaking summer temperatures and its mitigation, Climatic Change, 146, 363–375, https://doi.org/10.1007/s10584-016-1616-2, 2018.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Lenderink, G.: Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations, Clim. Res., 44, 151–166, https://doi.org/10.3354/cr00946, 2010.

Lorenz, P. and Jacob, D.: Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40, Clim. Res., 44, 167–177, https://doi.org/10.3354/cr00973, 2010.

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., Kröger, J., Takano, Y., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability, J. Adv. Model. Earth Sy., 11, 2050–2069, https://doi.org/10.1029/2019MS001639, 2019.

Martel, J.-L., Mailhot, A., Brissette, F., and Caya, D.: Role of Natural Climate Variability in the Detection of Anthropogenic Climate Change Signal for Mean and Extreme Precipitation at Local and Regional Scales, J. Climate, 31, 4241–4263, https://doi.org/10.1175/JCLI-D-17-0282.1, 2018.

McKinnon, K. A. and Deser, C.: Internal Variability and Regional Climate Trends in an Observational Large Ensemble, J. Climate, 31, 6783–6802, https://doi.org/10.1175/JCLI-D-17-0901.1, 2018.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges, Ann. NY Acad. Sci., 1436, 19–35, https://doi.org/10.1111/nyas.13912, 2019.

Nikiéma, O., Laprise, R., and Dugas, B.: Energetics of transient-eddy and inter-member variabilities in global and regional climate model simulations, Clim. Dynam., 51, 249–268, https://doi.org/10.1007/s00382-017-3918-0, 2018.

Ouranos: CRCM5-LE, ClimEx, available at: https://climex-data.srv.lrz.de/Public/CanESM2_driven_50_members/, last access: 16 November 2020.

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., and Sanderson, B. M.: Precipitation variability increases in a warmer climate, Sci. Rep., 7, 17966, https://doi.org/10.1038/s41598-017-17966-y, 2017.

Räisänen, J.: CO_2-Induced Changes in Interannual Temperature and Precipitation Variability in 19 CMIP2 Experiments, J. Climate, 15, 2395–2411, https://doi.org/10.1175/1520-0442(2002)015<2395:CICIIT>2.0.CO;2, 2002.

Rondeau-Genesse, G. and Braun, M.: Impact of internal variability on climate change for the upcoming decades: analysis of the CanESM2-LE and CESM-LE large ensembles, Climatic Change, 156, 299–314, https://doi.org/10.1007/s10584-019-02550-2, 2019.

Screen, J. A.: Arctic amplification decreases temperature variance in northern mid- to high-latitudes, Nat. Clim. Change, 4, 577–582, https://doi.org/10.1038/nclimate2268, 2014.

Seneviratne, S. I., Lüthi, D., Litschi, M., and Schär, C.: Land-atmosphere coupling and climate change in Europe, Nature, 443, 205–209, https://doi.org/10.1038/nature05095, 2006.

Sørland, S. L., Schär, C., Lüthi, D., and Kjellström, E.: Bias patterns and climate change signals in GCM-RCM model chains, Environ. Res. Lett., 13, 74017, https://doi.org/10.1088/1748-9326/aacc77, 2018.

Suarez-Gutierrez, L., Li, C., Müller, W. A., and Marotzke, J.: Internal variability in European summer temperatures at 1.5 °C and 2 °C of global warming, Environ. Res. Lett., 13, 64026, https://doi.org/10.1088/1748-9326/aaba58, 2018.

Sun, L., Deser, C., and Tomas, R. A.: Mechanisms of Stratospheric and Tropospheric Circulation Response to Projected Arctic Sea Ice Loss, J. Climate, 28, 7824–7845, https://doi.org/10.1175/JCLI-D-15-0169.1, 2015.

Tamarin-Brodsky, T., Hodges, K., Hoskins, B. J., and Shepherd, T. G.: Changes in Northern Hemisphere temperature variability shaped by regional warming patterns, Nat. Geosci., 13, 414–421, https://doi.org/10.1038/s41561-020-0576-3, 2020.

Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., and Phillips, A. S.: Quantifying the Role of Internal Climate Variability in Future Climate Trends, J. Climate, 28, 6443–6456, https://doi.org/10.1175/JCLI-D-14-00830.1, 2015.

Torma, C., Giorgi, F., and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain-Precipitation over the Alps, J. Geophys. Res.-Atmos., 120, 3957–3972, https://doi.org/10.1002/2014JD022781, 2015.

Vautard, R. and Yiou, P.: Control of recent European surface climate change by atmospheric flow, Geophys. Res. Lett., 36, 231, L22702, https://doi.org/10.1029/2009GL040480, 2009.

Vidale, P. L., Lüthi, D., Wegmann, R., and Schär, C.: European summer climate variability in a heterogeneous multi-model ensemble, Climatic Change, 81, 209–232, https://doi.org/10.1007/s10584-006-9218-z, 2007.

Vogel, M. M., Zscheischler, J., and Seneviratne, S. I.: Varying soil moisture–atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central Europe, Earth Syst. Dynam., 9, 1107–1125, https://doi.org/10.5194/esd-9-1107-2018, 2018.

von Trentini, F., Leduc, M., and Ludwig, R.: Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble, Clim. Dynam., 53, 1963–1979, https://doi.org/10.1007/s00382-019-04755-8, 2019.

Wang, L., Deng, A., and Huang, R.: Wintertime internal climate variability over Eurasia in the CESM large ensemble, Clim. Dynam., 52, 6735–6748, https://doi.org/10.1007/s00382-018-4542-3, 2019.

Xoplaki, E., González-Rouco, J. F., Luterbacher, J., and Wanner, H.: Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs, Clim. Dynam., 20, 723–739, https://doi.org/10.1007/s00382-003-0304-x, 2003.

Yettella, V., Weiss, J. B., Kay, J. E., and Pendergrass, A. G.: An Ensemble Covariance Framework for Quantifying Forced Climate Variability and Its Time of Emergence, J. Climate, 31, 4117–4133, https://doi.org/10.1175/JCLI-D-17-0719.1, 2018.

7 Natural variability as part of the overall uncertainty of climate projections

After it was shown in the first publication that the IAV of SMILEs well represents the observed IAV and IMV, they can be used for a relation study against a classical multi-model ensemble (MME), built by 22 EURO-CORDEX members. The variability of signals from the MME is compared to the purely internal variability of the CRCM5-LE to quantify the relation between the two metrics. This allows to assess the share of internal variability on the overall variability of the MME.

7.1 How large is the spread of signals from the CORDEX ensemble with special consideration of the GCM-RCM combinations?

Overall, the spread of signals in CORDEX can mostly be explained by the different GCMs used for long-term projections, as already described by Kendon et al. (2010). Especially for larger GCM samples (EC-EARTH and HadGEM2-ES), the GCM dominance can be observed more robustly. There have been attempts to separate out the noise components in climate model ensemble signals. For example, Saffioti et al. (2017) showed that the removal of atmospheric circulation variability largely decreases the spread of trends in an initial condition ensemble as well as a multi-model ensemble of GCMs. Since there is no better multi-model ensemble available so far (in terms of sampling different models and members), the CORDEX ensemble can be seen as a first order approximation of the uncertainty in a multi-GCM/RCM ensemble. To better assess the uncertainties between signals in RCMs and their driving GCMs over Europe further research is needed.

7.2 How much of the uncertainty in a multi-model ensemble can be explained by natural variability?

The multi-model ensemble variability is usually much higher than the CRCM5-LE variability for temperatures. However, both ensembles generally show an increase in variability of temperature towards more continental climates in the East, which suggests that this gradient in CORDEX is at least partly due to internal variability.

In general, the contribution of internal variability is much higher for precipitation signals than for temperature signals. Additionally, the influence of internal variability significantly decreases for later future periods. Nevertheless, in many regions the contribution lies between 0.25 and 0.5 for seasonal temperature, and between 0.5 and 1.0 for seasonal precipitation.

Deser et al. (2012) conducted a similar experiment with 21 CMIP3 models and 40 CCSM3 members, building a ratio between the standard deviations of trends from 2005-2060 globally. They also find ratios above 0.75 and 1.0 for large parts of Europe for annual temperature and precipitation - the latter generally showing much higher ratios.

7.3 What does that mean for the interpretation of multi-model ensembles?

These findings are of such importance, since climate modelers are often facing criticism for the large uncertainty of ensemble projections, with the criticism implying that the variety of model results is a consequence of the models' inability to correctly represent climate processes (model response uncertainty). If natural variability can explain a large part of the spread of models, then the differences in climate signals from different models might not only be a result of insufficient or competing models but might also be partly explained by natural variability. Following the idea that natural variability is inherent to the chaotic nature of the climate system and therefore cannot be diminished, a certain part of uncertainty of climate projections will be irreducible, even if scenario definition will become more precise and models will improve (see also Deser et al. 2012).

The implications for the interpretation of multi-model ensembles in cases with similar variabilities (e.g., precipitation in DJF), might become clearer as a short mind game: First we need to accept the natural variability of the CRCM5-LE as a fair approximation to adopt it for other models. Then two things become apparent:

A) We can add the CRCM5-LE variability as a "cloud" of internal variability around each of the dots for the CORDEX models (in Figures 2 and 3 of the publication). This is blurring the model response uncertainty dramatically in some cases. This means that only one realization, as in CORDEX usually available, does usually not depict the model response very well.
B) On the other hand, if the CORDEX ensemble might even be totally (or to a large part) explained by natural variability, model response uncertainty may be interpreted as neglectable in these cases.

These conclusions are of course very much depending on the length of the time period, variable, season and region considered, and are not meant as universally valid for multi-model ensembles like CORDEX.

Title

Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble

Authors

Fabian von Trentini, Martin Leduc, Ralf Ludwig

Journal

Climate Dynamics 53, 1963-1979

Publishing Date

16.04.2019

DOI

https://doi.org/10.1007/s00382-019-04755-8

Journal Impact Factor (Web of Science)

4.38

Author Contributions

FvT designed the concept of the study, performed the analysis, and created all figures. ML and RL helped improve the concept and analysis. FvT led the paper writing with input from all authors.

Citations (crossref)

39

Plain Language Summary

To address uncertainty in climate projections, the common procedure is building an ensemble of different models and analyze the range of results in a statistical manner (mean, standard deviation, min, max). It is usually assumed that differences in the results stem from the different models used. In this study, the spread of such a typical ensemble of regional climate models over Europe is compared with another ensemble of climate model data that stem from a so called single-model initial-condition large ensemble (SMILE). Here, the exact same model is started 50 times, driven by the same emission scenario. The only difference between the runs are minor changes in the initial conditions. This leads to an ensemble that generally shares the same path for future changes, but directly quantifies the amount of natural variability that the climate system in the model incorporates. The comparison of the multi-model ensemble and the SMILE reveals that the influence of natural variability decreases the further into the future we look and is generally rather small for temperature-based indicators (<25-50 %). For precipitation-based indicators, however, natural variability can still account for over 50 % of the multi-model variability. This stresses the important role that natural variability can play in climate projection uncertainty on a regional scale.

Climate Dynamics
https://doi.org/10.1007/s00382-019-04755-8

Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble

Fabian von Trentini[1] · **Martin Leduc**[2] · **Ralf Ludwig**[1]

Received: 11 July 2018 / Accepted: 1 April 2019

Abstract
Uncertainties in climate model ensembles are still relatively large. Besides scenario and model response uncertainty, natural variability is another important source of uncertainty. To study regional natural variability on timescales of several decades and more, observations are often too sparse and short. Regional Climate Models (RCMs) can be used to overcome this lack of useful data at high spatial resolutions. In this study, we compare a new 50-member single RCM large ensemble (CRCM5-LE) with an ensemble of 22 EURO-CORDEX models for seasonal temperature and precipitation at 0.11° grid size over Europe—all driven by the RCP 8.5 scenario. This setup allows us to quantify the contribution of natural/model-internal variability on the total uncertainty of a multi-model ensemble. The variability of climate change signals within the two ensembles is compared for three future periods (2020–2049, 2040–069 and 2070–2099). Results show that the single model spread is usually smaller than the multi-model spread for temperature. Similar variabilities can mostly be found in the near future (and to a lesser extent in the mid future) during winter and spring, especially for northern and central parts of Europe. The contribution of internal variability is generally higher for precipitation. In the near future almost all seasons and regions show similar variabilities. In the mid and far future only fall, summer and spring still show similar variabilites. There is a significant decrease of the contribution of internal variability over time for both variables. However, even in the far future for most regions and seasons 25–75% of the overall variability can be explained by internal variability.

Keywords Regional climate models · Natural variability · EURO-CORDEX · Uncertainty · Large ensemble · Climate change signals

1 Introduction: uncertainties in climate projections

So far, decades of research and development have gone into a better understanding of atmospheric processes and their interactions with other components of the climate system like land surface, oceans and ice. Yet, there are still large uncertainties about the future development of the climatic conditions, particularly at the regional scale. For well-planned adaptation measures, which can include the use of impact models driven by (regional) climate models, decision makers demand more precise projections of how the future might look like. Climate scientists have a hard time explaining why their models still cannot show a clearer picture of likely future changes, narrowing down the model spread of ensemble projections (Hawkins and Sutton 2009, 2011; Deser et al. 2012a).

In general, the overall uncertainty of such climate projections can be separated into three parts: (1) scenario uncertainty, corresponding to the different emission scenarios that can be used as external forcing to the climate models, (2) model response uncertainty, corresponding to the response of the different models, developed by different institutions around the globe and (3) natural variability, inherent to the chaotic nature of the climate system (Hawkins and Sutton

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00382-019-04755-8) contains supplementary material, which is available to authorized users.

Fabian von Trentini
fabian.trentini@lmu.de

[1] Department of Geography, Ludwig-Maximilians-University Munich, Luisenstr. 37, 80333 Munich, Germany

[2] Ouranos, 550 Rue Sherbrooke W., West Tower, 19th Floor, Montréal, QC H3A 1B9, Canada

Published online: 16 April 2019

2009). Scenario uncertainty is difficult to reduce, since this includes future knowledge about greenhouse gas and aerosol emissions and advances in technology (or a best guess of it), but might potentially become less in the future (Moss et al. 2010). Model uncertainty might be reduced by advances in the knowledge of natural processes, and the capabilities of models in representing those. For instance, the increase in spatial resolution of regional climate models (RCMs) from 0.44° to 0.11° was shown to have a significant effect on the model performance, often referred to as "added value" (Torma et al. 2015; Fantini et al. 2016; Prein et al. 2016). On the one hand, increasing computational capabilities available to modeling groups and further advances in describing processes more accurately has the potential to narrow down uncertainties, but adding more complexity to models may on the other hand also lead to higher uncertainties in the projections (Knutti and Sedláček 2013). Constraining models by observations can also reduce the uncertainty of future projections in some cases (Borodina et al. 2017; Lorenz et al. 2018). The third part of uncertainty in climate projections is natural variability. Natural variability can be associated with interactions between the internal components of the climate system, which can lead to inter-annual, decadal or even multi-decadal variability like ENSO, NAO, etc. (Solomon et al. 2010; Zheng et al. 2018).

Observational data of the earth's climate usually cover only several decades, and are thus not long enough to capture this mid- to long-term variability (Hawkins et al. 2016). Brisson et al. (2015) for example, used a weather generator to quantify uncertainties in estimating climatic conditions from a 30-year period as recommended by the World Meteorological Organization (WMO), and showed that these uncertainties can well exceed 15% for extreme precipitation. This lack of data makes it very difficult or nearly impossible to use observations for the quantification of natural variability on longer time scales.

A possible solution to deal with the lack of sufficiently long observational time series to assess long-term climate variability can be the use of climate models. The underlying assumption has to be of course that the climate models are capable of simulating the climate system, including its natural variability (represented by the internal variability of the models). Aalbers et al. (2018) showed for example that the interannual variability of an EC-EARTH-RACMO initial conditions large ensemble of 16 members is similar to the one of E-OBS. So, if we accept the concept of a climate model as a surrogate of the natural climate system, we can use the model to generate a huge amount of data to create plausible parallel modifications of current and future climate. Therefore, the ClimEx project (http://www.climex-project.org) provides 50 members of the Canadian global earth system model CanESM2 (Fyfe et al. 2017), dynamically downscaled by the Canadian regional climate model version 5 (CRCM5) over large parts of Europe and northeastern North America, running from 1955 to 2099. This leads to 50 parallel realizations of climate, just altered by very small differences in the initial conditions in the CanESM2 members, which makes all 50 members equally likely realizations of the same long-term climatic conditions. This unprecedented dataset, hereafter referred to as the ClimEx CRCM5 large ensemble (CRCM5-LE) is described in Leduc et al. (2019).

This ensemble can be used to better understand and quantify the uncertainties of ensemble approaches in regional climate change studies. In those studies usually an ensemble of different combinations of global climate models (GCMs) and regional climate models (RCMs) is used to cover the range of possible outcomes (Jacob et al. 2014; Vautard et al. 2014; Roudier et al. 2016; Smiatek et al. 2016; Rajczak and Schär 2017). Often, the derived uncertainties from these multi-model ensembles are relatively large, since usually all three components of uncertainty are incorporated: Different scenarios driving different GCM/RCM combinations and inherent natural variability. For regional studies using RCMs, another dimension is added, as the model uncertainty is now composed of the GCM and RCM response. Further constraint for assessing scenario uncertainty and model response uncertainty is that the "matrix" of all possible combinations of scenario, GCM and RCM is sparse and not balanced according to these three dimensions. Analysis is always restricted by the available simulations (often on an opportunity basis), thus making systematic comparisons on the influence of these factors difficult. Additionally, not all models are fully independent from each other, why the often claimed and applied model democracy (one model one vote) is increasingly challenged by some authors (Pennell and Reichler 2010; Leduc et al. 2016a; Knutti et al. 2017). In addition to the fact that multiple members are rarely available to assess internal variability, all these factors together induce large problems in distinguishing the three sources of uncertainty in climate projections in existing multi-model ensembles like EURO-CORDEX.

Therefore, several studies have produced and analyzed single model large ensembles with small alterations in initial conditions to address the uncertainty resulting from internal variability. On the GCM scale: The 40-member CCSM-LE (Deser et al. 2012a, b, 2014) and the 30-member CESM-LE (Kay et al. 2015). Another set of experiments with 90–100 members was conducted by a group of Japanese authors with a 60 km resolution global atmospheric model and a 20 km resolution regional model for a Japanese domain (Mizuta et al. 2017). The Netherlands Meteorological Institute (KNMI) performed an RCM experiment with 16 EC-EARTH members, downscaled by the Dutch regional model RACMO2 to a resolution of 0.11° for Western Europe. Aalbers et al. (2018) give an overview of this ensemble and investigate several aspects of natural variability for mean and extreme precipitation. A set of 21 members of the CESM was downscaled by the regional model COSMO-CLM at a resolution of 0.44° for Europe (Fischer et al. 2013). All these studies showed that a single model large ensemble setup can be very valuable to better quantify natural variability in climate projections. However, only very few comprehensive comparisons with multi-model ensembles have been conducted, and so far only with GCMs (Deser et al. 2012b; Kay et al. 2015; Leduc et al. 2016b).

In this study, we use the ClimEx CRCM5-LE to compare the inter-member spread due to natural variability in a single model large ensemble with the multi-model variability (which implicitly includes the effect of natural variability) of a corresponding EURO-CORDEX ensemble. This results in the following research questions:

1. Which part of the multi-model spread (model uncertainty and internal variability combined) of a 0.11° EURO-CORDEX ensemble can be attributed to internal variability?
2. What are the implications of this attribution for analyses of climate change signals from multi-model ensembles?

This study will first investigate the composition of the CORDEX ensemble in terms of GCM (member)/RCM combinations. This analysis is supposed to evaluate the ability of this ensemble of opportunity to represent the different sources of uncertainty of a multi-model ensemble in a satisfying manner for the following comparison: The signals of the CRCM5-LE are compared to the signals of the CORDEX ensemble to quantify the spread of each ensemble on a grid point and regional scale. Thus, the uncertainty of signals originating from model-internal variability (natural variability uncertainty) is compared against that originating from a multi-model ensemble (model response uncertainty and internal variability combined). After a discussion of the results, the implications for the analysis of multi-model ensembles are briefly discussed.

2 Data

Two different ensembles of RCP 8.5 driven models in daily and 0.11° resolution are used in this study:

(A) 22 models from EURO-CORDEX.
(B) 50 members of the ClimEx CRCM5-LE.

All available datasets that share the common EURO-CORDEX 0.11° grid in the EURO-CORDEX and ReKliEs-De datasets at the Earth System Grid Federation (ESGF) are pooled together and only refered to as 'CORDEX' in the following. This means that the two MPI-REMO2009 runs, the ALARO run and the ALADIN53 run are not considered because of their different grid specifications. The UHOH data cannot be used for the signal calculation since no historical data are stored for this run [see Supplementary Material (SM) Table S1 for all excluded runs]. During the analysis of the data, the model IPSL-WRF showed very high increases (quadrupling) in summer precipitation, especially for coastal parts of France and the Mediterranean and is therefore excluded from the CORDEX ensemble. Excluding a model in a variability study for its extreme results is of course crucial. However, other studies also excluded the IPSL-WRF model from their analysis (Kotlarski et al. 2014; Smiatek et al. 2016; Rajczak and Schär 2017), supporting this decision. Differences introduced by including the IPSL-WRF model into the analysis are shown in Fig. S1, Supplementary Material. The 22 resulting CORDEX models are listed in Table 1 as a combination of RCMs and driving GCMs. These models are used to analyze the composition of the

Table 1 Matrix of GCM (-member) and RCM used in the CORDEX ensemble

	Member	CCLM4-8-17	HIRHAM5	RACMO22E	RCA4	REMO2015	SUM
CanESM2	r1i1p1	1				1	2
CNRM-CM5	r1i1p1	1			1	1	3
EC-EARTH	r1i1p1			1[a]			1
	r3i1p1		1				1
	r12i1p1	1		1	1	1	4
CM5A-MR	r1i1p1				1		1
MIROC5	r1i1p1	1				1	2
HadGEM2-ES	r1i1p1	1	1	1	1	1	5
MPI-ESM-LR	r1i1p1	1			1		2
NorESM1-M	r1i1p1		1				1
SUM		6	3	3	5	5	22

[a]The EC-EARTH_r1_RACMO22E simulation is used for the composition analysis of the CORDEX ensemble but not for comparison with the CRCM5-LE

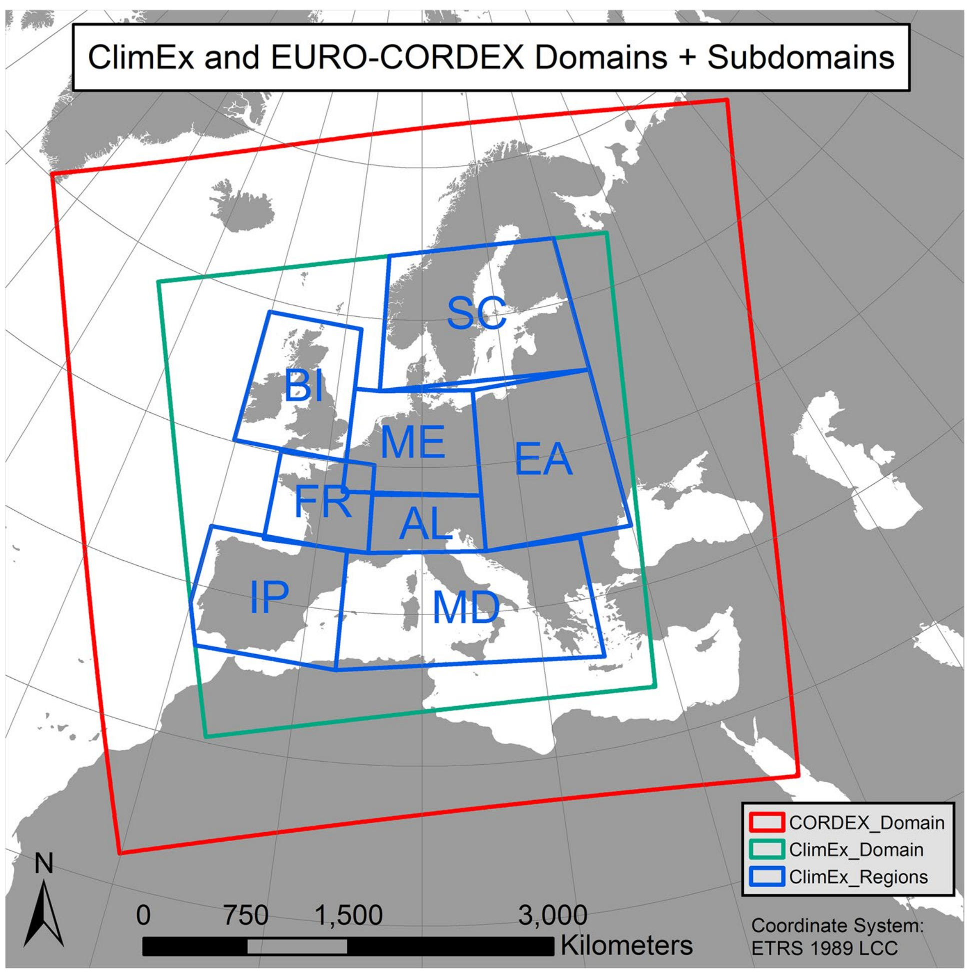
ensemble. Two members of EC-EARTH have been downscaled with RACMO22E (r1 and r12), but we only use r12 for the inter-comparison with the CRCM-LE for balancing reasons. Keeping both would double count this GCM-RCM pair (as the two EC-EARTH members are expected to be much more similar to each other than compared with other driving models), which would likely shrink the total variance of CORDEX (although probably by a small amount).

The second dataset is a 50-member ensemble of the Canadian earth system model CanESM2 (Fyfe et al. 2017), downscaled by the Canadian regional model CRCM5 (Version 3.3.3.1, Martynov et al. 2013; Šeparović et al. 2013) for two domains: Northeastern North America (not part of this study) and Europe. The 50 members originate from five families of simulations, each starting at different 50 year intervals of a preindustrial run with a stationary climate and run from 1850 to 1950. They are then seperated into ten members each by small atmospheric perturbations and run from 1950 to 2099. This 50 member CanESM2-LE was then dynamically downscaled within the ClimEx project with CRCM5 over a domain covering Europe (EU11d2) with a horizontal grid-size mesh of 0.11° on a rotated latitude–longitude grid, corresponding to a 12-km resolution, using 5-min time steps, which fits the common EURO-CORDEX grid specifications. Further information on the settings of the whole experiment, as well as a detailed description and analysis of the dataset (also for the American domain) can be found in Leduc et al. (2019). The stored variables and the terms of use can be found in the respective documentation on the project homepage (http://www.climex-project.org). The ClimEx 12-km grid equals the one used in EURO-CORDEX simulations, although the ClimEx domain is slightly smaller, still covering most of Europe (Fig. 1).

3 Methods

The subregions used for analysis are taken from the PRUDENCE project (Christensen and Christensen 2007) and were already used in several other studies in the context of European climate model analysis (Lenderink 2010; Lorenz and Jacob 2010; Kotlarski et al. 2014). The subregions cover almost all the ClimEx domain despite the Mediterranean

Fig. 1 The EURO-CORDEX and ClimEx domains with the subregions for analysis



parts of Morocco, Algeria and Tunisia, the Aegean Sea and small parts of Eastern Europe. Note that the Scandinavian subregion here is smaller than in other studies, since the ClimEx domain does not cover the northern parts of Scandinavia. An additional analysis domain is a combination of all subregions (‘TOT’). For the whole study, only land grid points are considered.

The data of the 22 CORDEX models is first cut to the ClimEx domain. Since the CORDEX and CRCM5-LE data share the same 0.11° grid, no interpolation is needed to compare the two datasets. Annual and seasonal means of temperature and precipitation sums are calculated at every grid point on an annual basis from 1980 to 2099. The analysis of the data comprises grid point analysis as well as spatially averaged results (arithmetic mean) for the subregions (TOT as an area weighted mean of: BI, SC, FR, ME, EA, IP, AL, MD). For the spatially averaged results, the temperature/precipitation values are aggregated before signals are computed. The change signals are calculated using a 30-year reference period (REF: 1980–2009) as a baseline for three future periods: The near future (FUT1: 2020–2049), the mid future (FUT2: 2040–2069) and the far future (FUT3: 2070–2099).

Seasonal means (temperature) and sums (precipitation) are calculated in each grid point for every year and member, followed by calculation of the temporal means for each 30-year period (REF, FUT1-3). This allows calculating the standard deviation of signals for both ensembles and every grid point. The variabilities are compared using the ratio of standard deviations:

$$\mathrm{SDR} = \sigma(\mathrm{CRCM5-LE})\,/\,\sigma(\mathrm{CORDEX}).$$

For this comparison the EC-EARTH_r1_RACMO22E run is not part of the CORDEX ensemble (see chapter 2). Additionally a two-sample F-Test is applied at the significance level of 5%. When the Null Hypothesis of this test cannot be rejected, we assume that the two distributions have similar variances. The test also accounts for the different sample sizes of the two ensembles (50 and 21). To see how similarities in spatial patterns change over time between the two ensembles, a simple pattern correlation is calculated for moving window 30-year periods.

4 Results

4.1 Composition of the CORDEX ensemble

The CORDEX ensemble in this study consists of 22 different combinations of GCM, GCM member and RCM, with eight GCMs and five RCMs in total. Not all combinations of these have been realized, leaving about half of the GCM (-member)/RCM matrix blank (Table 1). The composition is quite heterogeneous, with a slight dominance of EC-EARTH and HadGEM2-ES for the GCMs and CCLM4-8-17 and REMO2015 for the RCMs. The EC-EARTH is the only model with different downscaled members, and fortunately there is even a pair of two members downscaled with the same RCM (RACMO22E). Additionally, there are two simulations using the first member of the CanESM2 ensemble, downscaled with CCLM4-8-17 and REMO2015, giving us insight in the role of CRCM5’s contribution to the signals of the CanESM2-CRCM5-LE. Yet overall, the sampling is relatively random, which makes systematic analysis on the influence of GCM (-member) and RCM on the variability extremely difficult. Nevertheless, it seems valuable to take a look at the variabilities inside the CORDEX ensemble to better assess the capacity of the ensemble for the variance comparison with the CRCM5-LE.

To get an overview of the influence of the components (GCM/RCM) of each simulation, the climate change signals of temperature and precipitation for the far future FUT3 (2070–2099) are displayed in scatterplots for winter (DJF, Fig. 2) and summer (JJA, Fig. 3) for the TOT domain. A general clustering of the simulations sharing the same GCM can be observed, although there are large differences between the GCM cluster extents.

In winter, the differences between RCM simulations, driven by the same global model, range from 0.1 K in CanESM2 to 0.9 K in MIROC5 for temperature and from 3.6% points [pp] in MIROC5 to 5.1 pp in CanESM2 for precipitation (Fig. 2). These ranges are usually similar to the CRCM5-LE extent. The EC-EARTH members r1 and r3 are close again, as well as the two RACMO22E simulations (r1 and r12). The two CanESM2 simulations fit quite well into the CRCM5-LE, although being at the colder end of the cloud.

In summer, the spread of temperature signals of the same GCM downscaled by different RCMs range from 0.2 K in CNRM-CM5 to 1.5 K in HadGEM2-ES, while the spread for precipitation signals ranges from 4.5 pp in CNRM-CM5 to 36 pp in CanESM2 (Fig. 3). These GCM ranges are larger than in winter, due to the higher importance of large scale circulations, and these are mainly driven by the GCM. The RCM CCLM4-8-17 shows the strongest decreases in precipitation regardless the driving GCM—except CNRM-CM5, which generally seems to have a larger influence on the RCM output than other GCMs. The two CanESM2 simulations show large differences and span a larger range, both in temperature and precipitation, than the CRCM5-LE. The combination of the rather warm and dry CanESM2 with the also rather dry CCLM4-8-17 (usually the driest RCM, driven by the same GCM) results in an extreme decrease of precipitation, accompanied by a strong warming signal. The EC-EARTH is the only GCM with different members (1*r1, 1*r3, 4*r12), giving insight into the variability of

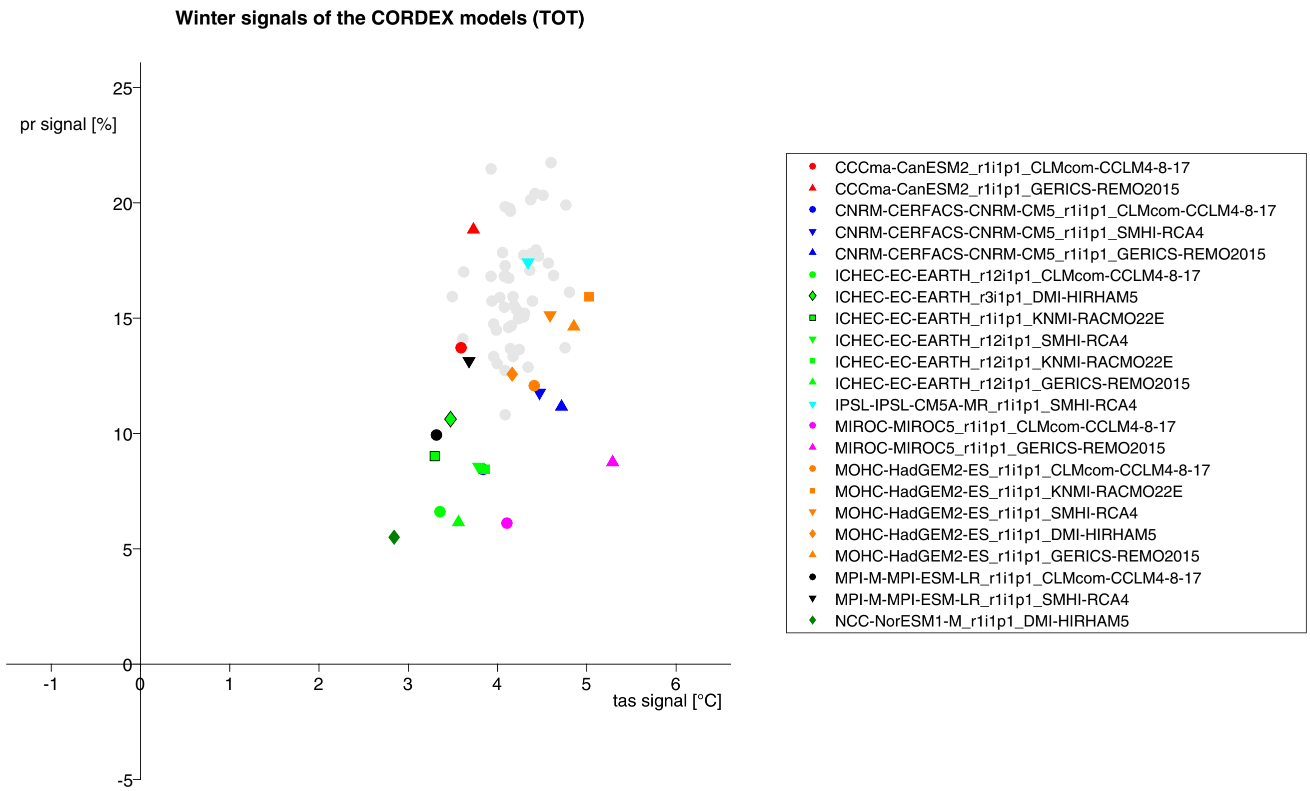


Fig. 2 Scatterplot of the signals (2070–2099 minus 1980–2009) of temperature and precipitation for winter (DJF) in the combined (area weighted mean of all other subregions) domain TOT (land grid points only). The color of the marker denotes to the GCM and the symbol denotes to the RCM. The members r1 and r3 of EC-EARTH have a black frame to distinguish them from member r12 simulations. The grey points show the signals of the 50 CRCM5-LE members

another multi-member ensemble. The two simulations of member r1 and r3 have colder and less dry signals than the r12 simulations, yet still at the edge of the r12 cluster. The two simulations using the same RCM and only different members of EC-EARTH (r1_RACMO22E and r12_RACMO22E) are very close. The CRCM5-LE cluster is at the very warm and dry end of the CORDEX ensemble, but the MIROC5_CCLM4-8-17, HadGEM2-ES_CCLM4-8-17 and CanESM2_CCLM4-8-17 simulations show similar or even stronger signals.

The GCMs usually dominate the signal of the simulations, but there are some cases where the RCM can significantly impact the resulting signal. These findings for the TOT domain can be found in a similar manner in the subdomains as well, although differences occur of course. For example, the positive winter precipitation signals mostly originate from Northern European subregions like Scandinavia, whereas the Mediterranean shows mostly negative signals (Figs. S2 and S3, SM). On the other hand, the same contrasting signals (mostly positive in SC, mostly negative in MD) result in a more or less balanced signal spread in TOT for summer (Figs. S4 and S5, SM).

In general, this CORDEX ensemble consists of a number of GCMs and RCMs with a wide range of signals. Although it is not a perfectly composed multi-model ensemble (which would be necessary for a real structured framework), the analysis suggests the assumption that its composition represents a fair assumption of the sources of uncertainty in a multi-model ensemble. It is therefore suited for the comparison with the 50 member single model large ensemble.

4.2 Comparison of variability in signals of CRCM5-LE and CORDEX

To better assess and quantify the fraction of internal variability in the CORDEX ensemble, we compare the standard deviations of the CRCM5-LE and the CORDEX ensemble. The variability between EURO-CORDEX models is analyzed on the grid point level, while other publications usually only mark areas where models agree on the sign of change and significant changes, without quantifying the uncertainty of the respective ensemble (Jacob et al. 2014; Vautard et al. 2014; Rajczak and Schär 2017).

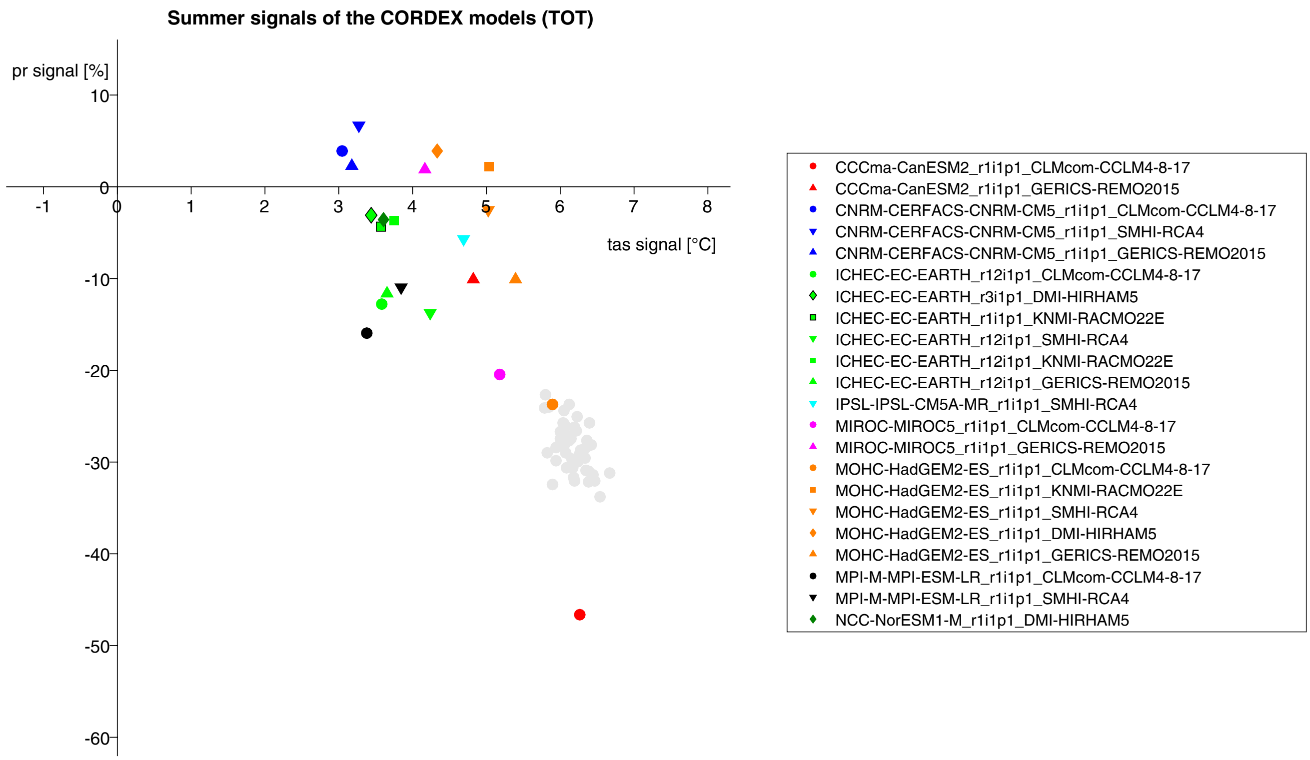


Fig. 3 Same as Fig. 2, but for summer (JJA)

4.2.1 Temperature

Figures 4 and 5 show the standard deviation of signals of CRCM5-LE and CORDEX for all three future periods and the respective SDR values for winter and summer temperature over Europe (respective figures for SON and MAM: Figs. S6 and S7, SM). The CRCM5-LE variability shows a large scale gradient from lower values in the West to higher values in the (North-) East, which might be associated with increasing continental climate as seen in Köppen-Geiger climate classifications (Beck et al. 2018). In winter CORDEX shows a similar, yet not so clear gradient, whereas in summer higher values seem to be found in southern parts of Europe. In both seasons, the most obvious gradient in CORDEX appears in mountainous regions like the Alps, the Pyrenees and the Scandinavian Mountains. Additionally, the variability increases from FUT1 to FUT3, especially in winter. The CRCM5-LE in contrast shows rather small variability in these mountainous regions.

The SDR mainly lies well below 1 in both seasons for most of Europe. This result suggests that for temperature, only a small part of the variability in the CORDEX ensemble can be explained by internal variability. For most parts, the SDR is smaller in summer than in winter. This is a result of two opposing effects: On the one hand, the overall CRCM5-LE variability is higher in winter; on the other hand, the CORDEX variability is smaller in winter. This mainly affects the British Islands, Scandinavia and other parts of Northern Europe, where SDR values can even exceed 1, especially in the early and mid-future. In these cases, the internal variability estimated from CRCM5-LE is larger than the CORDEX multi-model variability. While this result would be rather unexpected in a systematic framework, the current CORDEX imbalanced composition could lead to an underestimation of either (or both) RCM and GCM contributions to the total ensemble spread. In this context, it is not clear whether CRCM5-LE over- or underestimates the average internal variability of the CORDEX models. It is also not clear to which extent the true internal variability of the CORDEX ensemble is fully sampled by the available simulations.

A two-sample F-Test reveals the grid points with similar variances in both ensembles (Figs. 4 and 5, lowest rows). Empirical analysis shows that these are generally grid points with SDR values between 0.7 and 1.5 (similar to findings of Deser et al. 2012b). The share of grid points with similar variance decreases in both seasons for further future periods, with generally higher values in winter. Interesting to note is how the British Isles and parts of Norway fail the test for FUT1 because of the high variance in CRCM5-LE, showing similar variances for FUT2 and FUT3 with increasing CORDEX variability and relatively stable CRCM5-LE variability.

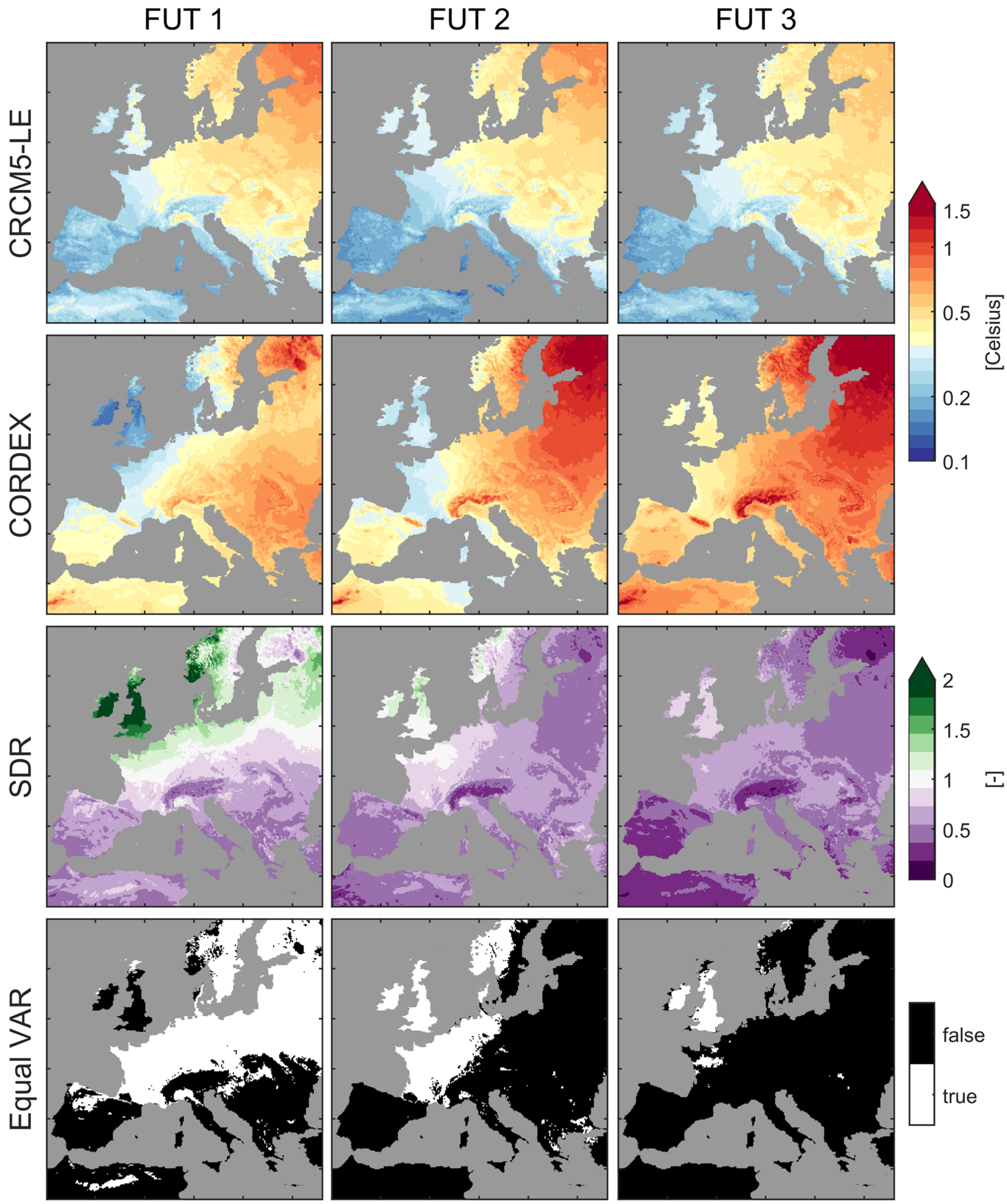


Fig. 4 Rows 1–2: standard deviation of the winter temperature (tasDJF) signals of all models in each ensemble for the three future periods. Note the logarithmic scale. Row 3: SDR (σ [CRCM5-LE]/σ [CORDEX]) for the respective periods. Row 4: Two-sample F-Test on equal variances at 5% significance level; For white grid points, the Null Hypothesis of equal variances cannot be rejected (Equal VAR = 'true')

4.2.2 Precipitation

The precipitation signals are calculated for each member individually in percent change, so the standard deviation over these members is also expressed in percent change. In winter, the general patterns of CRCM5-LE and CORDEX are quite similar with higher variability in southern parts of Europe (Fig. 6). Northern Africa shows the largest standard deviations of relative changes, because the absolute sums of precipitation are relatively small here. A remarkable band of high variability stretches from the southeastern parts of Spain over coastal France to the Italian Alps in both

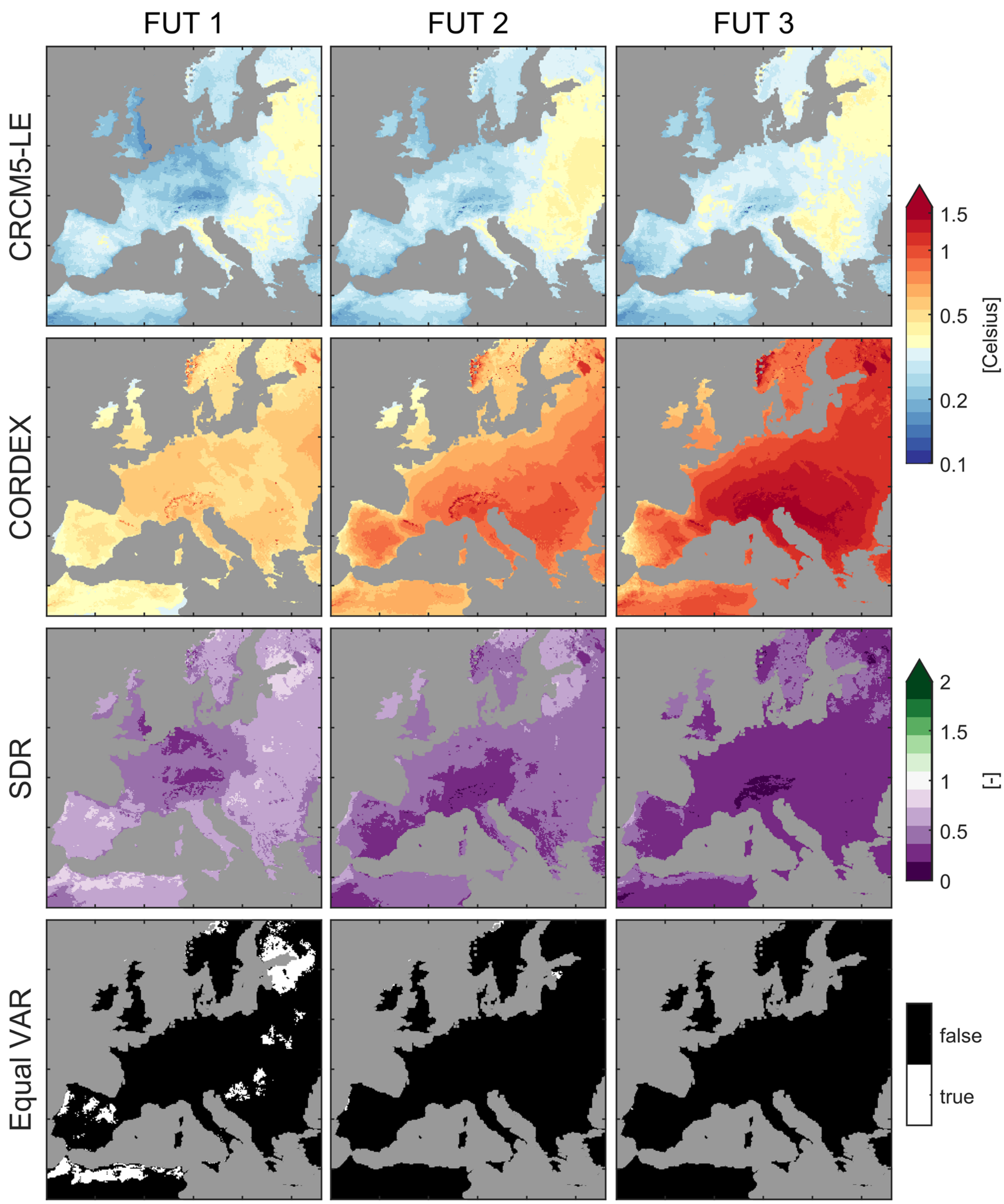


Fig. 5 Rows 1–2: standard deviation of the summer temperature (tas-JJA) signals of all models in each ensemble for the three future periods. Note the logarithmic scale. Row 3: SDR (σ [CRCM5-LE]/σ [CORDEX]) for the respective periods. Row 4: Two-sample F-Test on equal variances at 5% significance level; For white grid points, the Null Hypothesis of equal variances cannot be rejected (Equal VAR = 'true')

ensembles. The variability in CRCM5-LE on the Iberian Peninsula is higher than in CORDEX, leading to high SDR values in FUT1. Other high SDR values can be found in Northern Africa and some mainly coastal areas in FUT1 and FUT2, whereas in FUT3 most of Europe shows SDR values around 1 and below.

Some differences in the variability of signals can be observed in summer between the ensembles, despite a general increasing West–East and North–South gradient. The variability in CORDEX is larger in all future periods and almost all of Europe (Fig. 7). Topography does not seem to be a significant factor in CORDEX, while CRCM5-LE

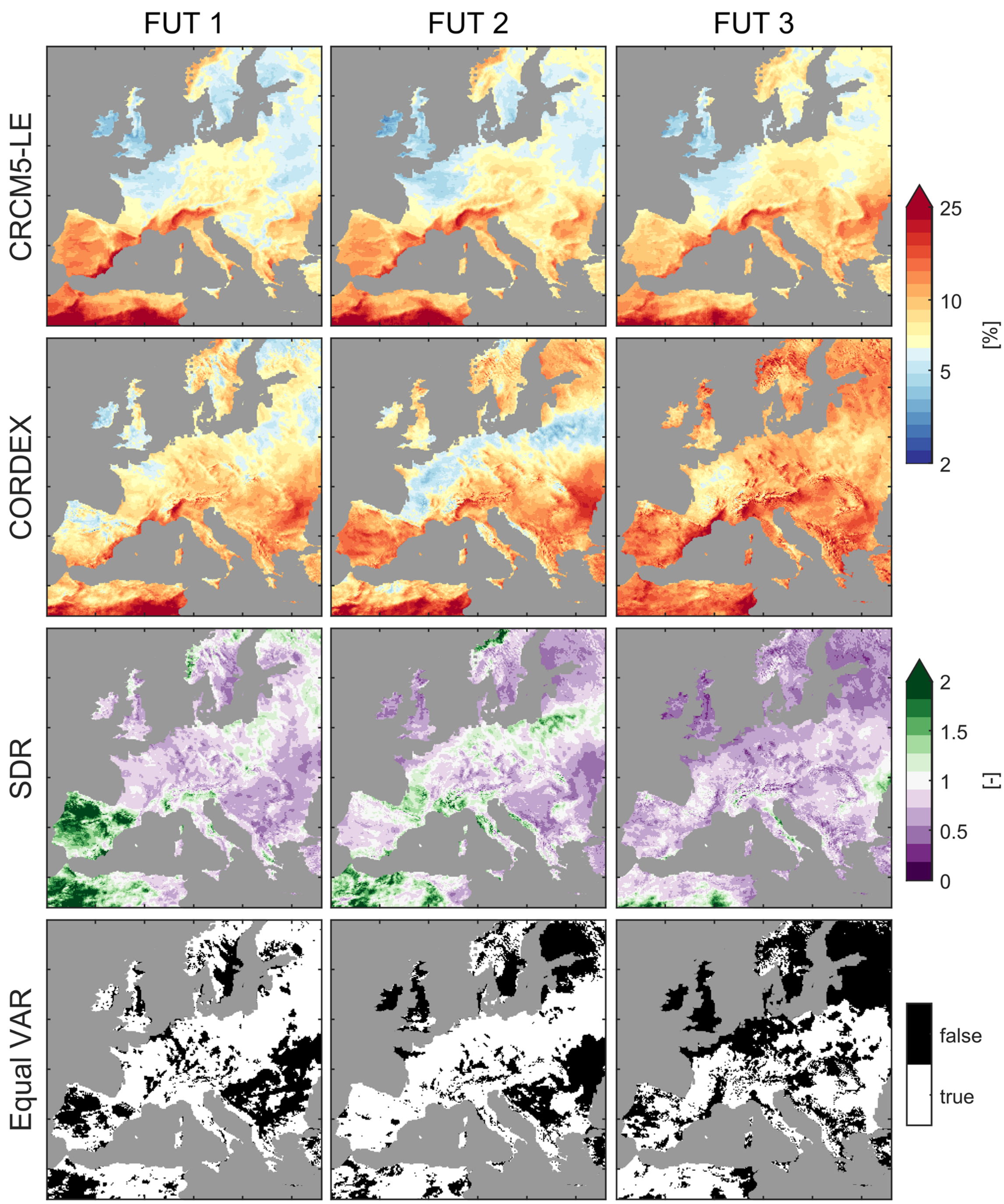


Fig. 6 Rows 1–2: standard deviation of the winter precipitation (pr-DJF) signals of all models in each ensemble for the three future periods. Note the logarithmic scale. Row 3: SDR (σ [CRCM5-LE]/σ [CORDEX]) for the respective periods. Row 4: Two-sample F-Test on equal variances at 5% significance level; For white grid points, the Null Hypothesis of equal variances cannot be rejected (Equal VAR = 'true')

displays especially low variability in mountainous regions, as already seen for temperature in winter. This results in SDR values around 1 in FUT1, decreasing to values well below 0.5 for FUT3. For FUT1 in summer, the number of grid points with similar variability is comparable to the number found in the map for winter (Fig. 6), but decreases significantly until FUT3 in contrast to the winter season. The respective figures for SON and MAM can be found in the SM, Figs. S8 and S9.

The pattern correlations as function of the time horizon between the standard deviation maps of both ensembles show two different behaviors for temperature and

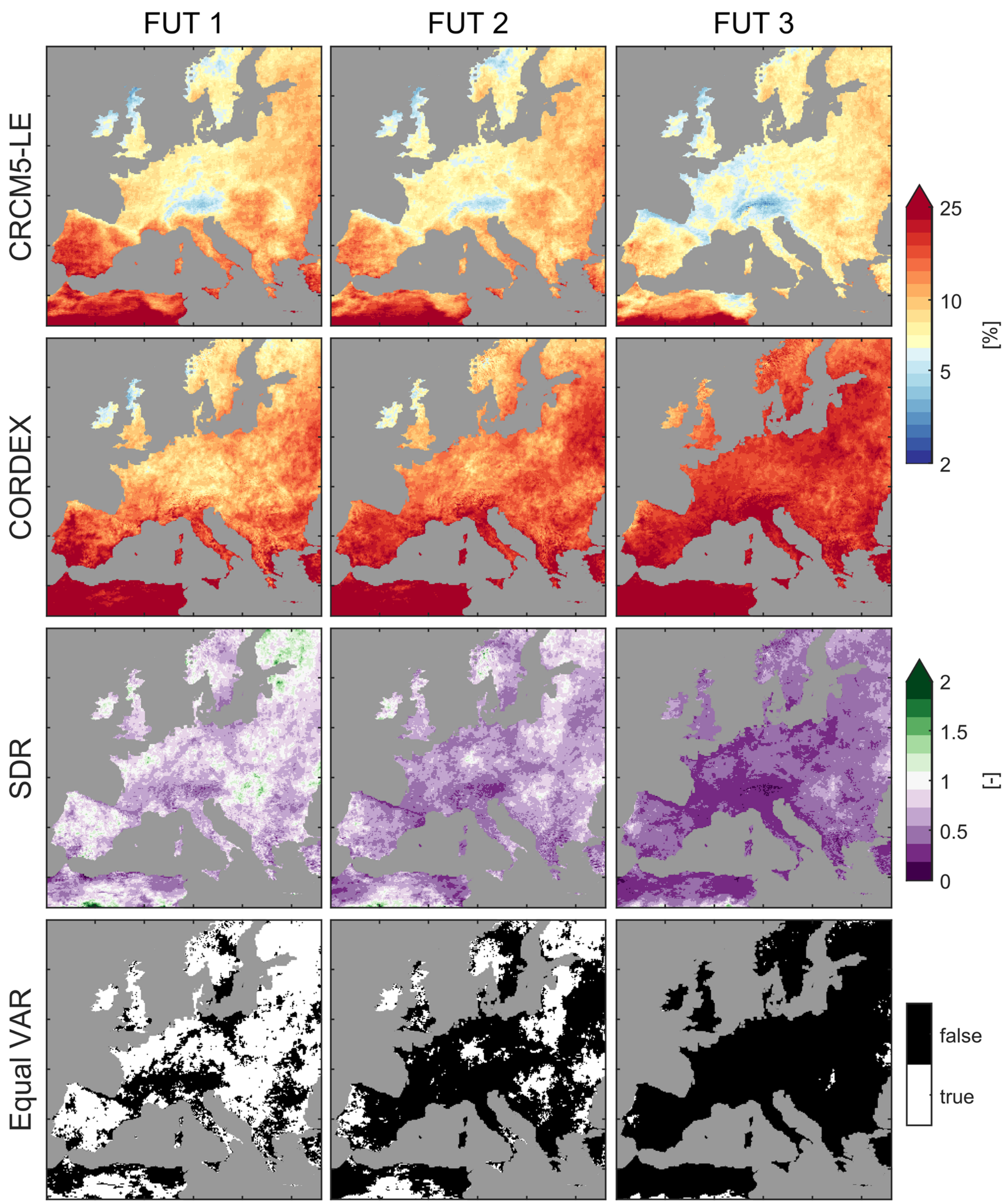


Fig. 7 Rows 1–2: standard deviation of the summer precipitation (pr-JJA) signals of all models in each ensemble for the three future periods. Note the logarithmic scale. Row 3: SDR (σ [CRCM5-LE]/σ [CORDEX]) for the respective periods. Row 4: Two-sample F-Test on equal variances at 5% significance level; For white grid points, the Null Hypothesis of equal variances cannot be rejected (Equal VAR = 'true')

precipitation (Fig. 8). Correlations are generally high for precipitation as well as for summer and fall temperatures. For all precipitation seasons, a decrease of correlation can be observed, which fulfills the expectation of a decreasing contribution of internal variability on the overall variability in the further future.

Temperature seasons show a remarkable behavior. The pattern correlation increases in all seasons during the first half of the twenty-first century, and remains more or less stable in tas-DJF and tas-MAM, while dropping significantly for tas-JJA and tas-SON afterwards. The patterns of temperature thus do not seem to follow the expectation

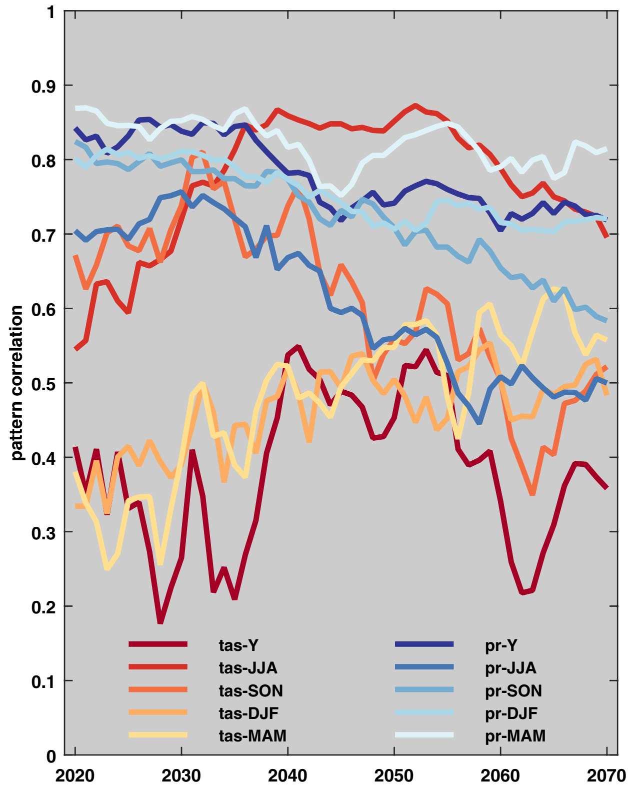


Fig. 8 Pattern correlations of the standard deviations of CRCM5-LE and CORDEX for all seasons with a 30-year moving window; the x-axis denotes to the 30-year period beginning with the respective year (2020–2049, 2021–2050,…)

of decreasing correlation with time, and even increase for winter and spring. Further research is needed in this direction, especially if these results can be reproduced with other initial condition ensembles in the future.

To give an overview, Fig. 9 shows the regionally averaged SDR grid point values for all seasons and future periods for the subregions from Fig. 1. In general, the contribution of internal variability is much higher for precipitation than for temperature, and it decreases significantly the further the future period lies ahead. Annual and summer temperature/precipitation and fall temperatures have significantly lower SDR values than the other seasons. The share of boxes with at least 2/3 of the grid points confirming the *F*-Test on equal variances is especially high for temperatures in FUT1 (12 boxes) and precipitation in FUT1 (32) and FUT2 (15). Ratios above 1 mainly appear in FUT1 for spring temperatures and fall precipitation. The threshold of 2/3 is chosen on the basis of similar existing concepts like robustness of change as a function of the numbers of climate models agreeing in the sign of change (Jacob et al. 2014).

5 Discussion

5.1 Composition of the CORDEX ensemble

The composition of the 22-member CORDEX ensemble is defined by the available datasets that match the preconditions described in the Data part of this study (chapter 2), leading to an ensemble of opportunity as a result. Although there have been efforts to fill the matrix of GCM-RCM-combinations, totally systematic analysis to separate contributions from GCMs and RCMs to the total ensemble spread is difficult—if not impossible—due to the sparsity of the matrix, the imbalance between GCM/RCM combinations, and the lack of several members for each combination to better discriminate model uncertainty from internal variability (except for EC-EARTH, all GCMs only have one member). The uncertainty of the CORDEX ensemble signals is thus a combination of model response uncertainty and internal variability.

Overall, the spread of signals in CORDEX can mostly be explained by the different GCMs used for long-term projections, as already described by Kendon et al. (2010). Especially for larger GCM samples (EC-EARTH and HadGEM2-ES), the GCM dominance can be observed more robustly. There have been attempts to separate out the noise components in climate model ensemble signals. For example, Saffioti et al. (2017) showed that the removal of atmospheric circulation variability largely decreases the spread of trends in an initial condition ensemble as well as a multi-model ensemble of GCMs. Since there is no better multi-model ensemble available so far (in terms of sampling different models and members), the CORDEX ensemble can be seen as a first order approximation of the uncertainty in a multi-GCM/RCM ensemble. To better assess the uncertainties between signals in RCMs and their driving GCMs over Europe further research is needed. For instance, considering several RCMs driven by the same GCM could help to better understand the uncertainty due to the choice of the RCM, while doing similarly over an RCM column (in Table 1) would allow to assess the RCM sensitivity to boundary conditions. Nevertheless, we tried to assess the contribution of internal variability to the CORDEX ensemble spread by comparing with CRCM5-LE, where the internal variability is sampled in a systematic manner.

5.2 Comparison of signals in CRCM5-LE and CORDEX

The comparison of a single model large ensemble, comprised of 50 initial condition members of the CanESM2-CRCM5 model chain (CRCM5-LE) with a multi-model ensemble of 21 different EURO-CORDEX models was

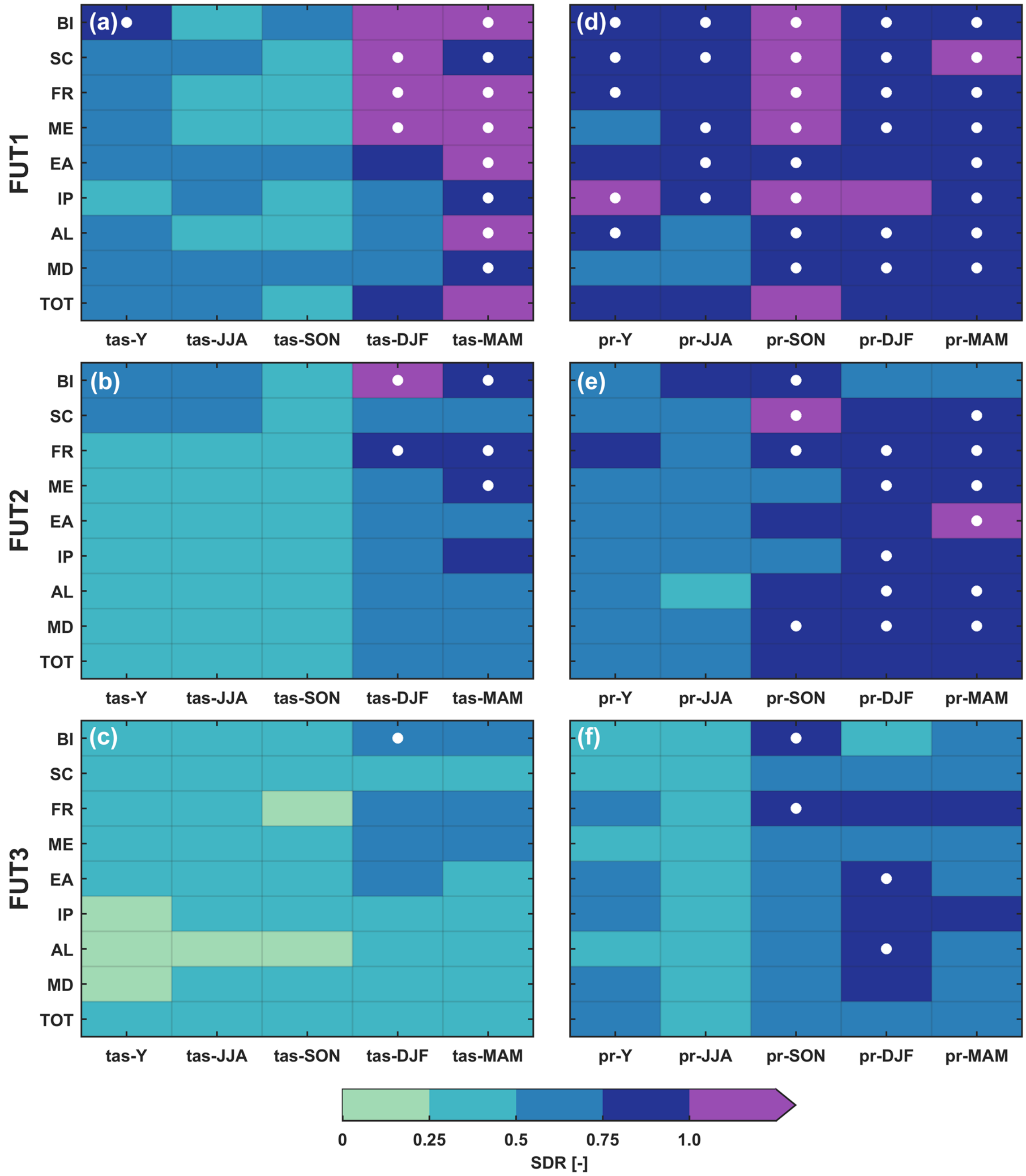


Fig. 9 SDR in all regions for annual and seasonal temperatures (**a**–**c**) and seasonal precipitation (**d**–**f**) in the three future periods. The white dots indicate no rejection of the Null Hypothesis (of equal variances) of the two-sample F-Test at a significance level of 5% in at least 67% of the grid points in this subregion

conducted to better quantify the contribution of natural variability in climate change signal uncertainty. In general, the CRCM5-LE ensemble shows stronger climate change signals for temperature than the CORDEX ensemble—especially for summer, where CRCM5-LE signals are very dry and warm. The negative precipitation signals in summer and fall are not due to a general dry bias of the CRCM5-LE data, since neither the CanESM2 nor the ERA-Interim driven CRCM5 simulation showed significant dry biases for most parts of Europe from 1980 to 2012, especially not for the Mediterranean, where the severe decreases are projected (Leduc et al. 2019). The CanESM2_CCLM4-8-17 simulation shows even drier signals for these seasons than the already relatively dry CRCM5-LE, while the CanESM2_REMO2015 simulation fits well into the CORDEX ensemble. The choice of RCM for downscaling CanESM2 thus seems to have a significant influence on the summer precipitation signals. A similar range can also be found for the five HadGEM2-ES simulations.

5.3 Signal variability comparison

Grid-point based analysis showed that the CORDEX variability is usually much higher than the CRCM5-LE variability for temperatures (Figs. 4 and 5). However, both ensembles generally show an increase in variability of temperature towards more continental climates in the East, which suggests that this gradient in CORDEX is at least partly due to internal variability. Only in the near future during winter larger parts of Europe show similar variabilities in both ensembles. For the British Isles and Norway, CRCM5-LE even shows higher variability in FUT1 than CORDEX. A significant difference between the two ensembles is that mountainous areas in CORDEX often show the highest standard deviations for temperature, while in CRCM5-LE they show rather low values. This is probably because of the different orographies and snow representations used in the different RCMs in CORDEX, since it is mostly visible during DJF and MAM, when the largest snow packs are present. For summer and winter precipitation the patterns of equal variances are way more scattered over all of Europe (Figs. 6 and 7). While in FUT3 for summer almost no grid points show similar variances, the internal variability can still reach a similar variance than the multi-model variance in large parts of Europe in winter precipitation.

The analysis of pattern correlations between signals of the two ensembles gave a two-folded result. While for precipitation, the expectation of decreasing pattern correlations was met, the temporal dynamics for temperature did not show clear indications (Fig. 8). Further research may be needed to clarify the relationships between patterns in this context.

In general, the contribution of internal variability is much higher for precipitation signals than for temperature signals. Additionally, the influence of internal variability significantly decreases for later future periods (Fig. 9). Nevertheless, in many regions the contribution lies between 0.25 and 0.5 for seasonal temperature, and between 0.5 and 1.0 for seasonal precipitation. SDR values around or even above one, seem to be implausible on the first glance. Even if internal variability plays an important role in the uncertainty, the sum of model response uncertainty and internal variability in the CORDEX ensemble should generally be higher than the internal variability alone in CRCM5-LE. These values probably occur when the CORDEX ensemble cannot capture the whole range of internal variability because of the limited sampling in this ensemble. Or the internal variability of the CRCM5-LE is not representative for other models (GCM/RCM combinations). Further research is needed on the comparison of the representations of internal variability in different large ensembles.

Deser et al. (2012b) conducted a similar experiment with 21 CMIP3 models and 40 CCSM3 members, building a ratio between the standard deviations of trends from 2005 to 2060 globally. They also find ratios above 0.75 and 1.0 for large parts of Europe for annual temperature and precipitation—the latter generally showing much higher ratios. The ratios found by Deser et al. (2012b) for Europe are usually higher than the ones calculated for the ensembles in this study. This might originate from the different models and methods. While we used dynamically downscaled regional models for Europe, they used GCMs. Additionally they quantified the precipitation trend variability in mm/day, while we used relative changes.

To evaluate the influence of spatial resolution on the regionalized SDR values, we conducted a small methodological experiment, which cannot directly clarify the differing results, but helps to identify possible sources better. For the values in Fig. 9, the SDR is calculated on a grid point scale and is averaged over the subregions afterwards (Method M1). Another method can be to average over the temperature/precipitation values as a first step and do the signal, variability and SDR computation with these spatially averaged values afterwards (Method M2). This is a possible way to “simulate” a coarser resolution of the underlying data, like it would come from a very coarse GCM. The differences between the two methods are shown in Fig. 10. A comparison of the results on annual time scale (Deser et al. 2012b only show annual results) reveal no difference between the two methods for tas-Y, and show higher SDR values by M1 for pr-Y. Thus, a coarser resolution data set will not produce higher ratios in this experiment. This contradicts the hypothesis that differences in the spatial resolution of the applied models could explain differing results. The differences between the two studies in terms of used models and applied methods make the identification of the reasons behind the differences even more difficult.

6 Conclusions

Natural variability (represented by the model-internal variability of a single model large ensemble) can play a major role in the variability magnitude of future climate projections, depending on the regarded variable, season, region and time horizon. These findings are of such importance, since climate modelers are often facing criticism for the large uncertainty of ensemble projections, with the criticism implying that the variety of model results is a consequence of the models’ inability to correctly represent climate processes (model response uncertainty). If natural variability can explain a large part of the spread of models, then the

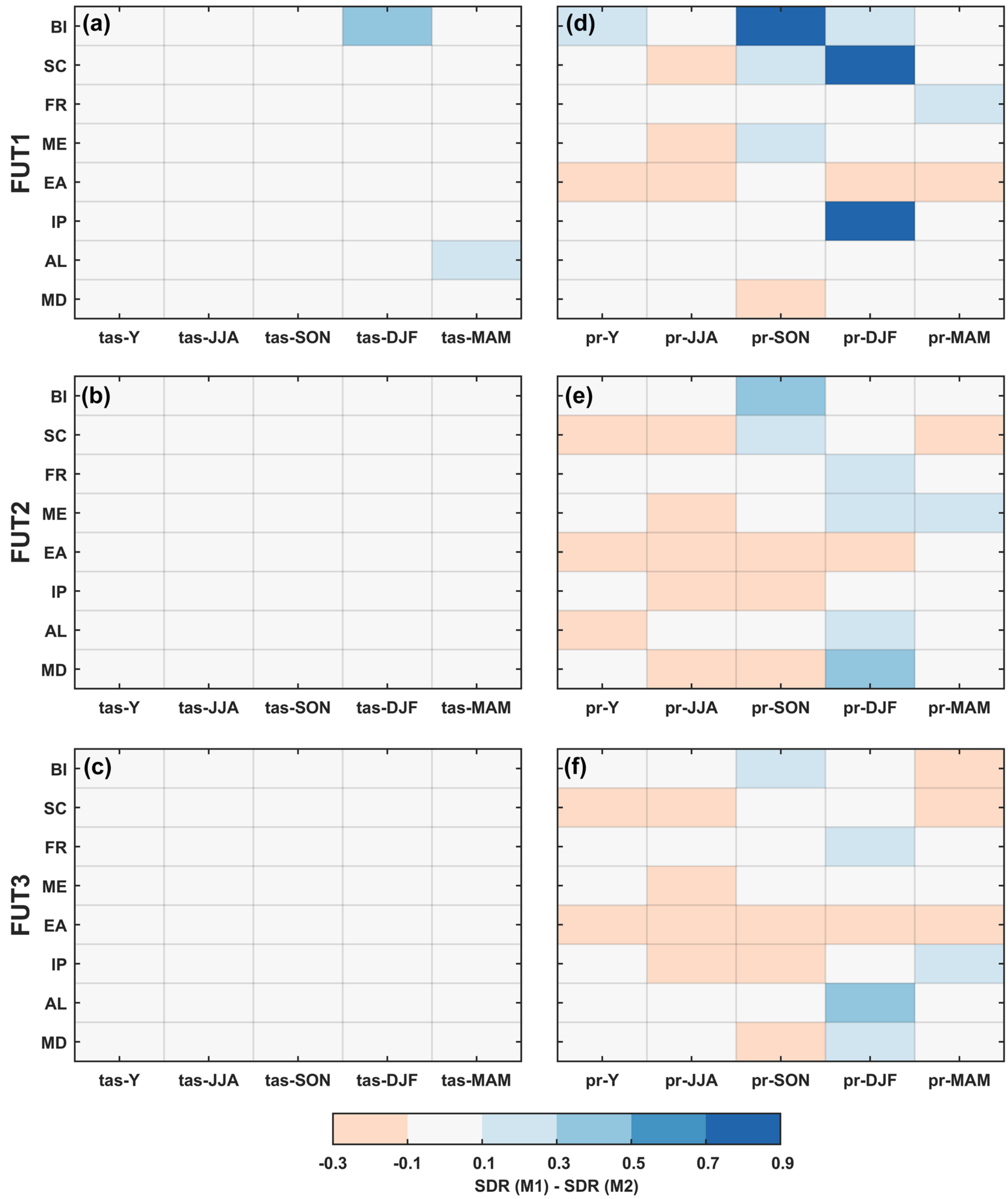


Fig. 10 Difference of regional SDR outcome between methods M1 and M2 (M1 minus M2); M1 results are shown in Fig. 9

differences in climate signals from different models might not only be a result of insufficient or competing models, but might also be partly explained by natural variability. Following the idea that natural variability is inherent to the chaotic nature of the climate system and therefore cannot be diminished, a certain part of uncertainty of climate projections will be irreducible, even if scenario definition will become more precise and models will improve (see also Deser et al. 2012b).

The implications for the interpretation of multi-model ensembles in cases with similar variabilities (e.g. precipitation in DJF), might become clearer as a short mind game: First we need to accept the natural variability of the CRCM5-LE as a fair approximation to adopt it for other models. Then two things become apparent:

(A) We can add the CRCM5-LE variability as a "cloud" of internal variability (gray dots in e.g. Fig. 3) around each of the dots for the CORDEX models. This is blurring the model response uncertainty dramatically in some cases. This means that only one realization, as in CORDEX usually available, does probably not depict the model response very well.
(B) On the other hand, if the CORDEX ensemble might even be totally (or to a large part) explained by natural variability, model response uncertainty may be interpreted as neglectable in these cases.

These conclusions are of course very much depending on the length of the time period, variable, season and region considered, and are not meant as universally valid for multi-model ensembles like CORDEX.

As a short outlook, the existing regional single model ensembles (Fischer et al. 2013; Aalbers et al. 2018) still need to be analyzed in more depth, since they show large potential for a better understanding of climate change uncertainty. Additionally, the previous results should be verified by more single model ensembles, and the differences between these kinds of ensembles need to be specified, e.g. to see if their representations of natural variability are similar (see also Xie et al. 2015).

Acknowledgements We thank all members of the ClimEx project working group for their contributions to produce and analyze the CanESM2-LE and CRCM5-LE. Special thanks go to Raul Wood and Florian Willkofer, helping to improve the work behind this paper by numerous discussions and by technical support. The ClimEx project is funded by the Bavarian State Ministry for the Environment and Consumer Protection. The CRCM5 was developed by the ESCER centre of Université du Québec a Montréal (UQAM; http://www.escer.uqam.ca) in collaboration with Environment and Climate Change Canada. We acknowledge Environment and Climate Change Canada's Canadian Centre for Climate Modelling and Analysis for executing and making available the CanESM2 Large Ensemble simulations used in this study, and the Canadian Sea Ice and Snow Evolution Network for proposing the simulations. Computations with the CRCM5 for the ClimEx project were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. The operation of this supercomputer is funded via the Gauss Centre for Supercomputing (GCS) by the German Federal Ministry of Education and Research and the Bavarian State Ministry of Education, Science and the Arts. We want to thank the CORDEX initiative and the ReKliEs-De project, as well as the ESGF for providing the multi-model ensemble data.

References

Aalbers EE, Lenderink G, van Meijgaard E, van den Hurk, Bart JJM (2018) Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? Clim Dyn 50:4745–4766. https://doi.org/10.1007/s00382-017-3901-9

Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future Köppen–Geiger climate classification maps at 1-km resolution. Sci Data 5:180214. https://doi.org/10.1038/sdata.2018.214

Borodina A, Fischer EM, Knutti R (2017) Emergent constraints in climate projections: a case study of changes in high-latitude temperature variability. J. Clim 30:3655–3670. https://doi.org/10.1175/JCLI-D-16-0662.1

Brisson E, Demuzere M, Willems P, Lipzig, van Nicole PM (2015) Assessment of natural climate variability using a weather generator. Clim Dyn 44:495–508. https://doi.org/10.1007/s00382-014-2122-8

Christensen JH, Christensen OB (2007) A summary of the PRUDENCE model projections of changes in European climate by the end of this century. Clim Change 81:7–30. https://doi.org/10.1007/s10584-006-9210-7

Deser C, Knutti R, Solomon S, Phillips AS (2012a) Communication of the role of natural variability in future North American climate. Nat Clim Change 2:775–779. https://doi.org/10.1038/nclimate1562

Deser C, Phillips A, Bourdette V, Teng H (2012b) Uncertainty in climate change projections: the role of internal variability. Clim Dyn 38:527–546. https://doi.org/10.1007/s00382-010-0977-x

Deser C, Phillips AS, Alexander MA, Smoliak BV (2014) Projecting North American climate over the next 50 years: uncertainty due to internal variability. J Clim 27:2271–2296. https://doi.org/10.1175/JCLI-D-13-00451.1

Fantini A, Raffaele F, Torma C, Bacer S, Coppola E, Giorgi F, Ahrens B, Dubois C, Sanchez E, Verdecchia M (2016) Assessment of multiple daily precipitation statistics in ERA-Interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. Clim Dyn 108:4257. https://doi.org/10.1007/s00382-016-3453-4

Fischer EM, Beyerle U, Knutti R (2013) Robust spatially aggregated projections of climate extremes. Nat Clim Change 3:1033–1038. https://doi.org/10.1038/nclimate2051

Fyfe JC, Derksen C, Mudryk L, Flato GM, Santer BD, Swart NC, Molotch NP, Zhang X, Wan H, Arora VK, Scinocca J, Jiao Y (2017) Large near-term projected snowpack loss over the western United States. Nat Commun 8:14996. https://doi.org/10.1038/ncomms14996

Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. Bull Am Meteor Soc 90:1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Hawkins E, Sutton R (2011) The potential to narrow uncertainty in projections of regional precipitation change. Clim Dyn 37:407–418. https://doi.org/10.1007/s00382-010-0810-6

Hawkins E, Smith RS, Gregory JM, Stainforth DA (2016) Irreducible uncertainty in near-term climate projections. Clim Dyn 46:3807–3819. https://doi.org/10.1007/s00382-015-2806-8

Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, Braun A, Colette A, Déqué M, Georgievski G, Georgopoulou E, Gobiet A, Menut L, Nikulin G, Haensler A, Hempelmann N, Jones C, Keuler K, Kovats S, Kröner N, Kotlarski S, Kriegsmann A, Martin E, van Meijgaard E, Moseley C, Pfeifer S, Preuschmann S, Radermacher C, Radtke K, Rechid D, Rounsevell M, Samuelsson P, Somot S, Soussana J-F, Teichmann C, Valentini R, Vautard R, Weber B, Yiou P (2014) EURO-CORDEX: new high-resolution climate change projections for

European impact research. Reg Environ Change 14:563–578. https://doi.org/10.1007/s10113-013-0499-2

Kay JE, Deser C, Phillips A, Mai A, Hannay C, Strand G, Arblaster JM, Bates SC, Danabasoglu G, Edwards J, Holland M, Kushner P, Lamarque J-F, Lawrence D, Lindsay K, Middleton A, Munoz E, Neale R, Oleson K, Polvani L, Vertenstein M (2015) The community earth system model (CESM) large ensemble project: a community resource for studying climate change in the presence of internal climate variability. Bull Am Meteor Soc 96:1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kendon EJ, Jones RG, Kjellström E, Murphy JM (2010) Using and designing GCM–RCM ensemble regional climate projections. J Clim 23:6485–6503. https://doi.org/10.1175/2010JCLI3502.1

Knutti R, Sedláček J (2013) Robustness and uncertainties in the new CMIP5 climate model projections. Nat Clim Change 3:369. https://doi.org/10.1038/nclimate1716

Knutti R, Sedláček J, Sanderson BM, Lorenz R, Fischer EM, Eyring V (2017) A climate model projection weighting scheme accounting for performance and interdependence. Geophys Res Lett 44:1909–1918. https://doi.org/10.1002/2016GL072012

Kotlarski S, Keuler K, Christensen OB, Colette A, Déqué M, Gobiet A, Goergen K, Jacob D, Lüthi D, van Meijgaard E, Nikulin G, Schär C, Teichmann C, Vautard R, Warrach-Sagi K, Wulfmeyer V (2014) Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci Model Dev 7:1297–1333. https://doi.org/10.5194/gmd-7-1297-2014

Leduc M, Laprise R, de Elía R, Šeparović L (2016a) Is institutional democracy a good proxy for model independence? J Clim 29:8301–8316. https://doi.org/10.1175/JCLI-D-15-0761.1

Leduc M, Matthews HD, de Elía R (2016b) Regional estimates of the transient climate response to cumulative CO_2 emissions. Nat Clim Change 6:474

Leduc M, Mailhot A, Frigon A, Martel J-L, Ludwig R, Brietzke GB, Giguére M, Brissette F, Turcotte R, Braun M, Scinocca J (2019) ClimEx project: a 50-member ensemble of climate change projections at 12-km resolution over Europe and northeastern North America with the Canadian Regional Climate Model (CRCM5). J Appl Meteorol Climatol 1:2. https://doi.org/10.1175/jamc-d-18-0021.1

Lenderink G (2010) Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations. Clim Res 44:151–166. https://doi.org/10.3354/cr00946

Lorenz P, Jacob D (2010) Validation of temperature trends in the ENSEMBLES regional climate model runs driven by ERA40. Clim Res 44:167–177. https://doi.org/10.3354/cr00973

Lorenz R, Herger N, Sedláček J, Eyring V, Fischer EM, Knutti R (2018) Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. J Geophys Res Atmos 123:4509–4526. https://doi.org/10.1029/2017JD027992

Martynov A, Laprise R, Sushama L, Winger K, Šeparović L, Dugas B (2013) Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian regional climate model, version 5: model performance evaluation. Clim Dyn 41:2973–3005. https://doi.org/10.1007/s00382-013-1778-9

Mizuta R, Murata A, Ishii M, Shiogama H, Hibino K, Mori N, Arakawa O, Imada Y, Yoshida K, Aoyagi T, Kawase H, Mori M, Okada Y, Shimura T, Nagatomo T, Ikeda M, Endo H, Nosaka M, Arai M, Takahashi C, Tanaka K, Takemi T, Tachikawa Y, Temur K, Kamae Y, Watanabe M, Sasaki H, Kitoh A, Takayabu I, Nakakita E, Kimoto M (2017) Over 5,000 Years of Ensemble Future Climate Simulations by 60-km Global and 20-km Regional Atmospheric Models. Bull Am Meteor Soc 98:1383–1398. https://doi.org/10.1175/BAMS-D-16-0099.1

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP, Carter TR, Emori S, Kainuma M, Kram T, Meehl GA, Mitchell JFB, Nakicenovic N, Riahi K, Smith SJ, Stouffer RJ, Thomson AM, Weyant JP, Wilbanks TJ (2010) The next generation of scenarios for climate change research and assessment. Nature 463:747. https://doi.org/10.1038/nature08823

Pennell C, Reichler T (2010) On the effective number of climate models. J Clim 24:2358–2367. https://doi.org/10.1175/2010JCLI3814.1

Prein AF, Gobiet A, Truhetz H, Keuler K, Goergen K, Teichmann C, Fox Maule C, van Meijgaard E, Déqué M, Nikulin G, Vautard R, Colette A, Kjellström E, Jacob D (2016) Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: high resolution, high benefits? Clim Dyn 46:383–412. https://doi.org/10.1007/s00382-015-2589-y

Rajczak J, Schär C (2017) Projections of future precipitation extremes over europe: a multimodel assessment of climate simulations. J Geophys Res Atmos 122:10773–10800. https://doi.org/10.1002/2017jd027176

Roudier P, Andersson JCM, Donnelly C, Feyen L, Greuell W, Ludwig F (2016) Projections of future floods and hydrological droughts in Europe under a + 2 °C global warming. Clim Change 135:341–355. https://doi.org/10.1007/s10584-015-1570-4

Saffioti C, Fischer EM, Knutti R (2017) Improved consistency of climate projections over Europe after accounting for atmospheric circulation variability. J Clim 30:7271–7291. https://doi.org/10.1175/JCLI-D-16-0695.1

Šeparović L, Alexandru A, Laprise R, Martynov A, Sushama L, Winger K, Tete K, Valin M (2013) Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. Clim Dyn 41:3167–3201. https://doi.org/10.1007/s00382-013-1737-5

Smiatek G, Kunstmann H, Senatore A (2016) EURO-CORDEX regional climate model analysis for the Greater Alpine Region: performance and expected future change. J Geophys Res Atmos 121:7710–7728. https://doi.org/10.1002/2015JD024727

Solomon A, Goddard L, Kumar A, Carton J, Deser C, Fukumori I, Greene AM, Hegerl G, Kirtman B, Kushnir Y, Newman M, Smith D, Vimont D, Delworth T, Meehl GA, Stockdale T (2010) Distinguishing the roles of natural and anthropogenically forced decadal climate variability. Bull Am Meteor Soc 92:141–156. https://doi.org/10.1175/2010BAMS2962.1

Torma C, Giorgi F, Coppola E (2015) Added value of regional climate modeling over areas characterized by complex terrain-precipitation over the Alps. J Geophys Res Atmos 120:3957–3972. https://doi.org/10.1002/2014JD022781

Vautard R, Gobiet A, Sobolowski S, Kjellström E, Stegehuis A, Watkiss P, Mendlik T, Landgren O, Nikulin G, Teichmann C, Jacob D (2014) The European climate under a 2 & #xB0;C global warming. Environ Res Lett 9:34006. https://doi.org/10.1088/1748-9326/9/3/034006

Xie S-P, Deser C, Vecchi GA, Collins M, Delworth TL, Hall A, Hawkins E, Johnson NC, Cassou C, Giannini A, Watanabe M (2015) Towards predictive understanding of regional climate change. Nat Clim Change 5:921–930. https://doi.org/10.1038/nclimate2689

Zheng X-T, Hui C, Yeh S-W (2018) Response of ENSO amplitude to global warming in CESM large ensemble: uncertainty due to internal variability. Clim Dyn 50:4019–4035. https://doi.org/10.1007/s00382-017-3859-7

8 SMILEs as statistically sound data base for extreme event analysis

While the first two publications dealt with the aspects of variability and uncertainty on a more theoretical level, the third paper can be considered a straightforward impact analysis using one SMILE. The main focus are meteorological droughts that have also been addressed in the former publications (dry period lengths). Here, the drought is analyzed with a return level approach that intends to detect changes in former 30-year return levels (RL30) of monthly/seasonal low precipitation frequencies.

8.1 How will dry period frequencies change over Europe in the future?

The most dramatic changes are projected to occur during the summer and fall months in large parts of Europe, where former RL30 months will occur much more frequently. For the most extreme changes in Spain, former RL30 will occur every year. Consistently, consecutive dry summers will also become more frequent on the 2-year level, while only a few areas will also show increases on the 3-to-4-year level, mostly in Southern Europe and parts of France. Consecutive dry winters will only occur in very few areas in Spain and Northern Africa and remain very unlikely in the vast parts of Europe.

8.2 How does that relate to the outstanding 2018 summer drought in Germany?

The probability of a 2018-like summer drought in Germany will dramatically increase from a low 1.4 % to 20 % in the far future, based on a quantile transfer approach from observations to the CRCM5-LE.

8.3 How does this meteorological signal propagate into a hydrological discharge signal?

The summer drought in precipitation will have similar effects on the RL30 of two hydrological gauges in Bavaria. These gauges also show higher frequency of low flows in the winter months, where the precipitation is showing less frequent drought situations. This shows the high non-linearity between precipitation and streamflow.

TITLE
The summer drought 2018 in Germany – a window to future developments of meteorological droughts in Europe?

AUTHORS
Fabian von Trentini, Ralf Ludwig

JOURNAL
Geophysical Research Letters

PUBLISHING DATE
Submitted

DOI
-

JOURNAL IMPACT FACTOR (WEB OF SCIENCE)
5.58

AUTHOR CONTRIBUTIONS
FvT designed the concept of the study, performed the analysis, and created all figures. RL helped improve the concept and analysis. FvT led the paper writing with input from RL.

CITATIONS (CROSSREF)
-

PLAIN LANGUAGE SUMMARY
The summer 2018 drought in Germany was the second most severe in the history of records for precipitation. The future changes in drought frequencies are assessed for all of Europe by using a specific data source: an ensemble of climate model data that stem from a so called single-model initial-condition large ensemble (SMILE). Here, the exact same model is started 50 times, driven by the same emission scenario. The only difference between the runs are minor changes in the initial conditions. Months that had a statistical occurrence probability of every 30 years will occur much more often in the future. This is mainly true for summer and fall over large parts of Europe. Hot spots for consecutive dry summers (2-4) will mainly appear in the Mediterranean region, especially for Spain and France. Linking the statistics from the 2018 summer drought in Germany with the climate model data, the probability of such an event rises from 1.4 % under historical conditions to 20 % for the end of the 21st century. The changes in drought frequency defined by low precipitation amounts is then compared to changes in low flows in two rivers in Germany. The river discharge shows similar increases in frequency as the precipitation during summer. However, during winter half year, the low flows will occur much more often, although the precipitation changes do not depict this change in a similar way. This non-linearity might be caused by changes in evapotranspiration and snow dynamics.

9 References

Aalbers, Emma E.; Lenderink, Geert; van Meijgaard, Erik; van den Hurk, Bart J. J. M. (2018): Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? In: *Climate Dynamics* 50 (11), S. 4745–4766. DOI: 10.1007/s00382-017-3901-9.

Addor, Nans; Fischer, Erich M. (2015): The influence of natural variability and interpolation errors on bias characterization in RCM simulations. In: *J. Geophys. Res. Atmos.* 120 (19), 10,180-10,195. DOI: 10.1002/2014JD022824.

Bengtsson, L.; Hodges, K. I. (2019): Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability? In: *Clim Dyn* 52 (5-6), S. 3553–3573. DOI: 10.1007/s00382-018-4343-8.

Bonan, Gordon B.; Doney, Scott C. (2018): Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. In: *Science (New York, N.Y.)* 359 (6375). DOI: 10.1126/science.aam8328.

Brönnimann, Stefan; Rajczak, Jan; Fischer, Erich M.; Raible, Christoph C.; Rohrer, Marco; Schär, Christoph (2018): Changing seasonality of moderate and extreme precipitation events in the Alps. In: *Nat. Hazards Earth Syst. Sci.* 18 (7), S. 2047–2056. DOI: 10.5194/nhess-18-2047-2018.

Christensen, Jens Hesselbjerg; Larsen, Morten A. D.; Christensen, Ole B.; Drews, Martin; Stendel, Martin (2019): Robustness of European climate projections from dynamical downscaling. In: *Clim Dyn* 419, S. 224. DOI: 10.1007/s00382-019-04831-z.

Copernicus (2021): Latest projections of future climate now available. Online verfügbar unter https://climate.copernicus.eu/latest-projections-future-climate-now-available.

Coppola, Erika; Nogherotto, Rita; Ciarlo', James M.; Giorgi, Filippo; van Meijgaard, Erik; Kadygrov, Nikolay et al. (2021): Assessment of the European Climate Projections as Simulated by the Large EURO-CORDEX Regional and Global Climate Model Ensemble. In: *J. Geophys. Res. Atmos.* 126 (4), e2019JD032356. DOI: 10.1029/2019JD032356.

Deser, Clara; Phillips, Adam; Bourdette, Vincent; Teng, Haiyan (2012): Uncertainty in climate change projections. The role of internal variability. In: *Climate Dynamics* 38 (3-4), S. 527–546. DOI: 10.1007/s00382-010-0977-x.

Eyring, Veronika; Cox, Peter M.; Flato, Gregory M.; Gleckler, Peter J.; Abramowitz, Gab; Caldwell, Peter et al. (2019): Taking climate model evaluation to the next level. In: *Nature Clim Change* 9 (2), S. 102–110. DOI: 10.1038/s41558-018-0355-y.

Fischer, E. M.; Beyerle, U.; Knutti, R. (2013): Robust spatially aggregated projections of climate extremes. In: *Nature Clim Change* 3 (12), S. 1033–1038. DOI: 10.1038/nclimate2051.

Fischer, E. M.; Schär, C. (2010): Consistent geographical patterns of changes in high-impact European heatwaves. In: *Nature Geosci* 3 (6), S. 398–403. DOI: 10.1038/ngeo866.

Fischer, Erich M.; Schär, Christoph (2009): Future changes in daily summer temperature variability: driving processes and role for temperature extremes. In: *Clim Dyn* 33 (7-8), S. 917–935. DOI: 10.1007/s00382-008-0473-8.

Flato, G.; J. Marotzke; B. Abiodun; P. Braconnot; S.C. Chou; W. Collins et al. (Hg.) (2014): Climate Change 2013. The Physical Science Basis. Cambridge: Cambridge University Press.

Fyfe, John C.; Derksen, Chris; Mudryk, Lawrence; Flato, Gregory M.; Santer, Benjamin D.; Swart, Neil C. et al. (2017): Large near-term projected snowpack loss over the western United States. In: *Nature communications* 8, S. 14996. DOI: 10.1038/ncomms14996.

Giorgi, Filippo (2019): Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next? In: *J. Geophys. Res. Atmos.* 92 (7), S. 365. DOI: 10.1029/2018JD030094.

Hausfather, Zeke; Drake, Henri F.; Abbott, Tristan; Schmidt, Gavin A. (2020): Evaluating the Performance of Past Climate Model Projections. In: *Geophysical Research Letters* 47 (1). DOI: 10.1029/2019GL085378.

Hawkins, Ed; Sutton, Rowan (2009): The Potential to Narrow Uncertainty in Regional Climate Predictions. In: *Bull. Amer. Meteor. Soc.* 90 (8), S. 1095–1107. DOI: 10.1175/2009BAMS2607.1.

Hawkins, Ed; Sutton, Rowan (2011): The potential to narrow uncertainty in projections of regional precipitation change. In: *Climate Dynamics* 37 (1), S. 407–418. DOI: 10.1007/s00382-010-0810-6.

Holmes, Caroline R.; Woollings, Tim; Hawkins, Ed; Vries, Hylke de (2016): Robust Future Changes in Temperature Variability under Greenhouse Gas Forcing and the Relationship with Thermal Advection. In: *J. Climate* 29 (6), S. 2221–2236. DOI: 10.1175/JCLI-D-14-00735.1.

Hourdin, Frédéric; Mauritsen, Thorsten; Gettelman, Andrew; Golaz, Jean-Christophe; Balaji, Venkatramani; Duan, Qingyun et al. (2017): The Art and Science of Climate Model Tuning. In: *Bull. Amer. Meteor. Soc.* 98 (3), S. 589–602. DOI: 10.1175/BAMS-D-15-00135.1.

IPCC (2018): What is a GCM? Online verfügbar unter https://www.ipcc-data.org/guidelines/pages/gcm_guide.html, zuletzt geprüft am 14.03.2023.

Jacob, Daniela; Petersen, Juliane; Eggert, Bastian; Alias, Antoinette; Christensen, Ole Bøssing; Bouwer, Laurens M. et al. (2014): EURO-CORDEX. New high-resolution climate change projections for European impact research. In: *Reg Environ Change* 14 (2), S. 563–578. DOI: 10.1007/s10113-013-0499-2.

Kendon, Elizabeth J.; Jones, Richard G.; Kjellström, Erik; Murphy, James M. (2010): Using and Designing GCM–RCM Ensemble Regional Climate Projections. In: *J. Climate* 23 (24), S. 6485–6503. DOI: 10.1175/2010JCLI3502.1.

Leduc, Martin; Mailhot, Alain; Frigon, Anne; Martel, Jean-Luc; Ludwig, Ralf; Brietzke, Gilbert B. et al. (2019): ClimEx project: a 50-member ensemble of climate change projections at 12-km resolution over Europe and northeastern North America with the Canadian Regional Climate Model (CRCM5). In: *Journal of Applied Meteorology and Climatology*. DOI: 10.1175/JAMC-D-18-0021.1.

Lehner, Flavio; Deser, Clara; Maher, Nicola; Marotzke, Jochem; Fischer, Erich M.; Brunner, Lukas et al. (2020): Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. In: *Earth Syst. Dynam.* 11 (2), S. 491–508. DOI: 10.5194/esd-11-491-2020.

Martynov, A.; Laprise, R.; Sushama, L.; Winger, K.; Šeparović, L.; Dugas, B. (2013): Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. In: *Climate Dynamics* 41 (11), S. 2973–3005. DOI: 10.1007/s00382-013-1778-9.

Medhaug, Iselin; Stolpe, Martin B.; Fischer, Erich M.; Knutti, Reto (2017): Reconciling controversies about the 'global warming hiatus'. In: *Nature* 545 (7652), S. 41–47. DOI: 10.1038/nature22315.

Prein, A. F.; Gobiet, A.; Truhetz, H.; Keuler, K.; Goergen, K.; Teichmann, C. et al. (2016): Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations. High resolution, high benefits? In: *Climate Dynamics* 46 (1-2), S. 383–412. DOI: 10.1007/s00382-015-2589-y.

Rajczak, Jan; Schär, Christoph (2017): Projections of Future Precipitation Extremes Over Europe. A Multimodel Assessment of Climate Simulations. In: *J. Geophys. Res. Atmos.* 122 (20), 10,773-10,800. DOI: 10.1002/2017JD027176.

Saffioti, Claudio; Fischer, Erich M.; Knutti, Reto (2017): Improved Consistency of Climate Projections over Europe after Accounting for Atmospheric Circulation Variability. In: *J. Climate* 30 (18), S. 7271–7291. DOI: 10.1175/JCLI-D-16-0695.1.

Šeparović, Leo; Alexandru, Adelina; Laprise, René; Martynov, Andrey; Sushama, Laxmi; Winger, Katja et al. (2013): Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. In: *Climate Dynamics* 41 (11), S. 3167–3201. DOI: 10.1007/s00382-013-1737-5.

Suarez-Gutierrez, Laura; Li, Chao; Müller, Wolfgang A.; Marotzke, Jochem (2018): Internal variability in European summer temperatures at 1.5 °C and 2 °C of global warming. In: *Environ. Res. Lett.* 13 (6), S. 64026. DOI: 10.1088/1748-9326/aaba58.

Suarez-Gutierrez, Laura; Milinski, Sebastian; Maher, Nicola (2021): Exploiting large ensembles for a better yet simpler climate model evaluation. In: *Climate Dynamics* 57 (9), S. 2557–2580. DOI: 10.1007/s00382-021-05821-w.

Vautard, Robert; Gobiet, Andreas; Sobolowski, Stefan; Kjellström, Erik; Stegehuis, Annemiek; Watkiss, Paul et al. (2014): The European climate under a 2 °C global warming. In: *Environ. Res. Lett.* 9 (3), S. 34006. DOI: 10.1088/1748-9326/9/3/034006.

Vautard, Robert; Kadygrov, Nikolay; Iles, Carley; Boberg, Fredrik; Buonomo, Erasmo; Bülow, Katharina et al. (2021): Evaluation of the Large EURO-CORDEX Regional Climate Model Ensemble. In: *J. Geophys. Res. Atmos.* 126 (17), e2019JD032344. DOI: 10.1029/2019JD032344.

Vidale, P. L.; Lüthi, D.; Wegmann, R.; Schär, C. (2007): European summer climate variability in a heterogeneous multi-model ensemble. In: *Climatic Change* 81 (S1), S. 209–232. DOI: 10.1007/s10584-006-9218-z.

von Trentini, Fabian; Aalbers, Emma E.; Fischer, Erich M.; Ludwig, Ralf (2020): Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe. In: *Earth Syst. Dynam.* 11 (4), S. 1013–1031. DOI: 10.5194/esd-11-1013-2020.

Yettella, Vineel; Weiss, Jeffrey B.; Kay, Jennifer E.; Pendergrass, Angeline G. (2018): An Ensemble Covariance Framework for Quantifying Forced Climate Variability and Its Time of Emergence. In: *J. Climate* 31 (10), S. 4117–4133. DOI: 10.1175/JCLI-D-17-0719.1.