To the Editor and Reviewers:

We would like to thank the reviewers for their thoughtful comments and efforts towards improving our manuscript. We have addressed their concerns and updated the manuscript accordingly. We have also amended the references as suggested by Christian Dieterich in the open discussion of the paper. Below, we include the major changes and the point-to-point response to the comments of both reviewers. We hope the revised version is now suitable for publication and look forward to hearing from you in due course.

Major changes:

1) The atmospheric model is now rigorously described, including details about the parametrization schemes and new references. This version also includes a better description of the lateral boundary data, the initial state of the models, as well as the spin up process of the oceans. Since more effort was addressed in describing the details of the set-up of the experiment, the acknowledgment to Dr. Akhtar changed into a co-authorship in this version. This new version also includes a detailed description of the CPUs configuration of the experiment and the performance analysis with the LUCIA tool.

2) Some figures were slightly modified:
     Fig. 2: label of x-axis (NEMO-NORDIC instead of NEMO-BALTIC).
     Fig. 4b: OISSTv2 Observations added.
     Fig. 11e: Thanks to one comment from the reviewer, we realized that this figure
     included points from a domain slightly bigger (starting from a slightly lower latitude,
     what implied more points from Africa) than the domain represented in Fig. 11c. We
     have corrected this, so all figures now represent points from the domain shown in Fig.
     11(a-d).
     Fig. 17.: Two extra lines were added in the first plot to help the reader with the
     interpretation regarding the intensity of the observed and simulated events.

3) The edit of text has been substantially improved.

We have decided to keep the analysis of the extremes cases since we consider it is of interest to show that the system is also stable estimating extreme indices and in addition it shows improvement regarding heat/cold waves.

Reviewer #1: Alistair Sellar

This new model is a useful contribution to the field of regional climate modelling. I congratulate the authors on its completion, as it must surely have been a difficult task. This manuscript does a good job of assessing the robustness of the coupled system and summarising its performance in key output variables.

For me, however, there is a gap in the interpretation of the precipitation results. Significant effort has been put into the calculation of indices which describe various aspects of precipitation variability, but the text contains very few conclusions which can be drawn from these indices beyond a description of the main state bias.

I would like to see some brief discussion of the likely causes of differences between the coupled and uncoupled systems. For example, why does the coupled system have a larger

seasonal temperature range in the Med, and why is the coupled system colder in the North and Baltic Seas?

Finding a unique reason that explains the sea surface temperature changes in the coupled compared to the uncouple is not a simple thing to do, since those changes may be induced by many factors: ocean initialization, different mixing layer depth, aerosols blocking radiation, internal NEMO dynamics, etc. We have put some effort in looking for explanations (e.g. we analyzed the mixing layer depth, but found no concrete explanation to the question). Continuing working on possible methods that could help us to find out more in this regard is part of our current research, but it may take long to understand the processes affected by the coupling that lead to the temperature changes. Nevertheless, we have extended the discussion about the results and also the interpretation of the extreme indices.
It is also worth mentioning that in the meantime we got accepted a publication that includes a sensitivity test to show how the coupled system presented in our manuscript reacts to changes in the sea surface temperature (Kelemen et al. 2019). The added value of atmosphere-ocean coupling in this century-long regional climate simulation is analysed there in more detail, to complement the work presented in this manuscript. This helps us to answer parallel questions of interest related to the coupled system that cannot be covered in our manuscript to avoid an extremely long paper, but helps us to have a broader picture about the system. This new reference has been included in the new version of the manuscript.

There should ideally be some description of the steps taken to tune the coupled model. Were some model parameters adjusted to achieve acceptable performance, or were the component models simply coupled together and required no further tuning (i.e. all parameters are the same as for the un-coupled equivalents)? Even if no coupled model tuning were required, it would be useful to state this. In global climate model development, there is a growing consensus that it is important (and helpful to other modellers) to document such tuning. See e.g. Hourdin et al (2017, doi:10.1175/BAMS-D-15- 00135.1), Schmidt et al (2017, doi:10.5194/gmd-10-3207-2017), Golaz et al (2019, doi:10.1029/2018MS001603). Such information would be equally useful for a regional modelling paper.

We have added more details in the new version to explain this part better. Page 4 line 12: "There was no tuning in the coupled version, thus the configuration of the atmospheric model was the same in the coupled and uncoupled versions". Future research could be addressed to investigate a possible tuning of the coupled system.

Minor corrections and suggestions included in the supplement PDF.

The reviewer has included his minor corrections and suggestions classified by colours:

Red:

All these minor corrections have been included in the new version.

Yellow:

Page 3: "with an atmospheric grid resolution..." (the ocean resolution is higher)

Changed (Page 3, line 3): *"RCSM over Europe with an atmospheric horizontal grid resolution of about 25km".*

Page 5: How many years of spinup were required before the start of the coupled historical simulation? What criteria were used to decide that the model was sufficiently spun up?

The NEMO-MED12 model was spun up in coupled mode during 20-years driven by randomly resampled MPI-ESM years in the period 1900-1910. The NEMO-NORDIC model was spin-up for the period 1900-1905. Surface variables SST and sea surface salinity following balanced initialization were analysed, and no drift was detectable. This explanation has been included in Section 2.2 (Page 4, lines 27-28; Page 5, lines 6-10).

Green:

Page 5: What is the source of this external forcing data - the same as the uncoupled atmosphere-only run? Does the atmosphere model exhibit any artefacts of the jump in forcing data above the ocean model boundary?

The external forcing data is the same as the uncoupled atmosphere-only run. Clarification added in Pag 5. Line 23. On average, the atmosphere temperature shows differences in the ocean model boundary of the North Sea below 2.5°C. See Fig.8 (a,b). Limits of the marginal seas are shown in Fig 1., but please note that Fig.1 is in rotated coordinates, while Fig. 8 in regular lat-lon.

Page 6: You should be explicit that this is 2m air temperature, here and below.

Done.

Page 7: I don't believe that the upscaling itself penalizes the higher resolution data. As a thought experiment, consider a perfect 0.22 degree and a perfect 0.25 degree model, the latter aligned with the 0.5 degree obs grid. The upscaled 0.25 deg model results would verify perfectly because 4 grid points would be correctly averaged to their 0.5 deg equivalent. The upscaled 0.22 deg model would suffer some smoothing due to the interpolation required because it does not sub-divide the 0.5 deg grid. So it is the interpolation rather than the upscaling per-se which may penalise the high-res model, and I would expect this to be a minor effect. It may be truer to say that the upscaling prevents us from validating the high-resolution information available in the model.

Modified.

Page 8: Inconsistent with figure legend which refers to global rather than uncoupled model. It would be helpful to remind the reader here than for the ocean, uncoupled model = global model.

In the new text it has been emphasized that the uncoupled SST comes from the global model.

Page 8: Why only for the Med and not the North or Baltic seas? If OISST is not appropriate for these seas, how about regional observation and reanalysis products which are available from the EU's Copernicus Marine Service?

Figure updated with OISST in all plots.

Page 8: It is worth pointing out that we would expect to outperform global GCMs in the Baltic and North seas, since the ocean models they contain are not well suited to shelf seas.

That doesn't detract from the usefulness of this model, but it does make this aspect of its good performance unsurprising.

This has been pointed out in the new text (Page 8, lines28-30):" Improving the quality of the averaged global model was expected since the global circulation models contain ocean models that are not well suited to shelf seas like the Baltic and North seas".

Page 9: I find the word "assimilated" misleading in this context because it implies data assimilation. I think you mean that 20CR is forced at the surface by HadISST, which in turn constrains the global ocean simulation? If so then perhaps say that HadISST indirectly constrains the global model.

Text changed.

Page 9: "2m air temperature", "and over land"

Corrected.

Page 10: the total monthly precipitation differences

Changed.

Page 10: We already know this from the maps in Fig 11. What else do we learn from Fig 12?

A better explanation about what we get from Fig.11 has been included in page 11, from line 18 onwards: "Compared to the complete domain (Fig. 11e), the spread of the distributions (box height) over Mid-Europe is smaller than over the whole domain, since over Mid-Europe the temperatures do not differ as much as they do, when comparing the temperatures over the southern and northern parts of the whole domain. In addition, the extreme cases (points) are also milder compared to the whole domain since the summers are not as hot as in southern regions like the Iberian Peninsula or north Africa, as well as winters are not as cold as in northern regions. Fig. 12 also points out winter outliers. In this case, for example, the coupled system estimates better the coldest temperatures in January over Mid-Europe. This is a result that was not appreciated when comparing the whole domain."

Page 11: Same comment as above: can you say anything beyond the bias? E.g. the width of the mid-Europe distribution seems slightly narrower in the models while for the wider domain the distribution is too wide. Is this easy to explain?

The histograms allow a direct comparison of the 2m-temperature values between coupled and uncoupled system. Therefore, we can also describe what happens in the tails of the distribution with more detail. The text has been modified to better explain it. Regarding the narrower width of the histograms compare to the whole domain, this is due to the heterogeneous temperatures over the whole domain, where the southern part is much warmer than the northern part. When the focus lies over Mid-Europe, these differences between the north and south part are not that big.

Page 11: This should be 4.5.

Corrected.

Page 11: simply say annual minimum temperature instead?

Done.

Page 11: I don't understand what is meant by this.

The percentiles calculated considering not only the day of interest, but a window of 5 days centered in that day. This is a way of smoothing the daily percentiles. More details can be found in Zhang et al., 2005.

Page 11: improves?

Changed.

Page 12: I don't understand what is meant here. Perhaps give an example.

We meant that the coupled system also shows a stable evolution. We have changed the text.

Page 12: These complex indices deserve more interpretation. For example, do they tell us about differences in frequency or intensity or precip events?

Model data and observed temporal series have the same length, that is, a particular percentile represents the same number of cases in all of datasets. This allows a comparison between observations and model estimations regarding the frequency and intensity of the total precipitation events. For example, if we focus in a particular intensity range, let's say 700 to 850 mm/year, the number of points lying between two vertical lines at 700 and 850 mm/year gives an idea about the observed frequency in which these intensities occur. The more points, the more cases. The same process but considering horizontal lines, gives an idea about the estimated frequency given by the models. Regarding the intensity of the event, if the line is above the diagonal (e.g. in Potsdam), it overestimates the observed total precipitation distribution, whereas if it is below the diagonal (e.g. Hohenpeißenberg), it underestimates the observed distribution. The further to the diagonal, the higher the over/under-estimation. Lines parallel to the diagonal represent a similar distribution shape than the observed distribution (e.g. TNn, TXx in Postdam). Those lines not always parallel to the diagonal, as it happens for example for R95p, show a distribution with different tails than the observed distribution tails. This happens for example in the maximum total precipitation values

Page 12: If this is the aim, then the manuscript needs to say more about extreme rainfall.

We have removed the sentence about the extreme events.

Page 21: Covering which years?

The whole period, that is 20th century. Information added to the legend.

Page 23: Make the sign of the difference explicit: model minus observations.

Done

Page 24: Be explicit about sign.

Done

Page 24: 2m air temperature

Corrected

Page 24: There is more than one convention for the location of whiskers in boxplots, so please state the method.

Clarified.

Anonymous Referee #2:

This paper adresses the development of a new Regional Climate Model focused on the EURO-CORDEX domain. The model couples together the COSMO-CLM land- atmosphere regional climate model with two different configurations of the NEMO ocean model designed to model 1) the Mediterranean Sea (NEMO-MED) and 2) the Nordic seas (NEMO-NORDIC). A 100 year long simulation is presented.

I think that the technical objective of showing the stability and robustness of the RCSM has been achieved. The comparison of a set of key variables between the coupled, uncoupled simulations and various observations is convincing and demonstrates the correct behaviour of the system. The technical work behind the development of such system is important and worth to be shared with the climate modelling community.

However, I have the feeling that the reader could be somehow frustrated on the scientific side. Overall, I think that the paper is mainly descriptive and lacks some attempt of explanation of the observed results in agreement with the objectives claimed in the abstract and introduction: better understanding of the european regional climate, added value of three marginal seas and better representation of extreme events.

Somehow, on the one hand I would try to give more details on the technical aspects so the paper could be used as a reference for other similar works. But on the other hand, I would be less ambitious regarding the claimed objectives and I would reduce the number of results figures but trying to propose some arguments explaining the difference between the two simulations. Again, I acknowledge the great amount of work behind such a simulation and the difficulty to produce a general and concise scientific content with a long simulation.

We would like to thank the reviewer for the recognition of the amount of work involved in this paper. The main goal of the paper is to show the stability and performance of the system. Regarding the stability, it is clearly shown that the system is stable. However, regarding the performance, it depends on many aspects of interest: variable, area, season, etc. Therefore, it is quite ambitious to include all possible analysis. We have tried to give an overview of the different aspects, with the penalty that concise answers to particular questions could not be analyzed in detail. Nevertheless, in this new version, we have given more technical details as suggested, and we have included some sentence to improve the scientific discussion.

As a complementary work, we refer to Kelemen et al. 2019, for a longer discussion about the added value of this system.

See below a list of questions and suggestions:

- In the atmospheric model, you mention a variety of parametrization scheme, could you be more precise about it ? micro-physics scheme ? convection scheme ? aerosols climatology ?

We chose a parametrization scheme suitable for the domain of interest (EU): (1) Micro-physics scheme: The scheme used in the COSMO-EU model (itype gscp=3 in the namelist of CCLM). This is similar to the original Kessler (1969) scheme, but including cloud ice as an additional prognostic variable (cloud ice scheme). The scheme allows for an explicit representation of ice clouds and a more complete simulation of precipitation formation in mixed phase clouds. (2) Convection scheme: Mass flux Tiedtke scheme. (3) Aerosols climatology: AeroCom Global AOD data (An AeroCom initial assessment – optical properties in aerosol component modules of global models, Atmos. Chem. Phys., 6, 1815–1834, 2006). The manuscript has been modified to have a better description of the atmospheric model (Section 2.1).

- There is no reference to the land component of the system, is there any routing of water ? what about the vegetation ? change of landuse throughout the 100 years ?

COSMO-CLM is internally coupled to the TERRA model, that is the land and vegetation component. There is no lateral routing of water but climatologies are used for discharge. The land use applied throughout the century does not change. This information has been added to the new manuscript (Section 2.1).

- Could you explained the motivations for the 2 step methods providing the lateral boundary conditions ? What is the motivation to do a 3D relaxation of the ocean to- wards a simulation performed with another atmospheric forcings ?

The assimilation of the 20CR reanalyses made by the global MPI model was needed since no ocean lateral boundary conditions are available from 20CR. We better motivated this in the manuscript (Sec. 2.1).

- The exchanged variables and the according coupling time-step should be listed

We have added more details in the coupling section about it. Regarding the variables exchanged, CCLM sends to NEMO Evaporation-Precipitation, solar fluxes, net solar flux, wind stress, and only to NEMO-Nordic Surface pressure. On the other hand, NEMO sends to CCLM the SST, and only NEMO-Nordic the Ice fraction (Page 5, lines 23-26). The exchange is made every three hours (page 5, line 15).

- An important point of such a climatic run is the initial state: is there any spin-up ? spin-up of the ocean only ? spin-up in coupled mode ?

The three marginal seas were spun up. NEMO-Nordic with a 5-year spin-up, and NEMO-Med with 20 years, both in coupled mode. This has been clarified in the new version (Page 4, lines 27-28; Page 5, lines 6-10).

- Regarding the system performances, instead of nodes could you state the number of cpu cores used? and gives some numbers about the cpu-time needed for the run.

Page 5, lines 13-15: "For the centennial simulation, we used 576 CPUs optimally distributed as follows: 24x13 CPUs were assigned to CCLM, 12x8 to NEMO-MED12 and 14x12 to NEMO-NORDIC. With this configuration, each simulated month required around one and a

half hours in total, what implied 78 days to run the complete centennial simulation (110 years)"

- the 3 models are running simulteanously and eventually wait for one another, so to be faster it would be interesting to optimize the slowest model i.e. NEMO-BALTIC (or NEMO-NORDIC should be consistent all along the manuscript). It looks like you could move some cores from NEMO-MED to NEMO-NORDIC to make it faster and at least decrease the waiting time in COSMO-CLM. Any comments on that? and on the scalability of the system?

Will et al. 2017 show that the optimized computational performance of the coupled system is weakly dependent on the computing architecture or on the individual model components but strongly depends on the coupling method. It is true that some NEMO-MED cores could be transferred to NEMO-NORDIC, but one has to pay attention to the number of cores that each node has, because the interconnexion of nodes is slower (and so it penalizes more) than the connection of cores within the same node. Therefore, it is optimal to choose per model a number of cores that is a multiple of the number of cores per node, like we did.

- I find surprising the 3.6 ratio between the atmosphere only and coupled since atmo- sphere is running on the same number of cores and there is not that much wainting time. How do you deal with the I/O? Could it be some latency due to the writing of the other models outputs?

The I/O is done in parallel and effectively. Adding the marginal seas implies 523x619x56 (=18.129.272) grid points for the Baltic and 264x567x75 (=11.226.600) grid points for the Mediterranean, compared to the 226x232x40 (=2.097.280) grid points of CCLM. The NEMO models have a lower output frequency than CCLM, since NEMO writes output only every 5 days.

- in the analysis of the SST, a comparison is done between the result of the RCSM simulation and the prescribed SST coming from a global similation at low resolution. In my opinion, the comparison not only shows the impact of retroaction through coupling of atmosphere and ocean models but also the impact of using differents models and resolution. Should we not compare the SST from an ocean-only simulation forced by a COSMO-CLM run (driven by a prescribed SST) and the SST from the RCSM? Could you state something about this point?

We agree that the prescribed SST coming from the global simulation has lower resolution. To know more about the impact only of the coupling (and not affected by the resolution), we would like to compare the SST coming from an ocean-only NEMO simulation and the SST of our coupled system. However, there is no available ocean-only simulation with NEMO covering the whole century. Nevertheless, we might have access to an ocean-only NEMO simulation over the North and Baltic Seas that run for a couple of decades, so we will start with a comparison there.

- Do you have any hypothesis regarding the various bias we observe?

The bias in the SST in the Mediterranean is seasonal dependent (colder in winter, warmer in summer), and many different factors could affect it (ocean initialization, higher mixing layer depth, aerosols blocking radiation, internal NEMO dynamics, etc.). We have analyzed the mixing layer depth looking for some answers, but did not succeed. To understand better the coupled system, we have run sensitivity tests changing the SST. Those results are summarized in a recently accepted publication (Kelemen et al. 2019). Nevertheless, we will continue looking for answers that explain those biases.

- about density histograms, at first it was not obvious to me that the dark pink was in fact the blue lying behind the light pink

This has been better explained in the legends of the new version.

- could you mention the convention used for the boxplots?

We represent the box-and-whisker plot computed with R using the default method. The box is defined by the 25th, 50th (median) and 75th percentiles. The whiskers are the lowest value still within the 1.5 IQR of the lower quartile, and the highest the value still within the 1.5 IQR of the upper quartile. The points are outliers.
Convention included in the new version of the manuscript.

- for figures of differences, could you precise the sign of the difference (simulation - observation, for instance)

Simulation-Observation referred to simulation minus observations, and differences between coupled and uncoupled referred to Coupled minus uncoupled. Nevertheless, we are more precise in the new version.

- I find interesting the analysis of extreme events, but you could perhaps show less figures or indices and make same assumptions explaining the better behaviour of the coupled model, and why on some indices and not on others?

There are 27 recommended indices, however, we had to choose some of them to avoid a very long paper. We considered monthly max and min values of the daily max and min temperature, as well as dry/wet spells, since those are relevant and describe well extreme events showing some impact of memory in the climate system and additionally having an impact on human lives. We have presented these indices for some German stations were the complete century of data was available. However, explaining the mechanism affected by the coupling that lead to changes in the local max and min temperature in those stations is not trivial, that is why the paper just focus on a description of how the indices behave, rather than in the explanation why this happens like this. Nevertheless, we improved the description and interpretation of the q-q plots.

A regional atmosphere-ocean climate system model (CCLMv5.0clm7-NEMOv3.3-NEMOv3.6) over Europe including three marginal seas: on its stability and performance

Cristina Primo[1], Fanni D. Kelemen[1], Hendrik Feldmann[2], Naveed Akhtar[3], Bodo Ahrens[1]

[1]Institute for Atmospheric and Environmental Sciences, Goethe University, Frankfurt am Main, Germany,
[2]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany,
[3]Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany,

Correspondence to: Cristina Primo (primoram@iau.uni-frankfurt.de)

Abstract. The frequency of extreme events has changed, having a direct impact on human lives. Regional climate models help us to predict these regional climate changes. This work presents an atmosphere-ocean coupled regional climate system model (RCSM, with the atmospheric component COSMO-CLM, and the ocean component NEMO) over the European domain, including three marginal seas: the Mediterranean, the North and the Baltic Seas. To test the model, we evaluate a simulation of more than one hundred years (1900-2009) with a spatial grid resolution of about 25km. The simulation was nested into a coupled global simulation with the model MPI-ESM in a low-resolution configuration, whose ocean temperature and salinity were nudged to the ocean-ice component of the MPI-ESM forced with the NOAA 20th Century Reanalysis (20CR). The evaluation shows the robustness of the RCSM and discusses the added value by the coupled marginal seas over an atmosphere-only simulation. The coupled system runs stable for the complete 20th century and provides a better representation of extreme temperatures compared to the atmosphere-only model. The produced long-term dataset will help us to better understand the processes leading to meteorological and climate extremes.

1 Introduction

Regional climate affects directly human lives and the socio-economic conditions. The natural variability of the climate system impacts local weather. Due to the recent changes in the frequency and intensity of local extreme events (Tebaldi et al., 2006; Hartmann et al., 2013; Casanueva et al., 2014), like storms or heavy rainfall, we aim at a better understanding of the climate system dynamics. The main components of the Earth climate system are the atmosphere, land, ocean and rivers. To have a better representation of the interactions between the atmosphere and the rest of components of the Earth climate system, it would be necessary to couple models representing all components. However, this is highly complex since it requires combining different numerical models, what may not only bring instabilities, but it also implies high computational costs, etc.. Therefore, current coupled climate systems focus only on a reduced number of these components. Since the oceans are the main boundary of the atmosphere (they cover 71% of the Earth's surface) and their critical role regulating energy flows (they have an enormous

heat storage and transport capacity), coupled ocean-atmosphere models have been developed to better understand the interactions between ocean and atmosphere. For example, the WCRP's Working Group on Coupled Modelling (WGCM) established the Coupled Model Intercomparison Project (CMIP) as a standardized experimental protocol for studying the output of coupled atmosphere-ocean general circulation models (AOGCMs) [https://cmip.llnl.gov/index.html]. However, the

5 coarse-resolution of these models does not resolve important physical processes that take place at local and regional scales and that are relevant to understand extreme events like warming and precipitation trends changes. For example, marginal seas are not well represented in general circulation models (Somot et al., 2008; Li et al., 2006). In addition, it has also been demonstrated that the simulated Sea Surface Temperature (SST) has a large spread when comparing an ensemble of AOGCMs (Dommenget, 2012) and that GCM simulations tend to underestimate the high precipitation intensities (Sun et al., 2006). On the other hand,

10 there are very high-resolution process-oriented models, like those used to forecast fog or winter storms (e.g. the Weather Research and Forecasting model, WRF, High Resolution Limited Area Models, HIRLAM, or the High-Resolution Window Forecast System, HIRESW) that resolve specific smaller-scale physical processes, but the computational cost is unaffordable to run a long simulation or they miss interactive coupling with some climate system compartments (especially the marginal seas). Therefore, Regional Climate Systems Models (RCSMs) present as an appropriate tool to improve the spatial scale

15 compared to global models, but keeping an affordable computational cost compared to high-resolution process-oriented models.

Within the European region, different atmosphere-ocean-ice coupled RCSMs have been already run for shorter periods (a few decades). For example, Schrum et al. (2003) coupled the regional model REMO (Jacob and Podzun, 1997) and the ocean model HAMSOM (Schrum, 1997) to analyse the North and Baltic Seas, showing the improvements compared to running the

20 uncoupled HAMSOM version. Pham et al. (2014) coupled the regional model COSMO-CLM (Rockel et al., 2008) to the ocean model NEMO (Madec, 2011) for the Baltic and North Seas to evaluate the impact of these seas on the climate of Europe. They showed that the 2m air temperature high biases presented when compared to observations were of the same magnitude as other COSMO-CLM studies and smaller than for the uncoupled version. Sevault et al. (2014) described and evaluated a fully coupled regional climate system model (CNRM-RCSM4) dedicated to study the Mediterranean climate variability over the period 1980

25 to 2012, showing a good agreement between the model and observations (e.g seasonal cycle and the interannual variability of SST, sea level, water budget, etc.). In a recent study, Obermann et al. (2018) coupled CCLM with the NEMO setup for the Mediterranean (NEMO-MED12) over the Med-CORDEX domain with ERA-Interim as the driving data. They showed that the coupled system was mostly able to simulate Mistral and Tramontane events with smaller biases than ERA-Interim. Akhtar et al. (2017) used that system to show the impact of the horizontal grid resolution and the dynamic ocean coupling of the

30 NEMO-MED in climate simulations with the COSMO-CLM during the period from 1979 to 2009. However, all these studies focus only on a few decades and extreme events have long return periods, so long term simulations are more appropriate to better represent and analyse them. So far, no long-term simulation of more than one hundred years with a regional climate coupled system is available. Hence, one of the goals of this work is to fill this gap.

Our aim is to improve our understanding about the regional climate change in Europe and what is the added value by coupling three marginal seas (the Mediterranean, the North and the Baltic Seas). Therefore, this work presents an atmosphere-ocean RCSM over Europe with an atmospheric horizontal grid resolution of about 25km, and tests its stability and performance with a simulation of more than one hundred years. The added value of the coupling is analysed by comparing our simulation with

5 a centennial atmosphere-only model run. A description about the extra costs due to the coupling compared to an atmosphere-only system is also included. We have particular interest in better understanding changes on extreme events, like heat/cold waves and extreme precipitation, therefore, special focus is placed to analyse the performance of the system representing extremes compared to the atmosphere-only model.

The paper is structured as follows: Section 2 presents the regional climate system models used in this work, namely an

10 atmosphere-only model and an atmosphere-ocean coupled model. Section 3 presents the methods and reference data used to show the stability and performance of the coupled model. Section 4 evaluates the models, distinguishing the impact of the coupling over the ocean and the European continent. Special attention is given to describe the evolution of climate change indices during the last century. Finally, Section 5 includes a summary with the main conclusions of the study.

2 Regional climate system models

15 This work presents an atmosphere-ocean coupled RCSM and compares it with an atmosphere-only version. This section describes the details about the different components of the RCSM: the atmospheric model, the ocean model, their set-up (lateral and boundary conditions) and how the coupling in the atmosphere-ocean system was done.

2.1 Atmospheric model

The three-dimensional non-hydrostatic limited-area atmospheric prediction model COSMO of the German Weather Service

20 has a climate version, the COSMO-CLM (CCLM; Rockel et al., 2008). This land-atmosphere regional climate model is based on primitive equations and accounts a variety of physical processes by parametrization schemes (see Doms et al., 2011). In our experiment, we used the Tegen et al. (1997) aerosol climatology, the Ritter and Geleyn (1992) radiation scheme, a turbulent kinetic energy (TKE) scheme for vertical turbulence (Raschendorfer, 2001), a reduced one-moment cloud scheme following Seifert and Beheng (2001) and a convection parameterization following Tiedtke (1989). COSMO-CLM includes the soil and

25 vegetation model TERRA, that provides soil temperature and water content (Schrodin and Heise, 2002). The atmospheric model version used in this study is the CCLM v5.0 clm7 with a numerical time step of 150 seconds and with a third order Runge-Kutta numerical integration scheme. A sub-grid scale sea ice mask was implemented in the CCLM coupled configuration over the North and Baltic Seas to have a better representation of the sea ice by accounting for partially sea ice covered grid boxes.

30 In this study's set up, the atmospheric lateral and top boundary condition were provided by a simulation with the earth system model of the Max-Plank Institute (MPI-ESM, version 6.1, Stevens et al., 2013). This MPI-ESM simulation was nudged (via

ocean temperature and salinity) to a simulation with MPI-ESM's ocean component MPIOM (Jungclaus et al., 2013) which was forced by NOAA's atmospheric 20th Century Reanalysis (20CRv2, Compo et al., 2011) as described in Müller et al. (2015). Müller et al. ran three members, and we considered the first one (as20ncep08_r1i1p1-LR). This indirect nesting of CCLM into the 20th Century Reanalysis was necessary because of need of consistent lateral boundary data for forcing the
5 marginal seas in the RCSM.

This work compares the atmosphere-only CCLM model with an atmosphere-ocean RCSM. In the coupled version, the prescribed SST of the CCLM over the regional oceans, as well as the fraction of sea ice in the Baltic and North Seas, were replaced by the SST and sea ice fraction as simulated by coupled marginal ocean models presented in the following section, whereas the ocean models received information from the atmospheric model about the momentum and freshwater (evaporation
10 minus precipitation), winds, solar energy and non-solar heat flux. The MPI-ESM simulation drove both, the atmosphere-only and the atmosphere-ocean RCSMs. Hence, the SST of the atmosphere-only system was prescribed with the nudged MPI-ESM SST. There was no tuning in the coupled version, thus the configuration of the atmospheric model was the same in the coupled and uncoupled versions.

Within the Coordinated Regional Downscaling Experiment (CORDEX), a choice of different domains covering the land
15 around the world were defined. Our study aimed to better understand the regional climate of central Europe, therefore, our simulations applied the so-called EURO-CORDEX domain (http://www.cordex.org/domains/cordex-region-euro-cordex/, see Figure 1 for a representation), with a horizontal grid-spacing of 0.22° x 0.22° (~25km, 226x232 = 52432 grid points) and 40 vertical levels.

2.2 Ocean model

20 The Nucleus for European Modelling of the Ocean (NEMO) is a flexible tool for studying the interactions of the ocean with the atmosphere over a wide range of space and time scales. Within NEMO, the ocean is interfaced with a sea-ice model (LIM or CICE), passive tracer and biogeochemical models (TOP). High-resolution configurations are available for the regional oceans in the European domain. For example, Beuvier et al. (2012) developed MED12, a regional version of the NEMO ocean engine on the Mediterranean Sea. In our simulation we used NEMO-MED12, based on NEMO version 3.6, with a resolution
25 of 1/12° (~0.083°~9km, 264x567 = 149688 grid points), 75 vertical levels and with a numerical time step of 720s. The initial conditions for three-dimensional potential temperature and salinity were provided by the MEDATLAS-II (Rixen, 2012) mean monthly climatology (1945-2002) in the Mediterranean Sea. The sea model was spun up in coupled mode during 20-years driven by randomly resampled MPI-ESM years in the period 1900-1910. The Black Sea and river runoff water input were prescribed from the climatological average of interannual data from Ludwig et al. (2009). Water exchange of the in good
30 approximation closed Mediterranean ocean basin with the Atlantic Ocean was relaxed to the Levitus et al. (2005) climatology prescribed in the buffer zone.

Another NEMO set-up has been adapted to reproduce the barotropic and baroclinic dynamics, as well as the thermohaline structure, of the Baltic and North Sea basins. This is the so-called NEMO-NORDIC (Hordoir et al., 2018), whose ocean component is coupled to the sea ice model LIM3 (Vancoppenolle et al., 2009). In our study we used a NEMO-NORDIC version based on NEMO 3.3 (Dieterich et al., 2019; Gröger et al., 2019), including the LIM3 sea ice model, with a resolution 2' (~0.03°~3km, 523x619 = 323737 grid points), 56 vertical levels and a numerical time step of 180s. The initial conditions for three-dimensional potential temperature and salinity were provided by Janssen et al. (1999) and further balanced by a spin-up simulation of the period 1900-1905. The lateral boundary conditions in the North Sea were derived from the MPI-ESM simulation. Freshwater river inflow was provided from daily time series of the E-HYPE model output (Lindström et al., 2010). Neither in the NEMO-MED12 nor in the NEMO-NORDIC simulation any drift in the surface variables SST and sea surface salinity following balanced initialization was detectable. Figure 1 presents the domains where the NEMO-MED and NEMO-NORDIC models run.

2.3 Coupling

In the atmosphere-ocean climate system, the atmospheric model CCLM was coupled with two configurations of NEMO: one adapted to the Mediterranean Sea (NEMO-MED12) and one to the Baltic/North Seas (NEMO-NORDIC). The coupling was done every three hours through a fully parallel communication between parallel models executed via the Model Coupling Toolkit library (MCT; Jacob et al., 2005), named the OASIS3 Model Coupling Toolkit (OASIS3-MCT; Craig et al., 2017), since this library has already been successfully used to couple the CCLM model with NEMO (Will et al., 2017). This is an interface included in CCLM based on the Message Passing Interface (MPI). It has been proved that including this library significantly improves the performance over the previous version OASIS3, because the bottleneck due to the sequential separate coupler is entirely removed (Gasper et al., 2014). During the coupling, the data on the ocean coupling grids were interpolated to the CCLM grid. At runtime, all CCLM ocean grid points located inside the interpolated area were filled with values interpolated from the ocean model and all CCLM ocean grid points located outside the interpolated area were filled with the same external forcing data as the uncoupled system. The coupled set up consisted of CCLM sending information to NEMO about the solar energy, non-solar heat, momentum and freshwater fluxes, whereas it received SST from NEMO. In addition, CCLM sent the sea level pressure to NEMO-NORDIC and received the sea ice fraction. A more detailed description of the coupling strategy and its implementation can be found in Will et al. (2017) and Akhtar et al. (2019).

In addition, OASIS3-MCT offers a performance analysis tool, the LUCIA tool (Maisonnave and Caubel, 2014), that measures how much time each system component spends doing its own calculations (incl. send and receive operations, as well as time needed for interpolation of fields) and how much time it waits for information coming from the other components. This tool allows an optimization of computing resources and of the scaling of each model of the coupled system. We used the LUCIA tool to find an optimal distribution of the available number of cores used for the computation, having in mind that the model with the highest number of grid points of our system is the NEMO-NORDIC (8.6 times more grid points than CCLM). Figure 2 shows an example of a configuration using 11 nodes with 36 CPUs each of MISTRAL, the High-Performance Computing

system for Earth system research (HLRE3) at the German High-Performance Computing Centre for Climate and Earth System Research, Germany. We assigned three nodes to CCLM, seven to NEMO-NORDIC and one to NEMO-MED. Therefore, we assigned 3.6 times more compute resources to the coupled system than to the non-coupled system. Like this, only NEMO-MED had to wait for the exchange of the other models, while the other two models required about similar times for the calculations. Figure 2 refers to the time used to send/receive operations and interpolation. To have a broader picture about the costs due to the coupling, we calculated how long did it take to run just one day (saving the same list of CCLM variables) considering two different alternatives: (a) assuming that the number of compute resources was fixed and (b) assuming that more compute resources could be used for the coupling. In the first case, both coupled and uncoupled simulations ran in 11 nodes (CCLM ran in 3 nodes in the coupled system). Like this, the coupled simulation ran in around five minutes whereas the uncoupled in around one minute. In the second case, CCLM ran in 3 nodes for both coupled and uncoupled simulations. In this case, it took around two minutes to run the uncoupled system. Therefore, for this example, the coupled system was around 5 times slower given the same number of available nodes, but around 2.5 times slower when more resources were used.

For the centennial simulation, we used 576 CPUs optimally distributed as follows: 24x13 CPUs were assigned to CCLM, 12x8 to NEMO-MED12 and 14x12 to NEMO-NORDIC. With this configuration, each simulated month required around one and a half hours in total, what implied 78 days to run the complete centennial simulation (110 years). To obtain the optimized computational performance of a coupled system, Will et al. (2017) show that the coupling method plays a higher role compared to the computing architecture or on the individual model components.

3 Methods and reference data

Our aim was to test whether the coupled system runs stable over the whole century and whether the coupled simulation including the hydrosphere component represented by the Mediterranean Sea and the North and Baltic Seas improves not only the global MPI-ESM-LM simulations, but performs at least as well as the atmosphere-only model.

3.1 Methods

The stability of the coupled atmosphere-ocean RCSM was tested with a spatio-temporal analysis of a centennial simulation (1900-2009). The analysis consisted of a study of the temporal series evolution, annual cycles, spatio-temporal density distributions and spatial patterns of three variables of interest: sea surface temperature, the 2m air temperature and the total precipitation. Results were compared to the same analysis obtained with an atmosphere-only (CCLM) version simulation over the same period, run within the national research project on climate prediction MiKlip ("Mittelfristige Klimaprognosen", Marotzke et al., 2016). The temporal series analysis helped us to detect any bias or drift of the atmosphere-ocean simulation compared to the atmosphere-only simulation, and the spatial analysis to detect if the system behaves differently according to the area of interest.

Regarding the quality of the coupled model, we compared our simulation with different reference datasets (see next section for more details). Rather than a point-by-point comparison with the reference data, we would like to know if the system

represents well the reference's value distributions. For this purpose, we compared the density distributions and box-plots of our system with those obtained from observational datasets. We analysed the marginal seas separately, distinguishing also among seasons. Regarding the land, different relevant areas have been used in the literature for regional climate studies over Europe, e.g. within the European project PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining

5 EuropeaN Climate change Risks and Effects; Christensen, 2005) eight regions were defined: British Isles, Iberian Peninsula, France, Mid-Europe, Scandinavia, Alps, Mediterranean and Eastern Europe. Since we aim to improve our understanding of the regional climate in Germany, we showed results in the Mid-Europe PRUDENCE region.

We are also interested in high-impact phenomena: heavy precipitation, dry spells and heat waves. The joint CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI) suggested a list of 27 core

10 climate change indices based on daily temperature values and daily precipitation amounts (Karl et al., 1999; Zhang et al., 2011). The definition of these indices can be found on the webpage of the project (etccdi.pacificclimate.org/list_27_indices.shtml). We computed these indices, using the free available R package RClimDex, developed, maintained and provided by Xuebin Zhang and Yang Feng at Climate Research Division of the Environment and Climate Change Canada, that in addition to the computation of the indices, also provides simple quality control of the daily

15 input data. We also analysed how the distributions of these indices were represented compared to the distribution of the indices obtained with the observed dataset.

3.2 Reference data

Two centennial reference datasets were available: the gridded Climatic Research Unit (CRU) observation Time-series (TS) produced at the University of East Anglia for the period January 1901 – December 2016 (Harris et al., 2014), which consists

20 of monthly data at high-resolution (0.5°x0.5°) grids. In this work, version 4.01 data (CRU, 2017) is used. Our simulations had a higher resolution (0.22°x0.22°), therefore a necessary upscaling prevents us from validating the high-resolution information available in the model when comparing to CRU. However, there is no available higher resolution gridded dataset covering the complete century. If we wanted to compare model data with gridded observations with similar spatial resolution, we would have to consider shorter periods. For example, the gridded data E-OBS dataset (Haylock et al., 2008) is available on a spatial

25 resolution 0.22°. However, it covers only half of our period of interest (from 1950 onwards). It is worthy to remark that in any case none of these observational datasets are perfect, and that they also differ from each other. For example, Fig. 3 shows a comparison of the monthly mean in January 1995, when a flood event happened over Germany. The figure illustrates the information loss regarding the event through upscaling compared to the 0.22° resolution. This fact can penalize our system when comparing it with CRU, especially for the first half of the century, in which E-OBS data are not available. Nevertheless,

30 this will not affect the coupled and atmosphere-only model inter-comparison.

To compare our coupled data with centennial observations with higher quality, we considered historical daily station observations. The Climate Data Center (CDC) of the German national weather service "Deutscher Wetterdienst" (DWD) provides free access to quality-controlled observations of DWD climate stations (DWD-CDC, 2017). We took nine stations

with less than 15% of missing values, covering the complete period (1900-2009), and well distributed over Germany. Figure 1 shows two of these stations, with no data gap, and located at two different altitudes and distances to the seas: Potsdam (circle, altitude: 81m) and Hohenpeißenberg (triangle, altitude, 977m). For the sake of brevity, this work presents a comparison of the RCSMs only for these two stations, but similar conclusions were reached with the other seven stations.

5   Over the ocean, unfortunately, there is no high resolution observed data set for the complete period. Hence, we used the Hadley Centre sea Ice and Sea Surface Temperature data set (HadISST; Rayner et al. 2003) developed by the Met Office Hadley Centre for Climate Prediction and Research with a 1° resolution as reference. In addition, we compared the sea surface temperature over the Mediterranean with the NOAA Optimum Interpolation Sea Surface Temperature V2 (OISSTv2; Reynolds et al., 2002) for the available decades (1981-2009). We used these data even though it only covers a few decades, because the

10   sea surface temperature of the CCLM atmosphere-only simulation is not independent from the HadISST observations (they were used to obtained the MPI-ESM driving simulation). Besides the observations, we also compared the coupled system over the marginal seas with a multi-model ensemble consisting of the first member ('r1i1p1') of eight CMIP5 models.

4 Evaluation results

We based our analyses on three variables: the sea surface temperature, the 2m air temperature and the total precipitation. The

15   behaviour over ocean and land is presented separately. We name the atmosphere-only model (CCLM) uncoupled and the atmosphere-ocean model (CCLM-NEMO) coupled.

4.1 Sea surface temperature

We analysed the temporal evolution of the SST over the marginal seas (Mediterranean and Baltic/North Seas) to see if there is any drift or evolving bias in the coupled system over the ocean (Figure 4). The SST of the coupled version used the simulated

20   NEMO SST, whereas the SST of the uncoupled version was from the global system MPI-ESM. The long-term SST time series of our coupled system shows a stable system, although the annual mean SST values are colder than the observations (HadISST) and also than in the uncoupled system (from the global system) in both basins. The global system MPI-ESM-LM simulation is not independent from the HadISST observations, therefore, additionally the NOAA OISSTv2 was also included in the comparison for the available last three decades (1981-2009). Despite the cold bias, the regional coupled system follows the

25   evolution of the observed SST values. In the Mediterranean, it even matches the ensemble mean of the CMIP5 global simulations, and in the Baltic, it is within the spread of this ensemble. Therefore, the SST values from the coupled system have at least as good quality as the values of an average global model, with the advantage of having higher resolution, which preferably improves the model results especially in the land/sea transition zone. Improving the quality of the averaged global model was expected since the global circulation models contain ocean models that are not well suited to shelf seas like the

30   Baltic and North seas.

| Deleted: and took the stations that |
| Deleted: ed |
| Deleted: obtaining nine stations distributed |
| Deleted: We chose |
| Deleted: where |
| Deleted: happened during the period |
| Deleted: in Fig.1 |
| Deleted: located at |
| Deleted: at Fig.1 |
| Deleted: located at |
| Deleted: by analysing |
| Deleted: ice |
| Deleted: sea |
| Deleted: surface |
| Deleted: temperature |
| Deleted: -LR |
| Deleted: |
| Deleted: models from the |
| Deleted: simulations |
| Deleted: The |
| Deleted: - |

| Deleted: compared to |
| Deleted: SST of the uncoupled system comes from the |
| Deleted: , whose simulation is not |
| Deleted: . Thus |
| Deleted: an independent SST observation, namely the NOAA Optimum Interpolation Sea Surface Temperature V2, is |
| Deleted: for the Mediterranean as a |
| Formatted: Not Highlight |
| Deleted: of |
| Deleted: time |
| Deleted: and also the global |

Figure 5 shows a good representation of SST's annual cycle for both basins. In the Mediterranean, the coupled system is colder in winter and warmer in summer than observations and global simulation. In the Baltic and North Sea basin, the coupled system is colder throughout the year.

The density histograms of the three marginal seas summarize both the spatial and the temporal distribution of the SST values (Figure 6). Since the Baltic and North Sea have different climatologies in winter due to the presence of ice (the Baltic Sea is colder and less salty than the North Sea), we have analysed their histograms separately. In the Mediterranean Sea, the distribution of model data and observations have similar shape, namely a double maximum representing summer and winter temperatures. Comparing the modelled and the observed histograms, both the coupled and uncoupled models capture the main aspects of the SST distribution. The regional coupled simulation has a wider distribution than the observations, which is made up by a well-fitting upper range and a shift towards cooler temperatures at the lower range. In comparison with the uncoupled version, the coupled system represents better the upper extremes but has a colder bias in the lower tail. The distribution shape of the uncoupled dataset is very similar to the observation's shape, which was expected with the forcing SSTs constrained by the 20CR reanalysis and thus the observed SSTs. Still, the uncoupled model has a cold bias in both tails.

The SST of the coupled version in the Baltic and North Seas comes from the ocean NEMO-NORDIC model, that includes the sea ice and the freezing/melting processes via a sea ice model. This leads to an improvement in the lower tail of the distribution over the Baltic, compared to the uncoupled system that provides much colder temperatures. Both systems have a cold bias in the upper tail. Regarding the North Sea, the coupled model shows a colder bias compared to the uncoupled model.

The spatial distribution of the regional coupled system's SST bias shows that the modelled seas, as previously seen, are cooler than observations in all seasons (Figure 7, spring and autumn are not shown). Nevertheless, during winter the basin of the Baltic Sea has a warm bias and during summer the Mediterranean Sea has a gradient in the bias field from south to north. Explaining the SST bias of the coupled system compared to the uncoupled SST, is not straight forward and it is out of the scope of this study. Many factors may have an impact in the coupled SST (internal dynamics of NEMO, salinity changes, initialization of the ocean, deeper mixing layer depth, etc.). Nonetheless, given that the coupled system was not retuned, the results of the transient RCSM simulation are promising.

4.2 2m air Temperature

This section analyses how the coupled system propagates the interactive SST information into the atmosphere and over land, in particular the impact on 2m air temperatures. Figure 8 shows the differences of the 2m air temperature monthly mean between the coupled and uncoupled systems averaged during winter (a) and summer (b) for the period 1901-2009. The plots show differences up to almost 2.4°C. The coupled system gives colder temperatures over the Mediterranean Sea during winter, and warmer temperatures during summer, with the exception of the French coast and north-east coast (regions influenced by cold wind systems like the Mistral and the Meltemi). Regarding the Baltic and North Seas, the summer and winter difference patterns are similar. The coupled model provides colder temperatures over the North Sea and western parts of the Baltic Sea, whereas warmer temperatures over the north and eastern parts of the Baltic Sea. The boxplots (c) represent the distribution

over the marginal seas separately, Baltic-North and Mediterranean Seas, as well as over the land, for each season, over the whole period. The plot shows that the spread of the differences on the Baltic-North is similar during the year, and the coupled system is mainly colder. The highest spread of the differences happens in the Mediterranean in summer, where the coupled version is warmer. In winter the spread is smaller and the coupled system is colder. Regarding the land, in summer the median is about zero and there is very small spread, showing mainly no difference between the systems. However, there are some outliers where the coupled system shows higher temperatures (with up to 2°C difference). In winter the differences are slightly more noticeable, and the main outliers are negative, showing that the coupled simulation allows for colder temperatures.

Figure 9 shows the temporal evolution of the annual 2m air temperature averaged over the marginal seas. The figure shows that both, the coupled and uncoupled systems, represent a similar positive trend and strongly intercorrelated time series. To better understand how the 2m air temperature of the coupled system responds to changes in the SST, Kelemen et al. (2019) ran a few sensitivity experiments using perturbed SST in the uncoupled system. They showed positively oriented impact of SST disturbance on 2m air temperatures.

4.3. Total precipitation

Figure 10 represents the winter and the summer precipitation differences between the coupled and the uncoupled system. The largest differences are in the Eastern Mediterranean in the winter season, with large areas with 20 to 50 mm/month less precipitation in the coupled than in the uncoupled simulation. In summer, however, the coupled system gives more precipitation in most Mediterranean areas. Regarding the North Sea, the uncoupled simulation gives in general more precipitation than the coupled. The differences in the Baltic are smaller, being slightly more appreciable in summer than in winter. The precipitation differences over the seas are in concordance with the differences of 2m air temperature. Boxplots (Fig. 10c) show the monthly difference distributions over the marginal seas and land separately. The spread in the Baltic-North is higher in summer, with more outliers, whereas in the Mediterranean is in winter (with generally small monthly precipitation amounts in summer), with more negative outliers (higher precipitation for the uncoupled system). The differences over land are smaller in general, with large outliers in winter. The latter emerged near the Mediterranean coast, where the coupled system is drier, and in the Alpine region, where the coupled system is wetter.

To better understand how the total precipitation of the coupled system responds to changes in the SST, Kelemen et al. (2019) did sensitivity studies showing a higher response in total precipitation than in the 2m air temperature. They also showed an added value in the seasonal precipitation sums of the coupled system during winter over the eastern part of the domain.

4.4 Model-Observations comparison

This section studies the performance of the coupled system in Europe. For this purpose, we compared the model data (coupled and uncoupled systems) with the observed CRU dataset. Data coming from coupled and uncoupled systems were interpolated to the 0.5°x0.5° CRU-grid, and only those grid points defined in all three data-sets were considered. Errors of the 2m air temperature of the coupled and uncoupled systems when compared to the CRU observations for winter and summer, the 2m

air temperature distributions and the distributions of the 2m air temperature errors are shown in Figure 11. In winter, there is not clear positive or negative bias (Fig11a-b). However, in summer the coupled and uncoupled systems are colder than the CRU observations, apart from the Alpine region and the south-east area, where both systems are warmer (Fig11c-d). Boxplots show a good representation of the observed distribution (Fig11e). In winter distributions are similar and the main differences appear in summer, where the systems show slightly colder values in the low temperature range. The differences also show similar distribution for coupled and uncoupled systems (Fig11f). As mentioned in Sec. 4.2, the 2m air temperature differences between coupled and not-coupled systems are below 2.5°C, however boxplots show that the differences compared to CRU are much higher, up to 10°C in winter. In summer more than 75% of the 2m air temperature given by the systems is colder than the observations. In winter there is no clear bias, and boxplots are centred around the zero value. Nevertheless, there are more extreme higher values (longer upper tail in winter showing higher temperatures for the systems compared to the observations).

We are interested in the impact that the coupling may have in the 2m air temperature performance over Europe, in particular, over the PRUDENCE region named Mid-Europe. Figure 12 represents boxplots corresponding to the annual cycle of the monthly 2m air temperature averaged over Mid-Europe for the 20th century (a) and the distributions of the differences of the model values minus the CRU observations (b). Both systems show similar distributions than the CRU data in winter, however colder distributions in summer. The differences are centred around the zero value in winter and below zero in summer, that is, on average the winter is better represented by the regional systems than the summer. However, the spread is much bigger in winter than in summer, that is, in those cases in which the regional systems differ from the CRU observed data, the differences are higher in winter than in summer. Compared to the complete domain (Fig. 11e), the spread of the distributions (box height) over Mid-Europe is smaller than over the whole domain, since over Mid-Europe the temperatures do not differ as much as they do, when comparing the temperatures over the southern and northern parts of the whole domain. In addition, the extreme cases (points) are also milder compared to the whole domain since the summers are not as hot as in southern regions like the Iberian Peninsula or north Africa, and winters are not as cold as in northern regions. Fig. 12 also points out winter outliers. In this case, for example, the coupled system estimates better the coldest temperatures in January over Mid-Europe. This is a result that was not appreciated when comparing the whole domain. However, the boxplots do not give us detailed information about the different 2m air temperature values. To analyse in more detail the differences in the tails of the distributions, Fig. 13 shows the density histograms of the 2m air temperature of the coupled and uncoupled systems compared to the CRU data over Mid-Europe. Blue bars represent the uncoupled system, red bars the coupled system, white bars the CRU observations and purple the intersection. As shown in the left tail of the winter distribution, the coupled system estimates well the values around -5°C although over estimates those around -8°C. Nevertheless, both systems show a good fit in winter although a colder bias in summer.

4.5 Extreme events

The effect of coupling the marginal seas has been shown to be a useful tool to simulate regional climate over Europe and study extreme events, like Vb events during the period 1979-2014 (Akhtar et al., 2019). In this section we will focus on the

Deleted: -

Deleted: 2m-temperature

Deleted: smaller

Deleted: 2m-temperature

Deleted: 2m-temperature

Deleted: To compare the annual cycle of both systems with the CRU data, f

Deleted: 2m-temperature

Deleted: the PRUDENCE region namely

Deleted: ¶
We have also analysed the spatio-temporal distribution to see if the behaviour is different from the averages. Figure

Deleted: 2m-temperature

Deleted: and

Deleted: B

Deleted: and

Deleted: 3

Deleted: cases

representation of climate indices during the whole 20th century. The monthly CRU data cannot be used to analyse extreme events like heat/cold waves or dry/wet spells. Instead, observed station data provided by the German Weather Service (DWD-CDC, 2017) were considered in this study. We computed core climate change indices suggested by the ETCCDI over the 20th century for the coupled and uncoupled systems, as well as for long term series of station data located in Germany. We chose indices having an impact on human lives, for example, Fig. 14 shows the temporal evolution of four climate change indices related to extreme temperatures: annual minimum temperature TNn, annual maximum temperature TXx, information on warm spells TX90p (defined as the percentage of days when the maximum temperature is above the calendar day 90th percentile centred on a 5-day window for the base period 1961-1990), and finally information on cold spells TN10p (defined as the percentage of days when the minimum temperature is below the calendar day 10th percentile centred on a 5-day window for the base period 1961-1990). To compute the 10th and 90th percentiles for each calendar day, a bootstrap procedure was used to avoid possible inhomogeneity across the in-base and out-base periods (Zhang et al., 2005). Linear trends of the indices are given too. Figure 14 shows the stable evolution of the indices in the coupled version, the capturing of the trends and the improvement of the uncoupled version for the TNn and TXx indices, especially for the higher station. The coupled system detects the increase of temperature during the century, as well as the increase of the percentage of days with maximum temperatures above the 90th percentile, and percentage of days with minimum temperature below the 10th percentile.

Figure 15 compares the distributions of these indices for the coupled and uncoupled systems against the observations based on the quantiles. The diagonal represents the perfect case, assuming that observations are perfect. The closer to the diagonal, the better the simulated statistics of the considered extremes. Lines parallel to the diagonal show similar distributions to the observed one (e.g. TNn, TXx), whereas lines not parallel show differences in the spread and tails (e.g. upper tail of the uncoupled TN10P in Potsdam and the coupled TN10P in Hohenpeißenberg). The panels show that the coupled system corrects the overestimation of minimum temperatures of the uncoupled system, as well as the underestimation of the maximum temperatures. Therefore, the coupling has a positive impact with respect to extreme temperatures. Regarding the percentage of days above the 90th percentile and below the 10th percentile, the coupled version fits the observed distribution similarly to the uncoupled version, but improves the extreme quantiles.

Regarding precipitation indices, we focused on the following annual indices (Fig. 16): total precipitation PRCPTOT, total precipitation R95p when the daily precipitation (RR) is above the 95th percentile of precipitation on wet days in the 1961-1990 period, maximum length of dry spell CDD (maximum number of consecutive days with RR < 1mm), and maximum length of wet spell CWD (maximum number of consecutive days with RR ≥ 1mm). For the precipitation indices, the uncoupled system proved being in general more skilful. The coupled version overestimated the precipitation in Potsdam, but underestimated it in Hohenpeißenberg. Nevertheless, the coupled system shows a stable evolution. Figure 17 shows the precipitation indices' quantile-quantile plots. The distribution of the uncoupled system's simulation show a better performance than the coupled simulation. Note that the simulated data were not bias corrected and the large evaluation uncertainties because of observational uncertainties and point-to-area comparison. In this example, lines are not as parallel to the diagonal as in Fig.15, showing wider total precipitation distributions than the observed one in Potsdam, but more localized than the total precipitation observed

distribution in Hohenpeißenberg. Since all temporal series (model simulations and observations) have the same length, then observed and model percentiles represent the same number of cases. Each point of the q-q plot represents a 10% of the total number of cases. Hence, we can also analyse and direct compare the frequency of total precipitation events over a particular intensity. Let us focus on the first plot (PRCTOT in Potsdam). The number of points above a horizontal line over the intensity of interest will indicate the frequency of the estimated cases above this threshold given by the model. The number of points on the right part of a perpendicular line over the intensity of interest will indicate the frequency of the observed cases above that threshold. Vertical and horizontal lines in the plot correspond to a threshold 700mm/year in Potsdam. The coupled model always estimates total precipitation above this threshold, the uncoupled model in 80% of the cases (8 points are above the horizontal line), whereas it was only observed only 20 % of the cases (only two points of the lines are on the right part of the vertical line).

5. Summary and Conclusions

To better understand how the Earth climate system evolves at local to regional scales, it is necessary to gain a better understanding of the interactions among the different components of the system. This work presents an atmosphere-ocean coupled regional climate system model (RCSM) over Europe including three marginal seas: the Mediterranean, the Baltic, and the North Seas. The coupled system was tested by evaluating a centennial simulation (1900-2009) over the EURO-CORDEX domain on a 0.22°x0.22°grid (grid spacing ~25km). The atmospheric component was given by the COSMO model in climate mode (COSMO-CLM) and the lower boundary conditions over the sea surfaces by coupling CCLM to two NEMO ocean model set-ups, one for the Mediterranean Sea (NEMO-MED12) and one for the Baltic and North Seas (NEMO-NORDIC). The coupling was made through the OASIS3-MCT coupler. For the lateral and top boundary conditions, the regional atmosphere was forced by the global Earth System Model MPI-ESM, whose ocean was nudged to an MPI-ESM's ocean-ice component simulation forced with NOAA 20th Century Reanalysis (20CRv2).

Our aim was to know if the atmosphere-ocean coupled RCSM runs stable within one hundred years and what the cost and benefits of coupling the marginal seas are. We first analysed the computing costs (in terms of resources and time consumed) due to the coupling, showing that 3.6 times more of the resources are required to run the same period or that the coupled version is 5 times slower using the same amount of resources compared to an atmosphere-only version (only CCLM). To test the stability of the system during the 20th century, we did an analysis on three variables: sea surface temperature (SST), 2m air temperature and precipitation. Results show that the system runs stable over the whole century, with no drift nor evolving bias. Finally, we evaluated the performance of the coupled RCSM compared to a centennial simulation of the atmosphere-only version, as well as to observations (CRU data and DWD-station observations). We cannot conclude that one system is better than the other, since the results depend on the variable, area, and season of interest, as explained below.

This study includes a spatiotemporal analysis of the sea surface temperature (SST) of the coupled system (provided by the NEMO ocean-model) over the Mediterranean, North and Baltic Seas, as well as a comparison with the SST of the atmosphere-only version (prescribed with the MPI-ESM SST) and SST observations (HadISST and in case of the Mediterranean Sea also

OISSTv2). Results show a stable and realistic evolution of the SST over the century, with a cold bias compared to observations, but performing similar to the ensemble mean of a global atmosphere-ocean coupled ensemble system. This means that the coupled system provides SSTs on a higher resolution with the added value of preserving the spatial and temporal dynamics (the ensemble mean is not a realization of the system, but an average). The SST annual cycle is well represented with a in

5 general larger amplitude than with the uncoupled system. In winter, the coupled system shows a cold bias in most of the Mediterranean Sea and in the North Sea, whereas a warmer bias in the Baltic Sea and the western part of the Mediterranean. In summer, it shows mostly a cold bias in the three marginal seas, except in the southern part of the Mediterranean sea, that shows a positive bias. It is not straightforward to isolate any causes of the SST biases, since many factors may affect the SST in the coupled system (e.g. internal dynamics of the ocean, mixing layer depth, ocean initialization, etc.). This is planned in

10 future studies. Nevertheless, given that the oceans in the coupled simulation are not constrained to SST observations as in the uncoupled simulation, the results shown in this manuscript are very promising.

Regarding the 2m air temperature, the biases over sea of the coupled RCSM follows the SST biases. Over land differences are smaller on average, but with larger near-coastal differences. Coupled and atmosphere-only systems show a negative bias in summer months, whereas a better representation over the winter, compared to the 2m air temperature of the CRU data.

15 However, even though in general these errors are smaller in winter, the most extreme errors also occur in winter. Hence, the spread in differences is higher in winter than in summer. A comparison of the 2m air temperature annual cycle and the spatio-temporal density distributions within the 20th century over the PRUDENCE area namely Mid-Europe is included to show that this behaviour occurs during the whole period.

Regarding the total precipitation, the same pattern as the 2m air temperature is shown: the higher the 2m-temperature, the more

20 precipitation given by the systems. Thus, the coupled system provides less precipitation in the Mediterranean Sea than the uncoupled system during winter, and more during summer. In the Baltic and North seas, the coupled system gives in general more precipitation than the uncoupled during both seasons. Over land, the differences are smaller than over sea, apart from those near the Mediterranean coastline in winter, where the coupled system is drier compared to the uncoupled, and in the Alpine region, where the coupled system is wetter compared to the uncoupled.

25 Special focus was given to the analysis of extreme events. Since this study requires higher temporal precision than the monthly values provided by the CRU data, DWD station observations with daily resolution were considered. The evolution of some climate change indices was presented and discussed, showing that over Germany the coupled system is stable and improves the values of the climate change indices related to extreme temperatures compared to the uncoupled version. However, the precipitation extremes at the studied German stations was better represented in the uncoupled system.

30 To conclude, the centennial atmosphere-ocean coupled simulation presented in this work provides valuable information about the local climate in Europe. Having such a long temporal series of a stable atmosphere-ocean coupled system, whose spatial resolution is higher than the global models, helps us to improve our knowledge of the local phenomena, especially for extreme events that have longer return periods. It has been shown that coupling the ocean improves the representation of heat and cold waves over some German stations. Our centennial run can also be used to investigate the interactions among different variables

Deleted: er

Deleted: er

Deleted: compared to the uncoupled system

Deleted: er

Deleted: ing

Deleted: -half

Deleted: ¶

Deleted: -

Deleted: increases the 2m-temperature over the Mediterranean Sea in summer and decreases it in winter compared to the atmosphere-only version. It also decreases the 2m-temperature over the North Sea for both seasons, whereas it increases it over the Baltic Sea.

Deleted: the extreme differences are as high as those found over the sea.

Deleted: 2m-temperature

Deleted: , whereas a better representation over the winter

Deleted: cases

Deleted: of the

Deleted: distribution

Deleted: 2m-temperature

Deleted: 2m-temperature

Deleted: the

Deleted: the

Deleted: happening in the

Deleted: observations coming from

Deleted: -

Deleted: s

Deleted: hourly

Deleted: i

Deleted: uncoupled

Deleted: represents in a

Deleted: way

Deleted: the climate change indices related to precipitation.

Deleted: proved

on a regional scale and to learn more about the atmospheric drivers that lead to extreme events. In addition, having in mind that the E-OBS dataset covers only from 1950 onwards and there is a lack of observations during the first half of the century, these downscaled data might help us to know more about this period, as well as to improve our knowledge about the advantages and deficiencies of our decadal predictions over Europe. Finally, the here investigated RCSM can be used to improve our

5   knowledge about the future climate change in Europe, e.g. to simulate decadal predictions or climate projection ensembles, especially, in those areas in or close to the marginal seas. Examples of similar studies are Pham et al. (2018), where the added skill by coupling the Baltic and North Seas in decadal predictions is analysed, or Damarkis et al. (2019), where the future evolution of marine heatwaves in the Mediterranean Sea is studied. The here presented results indicate that the coupled RCSM not only provides more information but also provides better regional projections after retuning than uncoupled RCMs which

10   have to rely on coarse-gridded global SST projections.

Code and data availability

This paper describes an atmospheric-ocean coupled CCLM-NEMO system. The atmospheric CCLM model source code is freely available for scientific usage by members of the CLM community (www.clm-community.eu), a network of scientists who accept the CLM community agreement. To become a member, please contact the CLM community coordination office at

15   DWD, Germany (clm-coordination@dwd.de). The ice-ocean component set up for the North Sea and Baltic Sea (NEMO-Nordic) is released under the terms of the CeCill license (http://www.cecill.info, last access: 3 May 2019). It uses NEMO 3.3.1 with some changes and its code is available in the zenodo archive (https:/dx.doi.org/10.5281/zenodo.2643477). The ocean component set up for the Mediterranean (NEMO-MED) uses NEMO 3.6. and its code is available at https://prodn.idris.fr/thredds/fileServer/ipsl_public/rron960/NEMO_MED_v3.6.tar. The OASIS3-MCT coupling library can

20   be downloaded at https://verc.enes.org/oasis/. Data presented in this work is also available for research purposes in the zenodo archive (10.5281/zenodo.2659205).

Author contributions

C.P. performed the coupled simulation. H.F. performed the uncoupled simulation. C.P. and F.D.K. evaluated the simulations. N.A. helped in the coupled system set-up and B.A. conceived the experiment and contributed to the coupled system design.

25   All authors discussed the results. C.P. took the lead in writing the manuscript with feedback and comments from all the authors.

Competing interests.

The authors declare that they have no conflict of interest.

Acknowledgements

The authors acknowledge support by the German Research Foundation ("Deutsche Forschungsgemeinschaft", DFG) in terms

30   of the research group FOR 2416 "Space-Time Dynamics of Extreme Floods (SPATE)", the German Federal Ministry of

Deleted: know

Deleted: the first half of the century

Deleted: , where there is a lack of observations

Deleted: and

Deleted: -C

Deleted: -C

Deleted: -C

Deleted:

Deleted: under

Education and Research (BMBF) under Grant MiKlip: FKZ01LP1518C/A and funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776661, entitled DownScaling CLImate ImPACTs and decarbonisation pathways in EU islands, and enhancing socioeconomic and non-market evaluation of Climate Change for Europe, for 2050 and Beyond. The opinions expressed are those of the author(s) only and should not be considered as representative of the European Commission's official position.

The authors also thank the Centre for Scientific Computing (CSC) of the Goethe University Frankfurt and the German High-Performance Computing Centre for Climate and Earth System Research ("*Deutsches Klimarechenzentrum"*, DKRZ) for supporting the calculations.

Special thanks to Dr. Christian Dieterich and other colleagues from SMHI for providing the NEMO-NORDIC model.

References

Akhtar, N., Brauch, J. and Ahrens, B.: Climate modeling over the Mediterranean Sea: impact of resolution and ocean coupling, Clim. Dyn., 51 (3), 933 –948, doi:10.1007/s00382-017-3570-8, 2017.

Akhtar, N., A. Krug, J. Brauch, T. Arsouze, C. Dieterich, B. Ahrens: European Marginal Seas in a regional atmosphere-ocean coupled model and their impact on Vb-cyclones and associated precipitation. Clim. Dyn., doi.org/10.1007/s00382-019-04906-x, 2019.

Beuvier, J., Lebeaupin Brossier, C., Béranger, K., Arsouze, T., Bourdallé-Badie, R., Deltel, C., Drillet, Y., Drobinski, P., Ferry, N., Lyard, F., Sevault, F., and Somot, S.: MED12, Oceanic component for the modeling of the regional Mediterranean earth system, Mercator Ocean Quarterly Newsletter, 46, 2012.

Casanueva, A., Rodríguez-Puebla, C., Frías, M. D., and González-Reviriego, N.: Variability of extreme precipitation over Europe and its relationships with teleconnection patterns, Hydrol. Earth Syst. Sci., 18(2), 709–725, doi:10.5194/hess-18-709-2014, 2014.

Christensen, J. H.: Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects (PRUDENCE), Final Report, DMI, 269p, 2005.

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff, and S.J. Worley: The Twentieth Century Reanalysis Project, Quarterly J. Roy. Meteorol. Soc., 137, 1-28, doi: 10.1002/qj.776, 2011.

Craig A., Valcke S., and Coquart L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, Geoscientific Model Development, 10, pp. 3297-3308, doi:10.5194/gmd-10-3297-2017, 2017.

Darmaraki, S., Somot, S., Sevault, F., Nabat, P., Cabos Narvaez, W. D., Cavicchia, L., Djurdjevic, V., Li, L., Sannino, G. and Sein, D.V.: Future Evolution of Marine Heatwaves in the Mediterranean Sea, Clim. Dyn., 45, 1–22, doi:10.1007/s00382-019-04661-z, 2019.

Dieterich, C., Wang, S., Schimanke, S., Gröger, M., Klein, B., Hordoir, R., Samuelsson, P., Liu, Y., Axell, L., Höglund, A., and Meier, H. E. M.: Surface Heat Budget over the North Sea in Climate Change Simulations, Atmosphere, 10, https://doi.org/10.3390/atmos10050272, 2019.

Dommenget, D.: Analysis of the model climate sensitivity spread forced by mean sea surface temperature biases, J. Clim. 25:7147–7162, 2012.

Doms, G., J. Förstner, E. Heise, H.-J. Herzog, D. Mironov, M. Raschendorfer, T. Reinhardt, B. Ritter, R. Schrodin, J.-P. Schulz, and G. Vogel: A description of the nonhydrostatic regional COSMO model. Part II: Physical parameterization. Deutscher Wetterdienst, O enbach, 154 pp. (Available at http://www.cosmo-model.org/), 2011.

DWD Climate Data Center (CDC): Historical daily station observations (temperature, pressure, precipitation, sunshine duration, etc.) for Germany, version v005, 2017.

Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., and Kollet, S.: Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1.0) in a massively parallel supercomputing environment – a case study on JUQUEEN (IBM Blue Gene/Q), Geosci. Model Dev., 7, 2531–2543, doi:10.5194/gmd-7-2531-2014, 2014.

Hartmann, D., Tank, A. K., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P.: Observations: atmosphere and surface - Climate Change 2013: the Physical Science Basis, EU-FP6 project ENSEMBLES & ECA&D project E-OBS gridded dataset, Website, accessed at 2016/1/22. http://www.ecad.eu/download/ensembles/download.php, 2016.

Gröger, M., Arneborg, L., Dieterich, C., Höglund, A., and Meier, H. E. M.: Summer Hydrographic changes in the Baltic Sea, Kattegat and Skagerrak projected in an ensemble of climate scenarios downscaled with a coupled regional ocean-sea ice-atmosphere, Climate Dynamics, accepted, 2019.

Harris, I., Jones, P.D., Osborn, T.J., and Lister, D.H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34(3): 623-642, doi: 10.1002/joc.3711, 2014.

Haylock, M.R., Hofstra, N., Tank, A. K., Klok, E., Jones, P. and New, M.: A European daily high-resolution gridded dataset of surface temperature and precipitation, J. Geophys. Res., 113, D20119, doi:10.1029/2008JD10201, 2008.

Hordoir, R., Axell, L., Höglund, A., Dieterich, C., Fransner, F., Gröger, M., Liu, Y., Pemberton, P., Schimanke, S., Andersson, H., Ljungemyr, P., Nygren, P., Falahat, S., Nord, A., Jönsson, A., Lake, I., Döös, K., Hieronymus, M., Dietze, H., Löptien, U., Kuznetsov, I., Westerlund, A., Tuomi, L., and Haapala, J.: Nemo-Nordic 1.0: A NEMO based ocean model for Baltic & North Seas, research and operational applications, Geoscientific Model Development Discussions,1-29, doi: 10.5194/gmd-2018-2, 2018.

Jacob, R., Larson, J., and Ong, E.: M x N communication and parallel interpolation in Community Climate System Model Version 3 using the model coupling toolkit, Int. J. High Perform. C., 19, 293–307, 2005.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D. and Storch, J. S.: Characteristics of the ocean simulations in MPIOM, the ocean component of the MPI-Earth system model, J. Adv. Model. Earth Syst., 5, 422–446, 2013.

Karl, T.R., Nicholls, N. and Ghazi, A.: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes: Workshop summary. Climatic Change, 42, 3-7, 1999.

Kelemen, F. D., Primo, C., Feldmann, H., Ahrens, B.: Added Value of Atmosphere-Ocean Coupling in a Century-Long Regional Climate Simulation, Atmosphere, 10(9), 537; doi:10.3390/atmos10090537, 2019.

5   Levitus, S., Antonov, J., Boyer, T.: Warming of the world ocean 1955–2003. Geophys Res Lett 32: L02604. doi:10.1029/2004GL021592, 2005.

Li, L., Bozec, A., Somot, S., Béranger, K., Bouruet-Aubertot, P., Sevault, F., and Crépon, M.: Regional atmospheric, marine processes and climate modelling, Developments in Earth and Environmental Sciences, Vol. 4, 373-397, doi:10.1016/s1571-9197(06)80010-8, 2006.

10   Lindström, G., Pers, C.P., Rosberg, R., Strömqvist, J., and Arheimer, B.: Development and test of the HYPE (Hydrological Predictions for the Environment) model–A water quality model for different spatial scales. Hydrol. Res., 41, 295–319, doi: 10.2166/nh.2010.007, 2010.

Ludwig, W., Dumont, E., Meybeck, M., and Heussner, S.: River discharges of water and nutrients to the Mediterranean and Black Sea: major drivers for ecosystem changes during past and future decades?, Prog. Oceanogr., 80, 199–217,

15   doi:10.1016/j.pocean.2009.02.001, 2009.

Maisonnave, E. and Caubel, A.: LUCIA, load balancing tool for OASIS coupled systems, TR-CMGC 14-63, CERFACS, 2014.

Madec, G.: NEMO ocean engine (version 3.3), Tech. Rep. 27, Note du Pole de modélisation, Institut Pierre- Simon Laplace (IPSL), France, 2011.

20   Marotzke, J., W.A. Müller, F.S. Vamborg, P. Becker, U. Cubasch, H. Feldmann, F. Kaspar, C. Kottmeier, C. Marini, I. Polkova, K. Prömmel, H.W. Rust, D. Stammer, U. Ulbrich, C. Kadow, A. Köhl, J. Kröger, T. Kruschke, J.G. Pinto, H. Pohlmann, M. Reyers, M. Schröder, F. Sienz, C. Timmreck, and M. Ziese: MiKlip: A National Research Project on Decadal Climate Prediction. Bull. Amer. Meteor. Soc., 97, 2379–2394, doi:10.1175/BAMS-D-15-00184.1, 2016.

Müller, W.A., Matei, D., Bersch, M., Jungclaus, J.H., Haak, H., Lohmann, K., Compo, P., Sardeshmukh, D., and Marotzke,

25   J.: A twentieth-century reanalysis forced ocean model to reconstruct the North Atlantic climate variations during the 1920s., Clim Dyn 44:1935–1955, doi:10.1007/s00382-014-2267-5, 2015.

Müller, W. A., Pohlmann, H., Sienz, F., Smith, D.: Decadal climate prediction for the period 1901-2010 with a coupled climate model. Geophys. Res. Lett., 41, pp 2100-2107, doi:10.1002/2014GL059259, 2014.

Obermann, A., Bastin, S., Belamari, S. Conte, D., Gaertner, M.A., Li, L. and Ahrens, B.: Mistral and Tramontane wind

30   speed and wind direction patterns in regional climate simulations, Clim. Dyn., 51(3), 1059-1076, doi:10.1007/s00382-016-3053-3, 2018.

Pham, T., Brauch, J., Dieterich, D., Früh, B., and Ahrens, B.: New coupled atmosphere-ocean-ice system COSMO-CLM/NEMO: On the air temperature sensitivity on the North and Baltic Seas, Oceanologia, 56, 167–189, doi:10.5697/oc.56-2.167, 2014.

Pham, T., Brauch, J., Dieterich, D., Früh, B., and Ahrens, B.: Added Decadal Prediction Skill with the Coupled Regional Climate Model COSMO-CLM/NEMO, Meteorologische Zeitschrift 27 (5), 391–99, doi:10.1127/metz/2018/0872, 2018.

Raschendorfer, M.: The new turbulence parameterization of LM. COSMO Newsletter, No. 1, Consortium for Small-Scale Modeling, Offenbach, Germany, 89–97. [Available online at http://www.cosmo-model.org/content/model/documentation/newsLetters/newsLetter01/newsLetter_01.pdf], 2001.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander L. V., Rowell, D. P., Kent, E.C. and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., Vol. 108, No. D14, 4407, doi:10.1029/2002JD002670, 2003.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. and Wang, W.: An improved in situ and satellite SST analysis for climate. J. Clim. 15, 1609–1625, 2002.

Ritter, B., Geleyn, J.-F.: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. Mon Weather Rev 120(2):303–325. doi: 10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2, 1992.

Rixen, M.: MEDAR/MEDATLAS-II, GAME/CNRM, doi:10.6096/HyMeX.MEDAR/MEDATLAS-II.20120112, 2012.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model CLM, Meteorol. Z., 17, 347–348, 2008.

Rummukainen, M., Raeisaenen, J., Bringfelt, B., Ullerstig, A., Omstedt, A., Willen, U., Hansson, U., and Jones, C.: A regional climate model for northern Europe: model description and results from the downscaling of two GCM control simulations, Clim. Dyn., 17, 339–359, 2001.

Schrodin, R., and Heise, E.: The Multi-Layer Version of the DWD Soil Model TERRA-LM, COSMO Tech. Rep. 2. Offenbach: Deutscher Wetterdienst, 2002.

Schrum, C.: Thermohaline stratification and instabilities at tidal mixing fronts. Results of an eddy resolving model for the German Bight, Continental Shelf Research,17,(6), 689-716, 1997.

Schrum, C., U. Hubner, D. Jacob, and R. Podzun.: A Coupled Atmosphere-Ice-Ocean Model for the North Sea and the Baltic Sea, Climate Dynamica, 21: 123–145, 2003.

Seifert, A., and Beheng, K. D.: A double-moment parameterization for simulating autoconversion, accretion and self-collection. Atmos. Res., 59–60, 265–281, doi:10.1016/ S0169-8095(01)00126-0, 2001.

Sevault, F., Somot, S., Alias, A., Dubois, C., Lebeaupin- Brossier, C., Nabat, P., Adloff, F., Déqué, M., and Decharme, B.: A fully coupled Mediterranean regional climate system model: design and evaluation of the ocean component for the 1980–2012 period, Tellus A, 66, 23967, doi:10.3402/tellusa.v66.23967, 2014.

Somot, S., Sevault, F., Déqué, M., and Crépon, M.: 21st century climate change scenario for the Mediterranean using a coupled Atmosphere-Ocean Regional Climate Model, Global and Planetary Change, 63, 112–126, doi:10.1016/j.gloplacha.2007.10.003, 2008.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T. and Roeckner, E.: Atmospheric

component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model. Earth Syst., 5, 1590–1601, doi:10.1002/jame.20015, 2013.

Sun, Y., Solomon, S., Dai, A., and Portmann, R.W.: How often does it rain?, J. Clim., 19, 916–934, doi: 10.1175/JCLI3672.1, 2006.

5 Tebaldi, C., Hayhoe, K., Arblaster, J. M., and Meehl, G. A.: Going to the extremes: an intercomparison of model simulated historical and future changes in extreme events, Climatic Change, 3–4, 185–211, 2006.

University of East Anglia Climatic Research Unit; Harris, I.C.; Jones, P.D.: CRU TS4.01: Climatic Research Unit (CRU)

Tegen, I., Hoorig, P., Chin, M., Fung, I., Jacob, D. and Penner, J.: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results. J Geophys Res 102, 23,895 – 23,915. doi: 10.1029/97JD01864,

10 1997.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large scale models. – Mon. Wea Rev. 117, 1779–1800, DOI: 10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016). Centre for Environmental Data Analysis, 04 December 2017. doi:10.5285/58a8802721c94c66ae45c3baa4d814d0, 2017.

15 Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G. and Morales Maqueda, M.A.: Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation. Ocean Modelling, 27, 33-53, doi:10.1016/j.oceamod.2008.10.005, 2009.

Will, A., Akhtar, N., Brauch, J., Breil, M., Davin, E., Ho-Hagemann, H., Maisonnaive, E., Thürkow, M. and Weiher, S.: The regional climate model and the model components coupled via OASIS-MCT: description and performance. Geoscientific

20 Model Development, Vol. 10 (4), pp. 1549–1586, 2017.

Zhang, X. B., G. Hegerl, F. W. Zwiers, and J. Kenyon: Avoiding inhomogeneity in percentile-based indices of temperature extremes, J. Climate, 18, 1641, 2005.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B., Zwiers, F.W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Clim Change, 2: 851-870.

25 doi:10.1002/wcc.147, 2011.

Figures
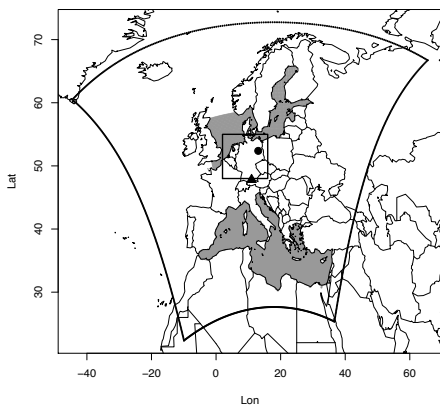
Figure 1: EURO-CORDEX domain where the coupled system runs, including the marginal seas (grey area), the Mid-Europe region from the PRUDENCE project (square) and two locations of German climate stations: Potsdam (circle) and Hohenpeißenberg (triangle).
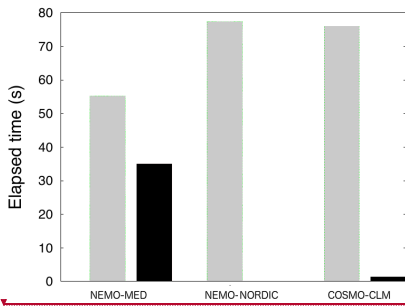
Figure 2: Computing time used for the exchange of each of the OASIS coupled model components. Grey bars show calculation time and black bars show waiting time.
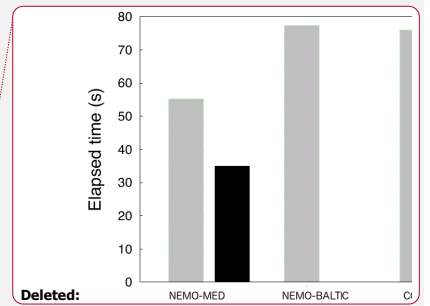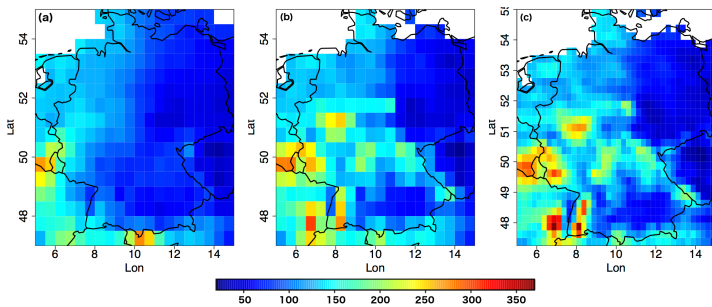
Figure 3: Gridded total precipitation observations for January 1995: (a) CRU data (0.5°x0.5°), (b) E-OBS data upscaled to 0.5°x0.5°,
5 **and (c) E-OBS data with original resolution (0.25° x 0.25°).**
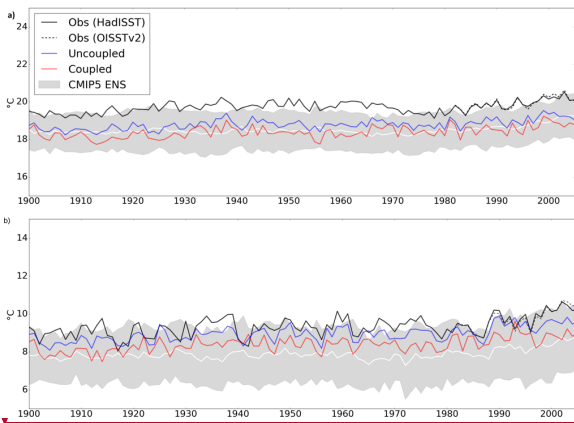
Figure 4: Sea surface temperature annual means over the marginal seas in our 20th Century coupled simulation (CCLM-NEMO)
10 **between 1900 and 2005, compared to observations (HadISST and OISSTv2), to the atmosphere-only CCLM simulation (with SSTs**
prescribed by the driving MPI-ESM, nudged to observations), and an ensemble mean (white line) and spread (shaded area) from
CMIP5 simulations. a) Mediterranean Sea and b) Baltic and North Sea.
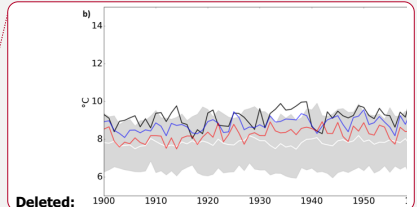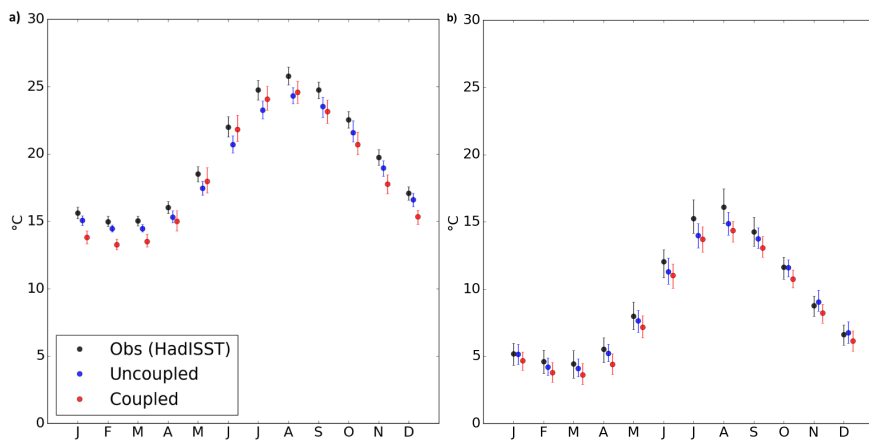
Figure 5: Annual cycle during the 20ᵗʰ century of the spatially averaged SST values in the marginal seas: a) Mediterranean Sea, b) Baltic and North Sea. Dots represent the mean monthly value and intervals show the 10ᵗʰ and 90ᵗʰ percentiles.

Formatted: Superscript

Deleted: mean

Deleted: averaged

Formatted: English (US)
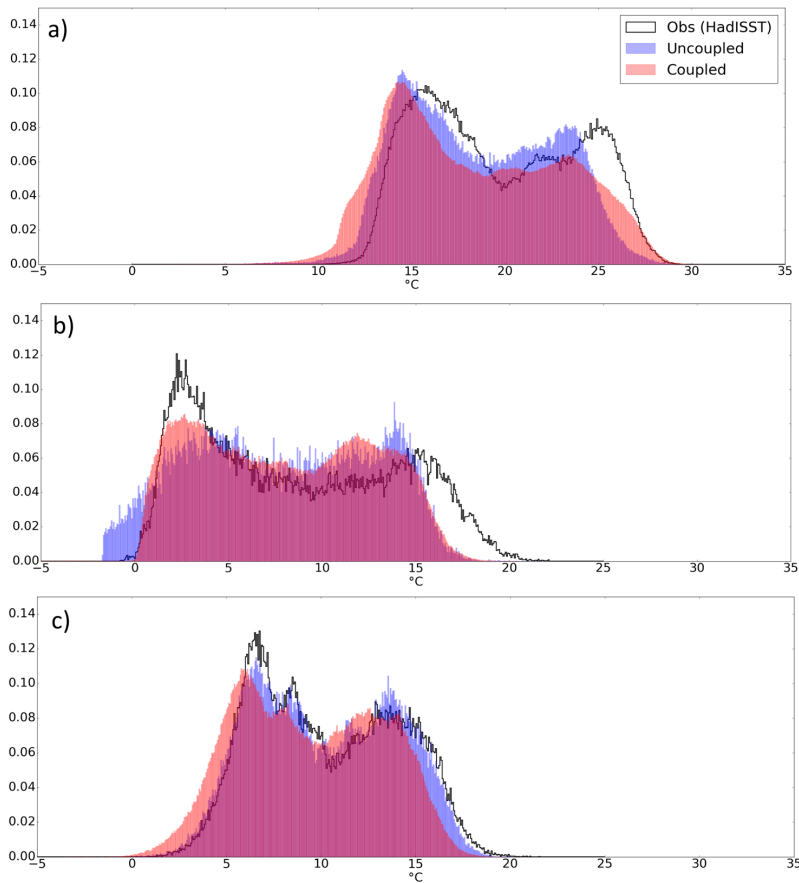
Figure 6: Density histograms of SST values in the marginal seas: a) Mediterranean Sea, b) Baltic Sea, and c) North Sea. Note that in case of the Baltic Sea the grid points containing sea ice were not considered. CRU is represented in white, the coupled system in red, the uncoupled in blue, and the data overlapping in purple.
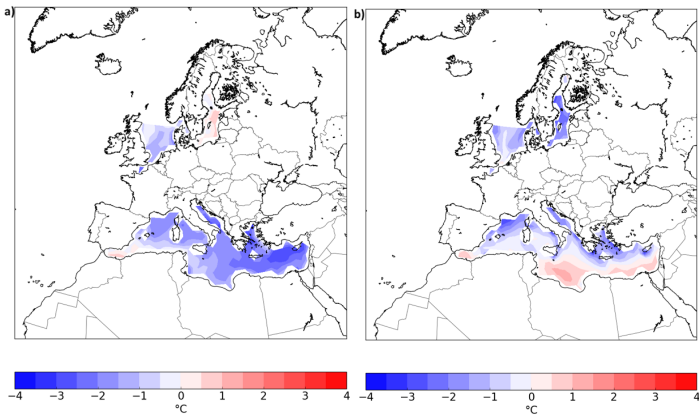
Figure 7: Spatial distribution of coupled simulation – observation SST differences in (a) winter and (b) summer.
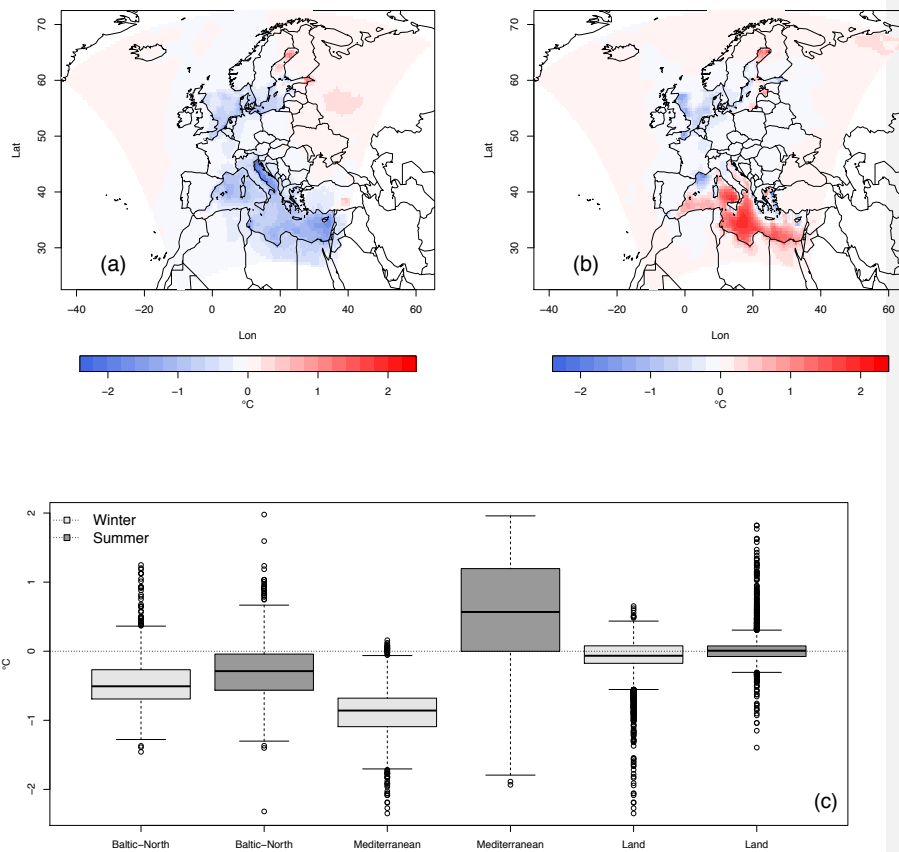
Figure 8: Coupled, uncoupled simulations 2m air temperature monthly mean (°C) difference averaged over the 20th century during winter (a) and summer (b). Boxplots represent the distributions of the monthly mean over the marginal seas and land separately (c). The box ends at quartiles, the horizontal line represents the median, and the points are values more than 3/2 the interquartile range from the end of the box.
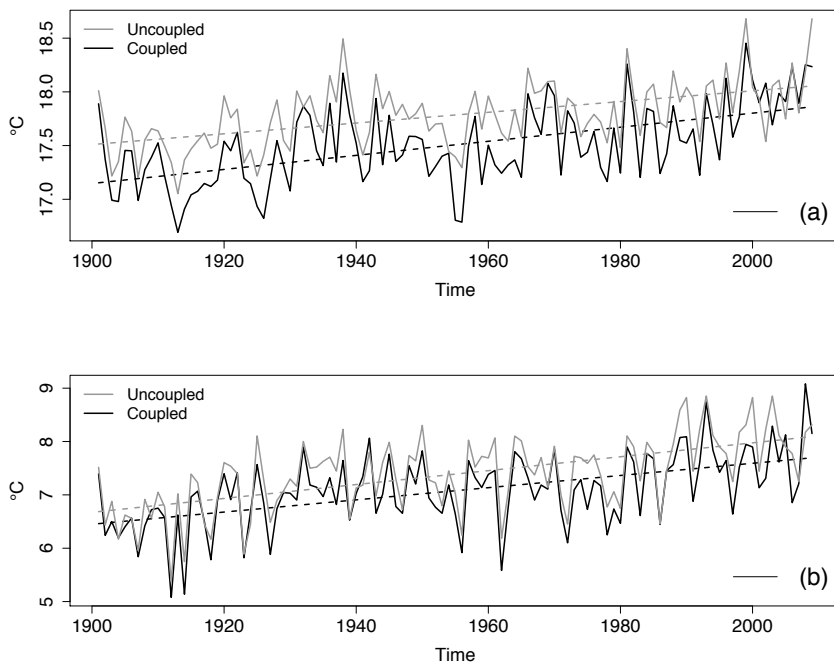
Figure 9: Temporal evolution of the annual 2m air temperature averaged over (a) the Mediterranean Sea and (b) the Baltic and North Seas. Dashed lines are linear fits.
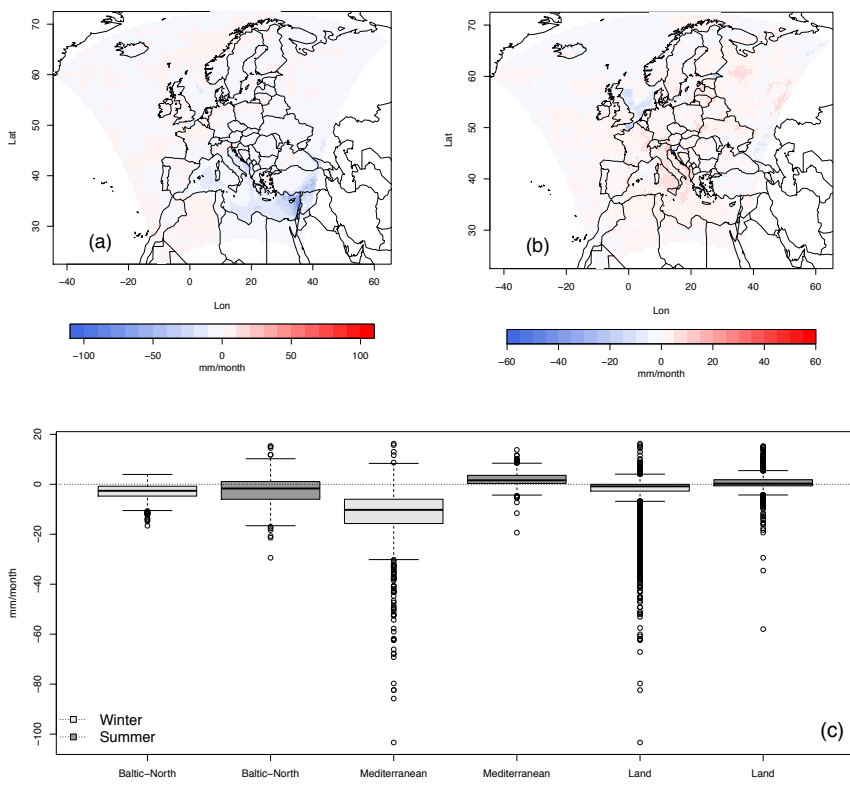
Figure 10: Total precipitation coupled - uncoupled systems monthly sum difference, averaged over (a) winter, and (b) summer, for the period 1901-2009. Boxplots (c) represent the distribution of these differences over the marginal seas and land separately. The box ends at quartiles, the horizontal line represents the median, and the points are values more than 3/2 the interquartile range from the end of the box.
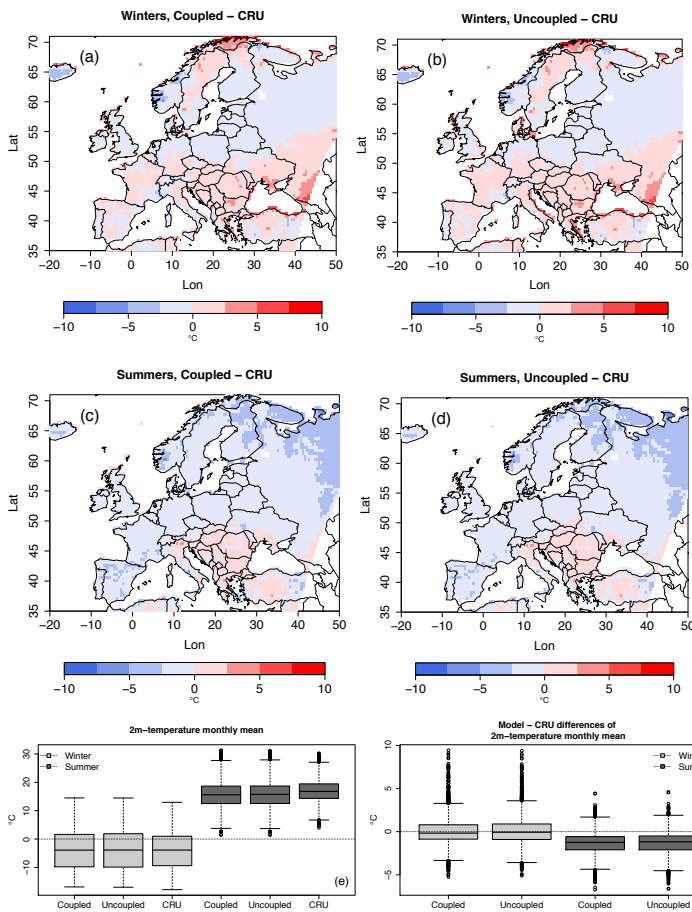
Figure 11: 2m air temperature coupled system - CRU data differences (a, c) and the uncoupled system and CRU data (b, d), during winter (a, b) and summer (c, d) for the 20th century. Boxplots of the 2m air temperature (e) and boxplots of the regional models - CRU difference (f). Boxplots follow the same criterion as in previous figures.

a)   b)
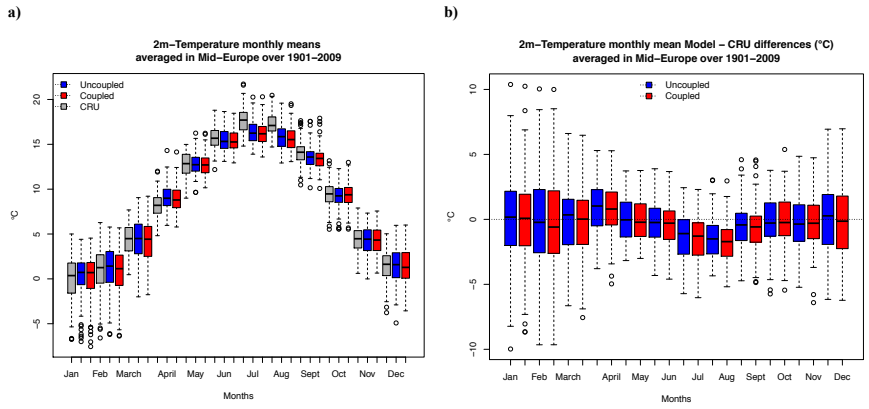
Figure 12: Boxplots showing the annual cycle during the 20th century of the 2m air temperature monthly means averaged over Mid-Europe for the CRU observations (grey), uncoupled system (blue) and coupled system (red) (a) and the annual cycle of the model errors compared to CRU observations over Mid-Europe (b).
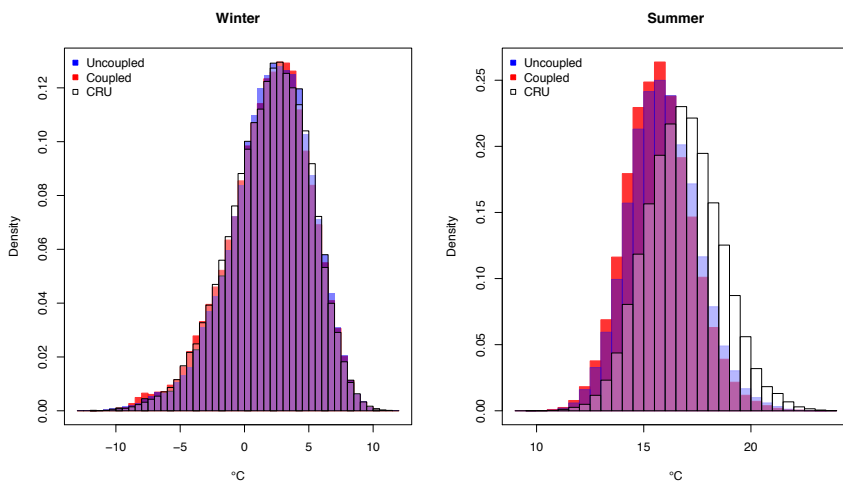
Figure 13: 2m air temperature density histograms of the coupled and uncoupled systems compared to the CRU data over the PRUDENCE area named Mid-Europe during the 20ᵗʰ century in winter (left) and summer (right). The coupled system is represented in red, the uncoupled in blue, CRU in white, and the data overlapping in purple.

Deleted: 2m-Temperature
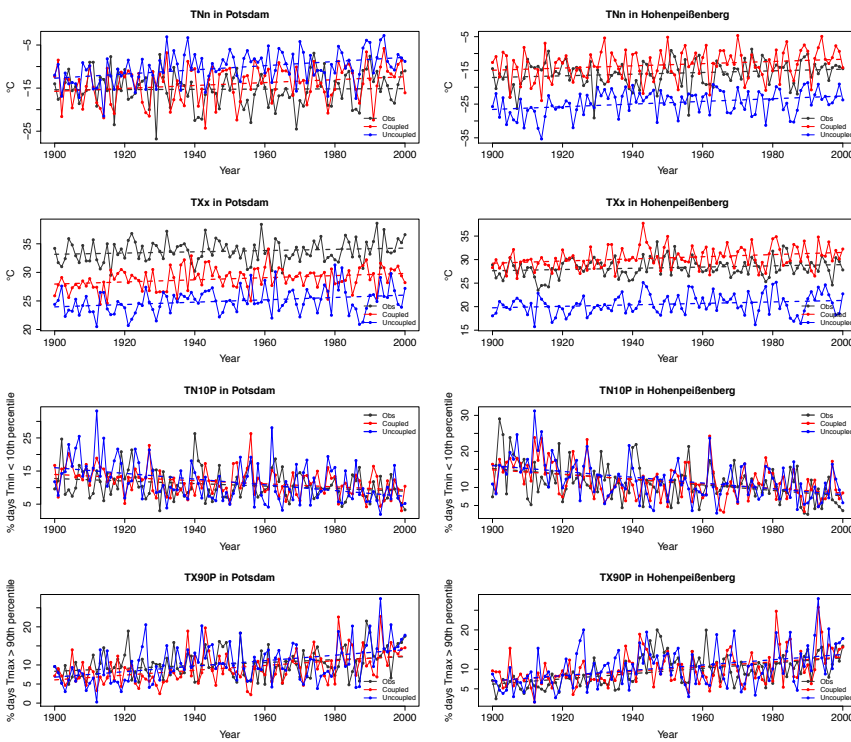
Figure 14: Temporal evolution of four climate change indices at two stations in Germany, Potsdam (81m) and Hohenpeißenberg (977m). TNn represents the annual minimum value of daily minimum temperature; TXx, the annual maximum value of daily maximum temperature; TX90p, the percentage of days when the maximum temperature is above the calendar day 90th percentile centred on a 5-day window for the base period 1961-1990; and TN10p, the percentage of days when the minimum temperature is below the calendar day 10th percentile centred on a 5-day window for the base period 1961-1990. Linear trends are shown as dashed lines.

Deleted: in
Deleted: of
Deleted: .
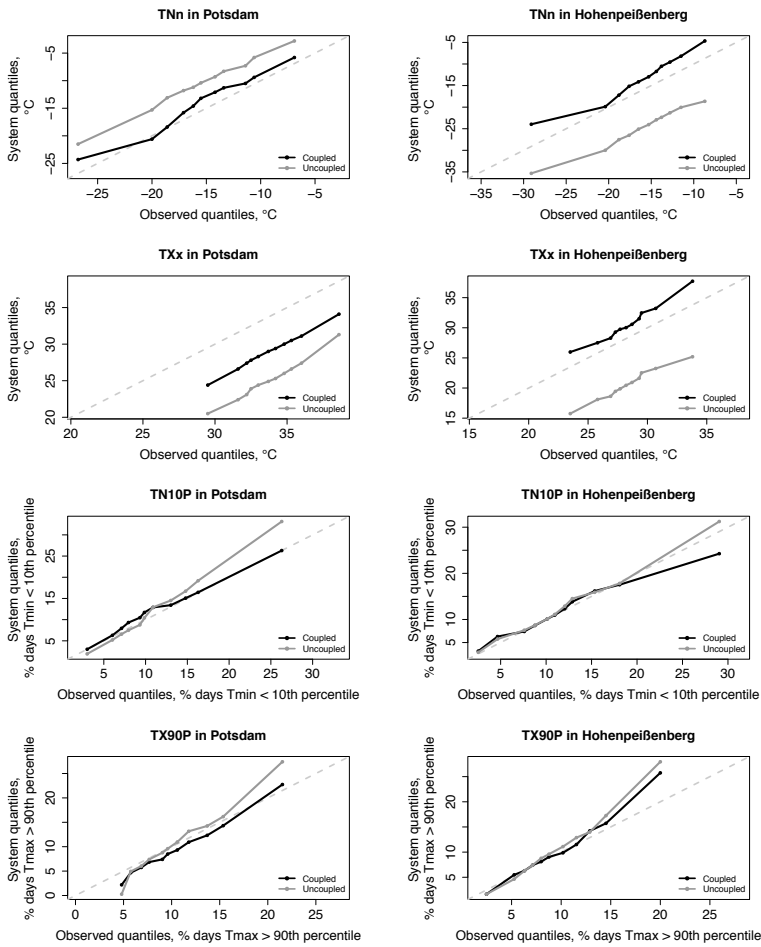
Figure 15: Quantile-quantile plots of the extreme indices shown in Fig.14. The diagonal represents the perfect case.
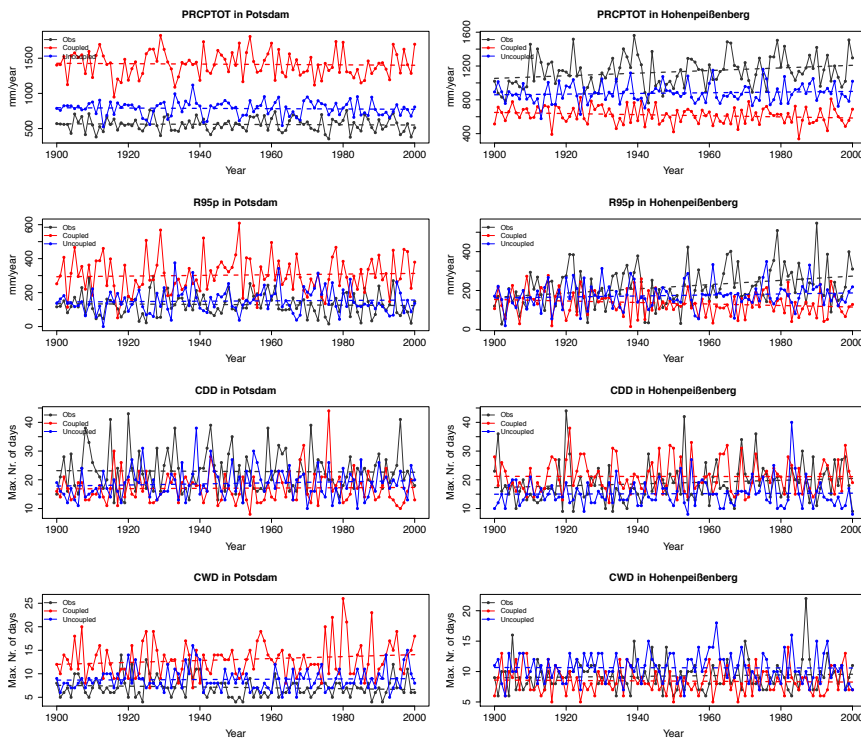
Figure 16: Temporal evolution of four climate change indices related to precipitation in two German stations: Potsdam (81m) and Hohenpeißenberg (977m). PRCPTOT represents the annual total precipitation in wet days, R95p is the annual total PRCP when the daily precipitation (RR) is above the 95th percentile of precipitation on wet days in the 1961-1990 period, CDD is the maximum length of dry spell (maximum number of consecutive days with RR < 1mm) and CWD is the maximum length of wet spell (maximum number of consecutive days with RR ≥ 1mm). Linear trends are shown as dashed lines.
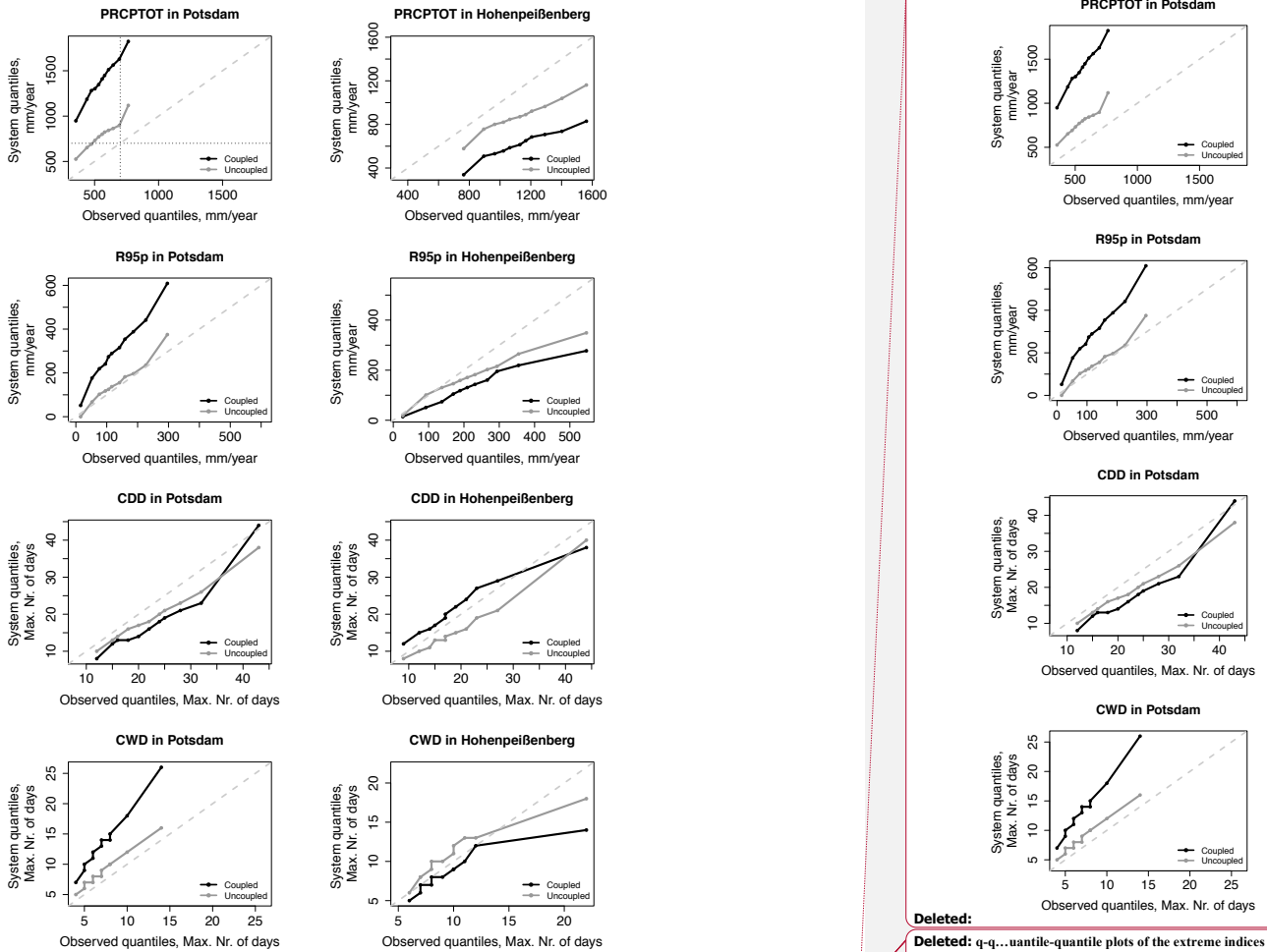
Figure 17: Quantile-quantile plots of the extreme indices shown in Fig. 16. The diagonal represents the perfect case.