Using a nested single-model large ensemble to assess the internal variability of the North Atlantic Oscillation and its climatic implications for Central Europe

Andrea Böhnisch[1], Ralf Ludwig[1], and Martin Leduc[2, 3]

[1]Department of Geography, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Ouranos, Montréal,Québec, Canada
[3]Centre ESCER, Université du Québec à Montréal, Montréal, Québec, Canada

Correspondence: Andrea Böhnisch (a.boehnisch@lmu.de)

Abstract. Central European weather and climate is closely related to atmospheric mass advection triggered by the North Atlantic Oscillation (NAO), which is a relevant index for quantifying internal climate variability on multi-annual time scales. It remains unclear, though, how large-scale circulation variability affects local climate characteristics when downscaled using a regional climate model. In this study, 50 members of a single-model initial-condition large ensemble (LE) of a nested regional climate model are analyzed for a NAO–climate relationship. The overall goal of the study is to assess whether the range of NAO internal variability is represented consistently between the driving global climate model (GCM; the CanESM2) and the nested regional climate model (RCM; the CRCM5). Responses of mean surface air temperature and total precipitation to changes in the index value are expressed for a Central European domain in both the CanESM2-LE and CRCM5-LE via Pearson correlation coefficients and the change per unit index change for historical (1981–2010) and future (2070–2099) winters. Results show that statistically robust NAO patterns are found in the CanESM2-LE under current forcing conditions. Reproductions of the NAO flow patterns in the CanESM2-LE trigger responses in the high-resolution CRCM5-LE that are comparable with reference reanalysis data. NAO–response relationships weaken in the future period, but their inter-member spread shows no significant change. The results stress the importance of single-model ensembles for the evaluation of internal variability. They also strengthen the validity of the nested ensemble for further impact modelling using RCM data only, since important large-scale teleconnections present in the driving GCM propagate properly to the fine scale dynamics in the RCM.

1 Introduction

One of the major sources of uncertainty regarding short-term future climate projections is internal climate variability, while model climate response and greenhouse gases concentrations scenarios become more important sources of uncertainty on a longer-term time horizon (Hawkins and Sutton, 2009, 2011). The term internal variability denotes variability which is not forced by external processes (either anthropogenic or natural), but arises from the chaotic properties of the climate system itself (Leduc et al., 2019; Deser et al., 2012), i.e. from varying sequences of weather events under identical external forcings. These sequences of weather events may be altered by global atmospheric modes of variability through the linking between

large-scale circulation and local weather characteristics (like surface air temperature and precipitation). Atmospheric modes can thereby establish periods of discernible states on multi-annual time scales.

Among these modes, the North Atlantic Oscillation (NAO) is particularly important for northern hemisphere climate. Its two states, positive and negative, are evoked by planetary wave-breaking in the polar front, leading to antagonistic pressure behaviour of two centres over the North Atlantic: one located within the subtropical high pressure belt ("Azores High", AH), the second in subpolar regions ("Icelandic Low", IL) (Benedict et al., 2004). The resulting pressure gradient, which is stronger during positive and weaker during negative phases, affects large-scale extra-tropical circulation, especially the strength and position of mid-latitude westerly winds connected to the jet stream, and air mass advection during boreal winter (Deser et al., 2017; Hurrell and Deser, 2009). Compared to neutral conditions, the positive NAO state leads to warmer and moister winters in northern Europe, but cooler and drier conditions in the south, and vice versa in the negative state (e.g., Hurrell and Deser, 2009; Pokorná and Huth, 2015; Woollings et al., 2015).

Commonly, the NAO is quantified with an index that makes use of the air pressure or geopotential height gradient between AH and IL. The index may be calculated as a normalized difference of station measurements, spatially averaged values of pre-set regions, or the region of highest variance is obtained by principal component analysis (PCA) (Pokorná and Huth, 2015; Hurrell and Deser, 2009; Stephenson et al., 2006; Hurrell, 1995; Rogers, 1984). Each method has its advantages and limitations. For example, station-based or fixed in space indices do not reproduce shifting NAO patterns and may be affected by micro-climatic noise and other teleconnection patterns (Hurrell and Deser, 2009; Osborn, 2004). Indices based on PCA on the other hand are dependent on the chosen data domain for calculation and on the data set itself (Osborn, 2004). The different approaches though lead to highly similar index time series (see e.g., Pokorná and Huth, 2015, for a detailed survey of various approaches).

While the typical NAO pattern and its impacts are usually correctly reproduced in global climate models (GCMs) (Stephenson et al., 2006; Ulbrich and Christoph, 1999; Reintges et al., 2017), its fidelity in a future climate remains uncertain: the NAO is found as intensifying, but also counteracting global warming in the northern hemisphere ("global warming hiatus", Iles and Hegerl, 2017; Deser et al., 2017; Delworth et al., 2016). Similarly, the findings regarding the prevalence of future positive or negative states lack unity: Some analyses of CMIP5 models, for example, suggest more positive phases under rising greenhouse gas concentrations until 2100 (e.g., Kirtman et al., 2013; Christensen et al., 2013), others favour an increase of negative phases (Cattiaux et al., 2013).

In most of these studies it was common to rely on one simulation per model and estimate the model's performance regarding the NAO by this single run. This approach allows for comparing different models (and observations). However, it is not possible to robustly evaluate the range of NAO index values, or whether the chosen simulation is a good representation of how this model simulates the phenomenon in question (Leduc et al., 2019). Relying on single realizations possibly deteriorates the assessment of a given model, as single realizations may vary considerably among themselves due to internal variability (and also deviate from the climate evolution observed in reality). One way to trigger internal variability in GCM simulations is to perturb the initial conditions of the model, leading to several realizations of weather sequences under identical external forcing which also allow to derive a robust distribution of NAO index values. That is why this study is investigating the NAO pattern in a single-model large ensemble (50 members) of a GCM.

However, when interested in NAO impacts on a regional scale, like Central Europe, the GCM is not sufficient for fine-scale responses. Due to their coarse spatial resolution, GCMs are poorly resolving land–water contrasts and topographic characteristics which may be highly relevant in climate impact studies over heterogeneous landscapes (Leduc et al., 2019). Thus, dynamical downscaling of the GCM members using a regional climate model (RCM) is advised (Leduc et al., 2019). Such downscaling of a GCM single-model large ensemble was performed within the Climate Change and Hydrological Extremes project (ClimEx, www.climex-project.org, Leduc et al., 2019).

The combination of the driving GCM and nested RCM large ensembles (LE) allows for analyzing the spread of NAO states and responses within one model chain, thus establishing the range of internal variability of the NAO, and finding robust NAO and response patterns by significantly reducing uncertainty associated with internal variability in the ensemble.

This study also targets the question, how global circulation variability, in this case the NAO teleconnection, affects local climate characteristics when downscaled using an RCM. It specifically aims at evaluating whether the range of internal variability is represented consistently between the driving GCM and the driven RCM. This issue may be important for impact modellers who work with RCM data without taking the driving GCM into account.

To answer these research questions, the study is focussing on four topics and related key questions:

(a) General performance of the model chain: Can the driving GCM resolve the NAO correctly and are climatic implications for Central Europe reproduced?

(b) Nesting approach: Does the RCM correctly incorporate the NAO pattern present in the driving data and produce realistic response patterns?

(c) Internal Variability: What is the range of possible NAO patterns and responses, expressed by the inter-member spread (IMS) among the 50 members?

(d) Climate change: How do (a), (b) and (c) change in transient climate simulations until 2099 using an RCP8.5 emissions scenario?

2 Data and Methodology

2.1 Data

Data from three different sources were employed in this study (Table 1). The major source was the LE data set of the ClimEx project which is described in detail in Leduc et al. (2019). The ClimEx project is conducted in a Québec-Bavarian cooperation and targets issues of hydrological extreme events in the time horizon of 1950–2099, using a nested high-resolution 50 member single-model initial-condition large ensemble with an RCP8.5 emissions scenario from 2006 onwards (Leduc et al., 2019). Five members of the Canadian Earth System Model version 2 (CanESM2 Large Ensemble, 2.8° spatial resolution, Fyfe et al., 2017) with different ocean inital conditions were slightly perturbed in 1950, leading to ten members per ocean family. The members are assumed to become independent about five years after their initialization in 1950 (spin-up-period) (Leduc et al., 2019).

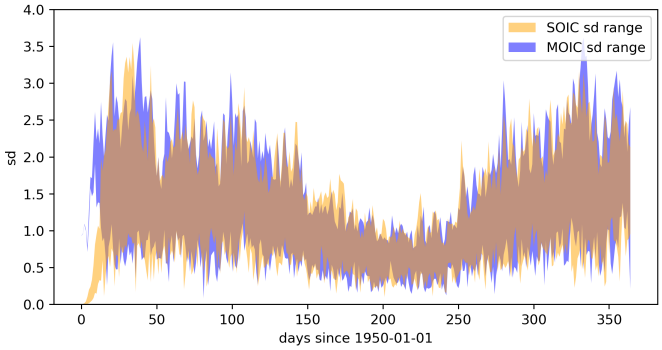

Figure 1. Inter-member standard deviation of a daily NAO index in the CanESM2-LE starting on 1 Jan 1950 as a function of time. The inter-member standard deviation was derived from ten groups of five members with the same ocean initial conditions (SOIC) and ten groups of five members with mixed ocean initial conditions (MOIC, following an approach in Leduc et al., 2019).

Regarding the atmospheric circulation, Fig. 1 shows that owing to the chaotic nature of the atmospheric system the daily NAO index seems to lose dependence from the initial conditions within the course of one month after initialization (see Leduc et al., 2019, for a similar presentation of member independence). As described in Leduc et al. (2019), these 50 GCM members were dynamically downscaled using the Canadian Regional Climate Model version 5 (CRCM5 Large Ensemble, 0.11° spatial resolution) over two domains covering Europe and north-eastern North America. During the nesting process, large-scale spectral nudging regarding the horizontal wind field was applied (Leduc et al., 2019). This single-RCM 50-member ensemble allows for internal variability and extreme events to be detected in high spatial and temporal resolution within a total of 7500 modelled years (Leduc et al., 2019). Comparing the internal variability of the CRCM5 members with the IMS of a subset of the multi-model EURO-CORDEX (Coordinated Regional climate Downscaling Experiment) ensemble regarding winter temperature and precipitation, von Trentini et al. (2019) showed that both ensemble spreads are of comparable magnitude. The CORDEX ensemble consists of several GCM-RCM combinations set up in a coordinated modelling framework and aims at evaluating uncertainty due to model configuration (Giorgi et al., 2009). The similarity of the single-model and multi-model spreads suggests that a large fraction of the CORDEX ensemble spread can be explained by internal variability, despite the fact that it was not explicitly sampled within the CORDEX framework (where most models provided a single simulation, von Trentini et al., 2019). Therefore, the GCM and RCM ClimEx ensemble can be expected to capture the range of winter temperature and precipitation internal variability despite the set up with a single model.

Model data is compared to the ERA-Interim (ERA-I) Reanalysis data set of the European Centre for Medium-Range Weather Forecasts (Dee et al., 2011, ECMWF) which serves as a reference (REF). Additionally, a CRCM5 run driven by ERA-I was used to evaluate the CRCM5 under "perfect" (as far as ERA-I can be assumed to represent reality) lateral boundary conditions (LBC), i.e. without the potential CanESM2 input data error.

The relevant variables for this study were:

- (mean) sea level air pressure (referred to as "SLP", converted to [hPa]) to obtain the NAO,

Table 1. Overview of used data sets, their spatial resolution, the number of members and the employed variables.

data name	model type	spatial resolution	members	model output variable names	institution
ERA-I	re-analysis	$0.75° \times 0.75°$	1	msl [Pa], t2m [K], tp [m]	ECMWF
CRCM5/ERA-I	RCM	$0.11° \times 0.11°$	1	tas [K], pr [$\mathrm{kgm^{-2}s^{-1}}$]	Ouranos
CanESM2	GCM	$2.8° \times 2.8°$	50	psl [Pa], tas [K], pr [$\mathrm{kgm^{-2}s^{-1}}$]	CCCma
CRCM5-LE	RCM	$0.11° \times 0.11°$	50	tas [°C], pr [mm]	Ouranos

CCCma – Canadian Centre for Climate Modelling and Analysis

- near surface air temperature (referred to as "nSAT", converted to [K]),
- total precipitation including liquid and solid precipitation from all types of clouds (referred to as "PR", converted to [mm]).

ERA-I variables t2m, tp and msl were chosen as they were assumed to most accurately represent the variables from the GCM and RCM models. As the variables derived from the three data sources were originally available with different temporal resolutions (three-hourly for tas in RCM, hourly for pr in RCM, daily in GCM for psl, pr and tas, 6-hourly for ERA-I t2m and msl analysis, and 12-hourly for ERA-I tp forecast data), they were all aggregated to daily values.

Figure A1 shows that the CRCM5 tends to underestimate (overestimate) mean winter nSAT mean in the northern (southern) part of the domain, regardless of the driving data (see first column for ERA-I and third column for CanESM2), whereas winter PR sums are overestimated in nearly the entire domain with strongest values in the south-eastern part. The GCM overestimates (underestimates) nSAT mean north (south) of the Alps. PR sum is underestimated in the entire domain apart from the western side of the Alps in the GCM. However, as this study will focus on changes in nSAT and PR induced by the NAO (see Section 2.2.4), biases are of no large relevance in general, but may show some influence when it comes to regions with particularly high PR sum values.

Commonly, NAO impact studies focus on seasonally aggregated values of the analyzed variables or extreme events (e.g., Stephenson et al., 2006). Yet the NAO, which accounts for variations in the mean zonal atmospheric flow towards Europe, can be assumed to not only influence winter mean values, but also their dispersion. So the following analyses were not confined to winter mean temperature (nSAT mean) and precipitation sums (PR sum), selected analyses were also performed on winter mean monthly standard deviations of daily mean temperature (nSAT std) as a measure of temperature variation.

2.2 Methodology

2.2.1 Regions of interest and time horizon selection

Analyses were performed on time series of spatially averaged information (nSAT mean, PR sum for response variables and SLP for index calculation) as well as on spatially explicit data (nSAT mean, nSAT std, PR sum). All data were provided as

netCDF and most pre-processing was performed using the Climate Data Operators (CDO) of the Max-Planck-Institute for Meteorology (Schulzweida, 2017).

The regions of interest used in this study are displayed in Fig. 2. The formation of the NAO over the North Atlantic (NAR, AH, IL regions, see annotations in Fig. 2) was analyzed in the ERA-I and CanESM2-LE dataset, while responses over Central Europe (CEUR, NE, BY, SE) were evaluated in ERA-I, CRCM5/ERA-I, CRCM5-LE and CanESM2-LE.

AH and IL regions are centered over Ponta Delgada/Azores and Reykjavik/Iceland, two commonly used stations for NAO index calculations. To avoid micro-climatic impacts and sampling uncertainties of a single gridcell and to account for moving SLP centres (see e.g., Moore et al., 2013), both NAO core regions were extended to 3×3 GCM grid cell matrices. The NAO index proved to be very robust towards the exact shape of the core regions in preliminary analyses.

The Central European domain (CEUR) was defined in the CanESM2-LE by selecting a 5×5 GCM grid cell matrix centered over Munich/Germany. This CEUR domain extends from Denmark in the north to mid-Italy in the south and from Poland to France in east–west direction. The corresponding CEUR region within the ClimEx European domain was used to quantify the impacts of the NAO in the CRCM5-LE dataset. It lies downstream of the westerly flows initiated by the NAO, so the following analyses will set a special focus on the incorporation of large-scale inflow from the western side into the nested RCM.

As the responses to the NAO were expected to vary over the CEUR domain, it seemed favourable to analyze spatial structures explicitly in addition to analyses of time series over several subset regions. These subset regions (see e.g., Déqué et al., 2007) denote small-scale sample areas inside the CEUR domain, sized one GCM grid cell each, with expected typical "northern European" (NE) and "southern European" (SE) NAO responses for a more detailed statistical analysis. A third GCM grid cell was chosen to represent the transition zone between NE and SE. Coincidentally, it closely represents the region of Bavaria which is why the name "BY" was assigned to it. REF and RCM data (3×4 and 26×26 grid cells, respectively) was spatially aggregated to GCM resolution for this part of the analysis.

This study focused on inter-annual analyses which were conducted for two time horizons covering 30 years each. The chosen period length was assumed to include major fluctuations, like internal climate variations or several solar cycles, which might affect NAO phases (Andrews et al., 2015). Thus their influence can be assumed to be represented by the sampled NAO time series.

Relationships between the NAO and response variables most probably vary on different time scales (Hurrell and Deser, 2009; Woollings et al., 2015; Xu et al., 2015; Hurrell and Van Loon, 1997). However, as 30 year periods are not long enough for analyses of multi-decadal (>30 years) NAO–response variability (Woollings et al., 2015), stationarity in NAO-impact relationships was assumed for simplicity reasons. The historical (hist; 1981–2010) period was used to establish reference statistics in the ERA-I data and the ERA-I driven CRCM5 run. These statistics were evaluated in GCM and RCM data to check the models' ability of depicting NAO responses. Links and relationships established for the historical period were also investigated in a far future horizon (fut; 2070–2099).

All data (spatially explicit and subset time series) was aggregated to the seasonal time scale (winter means for nSAT and winter sums for PR).

Since the NAO is known to be strongest in winter (Hurrell and Deser, 2009) and the connection between station-based indices

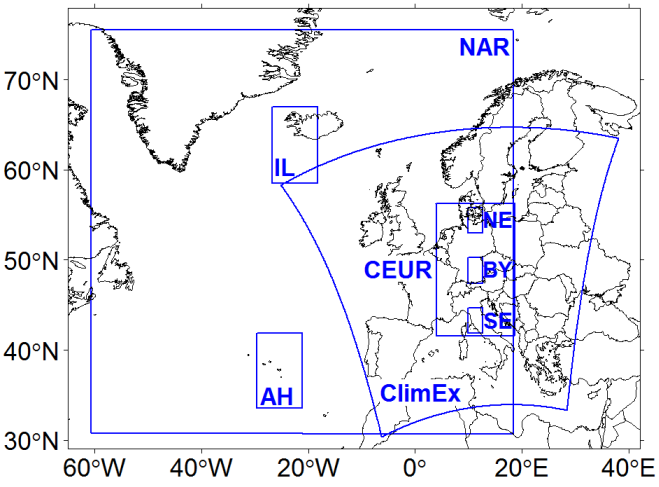

Figure 2. Regions of interest. Abbreviations and domain sizes in terms of GCM grid cells (2.8°) are as follows: AH – Azores High (3×3); IL – Icelandic Low (3×3); NAR – large-scale North Atlantic region (28×16); CEUR – Central Europe (5×5); NE – northern Europe (1); BY – Bavaria (1); SE – southern Europe (1); ClimEx – domain used in ClimEx project (extent approximately 22×12 after resampling to GCM grid).

and NAO responses tends to be best in winter (see Pokorná and Huth, 2015, for months DJF), analyses were performed for this season only. Preliminary tests had shown that correlations and links between the NAO index and the climate variables were more distinct from noise, if March was included as well. That is why an extended winter season was used here (DJFM, see also Iles and Hegerl, 2017; Hurrell, 1995; Osborn, 2004).

2.2.2 Deriving an NAO index

The NAO index was derived from ERA-I and CanESM2-LE data, resulting in 1 REF and 50 GCM realizations. The NAO is quantified in this study with an index which is closest to a station based or zonally averaged index. This allowed obtaining an index in a large data set (50 members during hist and fut time horizons) at justifiable computational time. Other than indices based on PCA, this index does not represent a "pure" NAO pattern, i.e. the variability of North Atlantic SLP without any other teleconnection patterns like the East Atlantic Pattern (EA) and the Scandinavian Pattern (SCA) (Moore et al., 2013). Instead, it directly represents the winter SLP gradient over the North Atlantic.

The time series of AH and IL originated from the temporally shortened and spatially averaged SLP time series of both grid cell matrices for REF and GCM data only. As the CRCM5 ClimEx domain does not cover the AH and IL regions (see Fig. 2), the index was not derived from this data source.

Daily SLP values were averaged to monthly means (Cropper et al., 2015) and scaled to obtain mean $\mu = 0$ and standard deviation $\sigma = 1$, as outlined in Osborn (2004) and Hurrell and Van Loon (1997), by subtracting the 1981–2010 seasonal mean

(overbar) and dividing by the 1981–2010 seasonal standard deviation (s_{IL}, s_{AH}):

$$\mathrm{NAOIndex} = \frac{AH - \overline{AH}}{s_{AH}} - \frac{IL - \overline{IL}}{s_{IL}} \tag{1}$$

Monthly indices were next averaged to DJFM means. This approach is similar to Woollings et al. (2015) and Jones et al. (2013).

To compare future with historical index values, the future time series of AH and IL were normalized with the present SLP standard deviations (see also Ulbrich and Christoph, 1999; Hansen et al., 2017) and mean values. The normalization of each GCM member was carried out individually, such that each member has specific normalization parameters.

2.2.3 Evaluation of the large-scale SLP pattern in RCM data

To estimate whether the NAO may be seen as being correctly represented in the nested RCM data, the reproduction of inter-annual SLP pattern variations in the CRCM5 data was verified. Therefore, monthly mean SLP data of the CRCM5 (both driving data sets) and ERA-I were linearly interpolated to GCM resolution over the ClimEx domain. During interpolation, small scales were automatically filtered such that the remaining large scales of GCM and RCM data may be compared. As a next step, the monthly difference between driving data and the RCM data was taken for each time step and member. From these differences, a root-mean-square difference over the hist and fut time periods was obtained which was later averaged over all ensemble members and the winter months:

$$\mathrm{RMS}^*(i,j) = \left\langle \left\langle \frac{\sqrt{\langle D_m(i,j,t,n)^2 \rangle_t}}{\sqrt{\mathrm{VarDrive}_m(i,j,n)}} \right\rangle_n \right\rangle_{m=12,1-3} \tag{2}$$

$$\mathrm{VarDrive}_m(i,j,n) = \left\langle \left(\mathrm{Drive}_m(i,j,t,n) - \langle \mathrm{Drive}_m(i,j,t,n) \rangle_t\right)^2 \right\rangle_t \tag{3}$$

where $\langle \cdot \rangle$ is the averaging operator over a given index, D_m is the difference between driving data and RCM data; Drive_m is driving SLP data; $\mathrm{VarDrive}_m$ is the variance of the driving data over the 30 year periods; i, j are spatial coordinates of the grid, m are months 12, 1–3, n are ensemble members 1–50 and t are years in 1981–2010 and 2070–2099. The normalization by the square root of the temporal variance of the driving data allows to derive a measure relative to the inter-annual variability of the SLP pattern on a given location. Low RMS^* values indicate a low error.

2.2.4 Climatic Changes Associated with NAO

All data sources (Table 1) were used to obtain response patterns of the variables nSAT and PR. Climatic changes associated with the NAO were evaluated using Pearson correlation coefficients and a slope parameter obtained by linear regression.

ERA-I and CRCM5/ERA-I nSAT and PR data were correlated with the ERA-I index, CanESM2 and CRCM5 members were correlated with the CanESM2 index calculated for the corresponding member.

The correlation analysis assumes (symmetric) linear relationships between the NAO index and nSAT or PR. So the associated

response of the variables to NAO changes may be quantified by a linear equation (Iles and Hegerl, 2017; Stephenson et al., 2006; Hurrell, 1995):

$$Y = \alpha_1 X + \alpha_0 + \varepsilon_Y \tag{4}$$

with Y being the (response) variable at a given grid cell that is partly explained by the NAO (X, the predictor) and by any other influences (ε_Y; Stephenson et al., 2006; von Storch and Zwiers, 2003). The coefficient α_1 was estimated on each grid cell using ordinary least squares regression with the R function lm (www.rdocumentation.org). It represents mean change in nSAT or PR that accompanies unit index change during the time period under consideration (Iles and Hegerl, 2017). The line offset α_0 in Eq. (4) equals the long-term mean. The α_1 coefficients may be computed with respect to normalized index series (von Storch and Zwiers, 2003), but in this study the non-normalized index time series was preferred in order to take into account the member-specific index units.

2.2.5 Addressing Internal Variability

Internal variability was understood as being represented by the oscillations around the long-term mean of the time series of a given variable (Hawkins and Sutton, 2011). In this point of view, IMS of the LE originates from the superposition of all 50 realizations with their respective inter-annual variability. As the climatic evolution of all 50 members is equally likely by construction of the ensemble, this spread represents an envelope of possible sequences of weather events at any given time step or location. This allows to sample internal variability at single points in time as the range of the members' values.
Therefore, the NAO–response relationship was analyzed individually for each GCM and RCM member (as is done e.g. in Woollings et al., 2015).
Aggregations to ensemble means (like in Deser et al., 2017) and standard deviations (sd, see also Leduc et al., 2019; Déqué et al., 2007), the latter representing the IMS in maps, were only performed for illustrating purposes in order to avoid masking model internal variability (Zwiers and von Storch, 2004).

3 Results

The result section is structured in two large parts: Section 3.1 deals with the representation of the NAO and climatic responses in the GCM and RCM and Section 3.2 targets internal variability in the GCM and RCM. Section 4 will follow the structure as defined by the four key questions.

3.1 NAO within the ClimEx Data Set

3.1.1 NAO index and SLP conditions

First, a reference NAO index was calculated from the ERA-I reanalysis. It is found to be in good accordance with often cited NAO indices like the time series of Hurrell (Pearson correlation of $r = 0.95$ with REF NAO index; index available at

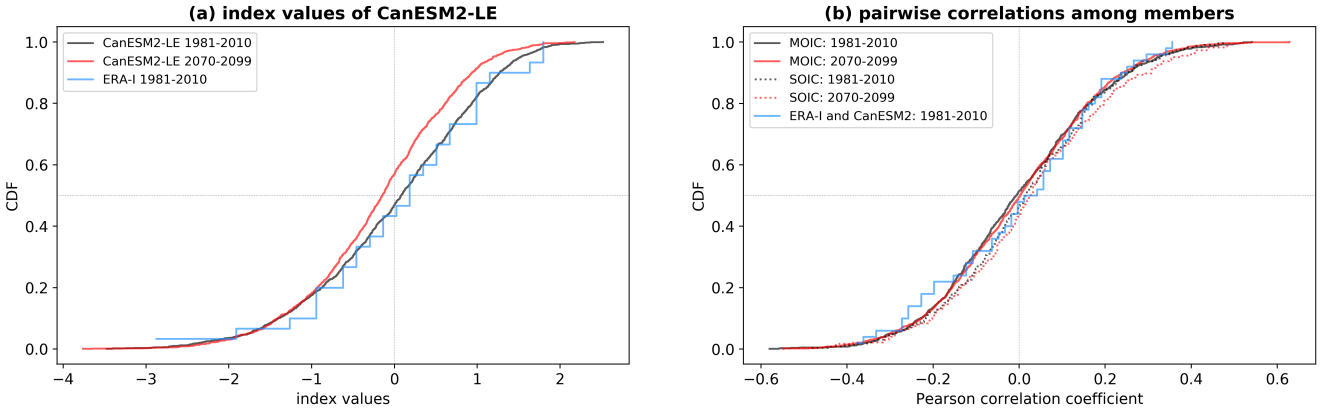

Figure 3. Cumulative density functions (CDFs) of NAO index values. (a) distribution of all CanESM2-LE ($n = 50 \times 30$ per period) and ERA-I ($n = 30$) NAO index values. (b) pairwise correlations among member NAO index time series from the same ocean families (SOIC – same ocean initial conditions, dotted lines, $n = 225$), from different ocean families (MOIC – mixed ocean initial conditions, solid lines, $n = 1000$) and between ERA-I and all CanESM2 members ($n = 50$). Black: 1981–2010 CanESM2-LE, red: 2070–2099 CanESM2-LE, blue: 1981–2010 ERA-I.

https://climatedataguide.ucar.edu/climate-data/hurrellnorth-atlantic-oscillation-nao-index-station-based). For further analyses it will therefore serve as a reference.

The CanESM2-LE produces NAO index values which follow a distribution similar to the ERA-I data (centered over zero, slight surplus of low positive NAO values, see Fig. 3 (a)). The CanESM2-LE distribution appears smoother due to a larger sample size ($n = 1500$ for CanESM2-LE and $n = 30$ for ERA-I). Maximum and minimum index values (x-axis in Fig. 3 (a)) of some of the 50 members exceed those of the REF realization; thus, the REF realization lies well within the ensemble IMS. The future NAO index shows a similar distribution of values, but with slightly less positive and more negative values (red curve in Fig. 3 (a)).

For further analyses on the IMS as a measure of internal variability, the independence of the 50 ensemble members is of high importance. To investigate independence among the ensemble members in both 30 year time frames, it seems favourable to analyse pairwise member correlations. Although zero correlations do not automatically imply independence, clear correlations among members would contradict the assumption of independence. In order to take into account the two perturbations during the production of the LE (1850 for 5 ocean families, 1950 for perturbations leading to 10 members per ocean family), these correlations were split in two groups like in Leduc et al. (2019): (i) correlations among the 10 members from the same ocean family (same ocean initial conditions in 1950, SOIC, $n = 225$, see dotted lines in Fig. 3(b)) and (ii) correlations between each member and the 40 members from the 4 other ocean families (mixed ocean initial conditions, MOIC, $n = 1000$, see solid lines in Fig. 3 (b)).

These correlations approximately follow a normal distribution with $\mu = 0$. There is a slight surmount of low positive correla-

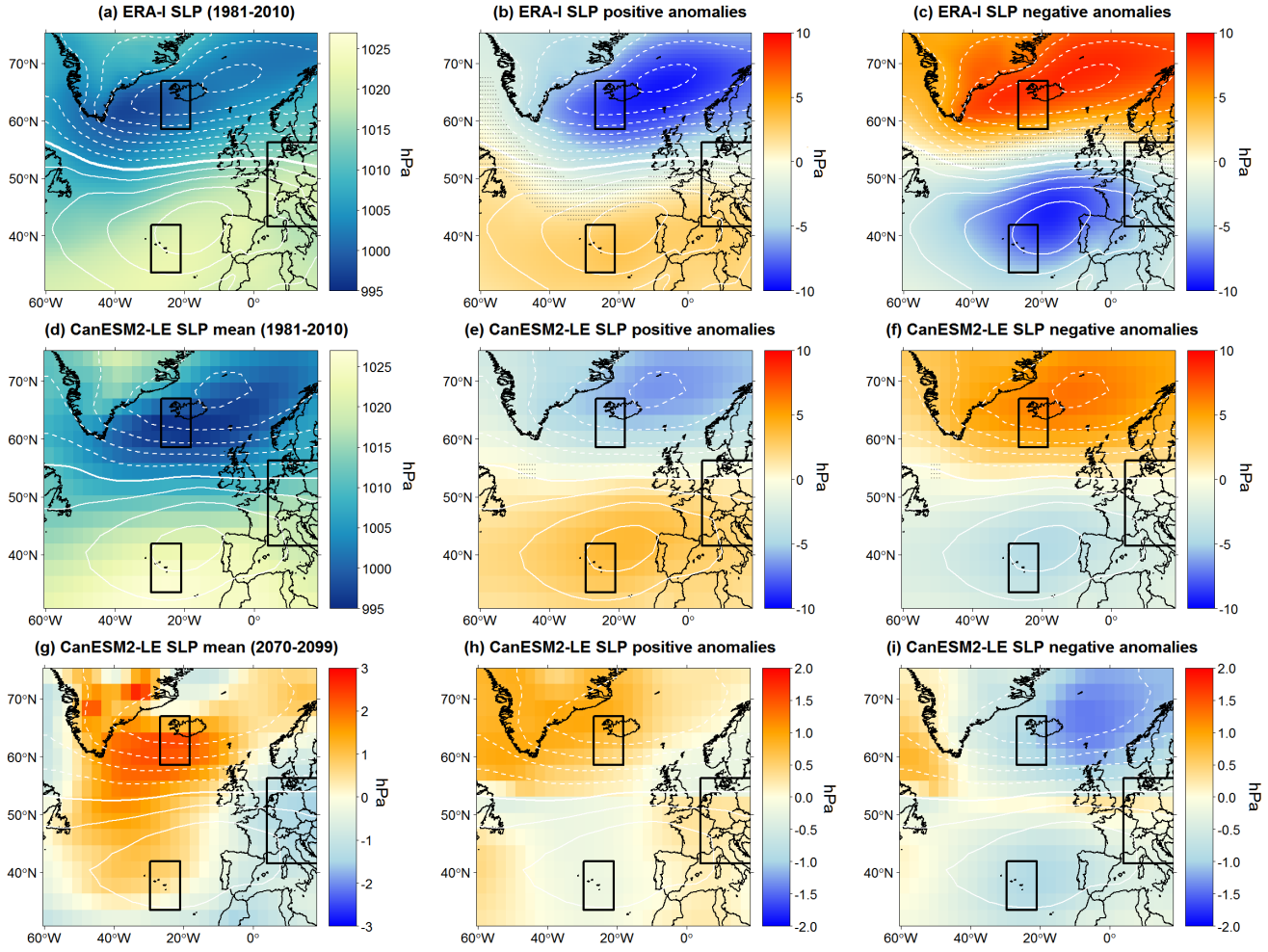

Figure 4. NAR winter mean SLP [hPa] composites in REF ((a)–(c)) and GCM ((d)–(i)) data showing long-term neutral conditions (left column), NAO positive (mid column) and negative anomalies (right column). (a)–(f): for 1981–2010. (g)–(i): 2070–2099 changes with respect to 1981–2010 in GCM data. White isolines: difference between positive and negative anomalies by a step of 2.50 hPa, as e.g. in Hurrell (1995), solid: positive, dashed: negative, bold line: zero. Stippling in subpanels (a)–(f): regions where the anomaly is smaller than the standard error of the composite samples. Black boxes: AH, IL and CEUR regions (see Fig. 2).

tions in the SOIC group compared to the MOIC group which is (not significantly) stronger in the fut time horizon (see red and
265 black dotted lines in Fig. 3 (b)). In general, the members are thus not seen as being dependent. As will be discussed below, the SLP pattern over the North Atlantic changes slightly in the future period. So the direct comparison between historical and future SOIC and MOIC correlations remains difficult. The members also show no systematic correlation with the REF NAO index despite similar statistics (see also Fig. 9). Thus, the ERA-I and GCM indices can be seen as not dependent realizations drawn from the same distribution.

In addition to the NAO index evaluation, Fig. 4 presents the large-scale SLP patterns in the NAR region during neutral, positive and negative NAO conditions. Positive (negative) index years are chosen, if the respective index value exceeds 1 (-1) as in Rogers (1984). The neutral conditions refer to the 30 year SLP average. Under neutral NAO conditions, the North Atlantic region is characterized by a pressure dipole. This structure is intensified and tilted clockwise in the CanESM2-LE ensemble mean (middle row of Fig. 4) compared to REF (top row). The mean SLP difference between the CanESM2-LE mean and REF reaches up to 10 hPa in both directions. SLP values are higher over Greenland and lower over the North Sea in the CanESM2-LE compared to ERA-I (panels (a), (d) in Fig. 4). Long-term neutral states of both data sources show robust signals in the entire NAR region (i.e. no stippling). This suggests that the different patterns in GCM and REF data are not singularly artefacts arising from different sample sizes. The GCM multi-member composites of positive and negative phases show less pronounced SLP anomalies than the REF data. This difference between GCM and REF may be due to the fact that REF composites were derived from $n = 3$ negative and $n = 4$ positive years whereas the GCM data provided $n = 264$ negative and $n = 263$ positive years during 1981–2010. Regions with strong sampling uncertainties, i.e. where the standard error is larger than the anomaly, are indicated with stippling in panels (a)–(f). These regions are mostly found in the transition region between the wider ERA-I AH and IL nodes, whereas the SLP anomalies at the NAO centres of action show less uncertainty. The GCM patterns are more robustly assessed (i.e. less prone to sampling uncertainty) as can be seen by the very small area with stippling in which the sign of the anomaly may not be assessed robustly.

The difference between SLP anomalies in positive and negative years representing the pressure variability is indicated by white lines. This difference is weaker in the CanESM2-LE mean than in ERA-I data, but located in similar regions. These NAO centres of action reach GCM (REF) SLP differences between positive and negative conditions of about 12.5 (17.5) hPa in the IL region and 7.5 (10.0) hPa in the AH region. They do not coincide with the highest and lowest SLP values in the neutral state, but are situated near the 3×3 GCM grid cell matrices used for index calculation. This is very promising as it supports the choice of these SLP centres for index calculation.

Under projected future climate conditions, SLP rises over large parts of the North Atlantic and shows less variability (see Fig. 4 (g)–(i)). Future positive phases tend to be weaker as SLP shows a marked increase in the northern NAO node region. Negative phases exhibit SLP decreases in both node regions, although with larger changes near IL, resulting in negative phases to become slightly weaker as well.

Having established a reasonably plausible representation of the NAO in the driving data, the next step is to evaluate the large-scale NAO pattern in the RCM data. Figure 5 maps the RMS* of the difference between driving data and RCM SLP during 1981–2010 for driving data ERA-I (a) and CanESM2 (b) and 2070–2099 (c). A value of RMS$^* \geq 1$ indicates that the root-mean squared error between the RCM and driving data is larger than the temporal variability in the driving data. In this case, the large-scale SLP pattern may not be seen as being correctly represented in the RCM data. The large-scale SLP pattern over the entire ClimEx domain, which also includes the CEUR, NE, BY and SE domains, is reasonably well represented: with RMS$^* < 1$ in most parts of the entire ClimEx domain for both driving data sets and both periods (significant at $p \leq 0.05$ using a t-test with a false detection rate < 0.1 to account for multiple hypothesis testing, see Wilks, 2016). All data sets show an RMS* increase towards the south, indicating that in these regions the control exerted by the lateral boundary conditions on the

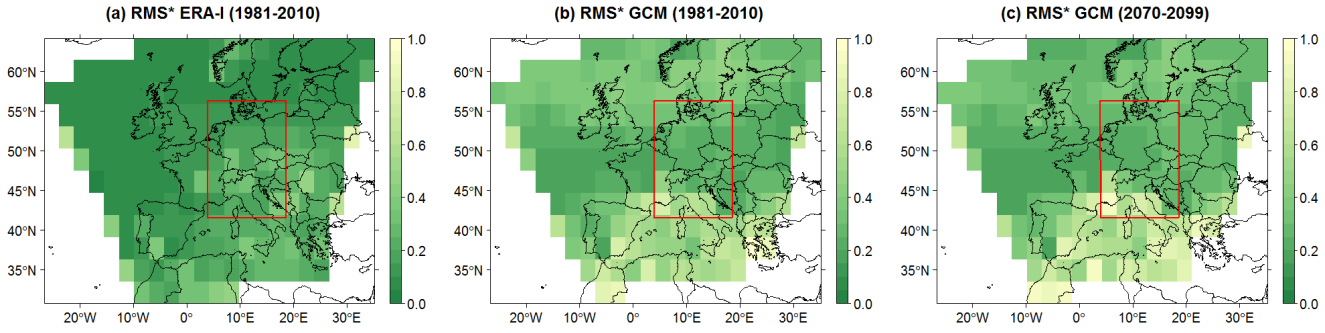

Figure 5. RMS* of monthly SLP differences between driving data and CRCM5 members, calculated following Eq. (2). Colouring: RMS* ≤ 1 significant at $p \leq 0.05$ with a false detection rate smaller than 0.1 (see Wilks, 2016). (a) for driving data set ERA-I (1981–2010), (b) for driving data set CanESM2 (1981–2010), (c) for driving data set CanESM2 (2070–2099). Red box: position of CEUR domain.

CRCM5 internal solution appears to be weaker. The RMS* is larger in the CanESM2/CRCM5 combination than in the ERA-I/CRCM5 combination, and slightly increases in the future period in the southern parts (see Fig. 5 (c)). In the CEUR domain (indicated as red box in Fig. 5), however, errors are low in general and therefore the NAO pattern of the driving data may be assumed to be correctly incorporated there. It is thus reasonable to continue with the evaluation of nSAT and PR responses in the CEUR domain.

3.1.2 Local climate response to the NAO

NAO nSAT and PR spatial responses as revealed in the ERA-I data are reproduced in their general properties under current climate conditions in the CanESM2-LE and CRCM5-LE (see Figs. 6–8). Responses with the highest α_1 magnitudes occur in the CRCM5/ERA-I run for all variables. Regarding the absolute α_1 values, the CRCM5-LE values meet the ERA-I values better than the CRCM5/ERA-I run.

Positive NAO conditions are accompanied by winters with warmer temperatures (up to $+2$ K per unit index change, see Fig. 6) and less day-to-day nSAT variability (see Fig. 7). The generally positive relationship between nSAT mean and NAO (see Fig. 6) is strongest in the north-eastern parts of the domain. Regionally, the NAO explains up to 40–60 % of nSAT mean variability (see also Fig. A2 where the nSAT mean α_1 share of the entire winter standard deviation of daily temperature values is shown). Explained variance is highest in the CRCM5/ERA-I run and lowest in the CanESM2-LE. The reduction of nSAT variability reaches up to 0.4–0.6 K in the northeastern continental section while it is near zero in the southern part of the domain.

In comparison to the neutral state, positive phases are also accompanied by more humid conditions in the north, and drier conditions in the south of the CEUR domain (see Fig. 8). The strength of the NAO–PR relationship, r, is not affected by topography in any of the models within the domain; only the pivotal line crossing Europe is following the Alpine ridges (see solid dark line in Fig. 8, panels (a)–(c) and (e)). The change between positive and negative r and α_1 occurs within a very narrow region. Within the CanESM2-LE, this zero-line is shifted northwards compared to ERA-I, CRCM5/ERA-I and CRCM5-LE. As is

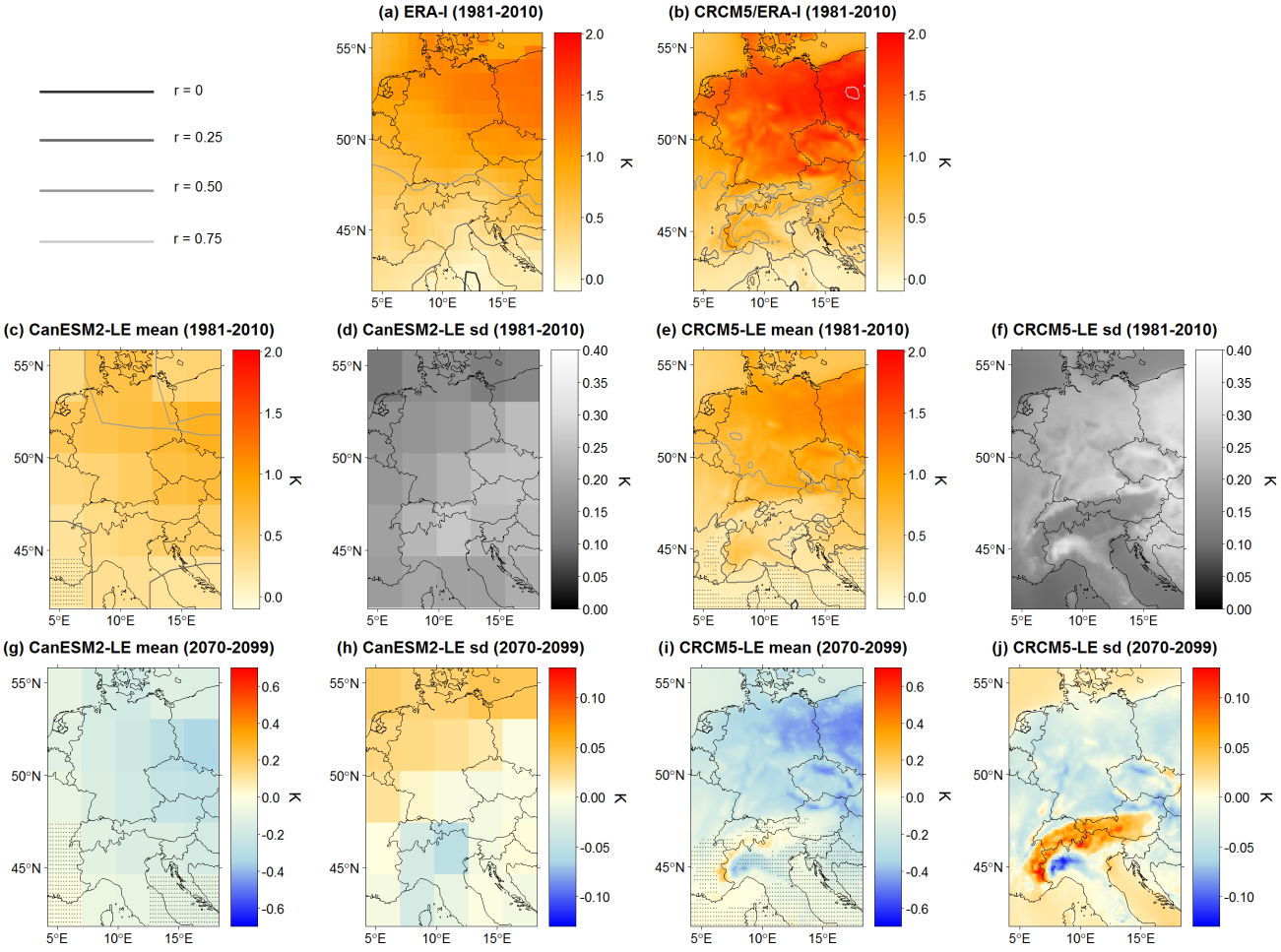

Figure 6. Spatial patterns of change in nSAT mean (in [K]) for a unit change in the NAO index for ERA-I, CRCM5/ERA-I, CanESM2-LE and CRCM5-LE in 1981–2010 ((a)–(f)) and the difference of 2070–2099 with respect to 1981–2010 ((g)–(j)). Both 50-member ensembles are represented with ensemble mean and sd representing the IMS. Grey lines in the ensemble mean maps represent the Pearson correlation between nSAT mean and the NAO index at an increment of 0.25; grey shadings see legend in upper left panel. Grey stippling in the ensemble mean maps show regions were $\mathrm{SNR} < 1$, SNR being the signal-to-noise ratio between the 30 year ensemble mean and sd of GCM and RCM LEs in both time periods.

visible in Fig. 8, higher α_1 values in mountainous regions indicate strong NAO responses related to orography. Regionally, the NAO accounts for 40–50 % of total PR sum variance, in both positively and negatively correlated regions. In the CRCM5-LE, single spots in mountainous regions (e.g., in the Dinaric Alps) show extremely high PR sum α_1 values (up to ± 220 mm per unit index change) where long-term mean PR sums are also very high. This stresses the more detailed production of geographical features, but also the tendency to evolve local extreme values in the high-resolution RCM (see similar results for local daily extreme precipitation in Leduc et al., 2019) which may even be noted in the (spatially aggregated) bias towards the GCM (see

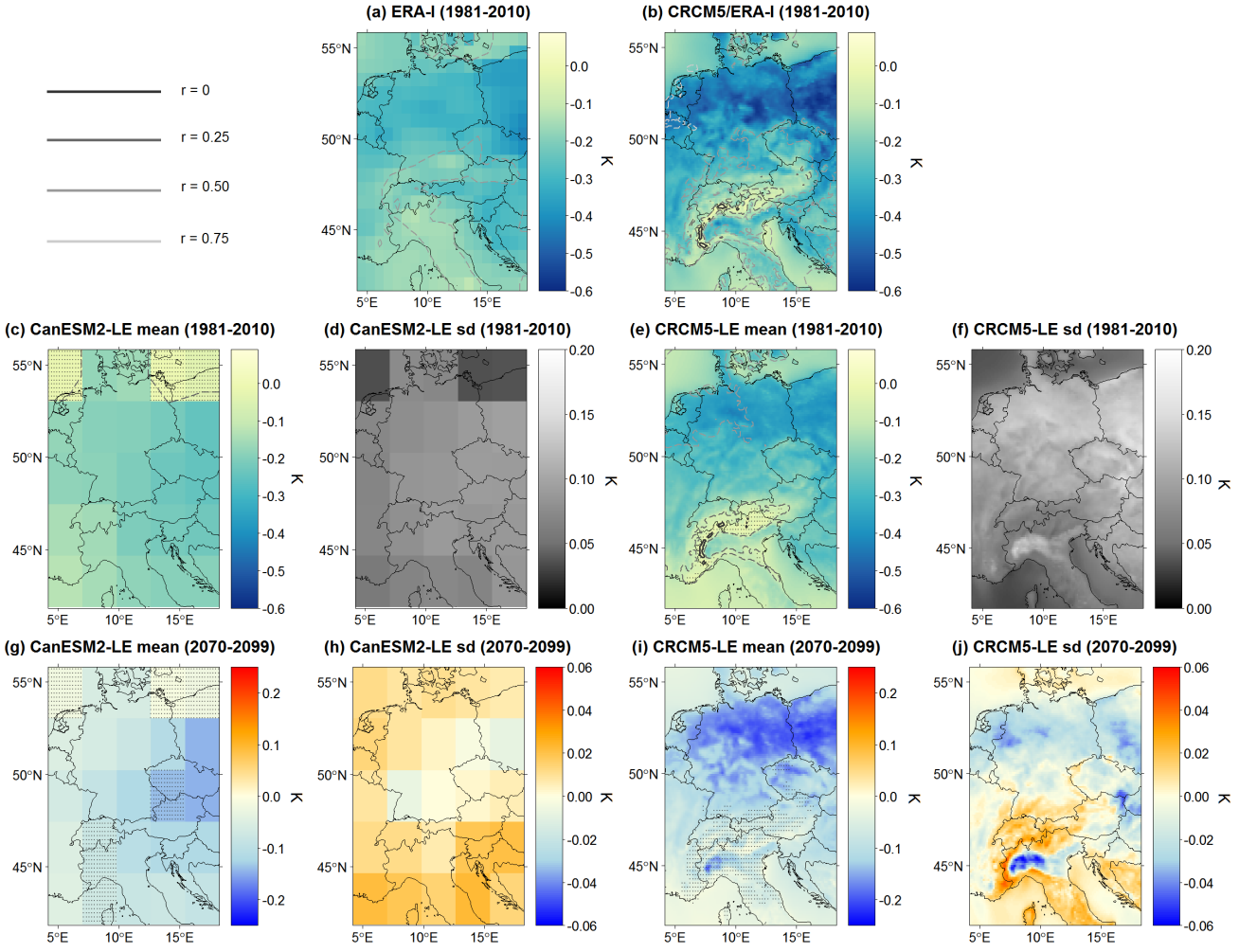

Figure 7. Like Fig. 6, but for nSAT std (in [K]). Note that the difference maps for CanESM2-LE and CRCM5-LE mean were calculated using absolute values.

Fig. A1 (f)). PR sum shows only weak correlations in the central region of the CEUR domain.

The mean state of nSAT and PR changes in the transient climate simulation towards warmer and moister conditions with less intra-seasonal variability of nSAT. For a detailed description of the future climate evolution (though for 2080–2099) in Europe within the CRCM5-LE see Leduc et al. (2019). Future NAO–climate relationships weaken in general compared to the historical ones for all variables. The spatial patterns of NAO-induced change do not change considerably between both periods. The response to the NAO, α_1, is clearly reduced in nSAT mean as is nSAT std, and there is also a reduction in PR sum change (panels (g), (i) in Figs. 6–8).

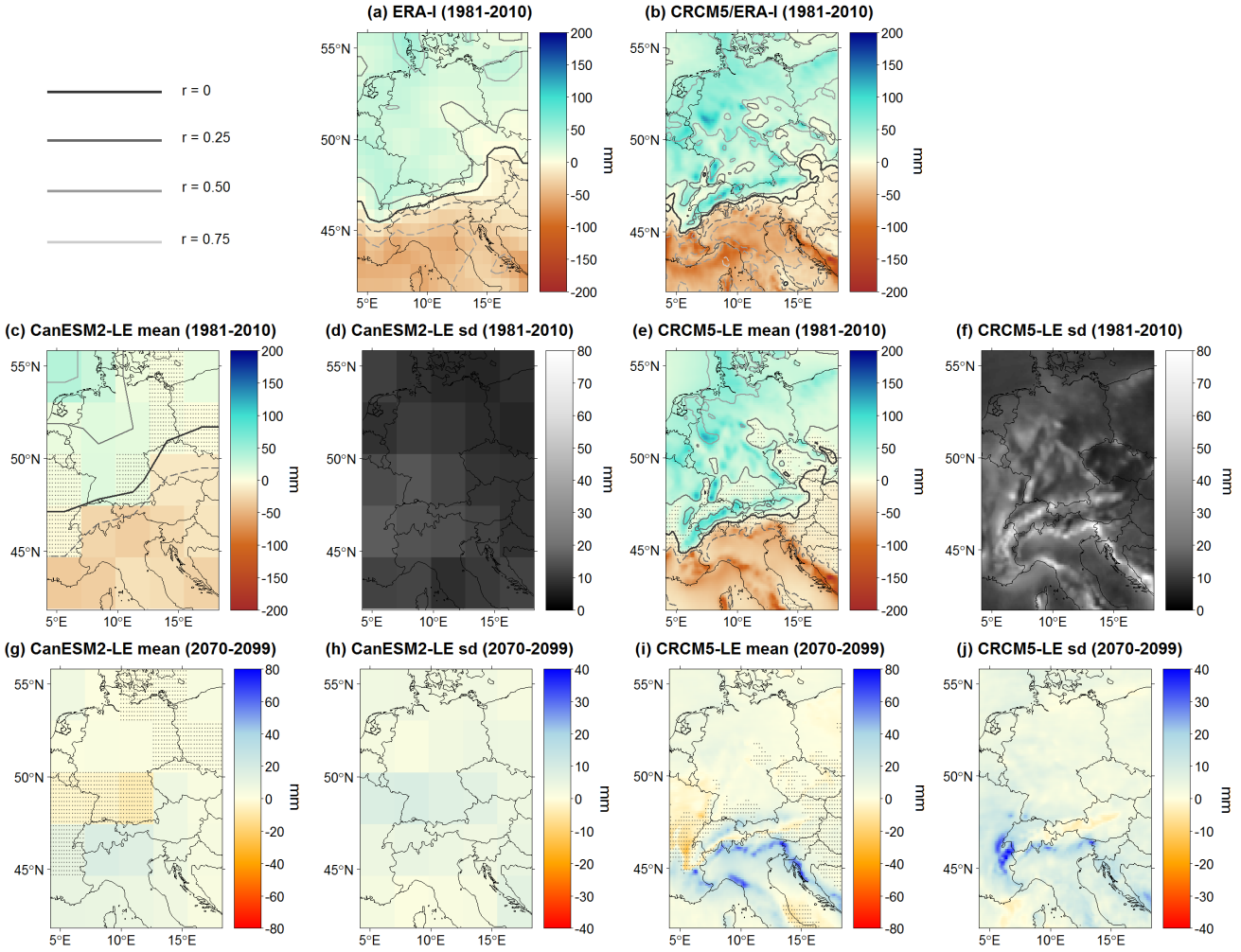

Figure 8. Like Figs. 6–7, but for PR sum (in [mm]). Note that the difference maps for CanESM2-LE and CRCM5-LE mean were calculated using absolute values and that the colour bar in the bottom row is flipped compared to Figs. 6–7.

3.2 Internal Variability at the GCM and RCM scale

The representation of internal variability in the GCM and RCM regarding the responses to the NAO in CEUR and subset regions NE, BY, SE is assessed via differences in the IMS of the CRCM5-LE compared to the CanESM2-LE.

3.2.1 Multi-member ensemble

The CanESM2-LE reproduces typical index characteristics: Fig. 9 summarizes several statistics for all 50 GCM members as multiples of the REF value. Generally, the ensemble meets the REF value in all aspects of the NAO index. However, some

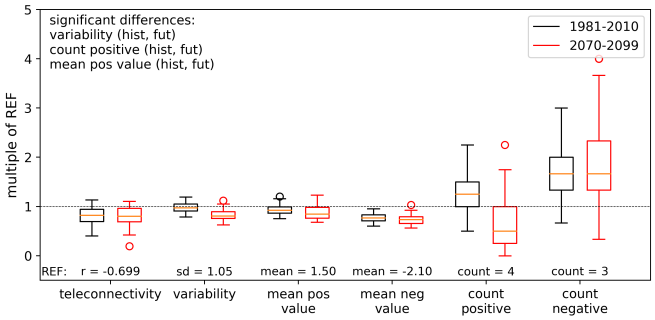

Figure 9. Several index statistics of all 50 CanESM2-LE members expressed as multiples of the respective ERA-I value (REF value set to 1.0): teleconnectivity (Pearson correlation between AH and IL time series), index variability (expressed as standard deviation in time of index time series), mean value of all positive (negative) phases and count of all positive (negative) phases in a single realization. Positive (negative) years are defined by an absolute index value exceeding 1. Text in upper left corner: significantly ($p \leq 0.05$, using an unpaired Mann-Whitney/U-test) different outcomes in the fut time frame.

GCM members only reach half of the REF teleconnectivity values (minimum: $r = -0.281$, not significantly different from zero at $p \leq 0.05$ using a t-test; REF $r = -0.699$). This finding is especially interesting as this metric quantifies the strength of the NAO within the individual members. The IMS of the teleconnection strength, though, does not change significantly over time, in spite of the SLP changes over the North Atlantic. The 2070–2099 NAO index exhibits less inter-annual variability, less positive phases, more neutral phases and a relative increase of negative phases but with reduced mean values (see also Fig. 3 (a)).

The spatial expression of NAO response internal variability in the form of diverging ensemble members can be derived from Figs. 6–8 (subplots (d), (f)) presenting spatially distributed ensemble sd as a measure for IMS. Largest deviations for nSAT mean are found in continental regions of CEUR, but they do not simply correspond to high or low α_1 (see also Fig. A3 (a)–(d)). Low IMS corresponds mostly to Alpine and sea regions. For nSAT mean, the signal-to-noise ratios (SNR) between ensemble mean and sd exceed 1 in most regions north of the Alps (see regions without stippling in Fig. 6). nSAT std shows SNR< 1 in the northern parts of the CanESM2-LE data (see Fig. 7 (c)) and in the Alpine region of the CRCM5-LE data (Fig. 7 (e)). This variable shows a strong linear relationship between LE mean and sd (Fig. A3 (e)–(h)). Regarding PR sum, RCM members vary most in regions with highest absolute α_1 values and altitudes, but there is no clear dependence in GCM (Fig. A3 (i)–(l)). For PR sum, there is an east-west corridor of SNR values below 1 which accompanies rather low α_1 values (see Fig. 8). In addition to future changes in the NAO responses ensemble means, there is also a change in the spatial distribution of the IMS expressed as ensemble sd (see subpanels (h), (j) in Figs. 6–8).

To further investigate the IMS, Fig. 10 illustrates the Pearson correlation coefficients r between the NAO index and subset regions nSAT mean or PR sum in GCM and RCM LEs separately. Both ensemble IMS generally envelope the REF value (dashed line) of the given region, apart from GCM hist in Fig. 10 (b). This finding does not change in the projected future climate: most boxes and whiskers keep their size, though GCM PR SE is characterized by a smaller range in the future and

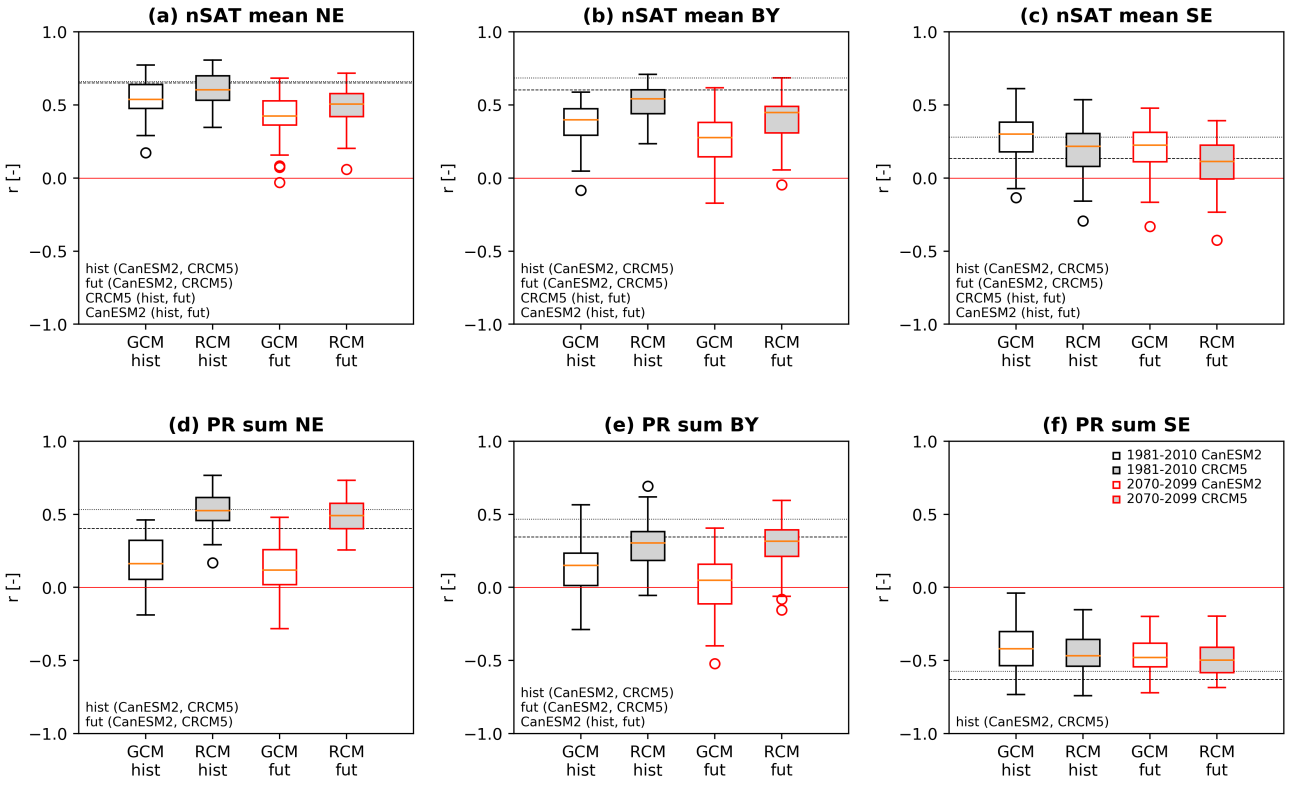

Figure 10. Boxplots of nSAT mean ((a)–(c)) and PR sum ((d)–(f)) showing Pearson correlation of 50 CanESM2-LE and CRCM5-LE (grey filling) realizations for three regions (NE, BY, SE) in historical (black outlines) and future (red outlines) time horizons. Dashed (dotted) horizontal lines indicate the ERA-I (CRCM5/ERA-I) value; text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann-Whitney/U-test for the comparison between hist and fut periods and a paired Wilcoxon test for the comparison between GCM and RCM LE.

GCM nSAT NE by a larger one (significant at $p \leq 0.10$ and $p \leq 0.05$, respectively, using an F-test for comparison of variances). Ensemble mean values exhibit a (sometimes significant) shift towards lower r values in the future for both models in some regions for nSAT mean and PR sum. An unpaired Mann-Whitney/U-test was applied here as the samples from hist and fut were seen as being drawn from different climates (using a χ^2-test, the null hypothesis of independence between hist and fut periods could not be rejected at $p \leq 0.05$).

3.2.2 Change of scales

Having analyzed GCM and RCM separately so far, it is now advised to compare both ensembles. A χ^2-test revealed that GCM and RCM samples of r can be seen as significantly dependent in both time frames. The amount of variance explained by the NAO is generally higher in REF than in the RCM ensemble mean, which in turn is higher than the GCM ensemble mean.

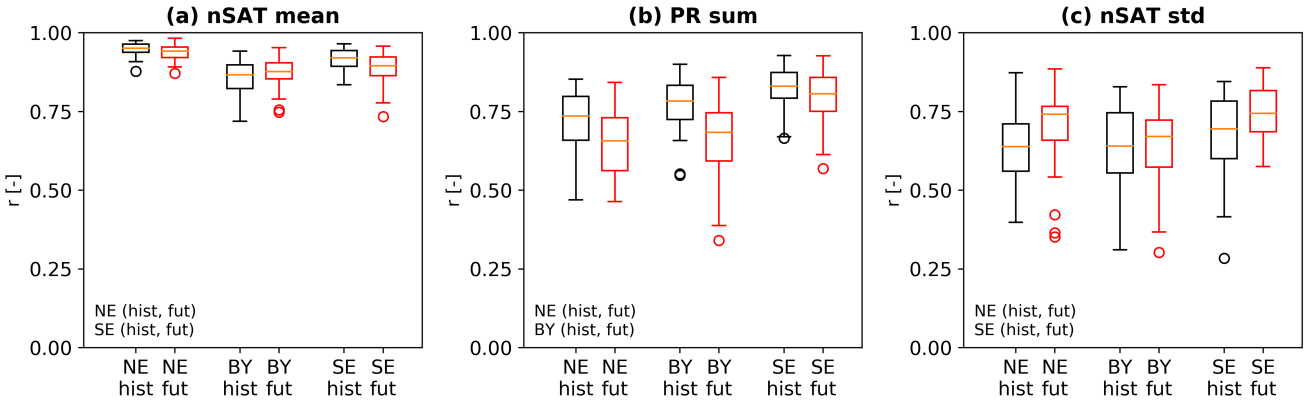

Figure 11. Temporal co-variability of CanESM2 and CRCM5 subset regions in all 50 members. Each boxplot represents 50 Pearson correlation coefficients between the time series of variables nSAT mean (a), PR sum (b) and nSAT std (c) in the subset regions of CanESM2 members and the corresponding CRCM5 members. Time periods used for correlations: 1981–2010 (hist, black), 2070–2099 (fut, red). For regions NE, BY, SE see Fig. 2. Text denotes combinations of which the differences are significant at $p \leq 0.05$ using an unpaired Mann-Whitney/U-test.

The CRCM5-LE enhances the relationship showing higher r and α_1 values than the CanESM2-LE (see Fig. 10 for r, where hist(CanESM2, CRCM5) or fut(CanESM2, CRCM5) is indicated; but also Figs. 6–8 for α_1). This enhancement by the CRCM5 is notably independent of the driving data: for both variables, the CRCM5/ERA-I r value (dotted lines in Fig. 10) is also found to be higher than the ERA-I value in most regions (dashed lines in Fig. 10). In all subset regions, the CRCM5/ERA-I r value lies in the upper part (stronger correlations) of the CRCM5-LE ensemble values.

Figure 10 shows that mean r values of RCM (grey filling) and GCM (hollow) members are significantly different in all subset regions for nSAT mean in both time horizons, but only in the NE and BY regions for PR sum; in SE, only weak differences between GCM and RCM PR sum r distributions are visible. In NE and BY this difference is expressed by higher r values in RCM data, whereas in the SE region lower r values are found in the RCM data (only for nSAT mean). Apart from PR sum in the NE region (both time horizons), no significant difference between the spread amplitudes of GCM and RCM is visible ($p \leq 0.05$, F-test). The IMS of the correlation between NAO and response variables is not generally altered during the nesting process.

To evaluate the co-variability of CanESM2 and CRCM5 data in the subset regions, time series of the response variables originating from both data sources were correlated member-wise (see Fyfe et al., 2017, for a similar approach). As can be seen in Fig. 11, highest accordance on average is reached for nSAT mean in both periods, indicating that CanESM2-LE and CRCM5-LE show very similar temporal variability for this variable. The co-variability of GCM and RCM time series is weaker for PR sum and nSAT std than for nSAT mean in both periods. Also, the IMS is larger for PR sum and nSAT std than for nSAT mean. This finding suggests that there is a larger discrepancy in portraying PR sum and nSAT std in the RCM with respect to the GCM compared to nSAT mean. The correlations between CanESM2 and CRCM5 subset regions are in general significantly lower under future climate conditions compared to the historical ones, apart from nSAT mean in BY, PR sum in SE and nSAT

std in BY (see text in Fig. 11). For nSAT std a shift of the distribution of r towards slightly larger values is visible. All variables exhibit a future IMS increase, though not all subset regions are affected (see e.g. nSAT mean BY or nSAT std SE in Fig. 11). This suggests that under future climate conditions a considerable reduction of GCM–RCM co-variability needs to be taken into account, at least for PR sum and (weaker) for nSAT mean.

4 Discussion

4.1 General performance of the model chain

The ClimEx climate data ensemble is able to reproduce an NAO-like pattern with realistic temporal and spatial characteristics over the North Atlantic and corresponding response patterns in CEUR. Ensemble mean information aggregates several realizations and thus differences towards the single REF realization may occur. However, results showed that the REF pattern may in general be seen as being "embedded" in the RCM or GCM IMS, implying that GCM, RCM and REF share comparable climate statistics.

Regarding temperature, Europe is commonly seen as divided into a region with positive NAO–response correlations in the north and negative correlations in the south (see e.g., Woollings et al., 2015). The first is found in the here presented results, the latter is not clearly visible in the chosen domain. nSAT std is correlated negatively with the NAO, pointing towards less temperature variability in winters with positive NAO phases, and a higher variability during negative phases. Correlations of PR sums and NAO are in accordance with the prevalence of large-scale (frontal) precipitation in winter which might be affected if the large-scale circulation is altered due to NAO impulses.

The strong SLP gradient under neutral NAO conditions over the North Atlantic noted in the CanESM2-LE though suggests an overestimation of the local atmospheric circulation with too strong westerlies. Similar model biases are widely reported (see e.g., Ruprich-Robert and Cassou, 2015; Stephenson et al., 2006; Reintges et al., 2017; Ulbrich et al., 2008). Since the NAO index was obtained from raw SLP data, it contains the contribution of the NAO, but possibly also of micro-climatic noise or other teleconnection patterns like the East Atlantic (EA) and the Scandinavian Pattern (SCA) which interact with the NAO and exert a notable control on the North Atlantic SLP gradient (Moore et al., 2013). Moore et al. (2013) investigated the contributions of the North Atlantic teleconnections NAO, EA and SCA in reanalysis data by separating them with empirical orthogonal functions. The authors found that the NAO accounts for about one third of winter SLP variability, and the second and third leading modes for roughly 20 % and 15 %, respectively (see also Comas-Bru and McDermott, 2014). Thus the results presented here may be seen as representing the superposition of these atmospheric modes.

The fidelity of NAO responses further depends on two aspects: (i) the goodness of representation of the large-scale NAO-related SLP pattern in CEUR and (ii) the strength of the linear relationship between the NAO and the response variables. The first point is addressed by a good representation of the SLP pattern in RCM data (see Fig. 5). The second point may be targeted by a combination of correlation analysis of the responses and the α_1 values: NAO responses in the CEUR domain of all data sets are most reliable in regions where a strong linear relationship between the NAO and the response variable may be as-

sumed. This may be the case if the correlation coefficient between the NAO index and the variable time series on the given grid cells is significantly different from zero. Linearity though does not apply under all conditions. For example particularly strong negative NAO phases with low-ice conditions in the Arctic coincide with cooling in Europe that is weaker than expected from a linear relationship due to an accompanying warming over Siberia (Screen, 2017). Low correlation values may also suggest that climate variability in these regions is only to a small fraction influenced by the NAO. In these cases, the NAO as expressed by the North Atlantic SLP gradient in this study is not the most important contributor and ε_Y in Eq. (4) is dominant.

Historical α_1 values (all data sources) are generally in accordance with observed composite anomalies (see also Fig. A4), but most so in regions with significant r. Thus, the future change of nSAT and PR per unit index change is most valid where r is high and where the signals of α_1 emerge from internal variability, i.e. SNR > 1. Of course, α_1 and composite maps are not identical, as on the one hand the average index value that accompanies nSAT and PR anomalies is not the same (± 1 for α_1, but $+1.498$ and -2.103 for REF composites, see Fig. 9). On the other hand, α_1 estimates a change which is singularly generated by the NAO index in a linear relationship, while composite maps originate from raw data which might include further influences (Eq. (4)).

4.2 Nesting approach

NAO response patterns are similar within the CanESM2-LE and CRCM5-LE, but some deviations remain due to differences in model parameterization and spatial resolution. Another possible explanation could be that the control exerted by CanESM2 through the CRCM5 LBC is insufficient, but this is unlikely given the relatively small CRCM5 domain implying stronger LBC control (Leduc and Laprise, 2009), in addition to the strong spectral nudging of large scales that was applied in the production of the CRCM5-LE (Leduc et al., 2019). Also, the large-scale SLP pattern over CEUR shows no large errors in the CRCM5-LE with respect to its driving data sources (see Fig. 5). Nevertheless, the influence of the lateral boundary conditions appears to vary over the CRCM5 domain, being a bit weaker in the southern part. It is worth noting that this feature is less pronounced when CRCM5 is driven by ERA-I as compared with CanESM2, highlighting the importance to investigate further the interactions between global atmospheric circulation, surface forcings (e.g., topography and land-sea contrasts) and local feedbacks. The CRCM5 reproduces the REF response structures much finer than the CanESM2 and adds some robust high resolution geographical features which are clearly visible within the ensemble mean.

Apart from the coarser pattern resolution, there is also a shift in the spatial climate patterns in the CanESM2-LE within the CEUR domain with respect to ERA-I data which is not found in the CRCM5-LE: for example, typical continental climate features, such as high nSAT variability (as indicated by Fig. 6), are shifted southwards in the CanESM2-LE with respect to CRCM5-LE data (or ERA-I). This shift may be explained by the fact that due to coarser spatial resolution the GCM topography shows land grid cells where the Mediterranean or the Baltic Sea extend in ERA-I and CRCM5; thus, in the GCM, the continent Europe also occupies a region which is sea in ERA-I. Assuming that the land–sea distribution affects the climate evolution, the GCM also experiences a geographical shift of climatic characteristics (such as continental properties) compared with the ERA-I and RCM data within the study domain. Another example is the dividing line for NAO–PR sum relations (see Fig. 8)

which shows a displacement in the GCM compared to the RCM. This displacement is related to the GCM orography which deviates due to the coarser spatial resolution in shape, position and height from the RCM orography. These findings suggest that similar responses of GCM and RCM to the NAO may not be visible at the same geographical location (i.e. coordinates), but under similar geographical conditions (exposition, altitude, distance to sea). Continuing this thought, the RCM reproducing the spatial climatic patterns in the "correct" location is another expression of the RCM added value for regional or local scale analyses. However, for general statements on this issue, analyses on a larger domain would be necessary.

On the regional scale, the correlations in the CRCM5 are significantly stronger in several regions than in the CanESM2 (see Fig. 6–8). These are not evened out by spatial aggregation (see Fig. 10). Thus, in the CRCM5-LE, more variance is explained by the NAO (i.e. by large-scale circulation) than in the CanESM2-LE. Explained variance is higher in the single realizations of ERA-I and CRCM5/ERA-I than in the ensemble mean of GCM and RCM.

4.3 Internal Variability

In general, the 50 NAO signals from the atmospheric "inflow" as given by the GCM boundary conditions are correctly translated into 50 regional responses of the RCM regarding the range of internal variability.

The large ensemble internal variability favours a smoothing of structures in the ensemble mean. However, as the ensemble mean (GCM and RCM) reproduces patterns very similar to the observed ones, the atmospheric dynamics behind can be regarded as correctly reproduced in all members.

When looking at spatially explicit ensemble sd maps (see Figs. 6–7 and A3), the RCM LE exhibits higher ensemble sd values than the GCM. This is in accordance with Giorgi et al. (2009) who stated that internal variability at finer scales tends to be larger compared to larger scales. However, the amplitude of the IMS of r in the aggregated RCM and GCM subset regions is similar. Thus, the range of internal variability regarding the strength of the NAO–response relationship is transferred during nesting and the CRCM5 added internal variability (Leduc et al., 2019) does not significantly alter it. However, the ensemble values are shifted towards significantly higher r values in the RCM compared to the GCM in both time frames, but not in the SE region.

When comparing present and future values, a vertical shift of the boxes in Fig. 10 indicates that r is reduced in the future, but the inter-quartile distance of the r distributions (box size) stays nearly the same for GCM and RCM. This shows that the uncertainty range of the signals does not change significantly in the future horizon.

Temporally constant or only negligibly varying internal variability was already found for global mean temperature in Hawkins and Sutton (2009) and assumed for global mean precipitation in Hawkins and Sutton (2011). With the here presented results, it can also be argued that internal variability of more complex parameters (such as the NAO–response relationship quantified via Pearson correlation) shows no siginificant changes between historical and future periods.

When looking at the spatial distribution of α_1 sd however, several regions show slight future increases or decreases which are not necessarily consistent between GCM and RCM. Also, the potential time-dependent evolution of the IMS in the course of the analyzed periods is not taken into account.

It has to be added that this study evaluated two 30-year blocks rather than continuous time series, treating the NAO–response relationship as stationary during these blocks such that the IMS of both periods represents generalized conditions for 1981–2010 and 2070–2099. According to Comas-Bru and McDermott (2014), potential non-stationarity in NAO–response relationships can at least partly be attributed to influences of the EA/SCA patterns on the NAO, and especially the geographical position of the North Atlantic SLP gradient.

4.4 Climate Change

The results showed that historical and projected future climate statistics deviate such that the comparison of relationships in both periods remains difficult: the NAO pattern changes, NAO index variability and nSAT and PR responses are reduced in the future climate simulation.

The relative prevalence of negative index phases occurs in correspondence to a generally strengthened high pressure ridge over the North Atlantic and especially Greenland (see Fig. 4 (g)). The latter feature is supposed to be related with the emergence of negative index phases (Hanna et al., 2015; Woollings et al., 2010; Gillett and Fyfe, 2013; Cattiaux et al., 2013; Screen, 2017). Another relationship ties the emergence of negative NAO index phases to reduced sea ice extents: Warner (2018) found that particularly October sea ice extent over the Barents/Kara Sea is positively correlated with the NAO in that it leads to strengthened IL and AH. Consequently, a reduced sea ice extent leads to negative NAO phases, but this relationship is not simply linear (Warner, 2018). For example, Screen (2017) note that negative NAO events tend to be stronger during winters with low sea ice extents. The NAO-sea ice relationship may follow from sea ice effects on the stratospheric polar vortex or from tropospheric Arctic amplification which reduces the meridional temperature gradient leading to a weakened, more wavy jetstream in the mid-latitudes (Warner, 2018). The CanESM2-LE is known to show a low bias regarding Arctic sea ice in all seasons compared to observations (Kushner et al., 2018), but it follows quite correctly the observed downward trend (Kirchmeier-Young et al., 2017) and leads to a clear reduction of sea ice in the 2070–2099 horizon compared to 1981–2010 in the entire Arctic and also the Barents/Kara Sea as was verified with the CanESM2 variable “sea ice concentration” (not shown).

An increasing frequency (relative to positive phases) of negative NAO events as noted in Fig. 9 favours more cold and harsh winters in theory due to the advection of continental Eurasian air masses (Screen, 2017) which is in great contradiction to projected future background conditions (warmer, moister, see Leduc et al., 2019) that would rather, likewise following from theory, accompany positive phases. On the other hand, the response to NAO impulses is clearly reduced for nSAT mean, PR sum and nSAT std. A coherent explication for this discrepancy might be that as correlations weaken, the Eurasian influence (advection of cold, dry airmasses) during negative phases may be repressed or weaker in its occurrence than now or, as indicated by Screen (2017), is actually increasing warmer air mass advection. As less nSAT and PR variance is explained by the NAO in the future climate projections than in the historical period, the influence of this climate mode on CEUR climate may be seen as potentially reduced.

5 Conclusions

In this study, a RCM single-model initial condition large ensemble was analyzed with a special focus on the downscaled responses to a teleconnection, the NAO, that is present in the driving data. For proper assessment, the driving GCM ensemble was also included in the study. Referring to the key questions raised in the introduction, it can be stated that:

(a) Both large ensembles within the ClimEx project climate model chain are able to depict a robust NAO pattern under current forcing conditions. Each member represents a distinct climate evolution while sharing comparable statistics with all other 49 realizations and producing NAO and response patterns that are more robust than patterns of single realizations. The ensemble also shows comparable climate statistics with the REF time series and patterns. The clearly visible connection of the NAO with nSAT mean and PR sum follows well-known patterns. The influence of the NAO on nSAT variability, as expressed by the analyses on nSAT std, is also remarkable.

(b) The RCM is able to reproduce the large-scale SLP pattern and realistic response patterns in the analyzed domain. Clearly more topographic features are visible in the CRCM5-LE than in the CanESM2-LE which suggests added value of the RCM regarding the evaluation of small-scale NAO impacts. Deviations of nSAT and PR responses between members vary spatially within the domain and are found mostly in regions with strongest NAO responses.

(c) Internal variability of the NAO pattern is expressed very well within the 50 member single-model ensemble, and easily spans the observations regarding various indicators. The range of NAO responses is represented consistently between the driving GCM and the nested RCM. The spread is shifted towards stronger NAO–nSAT/PR relations in the RCM compared to the GCM in both time horizons.

(d) Concerning climate change, several changes go hand in hand: the winter index variability is reduced, the overall winter variability of nSAT and PR and also the fraction of NAO-explained nSAT is reduced, the relationship between NAO and response variables is weakened, the RMS* error regarding the large-scale SLP pattern between GCM and RCM slightly increases, and the co-variability of CanESM2 and CRCM5 subset regions for all weather variables is reduced.

While these results are especially valid for the ClimEx data sets, they allow drawing some general conclusions. The results strengthen the validity of the climate module for further applications, as important large-scale teleconnections only present in the GCM propagate properly to the fine-scale dynamics in the RCM. The RCM does not alter the spread of driving GCM data which is a valuable information for impact modelling with a focus on internal variability. The results also stress the importance of single-model ensembles for evaluating and estimating internal variability since single realizations show considerable variations among themselves and also deviations from the ensemble mean. So the ensemble mean and the ensemble spread together are needed for robust assessment of climate modes and whether a given model is able to reproduce the phenomenon of interest.

Data availability. Data used in this study may be retrieved from the following sources:

CanESM2-LE data is available via https://open.canada.ca/data/en/dataset/aa7b6823-fd1e-49ff-a6fb-68076a4a477c.

CRCM5-LE data can be retrieved at https://climex-data.srv.lrz.de/Public/

ERA-Interim Reanalysis data set was obtained at https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/

565

Appendix A

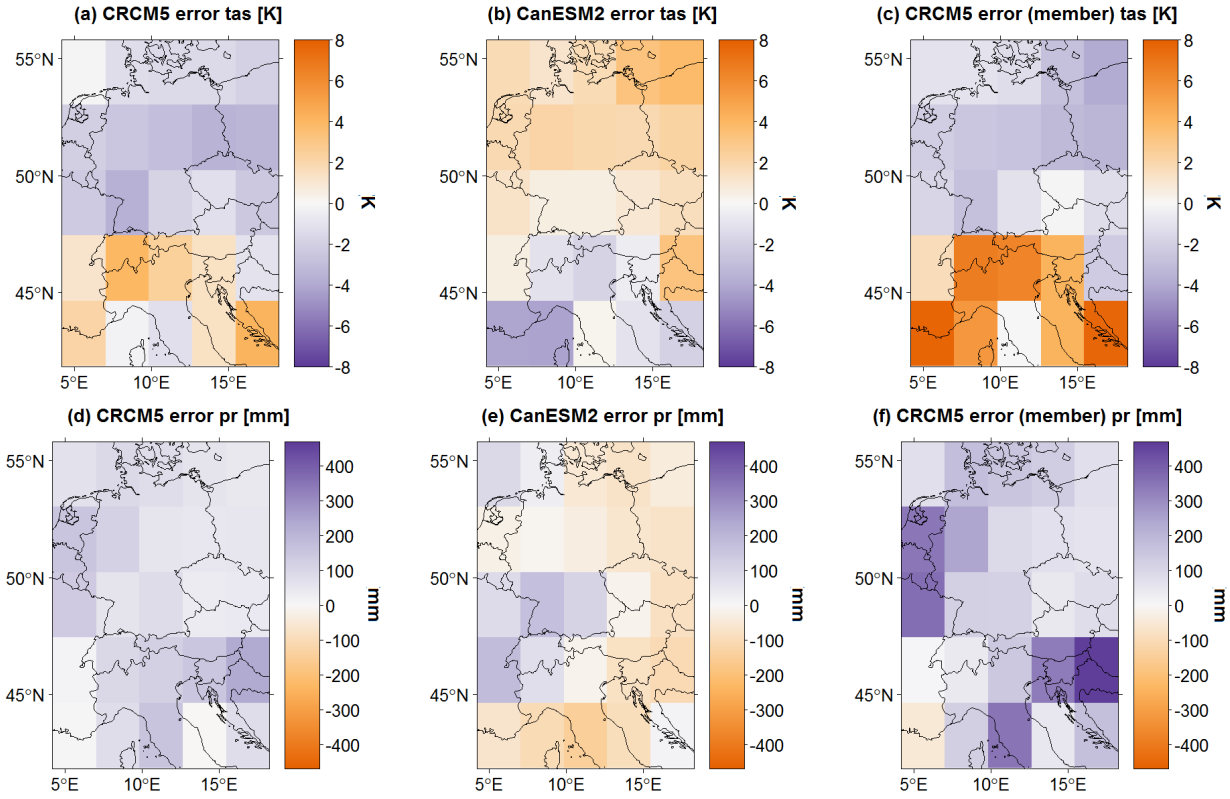

Figure A1. Model errors for 30 winter mean nSAT mean ((a)–(c)) and 30 winter mean PR sum ((d)–(f)) in GCM resolution (2.8°). First column: error of CRCM5 under "perfect"/ERA-I boundary conditions (difference between CRCM5/ERA-I and ERA-I). Second column: error of GCM towards ERA-I data (ensemble mean of differences between GCM members and ERA-I). Third column: CRCM5 error under GCM boundary conditions (ensemble mean of differences between CRCM5 members and corresponding CanESM2 members).

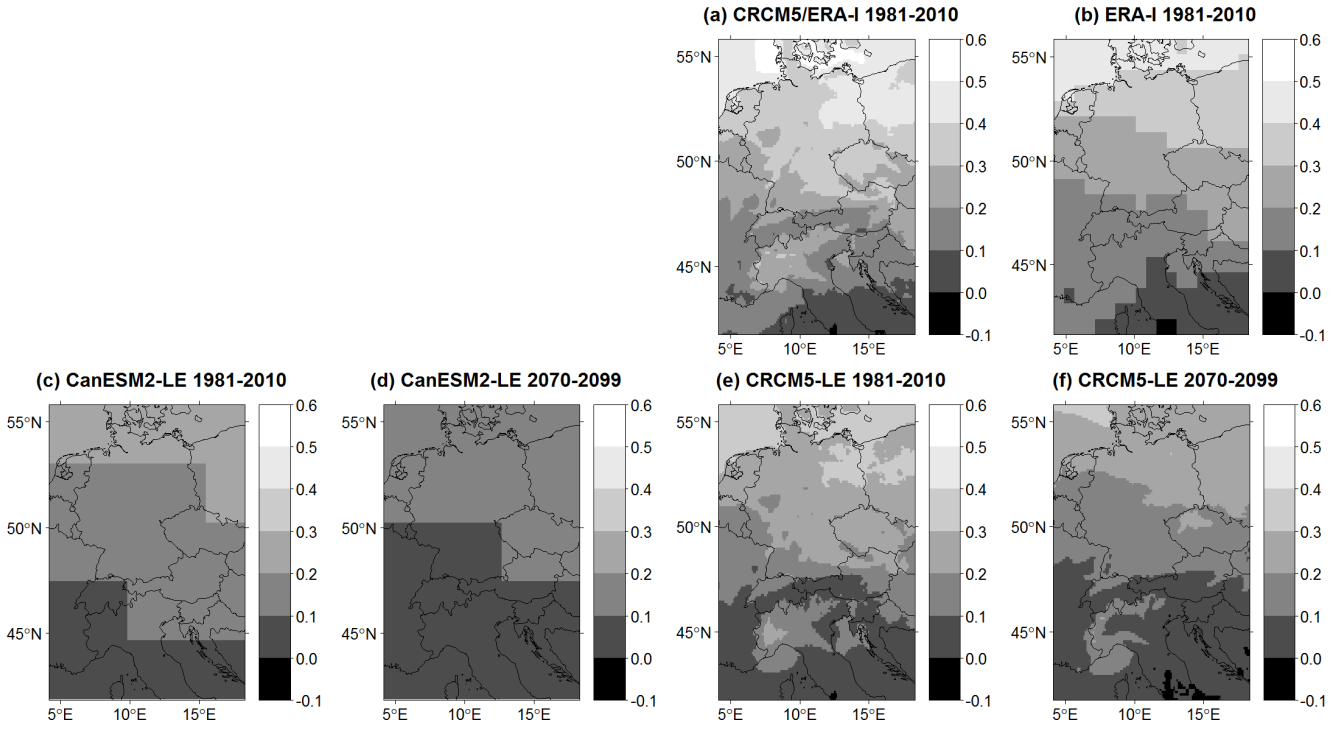

Figure A2. Ratio of nSAT α_1 and winter mean daily standard deviation of nSAT for CRCM5/ERA-I (a) and ERA-I (b) under historical conditions and CanESM2-LE mean ((c)–(d)) and CRMC5-LE mean ((e)–(f)) under historical and future climate conditions. The panels show the fraction of nSAT α_1 on winter mean daily standard deviation of nSAT.

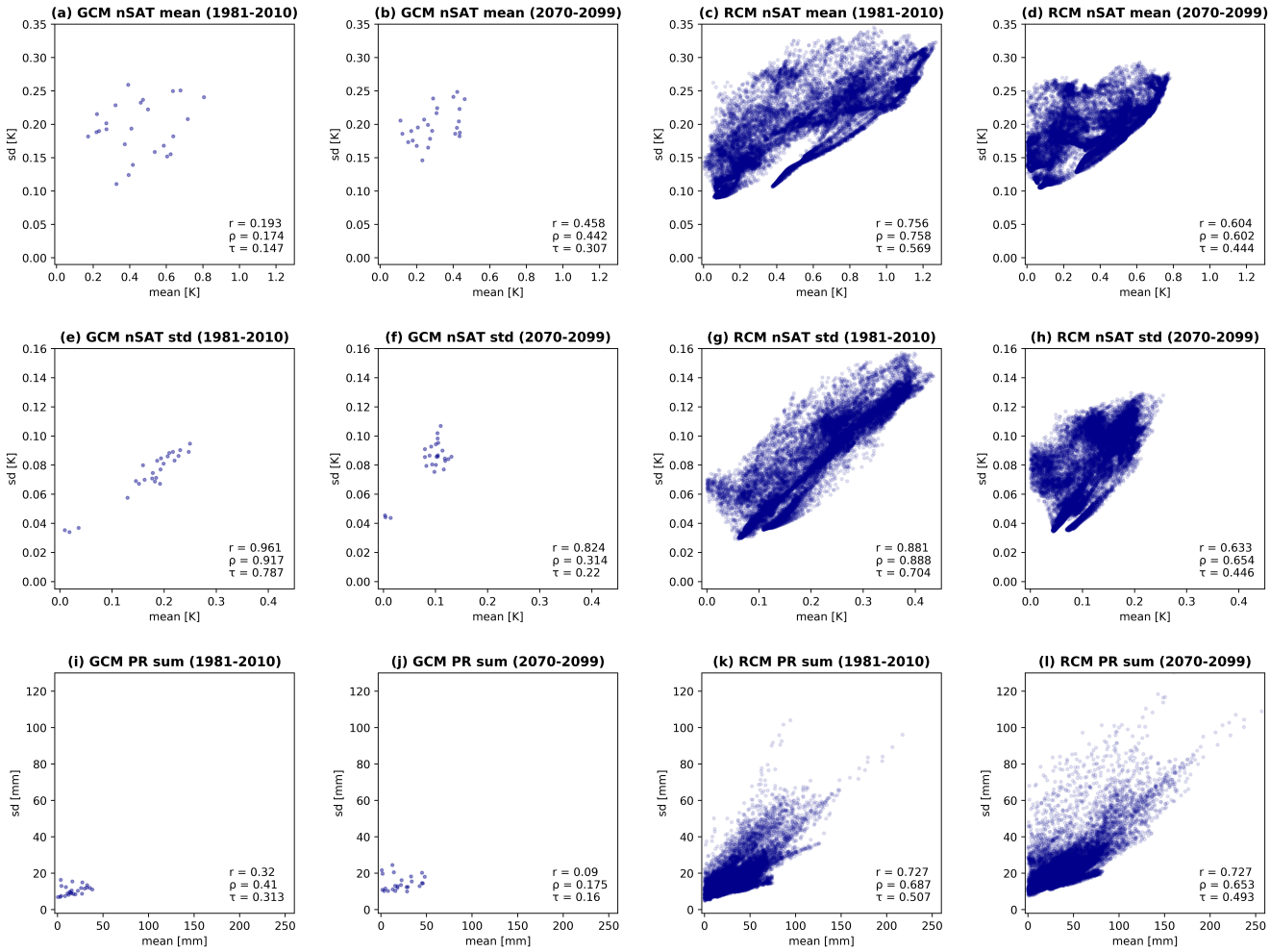

Figure A3. Relationship between LE mean and sd values for variables nSAT mean (a)–(d), nSAT std (e)–(h), PR sum (i)–(l) for hist and fut time frames. Upper right corner: r – Pearson correlation coefficient, ρ – Spearman rank correlation coefficient, τ – Kendall's Tau.

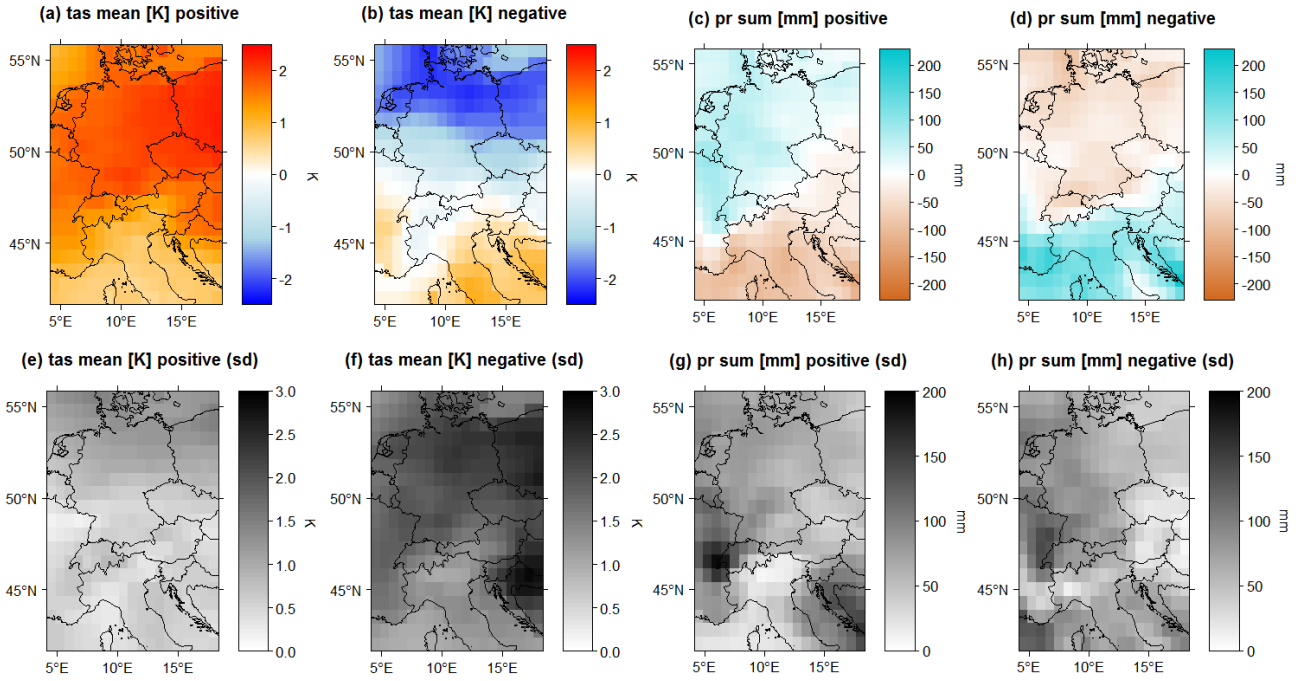

Figure A4. ERA-I anomalies from the long-term mean of nSAT mean in [K] and PR sum in [mm] in NAO positive (1989, 1990, 1994, 1995) and negative (1996, 2001, 2010) winters. Mean index value for positive (negative) NAO phases is $+1.498$ (-2.103).

Author contributions. This study was conceptualized by AB under supervision of RL. Formal analysis, visualization of results and writing of the original draft was performed by AB. All authors contributed to the interpretation of the findings and revision of the paper.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. The authors would like to thank two anonymous reviewers for their valuable input on this work. The production of ClimEx was funded within the ClimEx project by the Bavarian State Ministry for the Environment and Consumer Protection. The CRCM5 was developed by the ESCER centre of Université du Québec à Montréal (UQAM; www.escer.uqam.ca) in collaboration with Environment and Climate Change Canada. We acknowledge Environment and Climate Change Canada's Canadian Centre for Climate Modelling and Analysis for executing and making available the CanESM2 Large Ensemble simulations used in this study, and the Canadian Sea Ice and Snow Evolution Network for proposing the simulations. Computations with the CRCM5 for the ClimEx project were made on the SuperMUC supercomputer at Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities. The operation of this supercomputer is funded via the Gauss Centre for Supercomputing (GCS) by the German Federal Ministry of Education and Research and the Bavarian State Ministry of Education, Science and the Arts.

References

Andrews, M., Knight, J., and Gray, L.: A simulated lagged response of the North Atlantic Oscillation to the solar cycle over the period 1960–2009, Environmental Research Letters, 10, 054 022, https://doi.org/10.1088/1748-9326/10/5/054022, https://doi.org/10.1088%2F1748-9326%2F10%2F5%2F054022, 2015.

Benedict, J., Lee, S., and Feldstein, S.: Synoptic View of the North Atlantic Oscillation, Journal of the Atmospheric Sciences, 61, 121–144, https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2, https://doi.org/10.1175/1520-0469(2004)061<0121:SVOTNA>2.0.CO;2, 2004.

Cattiaux, J., Douville, H., and Peings, Y.: European temperatures in CMIP5: origins of present-day biases and future uncertainties, Climate Dynamics, 41, 2889–2907, https://doi.org/10.1007/s00382-013-1731-y, https://doi.org/10.1007/s00382-013-1731-y, 2013.

Christensen, J., Krishna Kumar, K., Aldrian, E., An, S.-I., Cavalcanti, I., de Castro, M., Dong, W., Goswami, P., Hall, A., Kanyanga, J., Kitoh, A., Kossin, J., Lau, N.-C., Renwick, J., Stephenson, D., Xie, S.-P., and Zhou, T.: Climate Phenomena and their Relevance for Future Regional Climate Change, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., pp. 1271–1308, Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2013.

Comas-Bru, L. and McDermott, F.: Impacts of the EA and SCA patterns on the European twentieth century NAO–winter climate relationship, Quarterly Journal of the Royal Meteorological Society, 140, 354–363, https://doi.org/10.1002/qj.2158, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2158, 2014.

Cropper, T., Hanna, E., Valente, M., and Jónsson, T.: A daily Azores-Iceland North Atlantic Oscillation index back to 1850, Geoscience Data Journal, 2, 12–24, https://doi.org/10.1002/gdj3.23, 2015.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Monge-Sanz, B., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Delworth, T., Zeng, F., Vecchi, G., Yang, X., Zhang, L., and Zhang, R.: The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere, Nature Geoscience, 9, 509–512, https://doi.org/10.1038/ngeo2738, https://doi.org/10.1038/ngeo2738, 2016.

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., de Castro, M., and van den Hurk, B.: An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections, Climatic Change, 81, 53–70, https://doi.org/10.1007/s10584-006-9228-x, 2007.

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

Deser, C., Hurrell, J., and Phillips, A.: The role of the North Atlantic Oscillation in European climate projections, Climate Dynamics, 49, 3141–3157, https://doi.org/10.1007/s00382-016-3502-z, https://doi.org/10.1007/s00382-016-3502-z, 2017.

Fyfe, J., Derksen, C., Mudryk, L., Flato, G., Santer, B., Swart, N., Molotch, N., Zhang, X., Wan, H., Arora, V., Scinocca, J., and Jiao, Y.: Large near-term projected snowpack loss over the western United States, Nature Communications, 8, https://doi.org/10.1038/ncomms14996, https://doi.org/10.1038/ncomms14996, 2017.

Gillett, N. and Fyfe, J.: Annular mode changes in the CMIP5 simulations, Geophysical Research Letters, 40, 1189–1193, https://doi.org/10.1002/grl.50249, https://doi.org/10.1002/grl.50249, 2013.

Giorgi, F., Jones, C., and Asrar, G. R.: Addressing climate information needs at the regional level: the CORDEX framework, WMO Bulletin, 58, 175–183, https://public.wmo.int/en/bulletin/addressing-climate-information-needs-regional-level-cordex-framework, 2009.

Hanna, E., Cropper, T., Jones, P., Scaife, A., and Allan, R.: Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index, International Journal of Climatology, 35, 2540–2554, https://doi.org/10.1002/joc.4157, https://doi.org/10.1002/joc.4157, 2015.

Hansen, F., Greatbatch, R., Gollan, G., Jung, T., and Weisheimer, A.: Remote control of North Atlantic Oscillation predictability via the stratosphere, Quarterly Journal of the Royal Meteorological Society, 143, 706–719, https://doi.org/10.1002/qj.2958, https://doi.org/10.1002/qj.2958, 2017.

Hawkins, E. and Sutton, R.: The Potential to Narrow Uncertainty in Regional Climate Predictions, Bulletin of the American Meteorological Society, 90, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hawkins, E. and Sutton, R.: The Potential to narrow uncertainty in projections of regional precipitation Change, Climate Dynamics, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, https://doi.org/10.1007/s00382-010-0810-6, 2011.

Hurrell, J. W.: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation, Science, 269, 676–679, https://doi.org/10.1126/science.269.5224.676, 1995.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: The role of the North Atlantic Oscillation, Journal of Marine Systems, 78, 28–41, https://doi.org/https://doi.org/10.1016/j.jmarsys.2008.11.026, 2009.

Hurrell, J. W. and Van Loon, H.: Decadal variations in climate associated with the North Atlantic Oscillation, Climatic Change, 36, 301–326, https://doi.org/10.1023/A:1005314315270, https://doi.org/10.1023/A:1005314315270, 1997.

Iles, C. and Hegerl, G.: Role of the North Atlantic Oscillation in decadal temperature trends, Environmental Research Letters, 12, 114 010, https://doi.org/10.1088/1748-9326/aa9152, https://doi.org/10.1088/1748-9326/aa9152, 2017.

Jones, P., Osborn, T., and Briffa, K.: Pressure-Based Measures of the North Atlantic Oscillation (NAO): A Comparison and an Assessment of Changes in the Strength of the NAO and its Influence on Surface Climate Parameters, in: The North Atlantic Oscillation: Climatic Significance and Environmental Impact, edited by Hurrell, J., Kushnir, Y., Ottersen, G., and Visbeck, M., pp. 51–62, American Geophysical Union (AGU), Washington, D.C., https://doi.org/10.1029/134GM03, 2013.

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

Kirtman, B., Power, S., Adedoyin, J., Boer, G., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A., Kimoto, M., Meehl, G., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G., Vecchi, G., and Wang, H.: Near-term Climate Change: Projections and Predictability, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., pp. 953–1028, Cambridge University Press, Cambridge, United Kingdom and New York, USA, 2013.

Kushner, P. J., Mudryk, L. R., Merryfield, W., Ambadan, J. T., Berg, A., Bichet, A., Brown, R., Derksen, C., Déry, S. J., Dirkson, A., Flato, G., Fletcher, C. G., Fyfe, J. C., Gillett, N., Haas, C., Howell, S., Laliberté, F., McCusker, K., Sigmond, M., Sospedra-Alfonso, R., Tandon, N. F., Thackeray, C., Tremblay, B., and Zwiers, F. W.: Canadian snow and sea ice: assessment of snow, sea ice, and related climate processes in Canada's Earth system model and climate-prediction system, The Cryosphere, 12, 1137–1156, https://doi.org/10.5194/tc-12-1137-2018, https://www.the-cryosphere.net/12/1137/2018/, 2018.

Leduc, M. and Laprise, R.: Regional climate model sensitivity to domain size, Climate Dynamics, 32, 833–854, https://doi.org/10.1007/s00382-008-0400-z, https://doi.org/10.1007/s00382-008-0400-z, 2009.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G., Giguère, M., Brissette, F., Turcotte, R., Braun, M., and Scinocca, J.: The ClimEx Project: A 50-Member Ensemble of Climate Change Projections at 12-km Resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5), Journal of Applied Meteorology and Climatology, 58, 663–693, https://doi.org/10.1175/JAMC-D-18-0021.1, https://doi.org/10.1175/JAMC-D-18-0021.1, 2019.

Moore, G., Renfrew, I., and Pickart, R.: Multidecadal Mobility of the North Atlantic Oscillation, Journal of Climate, 26, 2453–2466, https://doi.org/10.1175/JCLI-D-12-00023.1, https://doi.org/10.1175/JCLI-D-12-00023.1, 2013.

Osborn, T.: Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing, Climate Dynamics, 22, 605–623, https://doi.org/10.1007/s00382-004-0405-1, https://doi.org/10.1007/s00382-004-0405-1, 2004.

Pokorná, L. and Huth, R.: Climate impacts of the NAO are sensitive to how the NAO is defined, Theoretical and Applied Climatology, 119, 639–652, https://doi.org/10.1007/s00704-014-1116-0, https://doi.org/10.1007/s00704-014-1116-0, 2015.

Reintges, A., Latif, M., and Park, W.: Sub-decadal North Atlantic Oscillation variability in observations and the Kiel Climate Model, Climate Dynamics, 48, 3475–3487, https://doi.org/10.1007/s00382-016-3279-0, https://doi.org/10.1007/s00382-016-3279-0, 2017.

Rogers, J.: The Association between the North Atlantic Oscillation and the Southern Oscillation in the Northern Hemisphere, Monthly Weather Review, 112, 1999–2015, https://doi.org/10.1175/1520-0493(1984)112<1999:TABTNA>2.0.CO;2, https://doi.org/10.1175/1520-0493(1984)112<1999:TABTNA>2.0.CO;2, 1984.

Ruprich-Robert, Y. and Cassou, C.: Combined influences of seasonal East Atlantic Pattern and North Atlantic Oscillation to excite Atlantic multidecadal variability in a climate model, Climate Dynamics, 44, 229–253, https://doi.org/10.1007/s00382-014-2176-7, https://doi.org/10.1007/s00382-014-2176-7, 2015.

Schulzweida, U.: CDO User Guide. Climate Data Operators Version 1.9.1. October 2017, MPI for Meteorology. URL: https://code.mpimet.mpg.de/projects/cdo/embedded/cdo.pdf., access: 01.04.2018, 2017.

Screen, J.: The missing Northern European winter cooling response to Arctic sea ice loss, Nature Communications, 8, https://doi.org/10.1038/ncomms14603, https://doi.org/10.1038/ncomms14603, 2017.

Stephenson, D., Pavan, V., Collins, M., Junge, M., and Quadrelli, R.: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Climate Dynamics, 27, 401–420, https://doi.org/10.1007/s00382-006-0140-x, https://doi.org/10.1007/s00382-006-0140-x, 2006.

Ulbrich, U. and Christoph, M.: A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing, Climate Dynamics, 15, 551–559, https://doi.org/10.1007/s003820050299, https://doi.org/10.1007/s003820050299, 1999.

Ulbrich, U., Pinto, J. G., Kupfer, H., Leckebusch, G. C., Spangehl, T., and Reyers, M.: Changing Northern Hemisphere Storm Tracks in an Ensemble of IPCC Climate Change Simulations, Journal of Climate, 21, 1669–1679, https://doi.org/10.1175/2007JCLI1992.1, http://dx.doi.org/10.1175/2007jcli1992.1, 2008.

von Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 2003.

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

Warner, J. L.: Arctic sea ice – a driver of the winter NAO?, Weather, 73, 307–310, https://doi.org/10.1002/wea.3399, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/wea.3399, 2018.

Wilks, D. S.: "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It, Bulletin of the American Meteorological Society, 97, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1, https://doi.org/10.1175/BAMS-D-15-00267.1, 2016.

Woollings, T., Hannachi, A., Hoskins, B., and Turner, A.: A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing, Journal of Climate, 23, 1291–1307, https://doi.org/10.1175/2009JCLI3087.1, https://doi.org/10.1175/2009JCLI3087.1, 2010.

Woollings, T., Franzke, C., Hodson, D., Dong, B., Barnes, E., Raible, C., and Pinto, J.: Contrasting interannual and multidecadal NAO variability, Climate Dynamics, 45, 539–556, https://doi.org/10.1007/s00382-014-2237-y, https://doi.org/10.1007/s00382-014-2237-y, 2015.

Xu, T., Shi, Z., Wang, H., and An, Z.: Nonstationary impact of the winter North Atlantic Oscillation and the response of mid-latitude Eurasian climate, Theoretical and Applied Climatology, 124, https://doi.org/10.1007/s00704-015-1396-z, 2015.

Zwiers, F. and von Storch, H.: On the role of statistics in climate research, International Journal of Climatology, 24, 665–680, https://doi.org/10.1002/joc.1027, 2004.