

Evaluation of CMIP5 models over the northern North Atlantic in the context of forthcoming paleoclimatic reconstructions

Maria Pyrina¹  · Sebastian Wagner¹ · Eduardo Zorita¹

Received: 1 August 2016 / Accepted: 12 January 2017 / Published online: 18 February 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract We evaluated 11 coupled climate model simulations regarding the spatial structures of sea-surface temperature (SST) variability in the North Atlantic, during the second half of the twentieth century. The subset of models includes CCSM4, CSIRO, CanESM and MPI-ESM, participating in the fifth phase of the Climate Model Inter-comparison Project. The evaluation was performed to determine the potential of these models to be used at a later stage as test beds for the evaluation of climate field reconstruction methods that will use the extremely long-lived bivalve mollusk *Arctica islandica*, an outstanding paleoclimate archive for the boreal and temperate North Atlantic (Schöne, Glob Planet Change 111:199–225, 2013). Several validation metrics such as the mean bias, variance, spatial and temporal co-variability and trends of the North Atlantic summer SSTs showed that some of the models can be used to test paleoclimatic reconstructions. However, most models showed shortcomings in simulating the Atlantic Multidecadal Oscillation. Concerning the co-variability of summer SSTs between proxy sites and the whole North Atlantic SST field, we found that these proxy locations contain a SST signal that might represent a (basin-wide) signal for the north-eastern North Atlantic basin.

Keywords North Atlantic · Teleconnections · SST co-variability · *Arctica islandica* · CMIP5

Electronic supplementary material The online version of this article (doi:10.1007/s00382-017-3536-x) contains supplementary material, which is available to authorized users.

✉ Maria Pyrina
maria.pyrina@hzg.de

¹ Helmholtz Zentrum Geesthacht, Institute of Coastal Research, 21502 Geesthacht, Germany

1 Introduction

Climate reconstructions based on natural archives for which standard calibration and verification procedures are first developed on the inter-annual to decadal time scale (Jones et al. 2009), rely on statistical methods that link proxy records and local climatic variables (e.g. surface air temperature, precipitation, SST, sea level, sea level pressure, salinity). These methods are statistically calibrated using data from periods in which the proxy and instrumental records overlap. Whereas local climate reconstructions usually amount to a relatively simple linear re-scaling of one proxy record, Climate Field Reconstructions (CFRs) are based on a spatial network of proxy records and aim at spatially resolved climate reconstructions over a certain region. The statistical methods that have been applied so far for CFRs can attain a fair degree of sophistication (e.g. Smerdon, 2016 for a review of CFRs in the context of annually resolved proxy records). The properties and performance of the CFRs is sometimes difficult to establish, as the statistical methods are calibrated with data spanning the observational period—usually of the order of 100 years—which hampers a robust estimation of the skill of the reconstruction methods for multi-centennial or multi-millennial time scales. For the oceanic realm another level of complexity relates to the sparseness of observational data, especially prior to the pre-1950 period.

The CFR methods can, however, be tested using climate simulations as a virtual reality. Although in the ideal case all climate models should be physically consistent, the estimation of the skill of the CFR methods may depend on the climate model used to perform the climate simulation. Differences in the physical parameterizations, spatial resolution and processes incorporated in the model, may result in different assessment of the performance of CFRs.

For instance, Dee et al. 2016 found that structural model biases introduce uncertainties that systematically reduce the reconstruction skill of CFRs. Thus, it is advisable to use several climate simulations with different climate models and previously evaluate the performance of these climate models to realistically simulate the observed climate. A practical hurdle in estimating the overall performance of the models does not only include the definition of the evaluation metrics itself, but also the observational basis for the assessment. This issue is problematic even for continental areas in the latter half of the twentieth century (1950–1999) because of the scarcity of meteorological observations over vast continental areas, especially over high latitudes and the tropical and southern hemispheric regions (Hijmans et al. 2005). For the global oceans, even after 1982 when both in situ and satellite data are available (Reynolds et al. 2002), there are still only few observations in high latitudes and especially in regions covered, at least seasonally, with sea ice (Hirahara et al. 2014). This data issue has, therefore, also to be taken into account when evaluating climate models, especially over those regions where the observational basis is inhomogeneous and not extending far back in time.

There are three studies that have assessed CFR methods taking into account the influence of the model used when evaluating these methods. These studies have demonstrated how different model simulations affect the evaluation of the CFR performance, but none of these studies has tested the ability of the models used to realistically simulate the observed climate. The first study is that of Mann et al. (2007). These authors focused on hemispheric annual mean temperature and used two climate models. In the studies of Smerdon et al. (2011) and Smerdon et al. (2016), the influence of choosing a particular model simulation on the CFR results was investigated, but focusing on the spatially resolved surface temperature, covering global scales. They used a multiproxy network that consists mostly of terrestrial and a few oceanic proxies. The aim of the present study is to evaluate the representation of the modeled spatiotemporal characteristics in simulating SSTs of the North Atlantic (NA) Ocean. In a later study we will focus on specific models that can be considered realistic and evaluate CFRs that can be applied to reconstruct SSTs in the NA region, using annually resolved marine proxy records of *Arctica islandica*.

Arctica islandica is an extremely long-lived bivalve mollusk (225 to over 500 years: Butler et al. 2013; Wanamaker et al. 2008a, b), suitable for environmental and climate studies (Wanamaker et al. 2008a, b). It is found in the NA basin and lives in water depths ranging from approximately 4 to 500 m (Rowell et al. 1990; Nicol 1951). Compared to other existing records from extratropical oceans (i.e. sediment cores, coralline red algae), *Arctica islandica* can monitor environmental changes and ecological dynamics of

the NA ocean in seasonal to interannual time scales (Butler et al. 2013; Schöne 2013). Annually laminated marine sediments are rarely used to reconstruct high resolution paleoclimatology of the last millennium unless both their chronology and climate sensitivity is well understood (Jones et al. 2009). Moreover, until now, there is no network of marine sedimentary archives of annual resolution in the NA that could be potentially used for CFRs that are not limited to lower frequency domains of multi-decadal to centennial resolution (McGregor et al. 2015; Cunningham et al. 2013). Within the sectioned shell of *Arctica islandica* (see Mette et al. 2016, their Fig 1) distinct annual (Butler et al. 2009) and even daily (Schöne et al. 2005a, b) growth lines are apparent. The variability in the shells growth increment widths, which are the portions of shell between consecutive growth lines (Schöne 2013), and in the geochemical signature from the shell material (^{14}C , $\delta^{18}\text{O}$, $\delta^{13}\text{C}$) relates to changes in environmental conditions (Witbaard and Klein 1994; Schöne et al. 2011; Wanamaker et al. 2011). In addition to the high temporal resolution of *Arctica islandica*, the reconstructions derived from its records can be cross-validated, absolutely dated (Scourse et al. 2006; Butler et al. 2010) and offer significant advantages in evaluating long-term NA marine climate dynamics (Wanamaker et al. 2009). Records of *Arctica islandica* can be used to reconstruct sea water temperatures (Eagle et al. 2013; Wanamaker et al. 2016), salinity (Gillikin et al. 2006), major NA climate modes like the Atlantic Multidecadal Oscillation (AMO) (Mette et al. 2016), ocean dynamics (see for NAO in Schöne et al. 2003; see for AMOC in; Wanamaker et al. 2012), hydrographic changes and ecosystem dynamics (Witbaard 1996; Witbaard et al. 2003).

The second aim of this study is to perform an assessment of the capability of the network of *Arctica islandica* sites to provide a comprehensive and spatially resolved reconstruction of NA SSTs. The importance of testing already established locations of proxy archives lies in the further application of the local climate reconstruction into the broader concept of CFRs. As CFRs are co-variance-based approaches, we test whether the sites of *Arctica islandica* sufficiently co-vary with the NA basin. The information derived from the proxy archive could then be used to reconstruct the larger NA SST field using CFR methods. To assess the capability of the network of *Arctica islandica* sites to provide spatially resolved reconstruction of NA SSTs, we need to assess the capability of state-of-the-art climate models participating in the 5th phase of the Climate Model Intercomparison Project (CMIP5) to represent the co-variance of the spatially resolved NA SSTs and the SSTs at the *Arctica islandica* collection sites during the second half of the twentieth century.

The reasoning behind evaluating SST patterns of the anthropogenically forced period to reconstruct, in a second

step, changes prior to industrialization is supported by other studies. Rutherford et al. 2003 found that it is more important for the reconstruction skill to use a data-rich calibration period with increasing radiative forcing than a data sparse calibration period with relatively stationary forcing. Moreover, exploiting teleconnected locations implicitly assumes that the teleconnected relationship does not significantly depend on the external forcing (Batehup et al. 2015). Coats et al. 2013 found that atmospheric forcing cannot account for the non-stationary teleconnection between tropical Pacific SSTs and 200 mb geopotential height. Gallant et al. 2013 found significant variations through time in teleconnections on near-centennial timescales in model simulations forced by internal dynamics alone, but Batehup et al. 2015 found that using multiple teleconnected regions minimizes any effects of non-stationarities. As these relationships cannot be assessed within the instrumental record, it is crucial to first evaluate CMIP5 models in the twentieth century when model output and observations overlap, and additionally test the teleconnections of the proxy sites that will be used in CFRs.

The present study contributes to the evaluation of the CMIP5 models in several aspects. The spatial structure of SST variability simulated by CMIP5 models, with emphasis on the NA ocean, has been evaluated in previous studies (Perez et al. 2014; Liu et al. 2013). Patterns of interdecadal change have also been evaluated zonally (Carton et al. 1996; Kushnir 1994) and regionally (Qu and Huang 2014). However, only a few studies consider the northern part of NA, north of 60°N (Ruiz-Barradas et al. 2013; Jones et al. 2013), where one of the *Arctica islandica* sites tested in this study is located. In most studies, the structure of SST variability was studied using the ensemble of all CMIP5 models (Wang et al. 2015) or the ensemble mean (Jha et al. 2014), focusing on the mean response rather than the behavior of each individual model. Teleconnection patterns between SSTs of *Arctica islandica* sites (Dahlgren et al. 2000; Wanamaker et al. 2016) and the NA basin using CMIP5 models have not yet been investigated.

2 Data and methodology

We compare CMIP5 SST patterns with those derived from the Centennial in-situ Observation-Based Estimates COBE2 (Hirahara et al. 2014) to check the consistency over the NA, aiming to use this assessment of the most suitable models as a test-bed for assessing different climate reconstruction techniques used for *Arctica islandica* in a follow-up study. We focus on the latter half of the twentieth century (1950–1999), a period during which data coverage was substantially more complete than during the late nineteenth and early twentieth centuries. Based on the work

of Schleussner et al. (2014) and of Wang et al. (2015), 11 CMIP5 models were used in our analysis (Table 1) and then compared to the COBE2 data set (Table 2) for this 50 year period. For the selection of models we also took into account the horizontal resolution of the oceanic component of the respective Atmosphere–Ocean General Circulation Model (AOGCM). Additionally, we excluded CMIP5 models with known problems in their archived output (<http://cmip-pcmdi.llnl.gov/cmip5/errata/cmip5errata.html>).

The CMIP5 project design includes suites of simulations of past climates, future climates, and shorter-term hindcasts of the last few decades (<http://cmip-pcmdi.llnl.gov/cmip5/>). In this study we used the historical simulations, which are part of the long term coupled simulations and cover most of the industrial period (from the mid-nineteenth century to the beginning of the 21st century) and are sometimes referred to as “20th century” simulations. They are forced by changes in the total solar irradiance, observed atmospheric composition changes (reflecting both anthropogenic and natural sources) and include time-evolving land cover (Taylor et al. 2012). The models used in this study, as well as their original spatial and temporal resolution and other relevant information, are listed in Table 1. For a better comparison, the models’ original output was re-processed and re-gridded to a regular grid including the reference data sets. In this context the output was re-gridded onto a 1°×1° degree horizontal resolution covering the NA region, between 60°W–30°E and 40°N–70°N, because most ocean models have a resolution of the order of our target grid. In the following, a summary of the main characteristics of the models is presented.

2.1 Models

In CCSM4 (Gent et al. 2011) the atmosphere (CAM4/ Neale et al. 2013), the land (CLM4/ Lawrence et al. 2012) and the sea ice components (CICE4/Hunke and Lipscomb 2008) interchange both state information and fluxes through a coupler in every atmospheric time step. The fluxes between atmosphere and ocean (POP2/ Danabasoglu et al. 2012) are calculated in the coupler and communicated to the ocean component once a day. In CSIRO the ice model has been developed in conjunction with the atmospheric model (R21/ Gordon and O’Farrell 1997). The atmospheric fluxes are averaged over two steps and passed to the ocean model (modified MOM2.2/ Gordon et al. 2010). Land surface interactions are parameterized using a soil–canopy model (Kowalczyk et al. 1994). As described in detail in Arora et al. (2011), CanESM2 evolved from the first generation CanESM1. It is composed of atmosphere, ocean, sea ice and carbon cycle models. The calculation of energy and moisture fluxes at the land surface is carried out within the Canadian Land Surface Scheme (CLASS) module (von

Table 1 CMIP5 models used in this study

(1) Model name (2) Institution	Historical runs-tempo- ral resolution	(1) Atmosphere hori- zontal grid (2) Ocean horizontal grid	Aerosol component Type or name	Atmospheric chem- istry Component name	Ocean biogeochemistry Component name
(1) CSIRO Mk3.6 (2) Common Scientific and Industrial Research Organiza- tion	Monthly: Jan 1850– Dec 2005	(1) 1.875°×1.875° (2) 0.9°×1.875°	Interactive	Not implemented	Not implemented
(1) CanESM2 (2) Canadian Center of Climate Modelling and Analysis	Monthly: Jan 1850– Dec 2005	(1) spectral T63 (2) 256×192	Interactive	Included	CMOC
(1) HadCM3 (2) Met office Hadley Center, UK	Monthly: Jan 1859– Dec 2005	(1) 3.75°×2.5° (2) 1.25°×1.25° N144	Interactive	Not implemented	Not implemented
(1) MPI-ESM-P (2) Max Planck Insti- tute Earth System Model, Germany	Monthly: Jan 1850– Dec 2005	(1) 1.8° T63 (2) 1.5° GR15	Prescribed	Not implemented	HAMOCC
(1) INM-CM4 (2) Institute for Numerical Math- ematics, Russia	Monthly: Jan 1850– Dec 2005	(1) 2°×1.5° (2) 1°×0.5°	Prescribed	Not implemented	Included
(1) IPSL-CM5 (2) Institute Pierre Simon Laplace, France	Monthly: Jan 1850– Dec 2005	(1) 1.9°×3.75° (2) 2×2–0.5° ORCA2	Semi-interactive	Not implemented	PISCES
(1) MRI-CGCM3 (2) Meteorological Research Institute, Japan	Monthly: Jan 1850– Dec 2005	(1) 320×160 TL159 (2) 1°×0.5°	MASINGAR mk-2	Not implemented	Not implemented
(1) NorESM1-M (2) Norwegian Earth System Model	Monthly: Jan 1861– Dec 2005	(1) 1.9°×2.5° (2) 1.125° along the equator	CAM4-Oslo	CAM4-Oslo	Not implemented
(1) GFDL-CM2 (2) Geophysical Fluid Dynamics laboratory, USA	Monthly: Jan 1850– Dec 2005	(1) 0.9°×1.875° (2) 1° tripolar 360×200L50	Semi-interactive	Not implemented	Not implemented
(1) GISS-E2-H (2) Goddard Institute for Space Studies, USA	Monthly: Jan 1850– Dec 2005	(1) 2°×2.5° (2) 0.2° to 1°×1.875°	Interactive	G-PUCCINI	HYCOM Ocean
(1) CCSM4 (2) Community Cli- mate System Model, USA	Monthly: Jan 1850– Dec 2005	(1) 0.9°×1.25° (2) 1.25° in longitude, variable 0.27–0.64° in latitude	Interactive	Not implemented	Not implemented

Table 2 COBE2 data used in this study

Reanalysis data	Spatial resolution
COBE2_SST	1° × 1°

Salzen et al. 2013), while coupling with terrestrial ecosystem and ocean carbon models enables some important biogeochemical processes to be represented and feedback

to the physical climate (Yang and Saenko 2012). The SSTs and sea ice extent, are the central variables through which the atmospheric component (HadAM3/ Pope et al. 2000), the oceanic component (HadOM/ Collins et al. 2001) and the sea ice component (Cattle et al. 1995) are coupled in HadCM3 (Gordon et al. 2000). HadAM3 includes MOSES, a land surface scheme developed by Cox et al. 1999. The INM-CM4 climate model consists of two major blocks representing a model of general circulation of the atmosphere

and a model of general circulation of the ocean. When coupling the atmospheric and oceanic models, the heat, freshwater fluxes and wind stress are transmitted from the atmosphere to the ocean, and the surface temperature and sea-ice area are transmitted from the ocean to the atmosphere (Volodin et al. 2010). The IPSL-CM5A model is built around a physical core that includes the atmosphere (GCM LMDZ5A/ Hourdin et al. 2013), land-surface (ORCHIDEE/ Krinner et al. 2005), ocean and sea ice components (NEMOv3.2/ Madec 2008). The atmospheric model has a fractional land-sea mask and each grid point is divided into four sub-surfaces corresponding to land surface, free ocean, sea ice and glaciers, respectively. The OASIS coupler (Valcke et al. 2006) is used to interpolate and exchange the variables, and to synchronize the models (Dufresne et al. 2013). The MRI-CGCM3 model (Yoshimura and Yuki-moto 2008) consists of an atmosphere-land model (MRI-AGCM3) and the MRI.COM3 ocean and sea ice model. Each model component uses a simple coupler to exchange data with the other model components. NorESM is largely based on CCSM4 with the main differences being the isopycnic coordinate ocean module in NorESM. In addition, CAM4-Oslo substitutes CAM4 as the atmosphere module. The sea ice and land models in NorESM also include some differences regarding the aerosol calculations. The ocean component of GFDL-CM2 is the MOM4 code (Griffies et al. 2004) and the atmospheric component relates to the AM2p13 model. The land surface component allows for the simulation of the diurnal and seasonal cycles, while the sea ice component may produce five different ice thickness categories and open water at each grid point (Winton 2000). The new version of the GISS climate model used for CMIP5 simulations is called ModelE2. It is similar to the ModelE but with numerous improvements in physics (Schmidt et al. 2006). The atmosphere is coupled to a full dynamic ocean of the HYCOM model for the version GISS-E2-H that is used in this study (Sun and Bleck 2006). In the MPI-ESM model the coupling at the interfaces between atmosphere (ECHAM6/ Stevens et al. 2013) and land processes (JSBACH/ Reick et al. 2013), and between atmosphere and sea ice occurs at the atmospheric time step. The coupling between atmosphere and ocean (MPIOM/ Jungclaus et al. 2013), as well as land and ocean (the latter by river runoff) occurs once a day.

2.2 Data

Yasunaka and Hanawa 2011, performed an intercomparison of seven historical SST datasets including the extended reconstruction of global SST (ERSST/ Smith and Reynolds 2004) version 3, the second Hadley Center SST (HadSST/ Rayner et al. 2006), the optimal smoothing analysis by the Lamont-Doherty Earth observatory (LDEO/ Kaplan

et al. 2001), the SSTs by the authors at Tohoku University (TOHOKU/ Yasunaka and Hanawa 2002), as well as the Hadley Centre's sea ice and sea-surface temperature data set (HadISST/ Rayner et al. 2003) version 1, the release 2.1 of the International Comprehensive Ocean–Atmosphere Data Set (ICOADS/Worley et al. 2005) and the centennial in-situ observation-based estimate of SSTs (COBE/Ishii et al. 2005). They categorized the datasets into groups of fully interpolated (HadISST, COBE, ERSST, LDEO) and simply averaged data (ICOADS, HadSST, TOHOKU). They found that the latter group has many missing values and it includes extreme values, while the correlation and standard deviation of SST anomalies of ICOADS and HadSST are dependent on the number of observations. Furthermore, all datasets except ICOADS agree in the phases of the AMO index and on the SST global means. In the study of Loder et al. 2015 and in terms of Empirical Orthogonal Functions (EOFs) for the summer means during 1900–2011, it is found that for the data sets HadISST1, ERSST3 and COBE the leading summer EOF patterns show considerable similarities. Regarding the summer SST trends during the period 1950–2011, Loder et al. 2015 found that there are variations on local scales among the gridded datasets that will affect the trends at particular sites, indicating that caution is required regarding the spatial representativeness of the trend values from the gridded datasets. The large-scale patterns of the trends are generally similar for the period 1950–2011 amongst the three datasets, but there are differences particularly for the COBE dataset north of Iceland (see in Loder et al. 2015; Fig 7) due to a suspect abrupt jump in the COBE SSTs in that region around 1979. In our work, the SSTs are detrended prior to the analysis and therefore we do not expect these differences to affect our results.

We used the historical SST data set COBE2, developed at the Japanese Meteorological Agency (Hirahara et al. 2014). The COBE2 data set is a spatially complete SST product, covering the period 1850–2013 AD interpolated to a $1^\circ \times 1^\circ$ grid. It combines SST measurements from ICOADS release 2.0, the Japanese Kobe collection, and readings from ships and buoys. Data are gridded using optimal interpolation. Similar to HadISST, data up to 1941 were bias-adjusted using “bucket correction” (Hirahara et al. 2014). Prior to the interpolation analyses, data were also subject to quality control using combined a-priori thresholds and nearby observations (<https://climatedataguide.ucar.edu/climate-data/sst-data-cobe-centennial-situ-observation-based-estimates>).

2.3 Methodology

For the present analysis we considered the summer time (June–August) mean SST values for each year, covering

the period 1950–1999 AD. The summer time SSTs were motivated by the growing season of *Arctica islandica*, that in both sites that are tested in this study, growth occurs approximately from February to September but it is biased towards summer (Schoene et al. 2004, 2005a, b). For this reason we calculated the summer mean SSTs of the NA for the 11 CMIP5 models and the dataset COBE2. We then calculated the SST bias as the difference between the CMIP5 and COBE long-term summer mean and the climatological root mean square error (RMSE). The climatological RMSE was calculated from the long-term summer mean squared differences between the CMIP5 models and the COBE2 data. The temporal standard deviation was also calculated to infer information on the performance concerning the amplitude of the (inter-annual) temporal variability.

To assess the spatial co-variability of SST anomalies, EOF analysis (von Storch and Zwiers 1999) is applied. The leading EOF represents spatio-temporal patterns of variability that account for the maximum co-variance between the SST anomaly time series at all pairs of grid points in the data set. The remaining co-variability was subjected to the same decomposition with the additional constraint that the second EOF pattern is orthogonal (e.g., uncorrelated) in both time and space to the leading EOF pattern (Deser et al. 2010).

In a second step we correlated the simulated model time series (1950–1999) in two model regions co-located with proxy sites of *Arctica islandica* and the NA Ocean (40°–70°N and 30°E–60°W). The first region is located in the North Sea (NS) at 58.5°N, 0.5°E and the second north of Iceland (Icelandic Self-IS) at 66.5°N, 19.5°W. These sites were chosen based on previous collection sites and studies (Wanamaker et al. 2008a, b, 2012; Scourse et al. 2006; Butler et al. 2013). It is also important that the SST trends calculated by the models over NA share a similar distribution of the observed trends at a certain level of statistical significance. For both calculations of teleconnections and trends we tested the statistical significance at the 1% level, taking into account the effect of serial correlation (Zhang et al. 2000).

Moreover, the AMO index was computed from the modelled data, as it is associated with the NA dominant pattern of SST variability. To obtain the AMO index we linearly detrended the averaged NA SST (0°–70°N and 10°E–80°W), for the period 1850–2005 AD and calculated a 10-year running mean. AMO's spatial structures in both models and data are determined by linearly regressing the grid point SST onto the AMO index (Zhang and Wang 2013). Additional to the realization r1 of the MPI-ESM model, that has been used for all results presented in this work, we used two more realizations (r2, r3) of the MPI-ESM model to study the AMO. The r1 and r2 experiments are initialized with the same ocean state, but they

differ in the standard deviation of the assumed lognormal distribution of the volcanic aerosol size (1.2 μm in r1, 1.8 μm in r2 and r3). The simulations r2 and r3 used the same parameter setting, but are started from different initial conditions (Jungclaus et al. 2014).

3 Results

In each of the following figures, we choose to present in most cases the best performing models with respect to each of the validation metrics used in this study. The results regarding all models are shown in the Electronic supplementary material of Appendix.

3.1 RMSE, mean bias and variance

As an initial skill metric on the basis of RMSE (Table 3), the best performing models in terms of summer SST biases are shown in Fig. 1. The spatial distribution of the differences between the models and the observations share some similar characteristics (Appendix, Fig. 12A). Negative SST biases are shown for most models, with maximum values on the southwest of the study region east of Newfoundland (~4K), except from the GISS model, indicating a widespread overestimation of SSTs over most of the NA Ocean (Appendix, Fig. 1A). The simulation of the North Atlantic Current path could be the reason for this SST difference between the observations and the models. The systematic error is located near the tail of Grand Banks, where the Gulf Stream turns north. This phenomenon can be seen for all 11 CMIP5 models used in this study. Discrepancies between the models mainly occur over the northern NA and more specifically on the path of the east Greenland current, where some models underestimate and others overestimate the SSTs, showing a wide spread in SST biases. Finally, the climate models best simulating the climatological mean of the summer SSTs in the period 1950–1999, according to the aforementioned validation metrics, are CSIRO, CCSM4 and GFDL.

Table 3 RMSE of the 11 CMIP5 models used in this study

Model	RMSE	Model	RMSE
GFDL	1.77	INM-CM4	2.29
CCSM4	1.78	GISS	3.00
CSIRO	1.79	IPSL-CM5	3.10
NorESM	1.93	HadCM3	3.28
CanESM2	2.27	MRI-CGCM	2.64
MPI-ESM r1	2.28		

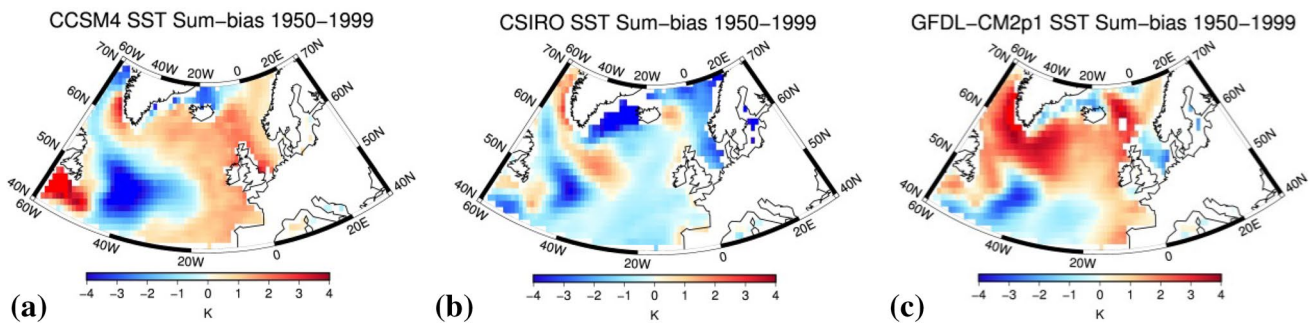


Fig. 1 Mean summer SST bias between COBE2 and the CMIP5 models **a** CCSM4, **b** CSIRO and **c** GFDL, for the period 1950–1999

Although SST biases can represent a measure of skill of the models, it is not the only metric for the evaluation of climate models. Additionally, models should also be able to reproduce the observed pattern of inter-annual variability of the SSTs for the 1950–1999 year period. The observations show (Fig. 2a) small standard deviation of SSTs, between 0.2 and 0.4 K, on the path of an arc connecting Greenland and Iberian Peninsula and largest variance, up to 1 K, on the southeast of Newfoundland. Most of the models indicate that the SSTs mainly vary in the west part of the NA (Appendix, Fig 12A), so there is large temperature range on the path of the NA current and of the Labrador Current further to the south (Appendix, Fig 2A). Therefore, it can be inferred that most of CMIP5 models are not able to simulate the correct position and evolution of the individual ocean currents (Willebrand et al. 2001). The model closest to the geographical distribution of the observed SSTs standard deviation is CCSM4 (Fig. 2b).

3.2 Leading modes of variability

In addition to investigating the variability pattern in terms of spatially resolved standard deviations, EOFs reveal the main patterns of co-variability in a given region. Therefore in terms of CFRs it is important for the models to simulate an appropriate spatial co-variability structure. Several studies already performed an evaluation related to the spatial

co-variability of NA SSTs (Cannaby and Hu 2009; Fan and Schneider 2010), but for our purposes regions north of 60° are also important. Those changes in the geographical domain will ultimately impact on the structure of the individual EOFs (Legates 1991) and therefore it is necessary to carry out a separate EOF analysis for our region of interest.

Figures 3 and 4 depict the first and second EOF of models and data, respectively. The third EOF pattern is shown in the Appendix (Fig. 5A). The first EOF (Fig. 3) represents around 30% of the variance in most simulations, except for the INM, HadCM3 (Appendix, Fig. 3A) and MPI models with reduced values around 20%. The corresponding eigenvector map of the COBE2 data describes a zonal dipole of the NA SST anomalies. The amplitude of this dipole is largest in the northern NA, at about 55°N and in the southeast of our study area close to the Iberian Peninsula, with SST anomalies of the opposite sign. The EOF maps derived from the models CanESM2 and CCSM4 also depict a zonal dipole, but are different to the one shown by COBE2. The EOF pattern of MPI-ESM shows some similarities to the COBE2 data, with cold SST anomalies north of 55°N and to the south and warm SST anomalies in some areas of the subpolar gyre. The EOF pattern of CCSM4 does not display warm anomalies in the area close to the Iberian Peninsula, but it depicts both the northern NA warm anomalies and the cold ones of the subtropical gyre. The CanESM2 model shows similar results as the CCSM4, while the rest of

Fig. 2 Standard deviation of summer SSTs for **a** COBE2 and **b** CCSM4, for the period 1950–1999

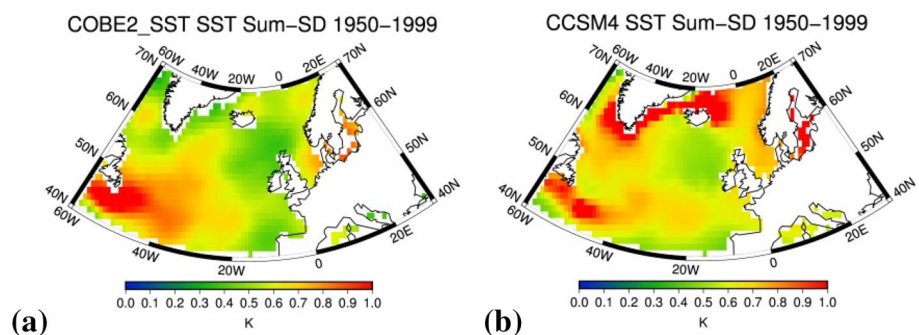


Fig. 3 1st EOF of the detrended summer SSTs of the period 1950–1999, for **a** COBE2 and the CMIP5 models, **b** CanESM2, **c** CCSM4 and **d** MPI-ESM-P

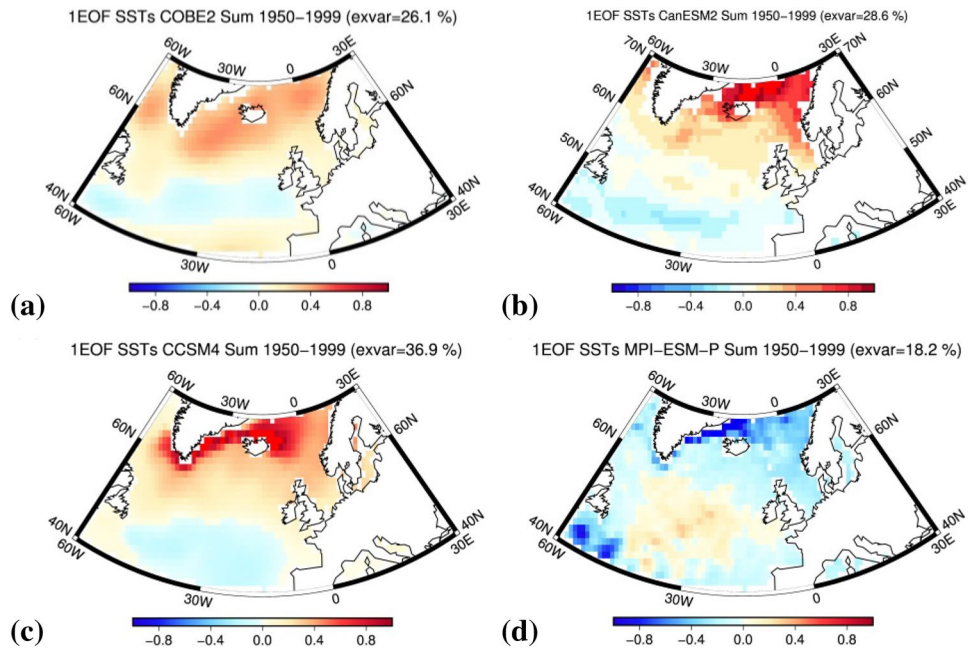
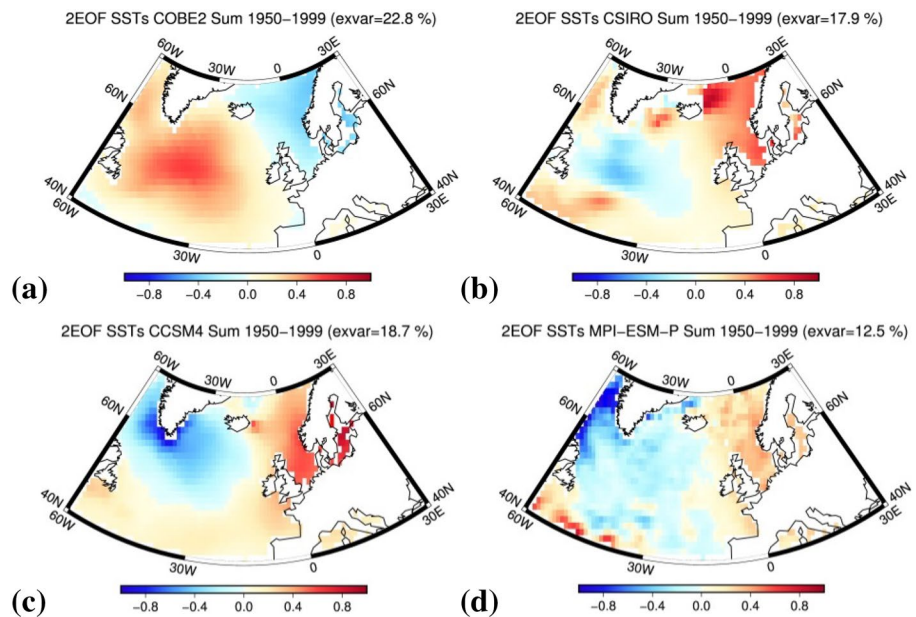


Fig. 4 2nd EOF of the detrended summer SSTs of the period 1950–1999, for **a** COBE2 and the CMIP5 models, **b** CSIRO, **c** CCSM4 and **d** MPI-ESM-P



the models show dipole or monopole patterns with different centers. Generally, AMO-like variability is found in the temporal evolution (PC1) of the 1st EOF patterns for both CMIP5 models and the COBE2 data for the period 1950–1999. Assuming that the AMO is externally forced, we expect significant correlations between the PC1 of the models and the observed AMO for the period 1950–1999. Taking into account the effect of serial correlation, significant correlation at the 5% level of $r = +0.3$ is found between the PC1 of the models CanESM2 and CCSM4 when correlated to the COBE2 AMO for the period 1950–1999. These results indicate that the

temporal evolution of the AMO in the models CCSM4 and CanESM2, is likely driven to some extent, although not totally, by the external forcing.

The second EOF of SST variability (Fig. 4) represents 20% of the variance in almost all models, with the MPI and HadCM3 models again showing reduced values of around 13%. This second EOF for COBE2 shows a dipole centered on the subpolar gyre and the Norwegian coast. This is also shown by the CCSM4 model, but the center of the east pole is located at the south coast of Greenland and of the west pole at the coast of Europe, while CSIRO and MPI-ESM models also show this relationship between the SSTs of

the west and the northeast of the study region. In general almost all models capture the strong center of variability over the south of Greenland, but with a northwestward shift of its center. Finally, the pattern shown by the third EOF (Appendix, Fig. 5A) is not realistically represented by any of the models.

3.3 Correlation patterns of collection sites of *Arctica islandica*

For an evaluation of the models related to their potential use for paleoclimatic applications and reconstructions using oceanic proxy data, we correlate SSTs in two model regions, co-located with two proxy sites of *Arctica islandica* in the NA, with all grid-cells SSTs in the study area simulated by the same model. The locations are in the Icelandic Shelf (IS, Fig. 5) and in the northern North Sea (NS, Fig. 6). The resulting correlation patterns are compared with the derived in the same way for the COBE2 data set. This comparison between COBE2 and the 11 initially chosen models, reveals that the correlation patterns regarding the IS are well represented by most models, with the largest differences over the southern part of the region where some models show anti-correlation with the IS SSTs (Appendix, Fig. 6A). CanESM2 and CCSM4 simulate well the spatial distribution shown by COBE2. Statistically significant correlation at 1% level is shown by COBE2 over the area surrounding the collection site of Iceland with positive values between $r = +0.7$ and $r = +1$, including the areas of the East Greenland Current and the northern coast of Norway. Even though the CSIRO model was one of the best performing models in terms of RMSE, mean bias, variance and leading modes of variability, it does not show the aforementioned

maximum correlation with the areas of the NA indicated by the analysis based on COBE2. A closer look at CSIRO's results, regarding the temperatures at the north of Iceland, shows that most of the values are covered by sea ice. Furthermore, the number of grid points between Iceland and Greenland is reduced due to the coarse resolution of CSIRO model (see Table 1) and therefore these areas are mostly represented by sea ice or represent land grid-cells.

According to COBE2, the correlation with the collection site at the North Sea (Fig. 6) reveals a dipole pattern between the positive correlated values around NS and the negative correlated values at the east of Newfoundland. The models that can replicate the COBE2 pattern, showing some grid points of negative correlation values in the central Atlantic are CCSM4, MPI-ESM, GISS, MRI and CSIRO. The dipole pattern shown by COBE2 between NS and the central NA is also one of the dominant patterns of SST co-variability, as indicated by the second EOF, but is not reproduced in any of the model simulations. Finally, all models show the influence (i.e. high correlations) of the surrounding waters on the NS location.

3.4 Long term trends

SSTs provided by the CMIP5 models are additionally tested on their long-term trends over the second half of the twentieth century. In this analysis we calculated the linear trends of the summer SSTs, for the 50 year period from 1950 to 1999 (Appendix, Fig. 8A). COBE2 shows that SSTs along the coast of Europe, along the north coast of Iceland and south of Newfoundland have undergone warming, while the SSTs south of Greenland and on the Norwegian coast show a negative temperature trend over the 50 year period

Fig. 5 Correlation patterns (one-point correlation maps) for the Icelandic Shelf (IS) summer SSTs, for the period 1950–1999, for **a** COBE2 and the CMIP5 models **b** CanESM2, **c** CCSM4 and **d** CSIRO. *Hatched areas* indicate values statistically significant at the 1% level according to a statistical test taking into account the effect of serial correlated data

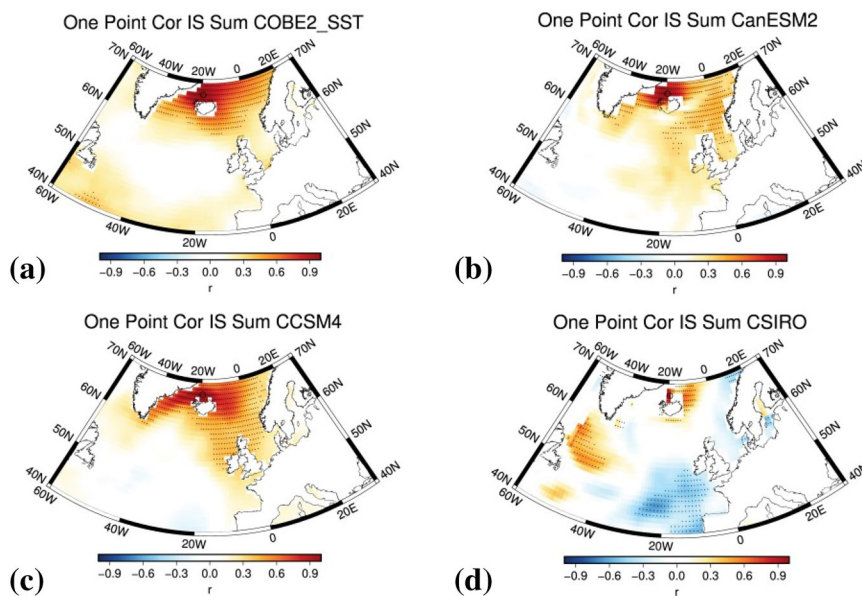


Fig. 6 Correlation patterns (one-point correlation maps) for the North Sea (NS) summer SSTs, for the period 1950–1999, for **a** COBE2 and the CMIP5 models, **b** GISS, **c** CSIRO, **d** CCSM4, **e** MPI-ESM-P and **f** MRI-CGCM3. *Hatched areas* indicate values statistically significant at the 1% level according to a statistical test taking into account the effect of serial correlated data

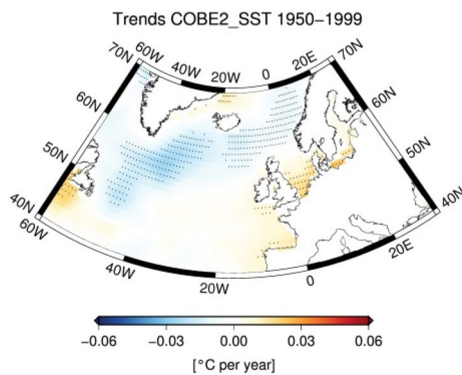
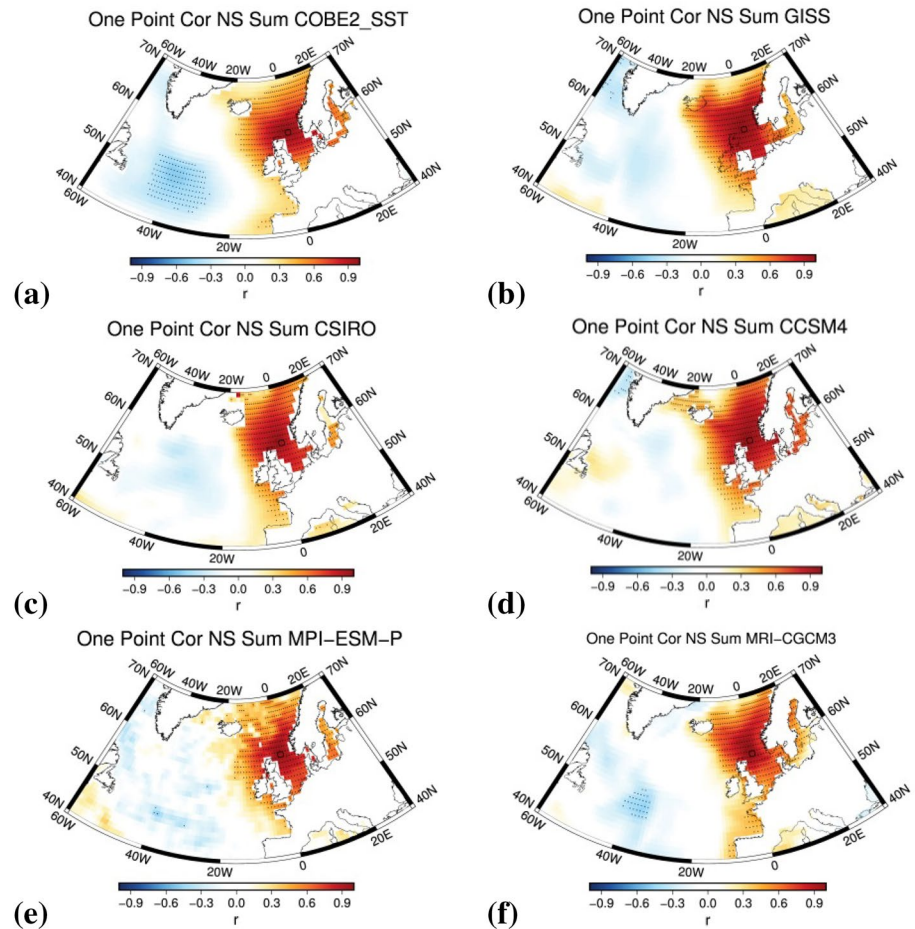


Fig. 7 Summer SST trends for the period 1950–1999 as calculated with COBE2 SSTs. *Hatched areas* indicate values statistically significant at the 1% level according to a statistical test taking into account the effect of serial correlated data

(Fig. 7). As shown in the Appendix in Fig. 8A, none of the 11 CMIP5 models are able to capture the SST trends shown by COBE2 over the NA for this period, but they can capture some of the individual characteristics in specific areas. Nor-ESM and MPI-ESM models show the negative trends along the coast of Norway reaching $0.4\text{ }^{\circ}\text{C}$ per decade,

while GFDL, CCSM4 and GISS the positive trends north of Iceland, with GFDL better approaching the magnitude shown by COBE2 ($\sim 0.4\text{ }^{\circ}\text{C}$ per decade). HadCM3, INM-CM4 and MRI-CGCM3 agree with the results of COBE2 only at the coast of Europe, while CSIRO and IPSL show warming to be dominant in the region of NA. As the external forcing applied in all simulations is similar, the way different processes are represented within each model could be the reason for the disagreement, not only between models and observations, but also amongst the individual models (Taboada and Anadón 2012).

3.5 Atlantic Multidecadal Oscillation

The model representation of the AMO, an important factor of decadal climate variability in the NA realm (Ruiz-Barraidas et al. 2013), was also investigated. The AMO was calculated for the eleven CMIP5 models used in this study and also for two additional realizations (r2, r3) of the MPI-ESM model (Appendix, Fig 10A). As seen in Fig. 8, the best performing models in terms of bias and variance do not follow the same AMO index that was calculated using the COBE2 SSTs. The AMO 10 year running mean was calculated and

is shown in Fig. 8, with the black line for COBE2-AMO and with the red line for the Models-AMO. If the AMO variability is externally induced, we can expect strong correlations between the COBE2 data and the CMIP5 models, as the external forcing applied in all simulations is similar. The Pearson correlation of the AMO running mean between data and models is shown in Table 4. Taking into account the effect of serial correlation, the models that exhibit statistically significant correlation at the 5% level are IPSL-CM5 and HadCM3, with correlation coefficients approximately equal to $r = +0.6$ and $r = +0.7$, respectively.

The r_1 , r_3 and r_2 , r_3 pairs of realizations of the MPI-ESM model are driven with the same forcing conditions (with a different aerosol forcing uncertainty between r_1 , r_2) and with different initial conditions. Therefore, the ratio of the forced AMO variability to total variability between the pairs of realizations can be estimated by the correlation of their respective AMO time series (Table 5). The correlation coefficient between the different realizations of the MPI-ESM model is not statistically significant at the 5% level, indicating that for the MPI-ESM model the AMO variability is to a large degree unaffected by changes in external forcings. However, this does not generally rule out that the real-world connection is different and it does not imply that other models will show the same behavior.

The effect of large volcanic eruptions ($AOD > 0.2$) can be seen in each modelled AMO, but prior to 1980 the AMO variability is largely model dependent. During the last 20 years it seems that the AMO index from most of the models is closer to observations, suggesting some role of the external forcing in pacing the AMO. Generally, the anomalies vary between $\pm 0.4^\circ\text{C}$ for COBE2 and for most CMIP5 models, indicating that the CMIP5 suite captures the range of observed amplitude of decadal-scale SST variations of the NA area. The spatial structure of AMO, for both models and data, is determined by linearly regressing the grid point SST onto the corresponding modelled or reanalyzed AMO index (Appendix, Fig. 9A). Despite the different temporal evolution in the CMIP5 models' AMO, almost all models are consistent in showing a warming across the entire NA as shown by COBE2. This is also shown by Ting et al. (2011) who found a well-defined spatial pattern for AMO in the NA, albeit with differing temporal behavior of the phenomenon between models and observations.

4 Discussion

4.1 RMSE, mean bias and variance

Most of the CMIP5 models analyzed in this work show a common cold SST bias east of Newfoundland, which is accompanied by a warm bias near the coast (Fig. 1).

Willebrand et al. (2001) showed that these biases are largely due to the early separation of the Gulf Stream, far too north from the coast of Cape Hatteras, and the turn of the NA Current northward near the mid-Atlantic ridge region and not at the Grand Banks of Newfoundland. MPI-ESM and CanESM2 show relatively large biases in the correct spatial simulation of the path of the Gulf Stream, while the INM models' colder temperatures are expanded to the north (Appendix, Fig. 1A). One reason for these dissimilarities might be the different spatial resolution in each model. Previous studies have shown that when ocean models are integrated at higher resolution, the representation of these currents is improved due to the better representation of meso-scale features (Smith et al. 2000; Bryan et al. 2007). This is, however, not necessarily the case for the INM model, which has the highest spatial resolution compared to the other models shown in Fig. 1. Accordingly, its cold bias is still dominant over most regions of the NA Ocean, and therefore other processes are obviously playing a more important role in explaining differences between these models. One reason might relate to the anthropogenic aerosol concentration changes that influence the simulated spatial response of the NA SSTs. Booth et al. (2012) found that the inclusion of aerosol indirect effects in model simulations allows for a better representation of the spatial structure of NA SST variability. Another cause of discrepancies between models could relate to a somewhat different representation of the Atlantic Meridional Overturning Circulation (AMOC). Wang et al. (2014) found common bias patterns among CMIP5 climate models and attributed them to AMOC and its associated northward heat transport. Negative SST anomalies are shown in the path of East Greenland Current by all models, except by the MPI model, which shows a warm bias compared to the COBE2 SSTs. Model biases are in the order of $\pm 4^\circ\text{C}$ and the model resolution and physical implementation might still not be realistic enough to represent certain features of this area (i.e. topography, meltwater) that are important for a good match between models and data. In addition to the coarse resolution, the magnitude and large scale patterns of SSTs are also influenced by additional factors such as surface and cloud albedo, and sea ice distribution (Hasumi 2014). Franco et al. (2012), found that most of the 5 km ice sheet topography of southeast Greenland can be reproduced at lower resolutions and that a model resolution of at least 10–15 km is needed to resolve the steep slopes in the vicinity of the ice sheet margin. Discrepancies could also arise by an overestimated transport of momentum, salt and heat diffusion around Iceland within the simulations (Logemann and Harms 2006).

The underestimation of the SST variability in the path of the East Greenland current by almost all models could also have its roots in the observational coverage related to the

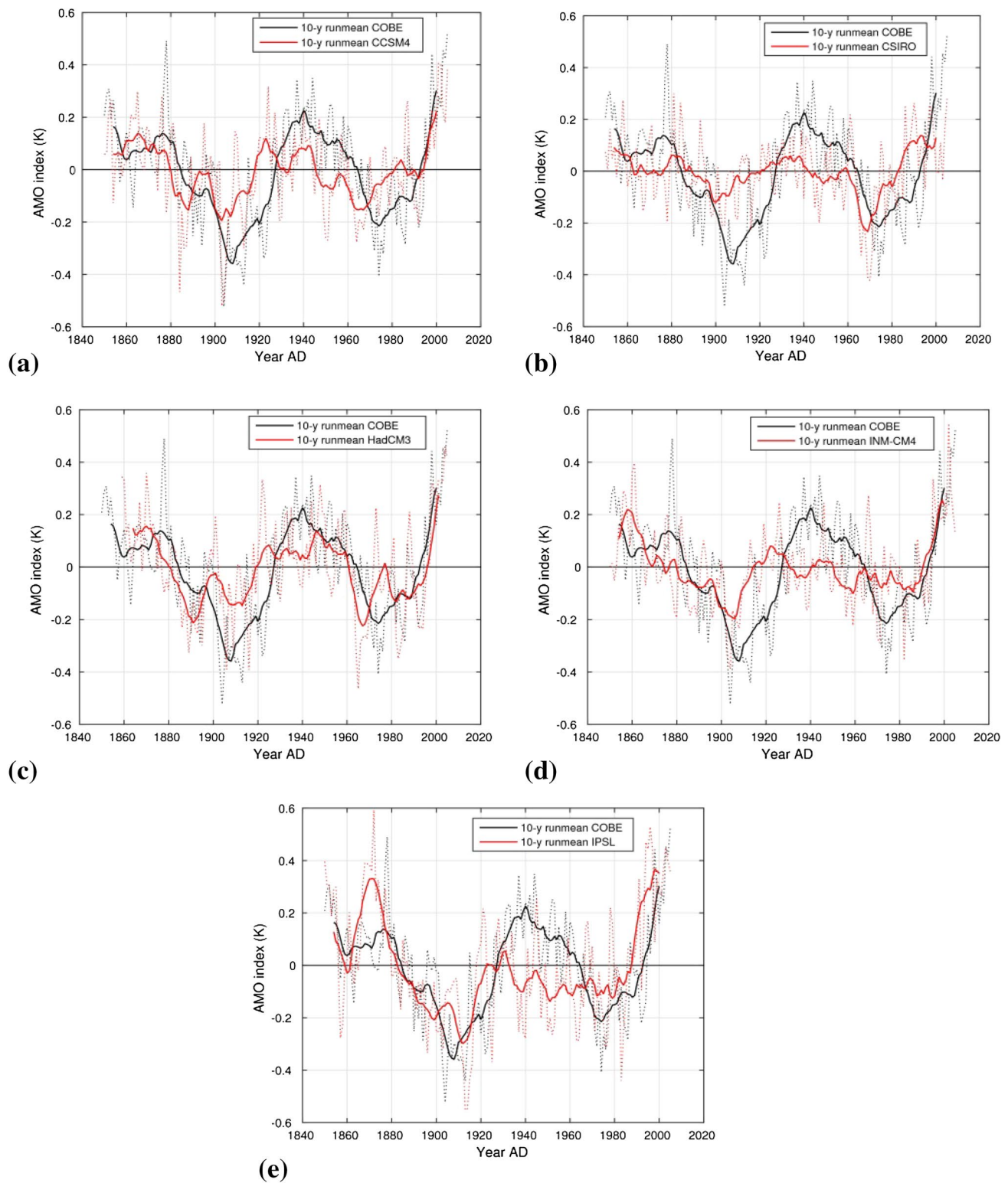


Fig. 8 AMO index as calculated from COBE2_SSTs (black dotted line) and from CMIP5 model SSTs (red dotted line). Black and red solid lines are the 10 years running mean for the COBE2 and models, respectively

Table 4 Pearson correlation between the 10 year running mean of the CMIP5 and COBE2 AMO

Model (years 1850–2005)	Correlation for the AMO 10 years running mean	Significance (denoted with 1) at the 5% significance level	Model (years 1850–2005)	Correlation for the AMO 10 years running mean	Significance (denoted with 1) at the 5% significance level
GFDL*	0.20	0	MPI-ESM r3	0.60	0
CCSM4	0.60	0	GISS	0.40	0
CSIRO	0.50	0	IPSL-CM5	0.60	1
NorESM	−0.30	0	HadCM3*	0.70	1
CanESM2	0.40	0	MRI- CGCM	0.40	0
MPI-ESM r1	0.30	0	INM-CM4	0.50	0
MPI-ESM r2	−0.03	0			

*The AMO for the models GFDL and HadCM3 was calculated for the years 1861–2005 and 1859–2005, respectively

Table 5 Pearson correlation of the AMO 10 year running mean between the different realizations of the MPI-ESM model

Realization number	Correlation for the AMO 10y running mean	Significance (denoted with 1) at the 5% significance level
r1–r3	0.29	0
r2–r3	0.16	0

COBE2 data. For instance, the study of Deser et al. (2010) claims that the observation coverage of SSTs over these areas has increased since 1960. However, approximately during 1950–1979 the coverage, is only around 20–50% compared to the number of measurements south of Iceland. Furthermore, there is a suspect abrupt jump in the COBE SSTs in that region around 1979 (Loder et al. 2015). Additionally, errors could result from a lack of atmospheric processes in the models i.e. an underestimation of low cloud cover is a cause of significant errors in radiative fluxes (Kauker 2003), as the SST response to shortwave radiative forcing is thermally direct (Cheng et al. 2013) and low-level stratiform liquid and mixed-phase clouds are found to be the most important contributors to the Arctic surface radiation balance (Shupe and Intrieri 2004). Another explanation could be the sea ice export and variation, so the changes of albedo and fresh water input of the area, during the 50 years of study (Halvorsen et al. 2015).

To quantify the magnitude of spatial distribution of the summer SST variability, Fig. 2 shows the standard deviation of the mean summer SSTs of the period 1950–1999 for the CMIP5 models and for COBE2. The COBE2 SSTs show their largest variability reaching $\pm 1^\circ\text{C}$ in the western boundary current region of the Gulf Stream, coinciding with a region of maximum north–south mean SST gradient (Deser et al. 2010). The large SST variability shown by the data in this region is captured well by the CCSM4 model, while the rest of the models show that the SSTs mostly vary in the west and south parts of the study

region. Along the path of the East Greenland Current, the SSTs vary by $\pm 0.4^\circ\text{C}$ according to COBE and by $\pm 1^\circ\text{C}$ according to the CMIP5 models. In this case, due to sea ice export from the Arctic Ocean and continental runoff (melt water is only added into the ocean during the summer half of the year), one could expect a larger variability in the observed summer SSTs than the one reproduced by the models. Many studies show that SSTs in the high-latitude Arctic Ocean are largely governed by sea-ice and continental runoff, rather than by evaporation and precipitation controlling low-latitude tropical oceanic variability. In addition, global satellite analyses and models incorporating remotely observed SSTs may be inaccurate due to lack of direct measurements for calibrating satellite data (Bai et al. 2015). On the other hand, due to lack of consistency in time, space and the number of SST measurements north of 60°N , the small variance shown by COBE2 on the east coast of Greenland and on the north coast of Iceland must be interpreted with care.

4.2 Leading modes of variability

Much of the spatial structure of SST variability is already highlighted by the variance, but it does not provide information on the spatial co-variability of SST variations. One common approach to investigate spatial co-variability is the EOF analysis. The EOF modes are pure geometric deconstructions of the domain, not considering any structure in the SST standard deviation (Wang et al. 2015). The differences between models and data may be a result of the decomposition into the linear combination of orthogonal spatial modes being driven by the large variability of SSTs simulated by the models (Fig. 2) along the path of the East Greenland current. Cannaby and Hu (2009) showed a zonal oscillatory mode of the NA SSTs, with amplitude of the eigenvector of the zonal pattern representing only 9% of the summer SST variability, being the 3rd mode derived from the SSTs in the winter months. Other studies found

that there are two dominant SST modes in the NA derived from observations in the twentieth century: a dipole mode on biennial and decadal time scales and a monopole on interdecadal time scales (Kushnir 1994; Deser and Blackmon 1993). The dipole becomes a part of a tripole if the domain is extended to the south (Fan and Schneider 2010). An important point when comparing EOF patterns calculated over a different spatial domain relates to the fact that the relative significance of each independent mode of variability is spatially dependent. Therefore, results of EOF analysis depend on the spatial domain of the data on which the analysis is performed. The length of the data set used could influence the ranking of the EOF patterns in terms of explained variance, i.e. the ranking of eigenvalues (Cannaby and Hu 2009). This is one of the reasons that could explain why the EOF results of our study area are different from those of other studies (e.g. Wang et al. 2015, for the central and southern NA).

The study by Marshall et al. (2001) shows that the leading pattern of SST variability in the NA is a tripole, with two poles in the NA above 40°N and the third pole near the Gulf of Mexico. These findings agree with our results (Fig. 4) regarding the second dominant pattern of SST variability in the NA, but due to the southerly limit of our study region, around 40°N, the 3rd pole cannot be seen. Marshall et al. (2001) relates this pattern to a direct response of the ocean to the anomalous air-sea fluxes controlled by the North Atlantic Oscillation (NAO). NAO is a large scale teleconnection pattern of atmospheric variability at sea level pressure, that refers to a reorganization of atmospheric mass between the subtropics of the Atlantic sector and the Arctic (Wallace and Gutzler 1981; Walker and Bliss 1932; Deser et al. 2010). One could argue that because NAO is more active from December to March we should not see this NAO related pattern in our results, as our study is restricted to the summer season. However, Gastineau and Frankignoul (2014) found a summer SST tripole, similar to the traditional SST tripole forced by the NAO during winter, that is lagged by an anticyclone over the subpolar NA.

4.3 Correlation patterns of collection sites of *Arctica islandica*

In Fig. 5 we present the one-point correlation patterns by correlating SSTs of the Icelandic Shelf for the period 1950–1999 with the entire central and northern NA Ocean. COBE2 shows strong and statistically significant correlations (at 1% level of significance) between the IS and the SSTs north of Iceland, while CanESM2 and CCSM4 models additionally show strong and statistically significant correlations along the coast of Europe ($r \sim 0.5$) and weak negative correlations over the southern domain of our study region. The positive correlation between these regions can

be expected because water mass transformation in the Iceland Sea produces Arctic Intermediate Water, which overflows the Greenland-Iceland and the Iceland-Faroe Ridges and contributes to the North Atlantic Deep Water (Swift and Aagaard 1981). Other models show statistically significant and strong correlations in the south and/or west part of our study region. As mentioned previously, NA SSTs are affected by the NAO and when correlation maps are displayed between SST anomalies and the NAO index they show positive anomalies in high latitudes and the subtropical area and negative anomalies over middle latitudes (Czaja and Frankignoul 2002; Bojariu and Gimeno 2003). NAO differently affects SSTs in different regions and that could be a reason for the negative correlation shown between the temperatures over the high and middle latitudes by some models in Fig. 5. Another interesting fact relates to the comparison of each graph of Figs. 3 and 5. Here we can see that for COBE2, and for most models, the dominant mode of variability of the NA summer SSTs is the spatial pattern shown by the correlation of SSTs with the ones of the Icelandic Shelf. The correlation between the summer SSTs of the NA and a location in the northern North Sea is shown in Fig. 6. COBE2, CSIRO, MPI-ESM and the GISS model show a dipole between the east and west NA SSTs. This dipole seems to represent the correlation pattern between SST anomalies and NAO shown by Wanner et al. (2001).

4.4 Long term trends

A calculation of the NA local SST trends was performed (Fig. 7) to compare the observed and model simulated temperature change for the period 1950–1999. CSIRO, GISS and the IPSL model show a basin wide SST warming trend that reaches 0.6 °C per decade, while the rest of the CMIP5 models show both cooling and warming trends in different areas. The COBE2 data show a statistically significant cooling, approximately equal to 0.4 °C per decade at high latitudes and a warming trend on the coast of Europe south of the Scandinavian Peninsula. In the work of Knutson et al. (2006) the annual regional surface temperature trends in observations and models were assessed for the period 1949–2000, using the GFDL model. Similar to our results, model and observations do not agree, as the ensemble mean trend map shows a warming trend around 0.1 °C per decade in the region of NA and the observed trend shows cooling approximately equal to 0.2 °C per decade. Possible reasons for the disagreement between the CMIP5 models and between models and data could be the models internal climate variability (Knutson et al. 2006), the sparse data in high latitudes before 1970 (Smith et al. 1996) that affect the overall trend of the regions in the northern NA of the second half of the twentieth century, and uncertainties in the

historical forcing, climate sensitivity and the rate of heat uptake by the ocean (Stott et al. 2000). Even though it is not a part of this study, the additional comparison of SSTs with available SST reconstructions could fill the gaps in the data sets based on space–time statistical methods and provide a better understanding of the differences between data and models.

4.5 Atlantic multidecadal oscillation

Positive AMO anomalies are shown by COBE2 (black line) during the 30 year periods 1850–1880, 1940–1970. All models exhibit AMO-like fluctuations (Appendix, Fig. 10A) and some of the models agree on the timing of the positive phases of AMO that the observed data show, but it is not clear whether that is internally generated by these models or they truly represent an externally forced signal. Several modeling studies have questioned the response of NA temperature variations to the ocean's internal variability. Lohmann et al. (2015) found the highest correlation coefficients between the AMO index and the NA SSTs in the tropical and subtropical regions, where the SSTs are mostly influenced by the external (volcanic, solar) radiative forcing (Otterå et al. 2010). The opposite picture is evident in several General Circulation Models (GCMs), which produce the AMO in the absence of external forcing (Knight et al. 2005). Ting et al. 2009 separated the externally forced component and the internally varying component of the NA SST variations and found that during the twentieth century the NA displayed an internal oscillation of considerable magnitude. Thus, the AMO cannot be fully explained by the radiative forcing (Ting et al. 2014; DelSole et al. 2011; Zhang and Wang 2013). The models' errors regarding the representation of AMO may suggest that their ability to simulate and predict at decadal time scales is compromised, because it could possibly mean that they do not incorporate the mechanisms associated to the generation of the AMO (or any other source of decadal variability like the PDO) and in turn incorporate or enhance variability at other frequencies (Ruiz-Barradas et al. 2013).

As the externally forced part of the AMO can be estimated from simulations and meaningfully compared to other empirically reconstructed AMO indices based on oceanic proxy data like *Arctica islandica*, we focused our analysis on the models ability to reveal the externally forced part of the AMO by comparing the observed and simulated AMO evolution, acknowledging the ongoing uncertainties regarding the forced and internal nature of the AMO. This analysis could give more confidence to the models' output regarding longer time scales such as the last millennium, as it could be assumed that if some of the models can reveal the externally forced part of the AMO during a period manifested by anthropogenic forcing, then it is plausible that

the same models can capture the externally forced part of the AMO during the preindustrial period.

5 Concluding remarks

The estimation of the skill of the CFR methods largely depends on the model used to evaluate the CFR method and the locations of the proxy network used to perform the spatially resolved CFR. Therefore, we first investigated the robustness of the CMIP5 simulated summer SSTs in the NA, compared to COBE2 data, for the second half of the twentieth century. Regarding the representation of the spatiotemporal characteristics of the NA SSTs we found that even though the second dominant pattern of NA SST co-variability is captured to a certain degree by most models evaluated in this study, the first dominant EOF pattern is well represented by the models CanESM2, CCSM4 and represented to a certain degree by the models MPI-ESM and HadCM3. We can, therefore, expect that these models will provide with a better representation of the NA region's co-variability that is important for CFRs. Uncertainties will be introduced in a potential application of CFR by any of the models, as most of them suffer biases that coincide with regions of maximum model inter-annual variability. The models with the highest climatological error distributions are found to be GISS, IPSL-CM5, HadCM3 and MRI-CGCM. The simulated AMO reveals similar evolutions in the 2nd half of the twentieth century within the models, but presents a prominent source of uncertainty for reconstructions.

To assess the capability of *Arctica islandica* collection sites to provide a good spatially resolved reconstruction, we tested the models ability to represent the co-variance of the spatially resolved NA SSTs and the SSTs at two of the *Arctica islandica* collection sites. Both the COBE2 data and the CMIP5 models showed that the IS and NS sites of *Arctica islandica* are promising sites that can be used to reconstruct the SSTs of the north-east Atlantic. The IS site does not provide any information about the central Atlantic, but it provides information about the northern NA, north of 60°N. A number of CMIP5 models, such as CanESM2, CCSM4, MPI-ESM and IPSL-CM5 can reproduce the IS teleconnection pattern shown by the COBE2 data. The site in the North Sea seems promising not only for the eastern NA basin but for the central Atlantic as well, as COBE2 shows a significant anti-correlation between the NS site and the central Atlantic. Some of the CMIP5 models can capture the NS teleconnection pattern shown by COBE2, while most of the models can capture the co-variance of the NS site SSTs to the surrounding waters. Therefore, we can expect that the given proxy record will contain a strong SST signal that might represent a basin-wide signal

for the north-east Atlantic, which is an important result in the context of using CMIP5 models to paleoclimate reconstructions based on the proxy sites of *Arctica islandica*. The models that are found to simulate well most of the spatiotemporal characteristics of the NA SSTs and the covariance of the SSTs of the two *Arctica islandica* collection sites are CanESM2, CCSM4 and MPI-ESM. In a forthcoming study we will assess the uncertainties relevant to paleoclimate reconstructions by using the best performing CMIP5 models, as evaluated in the context of this work, to test different CFR methods.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Arora VK, Scinocca JF, Boer GJ, et al. (2011) Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys Res Lett* 38: 3–8. doi:[10.1029/2010GL046270](https://doi.org/10.1029/2010GL046270)
- Bai Y, Cai WJ, He X, Zhai W, Pan D, Dai M, Yu P (2015) A mechanistic semi-analytical method for remotely sensing sea surface pCO₂ in river-dominated coastal oceans: a case study from the East China Sea. *J Geophys Res Oceans* 120:2331–2349. doi:[10.1002/2014JC010632](https://doi.org/10.1002/2014JC010632)
- Batehup R, McGregor S, Gallant AJE (2015) The influence of non-stationary teleconnections on palaeoclimate reconstructions of ENSO variance using a pseudoproxy framework. *Clim Past* 11: 1733–1749. doi:[10.5194/cp-11-1733-2015](https://doi.org/10.5194/cp-11-1733-2015)
- Bojariu R, Gimeno L (2003) Predictability and numerical modelling of the North Atlantic Oscillation. *Earth-Sci Rev* 63: 145–168. doi:[10.1016/S0012-8252\(03\)00036-9](https://doi.org/10.1016/S0012-8252(03)00036-9)
- Booth BBB, Dunstone NJ, Halloran PR, Andrews T, Bellouin N (2012) Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature* 484:228–232. doi:[10.1038/nature10946](https://doi.org/10.1038/nature10946)
- Bryan FO, Hecht MW, Smith RD (2007) Resolution convergence and sensitivity studies with North Atlantic circulation models. Part I: The western boundary current system. *Ocean Model* 16:141–159. doi:[10.1016/j.ocemod.2006.08.005](https://doi.org/10.1016/j.ocemod.2006.08.005)
- Butler PG, Richardson CA, Scourse JD, Wanamaker AD, Shammon TM, Bennell JD (2010) Marine climate in the Irish Sea: analysis of a 489-year marine master chronology derived from growth increments in the shell of the clam *Arctica islandica*. *Quat Sci Rev* 29:1614–1632. doi:[10.1016/j.quascirev.2009.07.010](https://doi.org/10.1016/j.quascirev.2009.07.010)
- Butler PG, Scourse JD, Richardson CA, Wanamaker AD, Bryant CL, Bennell JD (2009) Continuous marine radiocarbon reservoir calibration and the 13 C Suess effect in the Irish Sea: results from the first multi-centennial shell-based marine master chronology. *Earth Planet Sci Lett* 279:230–241. doi:[10.1016/j.epsl.2008.12.043](https://doi.org/10.1016/j.epsl.2008.12.043)
- Butler PG, Wanamaker AD, Scourse JD, Richardson CA, Reynolds DJ (2013) Variability of marine climate on the North Icelandic Shelf in a 1357-year proxy archive based on growth increments in the bivalve *Arctica islandica*. *Palaeogeogr Palaeoclimatol Palaeoecol* 373: 141–151. doi:[10.1016/j.palaeo.2012.01.016](https://doi.org/10.1016/j.palaeo.2012.01.016)
- Cannaby H, Hüsvrevoğlu YS (2009) The influence of low-frequency variability and long-term trends in North Atlantic sea surface temperature on Irish waters. *ICES J Mar Sci Journal du Conseil* 66:1480–1489. doi:[10.1093/icesjms/fsp062](https://doi.org/10.1093/icesjms/fsp062)
- Carton JA, Cao X, Giese BS, Da Silva AM (1996) Decadal and interannual SST variability in the Tropical Atlantic Ocean. *J Phys Oceanogr* 26:1165–1175. doi:[10.1175/1520-0485\(1996\)026<1165:DAISVI>2.0.CO;2](https://doi.org/10.1175/1520-0485(1996)026<1165:DAISVI>2.0.CO;2)
- Cattle H, Crossley J, Drewry DJ. (1995) Modelling arctic climate change [and discussion]. *Philos Trans R Soc Lond A: Math Phys Eng Sci* 352:201–213. doi:[10.1098/rsta.1995.0064](https://doi.org/10.1098/rsta.1995.0064)
- Cheng W, Chiang JCH, Zhang D (2013) Atlantic meridional overturning circulation (AMOC) in CMIP5 Models: RCP and historical simulations. *J Clim* 26:7187–7197. doi:[10.1175/JCLI-D-12-00496.1](https://doi.org/10.1175/JCLI-D-12-00496.1)
- Coats S, Smerdon JE, Cook BI, Seager R (2013) Stationarity of the tropical pacific teleconnection to North America in CMIP5/PMIP3 model simulations. *Geophys Res Lett* 40:4927–4932. doi:[10.1002/grl.50938](https://doi.org/10.1002/grl.50938)
- Collins M, Tett SFB, Cooper C (2001) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn* 17:61–81. doi:[10.1007/s003820000094](https://doi.org/10.1007/s003820000094)
- Cox PM, Betts RA, Bunton CB, Essery RLH, Rowntree PR, Smith J (1999) The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim Dyn* 15:183–203. doi:[10.1007/s003820050276](https://doi.org/10.1007/s003820050276)
- Cunningham LK, Austin WE, Knudsen KL, Eiríksson J, Scourse JD, Wanamaker AD, Butler PG, Cage AG, Richter T, Husum K, Hald M (2013) Reconstructions of surface ocean conditions from the northeast Atlantic and Nordic seas during the last millennium. *Holocene* 23:921–935. doi:[10.1177/0959683613479677](https://doi.org/10.1177/0959683613479677)
- Czaja A and Frankignoul C (2002) Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation. *J Clim* 15:606–623. doi: [10.1175/1520-0442\(2002\)015<0606:OIOASA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0606:OIOASA>2.0.CO;2)
- Dahlgren TG, Weinberg JR, Halanych KM (2000) Phylogeography of the ocean quahog (*Arctica islandica*): influences of paleoclimate on genetic diversity and species range. *Mar Biol* 137:487–495. doi:[10.1007/s002270000342](https://doi.org/10.1007/s002270000342)
- Danabasoglu G, Bates SC, Briegleb BP, et al. (2012) The CCSM4 ocean component. *J Clim* 25:1361–1389. doi: [10.1175/JCLI-D-11-00091.1](https://doi.org/10.1175/JCLI-D-11-00091.1)
- Dee SG, Steiger NJ, Emile-Geay J, Hakim GJ (2016) On the utility of proxy system models for estimating climate states over the common era. *J Adv Model Earth Syst*. doi:[10.1002/2016MS000677](https://doi.org/10.1002/2016MS000677)
- DelSole T, Tippet MK, Shukla J (2011) A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J Clim* 24:909–926. doi:[10.1175/2010JCLI3659.1](https://doi.org/10.1175/2010JCLI3659.1)
- Deser C, Alexander MA, Xie SP, Phillips AS (2010) Sea surface temperature variability: patterns and mechanisms. *Ann Rev Mar Sci* 2:115–143. doi:[10.1146/annurev-marine-120408-151453](https://doi.org/10.1146/annurev-marine-120408-151453)
- Deser C, Blackmon ML (1993) Surface climate variations over the North Atlantic Ocean during Winter: 1900–1989. *J Clim* 6:1743–1753. doi:[10.1175/1520-0442\(1993\)006<1743:SCVO TN>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1743:SCVO TN>2.0.CO;2)
- Dufresne JL, Foujols MA, Denvil S et al (2013) Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim Dyn* 40:2123–2165. doi:[10.1007/s00382-012-1636-1](https://doi.org/10.1007/s00382-012-1636-1)
- Eagle RA, Eiler JM, Tripathi AK, et al. (2013) The influence of temperature and seawater carbonate saturation state on 13 C-18O

- bond ordering in bivalve mollusks. *Biogeosciences* 10. doi: <https://doi.org/10.5194/bg-10-4591-2013>
- Fan M and Schneider EK (2010) Observed decadal North Atlantic Tripole SST variability. Part I: weather noise forcing and coupled response. *J Atmos Sci* 69: 35–50. doi: [10.1175/JAS-D-11-018.1](https://doi.org/10.1175/JAS-D-11-018.1)
- Franco B, Fettweis X, Lang C, Erpicum M (2012) Impact of spatial resolution on the modelling of the Greenland ice sheet surface mass balance between 1990–2010, using the regional climate model MAR. *Cryosphere* 6:695–711. doi: [10.5194/tc-6-695-2012](https://doi.org/10.5194/tc-6-695-2012)
- Gallant AJE, Phipps SJ, Karoly DJ, Mullan AB, Lorrey AM (2013) Nonstationary Australasian teleconnections and implications for paleoclimate reconstructions. *J Clim* 26: 8827–8849. doi: [10.1175/JCLI-D-12-00338.1](https://doi.org/10.1175/JCLI-D-12-00338.1)
- Gastineau G, Frankignoul C (2014) Influence of the North Atlantic SST variability on the atmospheric circulation during the twentieth century. *J Clim* 28:1396–1416. doi: [10.1175/JCLI-D-14-00424.1](https://doi.org/10.1175/JCLI-D-14-00424.1)
- Gent PR, Danabasoglu G, Donner LJ et al (2011) The community climate system model version 4. *J Clim* 24:4973–4991. doi: [10.1175/2011JCLI4083.1](https://doi.org/10.1175/2011JCLI4083.1)
- Gillikin DP, Dehairs F, Lorrain A, Steenmans D, Baeyens W, André L (2006) Barium uptake into the shells of the common mussel (*Mytilus edulis*) and the potential for estuarine paleo-chemistry reconstruction. *Geochimica et Cosmochimica Acta* 70: 395–407. doi: [10.1016/j.gca.2005.09.015](https://doi.org/10.1016/j.gca.2005.09.015)
- Gordon C, Cooper C, Senior CA et al (2000) The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn* 16:147–168. doi: [10.1007/s003820050010](https://doi.org/10.1007/s003820050010)
- Gordon HB, O'Farrell SP (1997) Transient climate change in the CSIRO coupled model with dynamic sea ice. *Mon Weather Rev* 125: 875–908. doi: [10.1175/1520-0493\(1997\)125<0875:TCCITC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0875:TCCITC>2.0.CO;2)
- Gordon HB, O'Farrell S, Collier M, et al. (2010) The CSIRO Mk3. 5 climate model. CSIRO and Bureau of Meteorology.
- Griffies SM, Harrison MJ, Pacanowski RC, Rosati A (2004) A technical guide to MOM4. GFDL Ocean Group Tech Rep 5:371
- Halvorsen MH, Smedsrud LH, Zhang R, Kloster K (2015) Fram Strait spring ice export and September Arctic sea ice. *Cryosphere Discuss* 9: 4205–4235. doi: [10.5194/tcd-9-4205-2015](https://doi.org/10.5194/tcd-9-4205-2015)
- Hasumi H (2014) A review on ocean resolution dependence of climate biases in AOGCMs. *CLIVAR Exch* 65: 7–9.
- Hijmans RJ, Cameron SE, Parra JL, Jones G, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978. doi: [10.1002/joc.1276](https://doi.org/10.1002/joc.1276)
- Hirahara S, Ishii M, Fukuda Y (2014) Centennial-scale sea surface temperature analysis and its uncertainty. *J Climate* 27:57–75. doi: [10.1175/JCLI-D-12-00837.1](https://doi.org/10.1175/JCLI-D-12-00837.1)
- Hourdin F, Foujols MA, Codron F et al (2013) Impact of the LMDZ atmospheric grid configuration on the climate and sensitivity of the IPSL-CM5A coupled model. *Clim Dyn* 40:2167–2192. doi: [10.1007/s00382-012-1411-3](https://doi.org/10.1007/s00382-012-1411-3)
- Hunke E, Lipscomb W (2008) CICE: the Los Alamos sea ice model user's manual, version 4, Tech. Rep, LA-CC-06-012. Los Alamos National Laboratory, Los Alamos, NM
- Ishii M, Shouji A, Sugimoto S, Matsumoto T (2005) Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. *Int J Climatol* 25:865–879. doi: [10.1002/joc.1169](https://doi.org/10.1002/joc.1169)
- Jha B, Hu ZZ, Kumar A (2014) SST and ENSO variability and change simulated in historical experiments of CMIP5 models. *Clim Dyn* 42:2113–2124. doi: [10.1007/s00382-013-1803-z](https://doi.org/10.1007/s00382-013-1803-z)
- Jones PD, Briffa KR, Osborn TJ et al (2009) High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *Holocene* 19:3–49. doi: [10.1177/0959683608098952](https://doi.org/10.1177/0959683608098952)
- Jones GS, Stott PA, Christidis N (2013) Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *J Geophys Res Atmos* 118:4001–4024. doi: [10.1002/jgrd.50239](https://doi.org/10.1002/jgrd.50239)
- Jungclaus JH, Fischer N, Haak H, et al. (2013) Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *J Adv Model Earth Syst* 5:422–446. doi: [10.1002/jame.20023](https://doi.org/10.1002/jame.20023)
- Jungclaus JH, Lohmann K, Zanchettin D (2014) Enhanced 20th-century heat transfer to the Arctic simulated in the context of climate variations over the last millennium. *Clim Past* 10: 2201–2213. doi: [10.5194/cp-10-2201-2014](https://doi.org/10.5194/cp-10-2201-2014)
- Kaplan A, Cane MA, Kushnir Y (2001) Reduced space approach to the optimal analysis of historical marine observations: accomplishments, difficulties and prospects. *WMO Guide to the Applications of Marine Climatology*
- Kauker F (2003) Variability of Arctic and North Atlantic sea ice: a combined analysis of model results and observations from 1978 to 2001. *J Geophys Res* 108. doi: [10.1029/2002JC001573](https://doi.org/10.1029/2002JC001573)
- Knight JR, Allan RJ, Folland CK, Vellinga M, Mann ME (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys Res Lett* 32:1–4. doi: [10.1029/2005GL024233](https://doi.org/10.1029/2005GL024233)
- Knutson TR, Delworth TL, Dixon KW, Held IM, Lu J, Ramaswamy V, Schwarzkopf MD, Stenchikov G, Stouffer RJ (2006) Assessment of twentieth-century regional surface temperature trends using the GFDL CM2 coupled models. *J Clim* 19:1624–1651. doi: [10.1175/JCLI3709.1](https://doi.org/10.1175/JCLI3709.1)
- Kowalczyk EA, Garratt JR, Krummel PB (1994) Implementation of a soil-canopy scheme into the CSIRO GCM-regional aspects of the model response. CSIRO
- Krinner G, Viovy N, de Noblet-Ducoudré N, Ogée J, Polcher J, Friedlingstein P, Ciais P, Sitch S, Prentice IC (2005) A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Glob Biogeochem Cycles* 19. doi: [10.1029/2003GB002199](https://doi.org/10.1029/2003GB002199)
- Kushnir Y (1994) Interdecadal variations in North Atlantic Sea surface temperature and associated atmospheric conditions. *J Clim* 7:141–157. doi: [10.1175/1520-0442\(1994\)007<0141:IVINAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<0141:IVINAS>2.0.CO;2)
- Lawrence DM, Oleson KW, Flanner MG et al (2012) The CCSM4 land simulation, 1850–2005: assessment of surface climate and new capabilities. *J Clim* 25:2240–2260. doi: [10.1175/JCLI-D-11-00103.1](https://doi.org/10.1175/JCLI-D-11-00103.1)
- Legates DR (1991) The effect of domain shape on principal components analyses. *Int J Climatol* 11:135–146. doi: [10.1002/joc.3370110203](https://doi.org/10.1002/joc.3370110203)
- Liu H, Wang C, Lee SK, Enfield D (2013) Atlantic warm pool variability in the CMIP5 simulations. *J Clim* 26:5315–5336. doi: [10.1175/JCLI-D-12-00556.1](https://doi.org/10.1175/JCLI-D-12-00556.1)
- Loder JW, Wang Z (2015) Trends and variability of sea surface temperature in the Northwest Atlantic from three historical gridded datasets. *Atmos-Ocean* 53: 510–528. doi: [10.1080/07055900.2015.1071237](https://doi.org/10.1080/07055900.2015.1071237)
- Logemann K, Harms I (2006) High resolution modelling of the North Icelandic Irminger Current (NIIC). *Ocean Sci* 2:291–304. doi: [10.5194/os-2-291-2006](https://doi.org/10.5194/os-2-291-2006)
- Lohmann K, Mignot J, Langehaug HR, et al. (2015) Using simulations of the last millennium to understand climate variability seen in palaeo-observations: similar variation of Iceland-Scotland overflow strength and Atlantic Multidecadal Oscillation. *Clim Past* 11:203–216. doi: [10.5194/cp-11-203-2015](https://doi.org/10.5194/cp-11-203-2015)

- Madec G, NEMO team (2008) NEMO ocean engine. Note du Pole de modelisation, Institut Pierre-Simon Laplace (IPSL), France
- Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstruction methods. *J Geophys Res Atmos* 112:1–18. doi:[10.1029/2006JD008272](https://doi.org/10.1029/2006JD008272)
- Marshall J et al (2001) North Atlantic climate variability: phenomena, impacts and mechanisms. *Int J Clim* 21:1863–1898. doi:[10.1002/joc.693](https://doi.org/10.1002/joc.693)
- McGregor HV, Evans MN, Goosse H, Leduc G, Martrat B, Addison JA, Mortyn PG, Oppo DW, Seidenkrantz MS, Sicre MA, Phipps SJ (2015) Robust global ocean cooling trend for the pre-industrial Common Era. *Nature Geosci* 8: 671–677. doi:[10.1038/ngeo2510](https://doi.org/10.1038/ngeo2510)
- Mette MJ, Wanamaker AD, Carroll ML, Ambrose WG, Retelle MJ (2016) Linking large-scale climate variability with Arctica islandica shell growth and geochemistry in northern Norway. *Limnol Oceanogr* 61:748–764. doi:[10.1002/lno.10252](https://doi.org/10.1002/lno.10252)
- Neale RB, Richter J, Park S et al (2013) The mean climate of the community atmosphere model (CAM4) in forced SST and fully coupled experiments. *J Clim* 26:5150–5168. doi:[10.1175/JCLI-D-12-00236.1](https://doi.org/10.1175/JCLI-D-12-00236.1)
- Nicol D (1951) Recent species of the veneroid pelecypod Arctica. *J Wash Acad Sci* 41:102–106
- Otterå OH, Bentsen M, Drange H, Suo L (2010) External forcing as a metronome for Atlantic multidecadal variability. *Nat Geosci* 3: 688–694. doi:[10.1038/ngeo955](https://doi.org/10.1038/ngeo955)
- Perez J, Menendez M, Mendez FJ, Losada IJ (2014) Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Clim Dyn* 43:1–18. doi:[10.1007/s00382-014-2078-8](https://doi.org/10.1007/s00382-014-2078-8)
- Pope VD, Gallani ML, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Clim Dyn* 16:123–146. doi:[10.1007/s003820050009](https://doi.org/10.1007/s003820050009)
- Qu X, Huang G (2014) The decadal variability of the tropical Indian Ocean SST—the South Asian High relation: CMIP5 model study. *Clim Dyn* 45:273–289. doi:[10.1007/s00382-014-2285-3](https://doi.org/10.1007/s00382-014-2285-3)
- Rayner NA, Brohan P, Parker DE, et al. (2006) Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: the HadSST2 dataset. *J Clim* 19: 446–469. doi: [10.1175/JCLI3637.1](https://doi.org/10.1175/JCLI3637.1)
- Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J Geophys Res Atmos* 108:4407. doi:[10.1029/2002JD002670](https://doi.org/10.1029/2002JD002670)
- Reick CH, Raddatz T, Brovkin V, Gayler V (2013) Representation of natural and anthropogenic land cover change in MPI-ESM. *J Adv Model Earth Syst* 5: 459–482. doi:[10.1002/jame.20022](https://doi.org/10.1002/jame.20022)
- Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. *J Clim* 15: 1609–1625. doi: [10.1175/1520-0442\(2002\)015<1609:AIHSAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIHSAS>2.0.CO;2)
- Rowell TW, Chaisson DR, McLane JT (1990) Size and age of sexual maturity and annual gametogenic cycle in the ocean quahog, Arctica islandica (Linnaeus, 1767), from coastal waters in Nova Scotia, Canada. *J Shellfish Res* 9:195–203
- Ruiz-Barradas A, Nigam S, Kavvada A (2013) The Atlantic Multidecadal Oscillation in twentieth century climate simulations: uneven progress from CMIP3 to CMIP5. *Clim Dyn* 41:3301–3315. doi:[10.1007/s00382-013-1810-0](https://doi.org/10.1007/s00382-013-1810-0)
- Rutherford S, Mann ME, Delworth TL, Stouffer RJ (2003) Climate field reconstruction under stationary and nonstationary forcing. *J Clim* 16: 462–479. doi:[10.1175/1520-0442\(2003\)016<0462:CFRUSA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0462:CFRUSA>2.0.CO;2)
- Von Salzen K, Scinocca JF, McFarlane NA, Li J, Cole JN, Plummer D, Verseghy D, Reader MC, Ma X, Lazare M, Solheim L (2013) The Canadian fourth generation atmospheric global climate model (CanAM4). Part I: representation of physical processes. *Atmos-Ocean* 51: 104–125. doi:[10.1080/07055900.2012.755610](https://doi.org/10.1080/07055900.2012.755610)
- Schleussner CF, Runge J, Lehmann J, Levermann A (2014) The role of the North Atlantic overturning and deep ocean for multi-decadal global-mean-temperature variability. *Earth Syst Dyn* 5: 103–115. doi:[10.5194/esd-5-103-2014](https://doi.org/10.5194/esd-5-103-2014)
- Schmidt et al (2006) Present-day atmospheric simulations using GISS ModelE: comparison to in situ, satellite, and reanalysis data. *J Clim* 19: 153–192. doi: [10.1175/JCLI3612.1](https://doi.org/10.1175/JCLI3612.1)
- Schöne BR (2013) Arctica islandica (Bivalvia): a unique paleoenvironmental archive of the northern North Atlantic Ocean. *Glob Planet Change* 111: 199–225. doi:[10.1016/j.gloplacha.2013.09.013](https://doi.org/10.1016/j.gloplacha.2013.09.013)
- Schöne BR, Castro AD, Fiebig J, Houk SD, Oschmann W, Kröncke I (2004) Sea surface water temperatures over the period 1884–1983 reconstructed from oxygen isotope ratios of a bivalve mollusk shell (Arctica islandica, southern North Sea). *Palaeogeogr Palaeoclimatol Palaeoecol* 212:215–232. doi:[10.1016/j.palaeo.2004.05.024](https://doi.org/10.1016/j.palaeo.2004.05.024)
- Schöne BR, Fiebig J, Pfeiffer M et al (2005a) Climate records from a bivalved Methuselah. *Palaeogeogr Palaeoclimatol Palaeoecol* 228:130–148. doi:[10.1016/j.palaeo.2005.03.049](https://doi.org/10.1016/j.palaeo.2005.03.049)
- Schöne BR, Houk SD, Castro ADF et al (2005b) Daily growth rates in shells of Arctica islandica: assessing sub-seasonal environmental controls on a long-lived bivalve mollusk. *Palaios* 20:78–92. doi:[10.2110/palo.2003.p03-101](https://doi.org/10.2110/palo.2003.p03-101)
- Schöne BR, Oschmann W, Rössler J et al (2003) North Atlantic oscillation dynamics recorded in shells of a long-lived bivalve mollusk. *Geology* 31:1037–1040. doi:[10.1130/G20013.1](https://doi.org/10.1130/G20013.1)
- Schöne BR, Zhang Z, Radermacher P, et al. (2011) Sr/Ca and Mg/Ca ratios of ontogenetically old, long-lived bivalve shells (Arctica islandica) and their function as paleotemperature proxies. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 302: 52–64. doi: [10.1016/j.palaeo.2010.03.016](https://doi.org/10.1016/j.palaeo.2010.03.016)
- Scourse J, Richardson C, Forsythe G et al (2006) First cross-matched floating chronology from the marine fossil record: data from growth lines of the long-lived bivalve mollusc Arctica islandica. *Holocene* 16:967–974. doi:[10.1177/0959683606h1987rp](https://doi.org/10.1177/0959683606h1987rp)
- Shupe MD, Intrieri JM (2004) Cloud radiative forcing of the Arctic surface: the influence of cloud properties, surface albedo, and solar zenith angle. *J Clim* 17:616–628. doi:[10.1175/1520-0442\(2004\)017<0616:CRFOTA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0616:CRFOTA>2.0.CO;2)
- Smerdon JE, Coats S, Ault TR (2016) Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era. *Clim Dyn* 6:1921–1942. doi:[10.1007/s00382-015-2684-0](https://doi.org/10.1007/s00382-015-2684-0)
- Smerdon JE, Kaplan A, Zorita E, González-Rouco JF, Evans MN (2011) Spatial performance of four climate field reconstruction methods targeting the Common Era. *Geophys Res Lett* 38:L11705. doi:[10.1029/2011GL047372](https://doi.org/10.1029/2011GL047372)
- Smith RD, Maltrud ME, Bryan FO, Hecht MW, Alamos L (2000) Numerical simulation of the North Atlantic Ocean at 1/10 degrees. *J Phys Oceanogr* 30:1532–1561. doi:[10.1175/1520-0485\(2000\)030<1532:nsotma>2.0.co;2](https://doi.org/10.1175/1520-0485(2000)030<1532:nsotma>2.0.co;2)
- Smith TM, Reynolds RW (2004) Improved extended reconstruction of SST (1854–1997). *J Clim* 17: 2466–2477. doi: [10.1175/1520-0442\(2004\)017<2466:IEROS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2466:IEROS>2.0.CO;2)
- Smith TM, Reynolds RW, Livezey RE, Stokes DC (1996) Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J Clim* 9:1403–1420. doi:[10.1175/1520-0442\(1996\)009<1403:ROHSST>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<1403:ROHSST>2.0.CO;2)

- Stevens B, Giorgetta M, Esch M, et al. (2013) Atmospheric component of the MPI-M earth system model: ECHAM6. *J Adv Model Earth Syst* 5:146–172. doi:[10.1002/jame.20015](https://doi.org/10.1002/jame.20015)
- Von Storch H, Zwiers FW (1999) *Statistical analysis in climate research*. Cambridge University Press, Cambridge. doi:[10.1017/CBO9780511612336](https://doi.org/10.1017/CBO9780511612336)
- Stott PA, Tett SFB, Jones GS, Allen MR (2000) External control of 20th century temperature by natural and anthropogenic forcings. *Science* 290:2133–2137. doi:[10.1126/science.290.5499.2133](https://doi.org/10.1126/science.290.5499.2133)
- Sun S, Bleck R (2006) Multi-century simulations with the coupled GISS-HYCOM climate model: control experiments. *Clim Dyn* 26:407–428. doi:[10.1007/s00382-005-0091-7](https://doi.org/10.1007/s00382-005-0091-7)
- Swift JH, Aagaard K (1981) Seasonal transitions and water mass formation in the Iceland and Greenland seas. *Deep Sea Research Part A. Oceanographic. Res Pap Hist Med Assoc* 28:1107–1129. doi:[10.1016/0198-0149\(81\)90050-9](https://doi.org/10.1016/0198-0149(81)90050-9)
- Taboada FG, Anadón R (2012) Patterns of change in sea surface temperature in the North Atlantic during the last three decades: beyond mean trends. *Clim Change* 115:419–431. doi:[10.1007/s10584-012-0485-6](https://doi.org/10.1007/s10584-012-0485-6)
- Taylor KE, Stouffer RJ, A. MG (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 3:485–498. doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
- Ting M, Kushnir Y, Li C (2014) North Atlantic multidecadal SST oscillation: External forcing versus internal variability. *J Mar Syst* 133: 27–38. doi:[10.1016/j.jmarsys.2013.07.006](https://doi.org/10.1016/j.jmarsys.2013.07.006)
- Ting M, Kushnir Y, Seager R, Li C (2009) Forced and internal twentieth-century SST trends in the North Atlantic*. *J Clim* 22: 1469–1481. doi:[10.1175/2008JCLI2561.1](https://doi.org/10.1175/2008JCLI2561.1)
- Ting M, Kushnir Y, Seager R, Li C (2011) Robust features of Atlantic multi-decadal variability and its climate impacts. *Geophys Res Lett* 38:1–6. doi:[10.1029/2011GL048712](https://doi.org/10.1029/2011GL048712)
- Valcke S, Craig T, Coquart L (2006) OASIS3-MCT User Guide OASIS3-MCT 3.0. CERFACS/CNRS SUC URA No1875.
- Volodin EM, Galin VY, Gusev AV, Diansky NA, Smyshlyayev SP, Yakovlev NG (2010) Earth system model of INM RAS3. *Russ J Numer Anal Math Model* 25:531–545.
- Walker GT, Bliss EW (1932) World weather. *V Mem R Meteor Soc* 4: 53–84.
- Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon Weather Rev* 109: 784–812. doi:[10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2)
- Wanamaker AD Jr, Scourse JD, Richardson CA, Butler PG, Reynolds DJ, Ridgeway I (2008a) Absolute chronologies from the ocean: Records from the longest-lived, non-colonial animals on earth. *PAGES News* 16:4–6
- Wanamaker AD Jr, Heinemeier J, Scourse JD et al (2008b) Very long-lived mollusks confirm 17th century ad tephra-based radiocarbon reservoir ages for North Icelandic shelf waters. *Radiocarbon* 50:399–412. doi:[10.2458/azu_js_rc.50.3222](https://doi.org/10.2458/azu_js_rc.50.3222)
- Wanamaker AD Jr, Kreutz KJ, Schöne BR et al (2009) A late Holocene paleo-productivity record in the western Gulf of Maine, USA, inferred from growth histories of the long-lived ocean quahog (*Arctica islandica*). *Int J Earth Sci* 98:19–29. doi:[10.1007/s00531-008-0318-z](https://doi.org/10.1007/s00531-008-0318-z)
- Wanamaker AD Jr, Kreutz KJ, Schöne BR, Introne DS (2011) Gulf of Maine shells reveal changes in seawater temperature seasonality during the Medieval Climate Anomaly and the Little Ice Age. *Palaeogeogr Palaeoclimatol Palaeoecol* 302:43–51. doi:[10.1016/j.palaeo.2010.06.005](https://doi.org/10.1016/j.palaeo.2010.06.005)
- Wanamaker AD Jr, Butler PG, Scourse JD et al (2012) Surface changes in the North Atlantic meridional overturning circulation during the last millennium. *Nat Commun* 3:899. doi:[10.1038/ncomms1901](https://doi.org/10.1038/ncomms1901)
- Wanamaker AD, Mette MJ, Whitney NM (2016) CLIVAR Newsletter: the potential for the long-lived bivalve *Arctica islandica* to contribute to our understanding of past AMOC dynamics. *US CLIVAR Var* 14:13–19
- Wang C, Zhang L, Lee S (2014) A global perspective on CMIP5 climate model biases. *Nat Clim Chang* 4:201–205. doi:[10.1038/NCLIMATE2118](https://doi.org/10.1038/NCLIMATE2118)
- Wang G, Dommeneget D, Frauen C (2015) An evaluation of the CMIP3 and CMIP5 simulations in their skill of simulating the spatial structure of SST variability. *Clim Dyn* 44:95–114. doi:[10.1007/s00382-014-2154-0](https://doi.org/10.1007/s00382-014-2154-0)
- Wanner H, Brönnimann S, Casty C, Gyalistras D, Luterbacher J, Schmutz C, Stephenson DB, Xoplaki E (2001) North Atlantic oscillation—concepts and studies. *Surv Geophys* 22:321–381. doi:[10.1023/A:1014217317898](https://doi.org/10.1023/A:1014217317898)
- Willebrand J, Barnier B, Böning C et al (2001) Circulation characteristics in three eddy-permitting models of the North Atlantic. *Prog Oceanogr* 48:123–161. doi:[10.1016/S0079-6611\(01\)00003-9](https://doi.org/10.1016/S0079-6611(01)00003-9)
- Winton M (2000) A reformulated three-layer sea ice model. *J Atmos Ocean Technol* 17:525–531. doi:[10.1175/1520-0426\(2000\)017<0525:ARTLSI>2.0.CO;2](https://doi.org/10.1175/1520-0426(2000)017<0525:ARTLSI>2.0.CO;2)
- Witbaard R (1996) Growth variations in *Arctica islandica* L. (Mollusca): a reflection of hydrography-related food supply. *ICES J Mar Sci—Journal du Conseil* 53:981–987. doi:[10.1006/jmsc.1996.0122](https://doi.org/10.1006/jmsc.1996.0122)
- Witbaard R, Jansma E, Klaassen US (2003) Copepods link quahog growth to climate. *J Sea Res* 50: 77–83. doi: [10.1016/S1385-1101\(03\)00040-6](https://doi.org/10.1016/S1385-1101(03)00040-6)
- Witbaard R, Klein R (1994) Long-term trends on the effects of the southern North Sea beamtrawl fishery on the bivalve mollusc *Arctica islandica* L. (Mollusca, bivalvia). *ICES J Mar Sci—Journal du Conseil* 51:99–105. doi:[10.1006/jmsc.1994.1009](https://doi.org/10.1006/jmsc.1994.1009)
- Worley SJ, Woodruff SD, Reynolds RW, Lubker SJ, Lott N (2005) ICOADS release 2.1 data and products. *Int J Clim* 25:823–842. doi:[10.1002/joc.1166](https://doi.org/10.1002/joc.1166)
- Yang D, Saenko O a (2012) Ocean heat transport and its projected change in CanESM2. *J Clim* 25:8148–8163. doi:[10.1175/JCLI-D-11-00715.1](https://doi.org/10.1175/JCLI-D-11-00715.1)
- Yasunaka S, Hanawa K (2002) Regime shifts found in the Northern Hemisphere SST field. *J Meteorol Soc Jpn* 80:119–135
- Yasunaka S, Hanawa K (2011) Intercomparison of historical sea surface temperature datasets. *Int J Climatol* 31:1056–1073. doi:[10.1002/joc.2104](https://doi.org/10.1002/joc.2104)
- Yoshimura H, Yukimoto S (2008) Development of a simple coupler (Scup) for earth system modeling. *Pap Meteor Geophys* 59: 19–29. doi:[10.2467/mripapers.59.19](https://doi.org/10.2467/mripapers.59.19)
- Zhang X, Vincent LA, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean* 38: 395–429. doi:[10.1080/07055900.2000.9649654](https://doi.org/10.1080/07055900.2000.9649654)
- Zhang L, Wang C (2013) Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations. *J Geophys Res Ocean* 118:5772–5791. doi:[10.1002/jgrc.20390](https://doi.org/10.1002/jgrc.20390)