HAL open science

Propagation of Thermohaline Anomalies and Their Predictive Potential along the Atlantic Water Pathway

H. R. Langehaug, Pablo Ortega, F. Counillon, Daniela Matei, E. Maroon, N. Keenlyside, Juliette Mignot, Y. Wang, Didier Swingedouw, Ingo Bethke, et al.

▶ **To cite this version:**

H. R. Langehaug, Pablo Ortega, F. Counillon, Daniela Matei, E. Maroon, et al.. Propagation of Thermohaline Anomalies and Their Predictive Potential along the Atlantic Water Pathway. Journal of Climate, 2022, 35, pp.2111-2131. 10.1175/JCLI-D-20-1007.1 . insu-03678655

HAL Id: insu-03678655

https://insu.hal.science/insu-03678655

Submitted on 25 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Propagation of Thermohaline Anomalies and Their Predictive Potential along the Atlantic Water Pathway

H. R. Langehaug,[a] P. Ortega,[b] F. Counillon,[a,c] D. Matei,[d] E. Maroon,[e] N. Keenlyside,[c] J. Mignot,[f] Y. Wang,[a] D. Swingedouw,[g] I. Bethke,[c] S. Yang,[h] G. Danabasoglu,[i] A. Bellucci,[j,k] P. Ruggieri,[j,l] D. Nicolì,[j] and M. Årthun[c]

[a] *Nansen Environmental and Remote Sensing Center, and Bjerknes Centre for Climate Research, Bergen, Norway*
[b] *Barcelona Supercomputing Center, Barcelona, Spain*
[c] *Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway*
[d] *Max Planck Institute for Meteorology, Hamburg, Germany*
[e] *University of Wisconsin–Madison, Madison, Wisconsin*
[f] *UMR LOCEAN, Sorbonne Université/IRD/MNHN/CNRS, IPSL, Paris, France*
[g] *Environnements et paléoenvironnements océaniques et continentaux (EPOC), UMR CNRS 5805 EPOC, Université de Bordeaux, Allée Geoffroy-Saint-Hilaire, Pessac, France*
[h] *Danish Meteorological Institute, Copenhagen, Denmark*
[i] *National Center for Atmospheric Research, Boulder, Colorado*
[j] *Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy*
[k] *Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche, Bologna, Italy*
[l] *Department of Physics and Astronomy, University of Bologna, Bologna, Italy*

(Manuscript received 30 December 2020, in final form 20 December 2021)

ABSTRACT: We assess to what extent seven state-of-the-art dynamical prediction systems can retrospectively predict winter sea surface temperature (SST) in the subpolar North Atlantic and the Nordic seas in the period 1970–2005. We focus on the region where warm water flows poleward (i.e., the Atlantic water pathway to the Arctic) and on interannual-to-decadal time scales. Observational studies demonstrate predictability several years in advance in this region, but we find that SST skill is low with significant skill only at a lead time of 1–2 years. To better understand why the prediction systems have predictive skill or lack thereof, we assess the skill of the systems to reproduce a spatiotemporal SST pattern based on observations. The physical mechanism underlying this pattern is a propagation of oceanic anomalies from low to high latitudes along the major currents, the North Atlantic Current and the Norwegian Atlantic Current. We find that the prediction systems have difficulties in reproducing this pattern. To identify whether the misrepresentation is due to incorrect model physics, we assess the respective uninitialized historical simulations. These simulations also tend to misrepresent the spatiotemporal SST pattern, indicating that the physical mechanism is not properly simulated. However, the representation of the pattern is slightly degraded in the predictions compared to historical runs, which could be a result of initialization shocks and forecast drift effects. Ways to enhance predictions could include improved initialization and better simulation of poleward circulation of anomalies. This might require model resolutions in which flow over complex bathymetry and the physics of mesoscale ocean eddies and their interactions with the atmosphere are resolved.

SIGNIFICANCE STATEMENT: In this study, we find that dynamical prediction systems and their respective climate models struggle to realistically represent ocean surface temperature variability in the eastern subpolar North Atlantic and Nordic seas on interannual-to-decadal time scales. In previous studies, ocean advection is proposed as a key mechanism in propagating temperature anomalies along the Atlantic water pathway toward the Arctic Ocean. Our analysis suggests that the predicted temperature anomalies are not properly circulated to the north; this is a result of model errors that seems to be exacerbated by the effect of initialization shocks and forecast drift. Better climate predictions in the study region will thus require improving the initialization step, as well as enhancing process representation in the climate models.

KEYWORDS: Ocean circulation; Climate prediction; Decadal variability

1. Introduction

Slow variations in heat transport of the Gulf Stream's extension toward the Arctic Ocean influence western European climate, Arctic sea ice conditions, and northern fisheries (e.g., Yeager et al. 2015; Årthun et al. 2017, 2019). It would

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WCAS-D-20-0160.s1.

Corresponding author: Helene R. Langehaug, helene.langehaug@nersc.no

This article is licensed under a Creative Commons Attribution 4.0 license (http://creativecommons.org/licenses/by/4.0/).

DOI: 10.1175/JCLI-D-20-1007.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

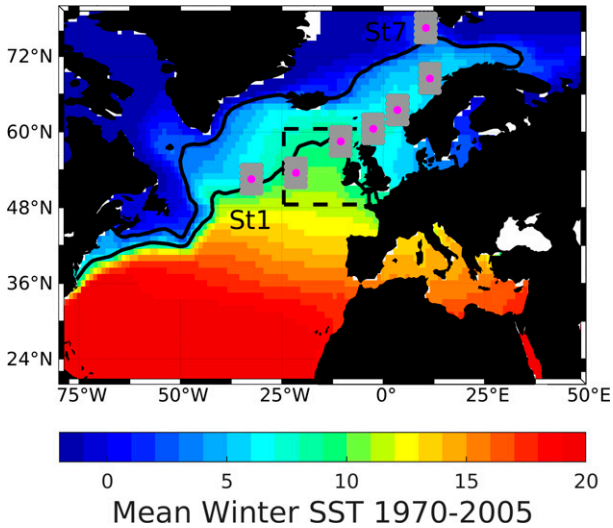

FIG. 1. The focus region of this study is the Atlantic water pathway, from the subpolar North Atlantic to the Fram Strait, dotted by seven stations (magenta circles). Station 1 (St1) is located in the subpolar region and Station 7 (St7) in the Fram Strait. Time series of temperature and salinity from these seven stations are used to track northward propagation of hydrographic anomalies in the Atlantic Water. When we refer to the stations in this study, we use an average value of grid points close to the stations (within a box with a size of approximately 5° × 5°; gray boxes). Colors indicate the mean winter (January–April) sea surface temperature (SST) from HadISST2 in the time period 1970–2005. The two black lines are aligned with temperature isolines for 9° and 3°C, and the dashed black box shows the region called the eastern subpolar North Atlantic.

accordingly be beneficial if one could skillfully predict the state of the ocean a few years in advance. Indeed, several studies have shown the capability of predicting changes in the subpolar North Atlantic several years in advance (e.g., Robson et al. 2012; Yeager et al. 2012; Persechino et al. 2013; Msadek et al. 2014). On interannual-to-decadal time scales, a recent study shows that there is little consistency in the predictive skill of three dynamical prediction systems in the Nordic seas (Langehaug et al. 2017). This is in contradiction to an observation-based study, demonstrating that it is possible to predict changes in the sea surface temperature (SST) and salinity a few years in advance in the eastern Nordic seas (Årthun et al. 2017; Fig. 1). These observed changes or anomalies in temperature and salinity have been found to be predictable, as they travel from the Gulf Stream region in the North Atlantic Ocean toward the Arctic Ocean within a time frame of about 10 years (e.g., Holliday et al. 2008; Chepurin and Carton 2012; Årthun et al. 2017), and have a periodicity of about 14 years (Årthun et al. 2017). Moreover, the mechanism associated with these thermohaline anomalies, such as how they are formed and how they develop on their way north, has recently received increased attention (e.g., Årthun and Eldevik 2016; Asbjørnsen et al. 2019; Årthun et al. 2021).

In this study we differentiate between two types of mechanisms: 1) the mechanisms that generate decadal-scale SST variability in the North Atlantic, and 2) the mechanisms responsible for propagating decadal-scale SST anomalies along the Atlantic water pathway to the Arctic Ocean. In the study presented here, the Atlantic water pathway is defined as the route where the Atlantic Water dominates and circulates (Fig. 1), following the poleward extension of the Gulf Stream, via the North Atlantic Current and the Norwegian Atlantic Current. A number of studies have focused on understanding the first type, and several mechanisms have been suggested to play an important role in generating decadal-scale SST anomalies in the North Atlantic (e.g., Ortega et al. 2015; Marshall et al. 2001; Muir and Fedorov 2017; Årthun et al. 2021). Our study is however related to the second type of mechanism, focusing on the subsequent propagation of SST anomalies along the Atlantic water pathway on interannual-to-decadal time scales. This northward propagation is clearly seen in Hovmöller diagrams of SST (and salinity) anomalies, as alternating cold and warm anomalies, and with a time lag between the subpolar North Atlantic and the Fram Strait (at the entrance to the Arctic Ocean; Furevik 2001; Holliday et al. 2008; Yashayaev and Seidov 2015; Årthun et al. 2017). The time lag suggests that anomalies propagate slowly poleward, using several years to travel along the Atlantic water pathway. This spatiotemporal SST pattern is at the core of this study, where our aim is to assess how model simulations (climate predictions and historical runs) represent the observation-based spatiotemporal variability of SST anomalies along the Atlantic water pathway.

Several mechanisms have been proposed to explain the decadal-scale propagation and along-path modification of SST anomalies along the Atlantic water pathway. This includes advection along the major currents (e.g., Furevik 2000; Krahmann et al. 2001; Årthun et al. 2017), local air–sea interaction (e.g., Saravanan and McWilliams 1998), Rossby waves (e.g., Liu 1999), boundary waves (Marshall and Johnson 2013), and shear-dispersion effects (Broomé and Nilsson 2018). This study does not aim to assess the relative importance of these mechanisms. In the discussion, however, we further elaborate on some of the mechanisms or factors that can impact the propagation of SST anomalies along the Atlantic water pathway and discuss to what extent these mechanisms can shed light on model differences.

Langehaug et al. (2019) tested to what degree a hindcast ocean model (forced with realistic atmospheric datasets) is able to represent the poleward propagation of SST anomalies and upper-ocean salinity anomalies. They found that simulations at 1° and 1/4° horizontal resolutions were both able to show the repeated propagation of surface anomalies, with the higher-resolution simulation showing a better timing of the anomalies along the Atlantic water pathway compared to observation-based data. The spatiotemporal SST variability described above has nevertheless not yet been evaluated in dynamical prediction systems. Thus, our main objective in this study is to investigate the predictive skill of SST along the Atlantic water pathway in seven different dynamical prediction systems. To capture potential skill related to the poleward propagation of SST anomalies, we

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

focus on multiyear to decadal time scales. We hypothesize that predictive skill or lack thereof in the prediction systems along the Atlantic water pathway is related to models' ability to represent the poleward propagation of these SST anomalies. To test this hypothesis, we compare the climate models of the prediction systems against the observation-based benchmark of poleward propagating anomalies—a mechanism connecting the Atlantic Ocean and the Arctic Ocean—as articulated in Årthun et al. (2017). Furthermore, for each forecast year, the spatiotemporal SST variability is assessed in the different prediction systems. This study therefore demonstrates a different way of assessing forecast skill, in addition to the traditional way of calculating anomaly correlation coefficients, which focuses on how well the prediction systems are representing a specific spatiotemporal pattern for each forecast time.

The manuscript is organized as follows. In section 2, we describe the observation-based data, prediction systems, and methods used. In section 3, we present the main results, and finally we discuss and conclude our study in sections 4 and 5, respectively.

2. Observation-based data, model data, and methods

a. Observation-based data

For observation-based SST analyses both version 1.1 and version 2 of the Hadley Centre Sea Ice and SST dataset (HadISST1 and HadISST2) are used. These datasets provide monthly global SST on a 1° latitude–longitude grid from 1870 to the present. A detailed description of HadISST1 and its production process is given in Rayner et al. (2003). It was shown in Hirahara et al. (2016) that HadISST2 is suitable for representing large-scale SST variability.

Predictive skill can vary substantially depending on the verification time period used (e.g., Brune et al. 2018; Borchert et al. 2019). In addition, the observation-based dataset used for verification (e.g., HadISST1, HadISST2) can affect the predictive skill. For instance, comparing the predictive skill resulting from using HadISST1 and HadISST2, there is a large skill discrepancy in the center of the North Atlantic at ~50°N (with a difference of about 0.4 in correlation; not shown). The two observation-based datasets are only used here to capture uncertainty in the predictive skill arising from observations. For forecast verification purposes, on the other hand, we use only the latest product, HadISST2, as our reference for all models, focusing on the 1970–2005 period.

In this study we analyze late-winter and early-spring data (January–April), hereafter simply referred to as "winter" data. The reason for analyzing winter SST is that this is assumed to represent the temperature of the winter mixed layer, and thereby a larger portion than only the surface. In summertime, typically a shallow mixed layer is present due to solar heating. In wintertime, the larger vertical extent of the mixed layer implies that the SST is expected to better reflect the upper-ocean heat content (e.g., Chepurin and Carton 2012). However, the poleward propagation of anomalies is also seen when using annual data, and also when analyzing annual salinity, both at surface and subsurface (at 200 m) (Årthun et al. 2017; Langehaug et al. 2019).

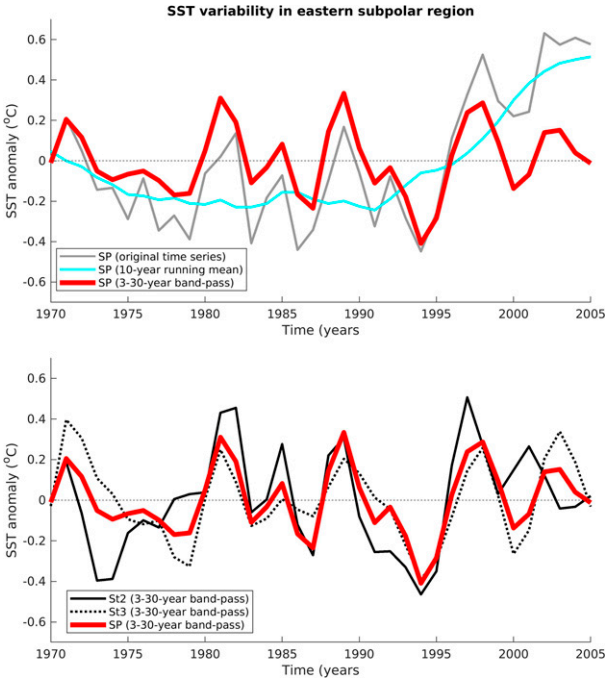

Fig. 2. (top) Winter SST for the eastern subpolar (SP) North Atlantic using HadISST2 original time series and filtered time series (3–30-yr band pass and 10-yr running mean). SST is averaged over the domain shown in Fig. 1 (dashed black box; 48°–61°N, 3°–25°W). (bottom) Winter SST from St2, St3, and the box domain. All time series are band-pass filtered.

b. Time filtering

Because we are interested in multiyear to decadal variability and not in long-term trends [e.g., global SST change and Atlantic multidecadal variability (AMV)], we apply a third-order 3–30-yr bandpass Butterworth filter to all time series, prior to the calculation of the correlation. This filter removes both the high-frequency variability and the lower-frequency signals, including long-term trends, highlighting decadal pulses as in Nigam et al. (2018). Figure 2 shows that the filter is phase-preserving. We have also applied the 10-yr running mean (using HadISST2 data from 1960 to 2010), which represents multidecadal variability in the time series (Fig. 2). The variability in this time series clearly differs from the variability in the bandpass filtered time series. The former (10-yr running mean) is not a focus of this study, but is only added here for comparison purposes. The observation-based SST in Fig. 2 is averaged over the box domain shown in Fig. 1 showing the eastern subpolar North Atlantic.

We demonstrate the robustness of the filter choice (3–30-yr bandpass Butterworth) by also applying a 5–40-yr bandpass filter (Figs. 4 and 6). This filter gives overall similar results, but with more smoothing of the peaks in the time series. The

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

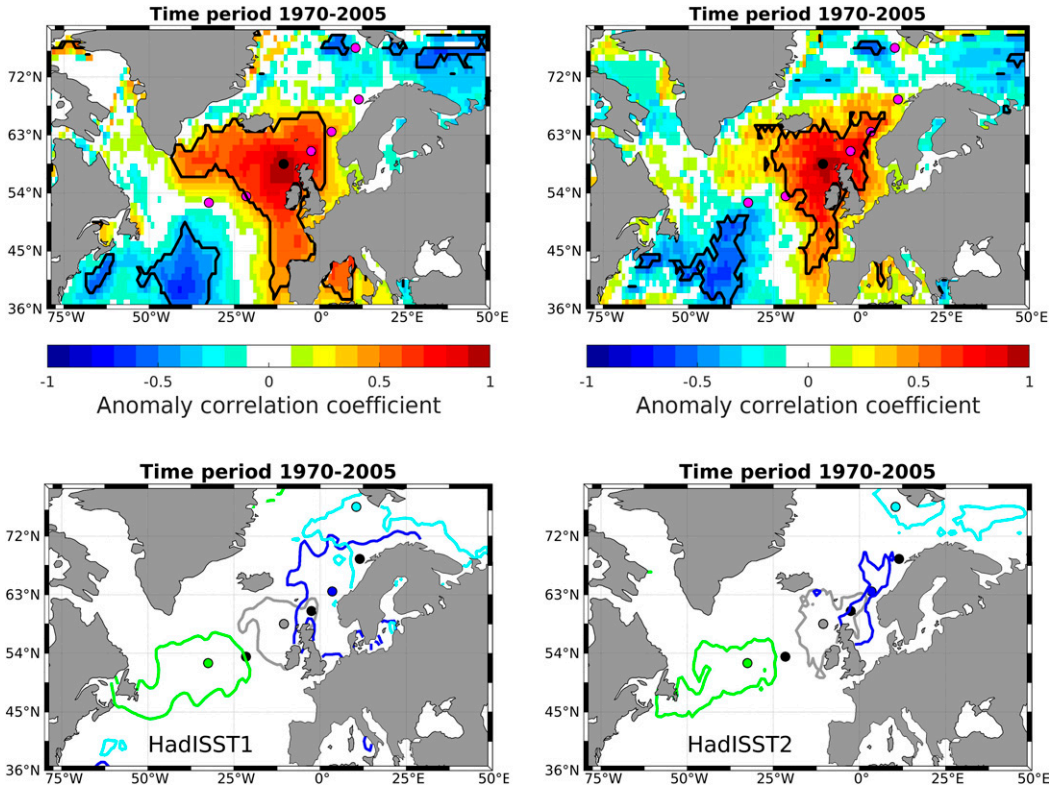

FIG. 3. (top) Estimating extent of observation-based SST anomalies by correlating the time series at station 3 (here using a single grid point) with the time series in all other grid points for two different data sets, (left) HadISST1 and (right) HadISST2. The black line shows significant correlation at the 90% level by the standard two-sided Student's *t* test and taking into account autocorrelation. (bottom) As at in the top row, but considering stations 1, 3, 5, and 7 and contour is drawn where the correlation equals 0.7. All time series are 3–30-yr band-pass filtered prior to correlation.

paper describing the observation-based poleward propagation of anomalies (Årthun et al. 2017), applied a 5-yr Butterworth low-pass filter, and additionally a 40-yr Butterworth high-pass filter when removing the AMV.

c. Poleward propagation of SST anomalies in observation-based data

Observation-based data have previously shown that there exists a lagged correlation for SST anomalies, and also for subsurface salinity, between the stations along the Atlantic water pathway (purple dots in Fig. 1; Årthun et al. 2017; Langehaug et al. 2019). We estimate the horizontal extent of the observation-based SST anomalies by calculating the simultaneous correlation between SST in each grid point with the SST at one of the stations along the Atlantic water pathway [for station 3 (St3) in Fig. 3, top panel]. In this way, we see the extent of the surrounding area that covaries with each station. HadISST1 shows that a region stretching from Iceland toward Spain has significant positive correlations with St3, and a similar region is also seen for HadISST2, but with its southern extent limited to the British Isles.

The same analysis is repeated for every second station (St1, St3, St5, and St7), and only locations with a correlation of 0.7 are shown (Fig. 3, bottom panels). We find high and significant correlations surrounding each station by up to hundreds of kilometers in HadISST1 (similar, but smaller regions for HadISST2). At the same time, there is little overlap in the extent of the SST anomalies for these four stations. The relatively large horizontal extent of the anomalies in Fig. 3 supports the decision of averaging the observational-based data and model data over a 5° × 5° box in the analysis with lagged correlations described below.

We calculate the lagged correlations (or cross-correlations) up to 10 years between the different stations and the northernmost one (St7 in the Fram Strait) for the time period 1970–2005 for both HadISST1 and HadISST2 (Fig. 4), with St7 always lagging the other ones. Due to the relatively short observational record (fewer samples increase the significance level), we show significant correlations at both the 90% and 95% level in Fig. 4. As mentioned previously, we have expanded each station to a 5° × 5° region instead of using one single grid point (Fig. 1, gray boxes), which means that several grid points have been averaged for each station. This analysis reveals a dipole pattern with positive (negative) correlations mostly at short (long) time lags. The positive correlations with St7 decrease upstream (i.e., when correlating with stations farther south) and also the time lag for the maximum correlations increases upstream. For instance, in HadISST2 with the

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

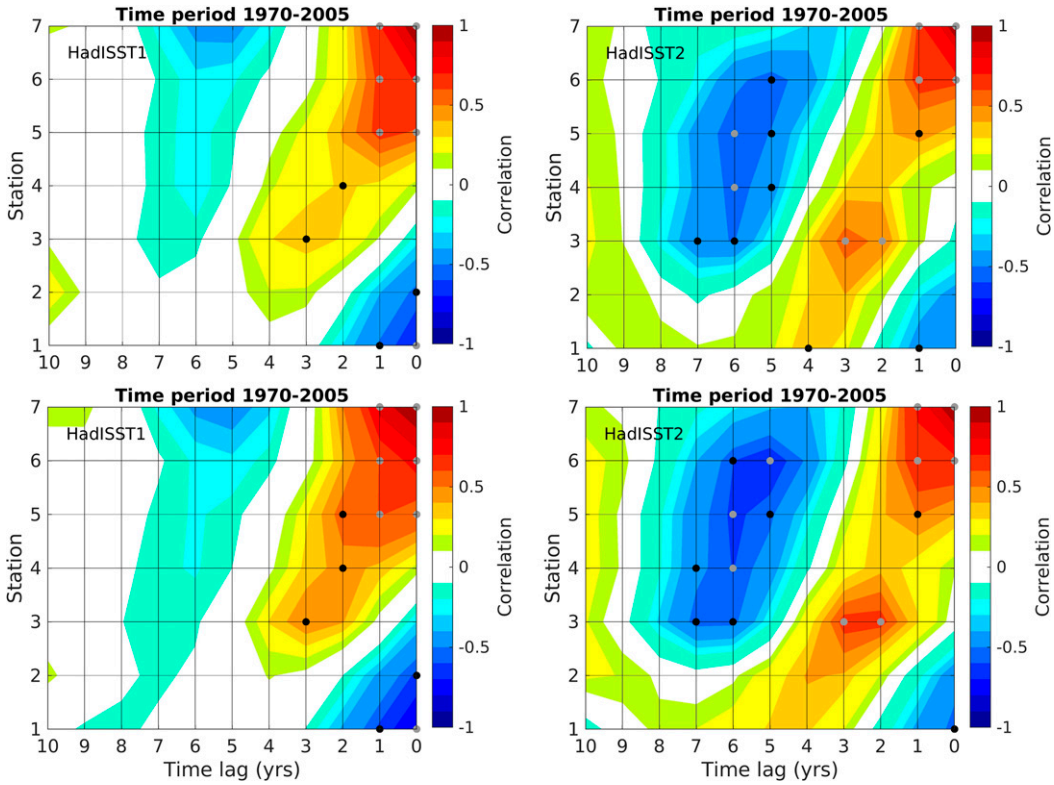

FIG. 4. Cross-correlation of observation-based winter SST between the northernmost station and all seven stations (Fig. 1). Time series are filtered by a (top) 3–30-yr and (bottom) 5–40-yr Butterworth band-pass filter. Significant correlations are marked by black (gray) circles, calculated at the 90% (95%) level by the standard two-sided Student's *t* test taking into account autocorrelation.

3–30-yr bandpass filter, a significant positive correlation exists between St7 and St3 with a time lag of 2–3 years (max correlation is 0.63), suggesting that a signal propagates from the entrance at the Greenland–Scotland Ridge toward the Fram Strait during that time. The significant negative correlations at longer time lags show that SST anomalies at St7 are opposite about 6–7 years earlier (correlation is −0.55 between St7 and St5 at a time lag of 6 years).

There are some differences in the cross-correlation pattern when comparing HadISST1 and HadISST2 (Fig. 4). In HadISST1, the correlation between St7 and the two stations in the subpolar region (St1 and St2) hints at a connection, but it is weak and not significant. In HadISST2, the correlation with the stations in the subpolar region is higher and suggests a link with anomalies in the edge of the subpolar region and the Nordic seas. On the other hand, the correlation between St7 and St4 (close to Shetland) is significant in HadISST1, but weak and not significant in HadISST2 (Fig. 4). Using a 5–40-yr bandpass filter instead of a 3–30-yr bandpass filter shows overall the same pattern but somewhat higher correlations.

d. Dynamical prediction systems

Dynamical prediction systems are fully coupled global climate models that are initialized with the observed state of the ocean through different strategies, to start from a realistic initial state (with respect to the observed phase and amplitude of internal climate variability) and then simulate future ocean circulation and climate. For instance, unusually warm subpolar North Atlantic Ocean conditions that are represented at the correct time could then circulate poleward due to the modeled ocean dynamics, potentially leading to consistently higher SST in the Fram Strait some years later.

The decadal hindcasts start at a specific date in each prediction system and the number of ensemble members varies, but all members are run for 10 years (Table 1). When computing the skill in the decadal prediction systems we use the same verification time window for all the forecast times (1970–2005). In this way, all forecast ranges are estimated consistently, and differences can be attributed to the differences in the systems (models, initialization methodology, forecast drift/shock, external forcing), without any contamination due to different verification windows. This approach implies that not all forecast times will be assessed for all start years (e.g., for start year 1961 we only use data for the 10-yr forecast time, and for start year 2005 we only use data for the first year forecast). For each start date, several ensemble members have been run, depending on the prediction system, and herein the main focus is on the respective ensemble means for each system. More details on each of the prediction systems are given in Table 1.

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

TABLE 1. Model and data assimilation specifications of the dynamical prediction systems used in this study. We verify hindcasts for the period 1970–2005 in the forecast range 1–10 years and focus on the winter season (January–April mean), i.e., we use hindcasts initialized from 1960 to 2005. In addition, hindcasts are verified at zero lead time, i.e., for the period 1969–2004 for the specific months given below (except the two systems initialized in January that are verified for the period 1970–2005). Individual members are analyzed, and the number of members (mem) are given below (a few members are not included in Fig. 12 due to a few years missing). We have also selected three members from the historical runs that are analyzed for longer time periods (as given in the table). Note that only one member is available from CMCC-CM2-SR5.

Prediction system	Ocean resolution (grid points, vertical levels)	Ocean resolution (nominal)	Ens. size (mem)	Years initialized	Full-field (F) or anomaly (A)	Date initialized	Initialization strategy	Zero lead time (month)	Historical runs (model/time period/mem)	Relevant references
NorCPM1 dcppA-hindcasts-i1	320 × 384, 51 + 2 layers (isopycnal)	1° × 1°	10	1960–2018	A	15 Oct	Initialized with EnKF using SST (HadISST2) and TS-profiles (EN4)	October, first month of hindcasts	NorESM (CMIP6)/ 1950–2014/3	Bethke et al. (2021)
CESM1-DPLE	320 × 384, 60 levels	1° × 1°	40	1954–2015	F	1 Nov	Initialized with CESM1.1 forced with historical atmospheric state and flux fields	November, data from forced CESM1.1	CESM-LE (CMIP5)/ 1920–2005/3	Yeager et al. (2018), Kay et al. (2015)
IPSL-CM5A-LR (DEC3)	182 × 149, 31 levels	2° × 2°	3	1961–2013	A	1 Jan	Surface nudging of SST (ERSST)	January–April, data from initialized simulation through nudging	IPSL-CM5, (CMIP5)/ 1850–2005/3	Mignot et al. (2016), Swingedouw et al. (2013)
EC-EARTH ano (v2.3) SMHI/DMI	362 × 292 (hindcasts interpolated to 320 × 160), 42 levels	1° × 1°	10	1960–2005	A	1 Nov	NEMOVAR-S4 + model climatology for ocean, initialized with ERA-40 + ERA-Interim for atmosphere and land	November, first month of hindcasts	EC-EARTH (CMIP5)/ 1850–2005/3	Hazeleger et al. (2013b)
EC-EARTH full (v2.3) BSC	362 × 292, 42 levels	1° × 1°	5	1960–2010	F	1 Nov	NEMOVAR-S4 for ocean, initialized with ERA-40 + ERA-Interim for atm and land	November, first month of hindcasts	EC-EARTH (CMIP5)/ 1850–2005/3	Hazeleger et al. (2013b)
MPI CMIP5 MPI-ESM-LR	256 × 220, 40 levels	1.5° × 1.5°	10	1961–2011	A	1 Jan	3D T&S from ORAs4 ocean reanalysis, full-field for atmosphere from ERA-40/ERA-Interim	January–April, data from assimilation experiment	MPI-ESM-LR, (CMIP5)/ 1850–2005/3	Giorgetta et al. (2013)
CMCC-CM2-SR5	362 × 292, 50 levels	1° × 1°	7	1960–2010	F	1 Nov	Ensemble of ocean-ice reanalyses (assimilating SST + TS profiles), land surface analyses, ERA-40/ERA-Interim for atmosphere	November, first month of hindcasts	CMCC-CM2-SR5 (CMIP6)/ 1960–2014/1	Bellucci et al. (not yet published)

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

Five of the models providing decadal hindcasts are from the EU Blue-Action project partners (CESM1-DPLE, IPSL-CM5A-LR, MPI-ESM-LR, NorCPM1, CMCC-CM2). The EC-EARTH (v2.3) hindcasts are available from two prediction systems, with the anomaly initialized system (referred to herein as EC-EARTH ano) performed by the Swedish Meteorological and Hydrological Institute (SMHI) and the full-field initialized system (herein called EC-EARTH full) performed by Barcelona Supercomputing Center, both within the EU project Seasonal-to-decadal climate Prediction for the improvement of European Climate Services (SPECS; IPSL-CM5A-LR was also part of this project). The prediction systems used herein represent versions that have been developed after CMIP5 and some of these have been submitted to CMIP6. We also use uninitialized historical runs from each of the fully coupled global climate models (Table 1).

e. Anomaly correlation coefficient

To calculate the anomaly correlation coefficient, we construct a time series from the hindcasts for each forecast time, remove the winter (January–April) climatological mean over the 1970–2005 period of this time series, and correlate it with the corresponding observation-based anomaly time series. To account for consistent model drift, the climatological mean that we subtract is dependent on the forecast time; for example, if the SST time series at forecast time = 8 years shows in general warmer conditions than at forecast time = 2 years, then the climatological mean value would be higher at forecast time = 8 years. The anomaly correlation coefficient is calculated for the ensemble mean. When calculating the correlation skill for averaged forecast times, we average the bandpass filtered time series for different forecast times before calculating the correlation.

When assessing skill along the Atlantic water pathway, the model SST is averaged over specific regions dependent on latitude and the model's own SST climatology. The exact same regions are used to average observation-based SST (this is possible as the HadISST2 data have been interpolated to each of the specific model grids). The regions vary to some extent from one model to another due to their different SST climatology.

The statistical significance level is tested by the standard two-sided Student's t test (e.g., O'Mahony 1986). The effective degrees of freedom are calculated after the filtering and we use the method by Chelton (1983), taking into account the autocorrelation of the time series. It is thus different for each forecast time and each averaging region (three regions along the Atlantic water pathway; Fig. 5). In Fig. 6 (left panel), for each model we only show the significance level for the *averaged* effective degrees of freedom N_{mean}; N_{mean} is averaged across the different forecast times and regions. For the multimodel mean (Fig. 6, right panel), we allow the effective degrees of freedom to vary with the three regions (but not with forecast time).

f. Comparing the prediction systems and climate models with the observed propagation

In this study we compare the lag-correlation pattern identified in observation-based data with its counterpart in each of the dynamical prediction systems. This means that we treat the data for each forecast time as a continuous dataset across hindcasts. In this way, we assess how the SST anomalies at each station covary with the northernmost station for different forecast ranges, which can indicate how the relation evolves with forecast time.

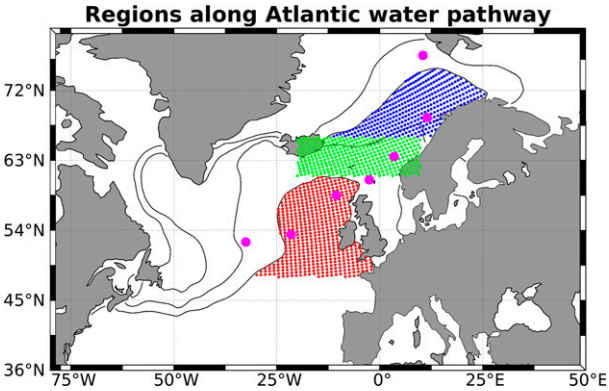

Fig. 5. SST skill is assessed in three different regions along the Atlantic water pathway: 1) the eastern subpolar region (red), 2) the inflow region (green), and 3) the Norwegian Sea (blue). These regions vary slightly from one model to another, as the regions are based on the models' mean SST: red corresponds to SST > 9°C and between 48° and 61°N; green is the same for all models, within a specific latitude and longitude interval; and blue corresponds to SST > 3°C and between 66° and 75°N. Note that IPSL-CM5 (MPI-ESM) has a cold (warm) bias, and thus, the SST limits are 6° and 0°C (10° and 4°C) for the subpolar and Norwegian Sea region, respectively. CMCC-CM2 has a cold bias in the subpolar region (SST limit is therefore 7°C), but not in the Norwegian Sea. See Fig. 1 in the supplemental material for the model-dependent regions.

In addition, we analyze the respective climate models and their uninitialized historical runs. Single members have been analyzed to test to which extent a physical propagation of thermohaline surface anomalies are realistically simulated in the climate models. In this way, our analysis is twofold: 1) we assess the skill in reproducing a realistic spatiotemporal SST pattern in the prediction systems, and 2) we investigate the existence of a physical propagation in the historical runs.

When we calculate lagged correlations for forecast time of 1 year, we use the following procedure. For each station, a time series is constructed using the first forecast year in each hindcast. Then, these time series from stations 1–7 are cross-correlated with the time series from St7. For longer forecast years the same procedure is repeated. Before these cross-correlations are calculated, the time series are bandpass filtered in the same way as for the observations. With this procedure, we are testing the skill of the prediction system to accurately predict the spatiotemporal pattern identified in observations. This analysis is also expanded to the individual members from the prediction systems, in order to assess the reliability of the spatiotemporal pattern across the ensemble.

As done with the observational-based data, we have expanded each station to a 5° × 5° region instead of a single grid point to produce more robust results. Each model has a

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

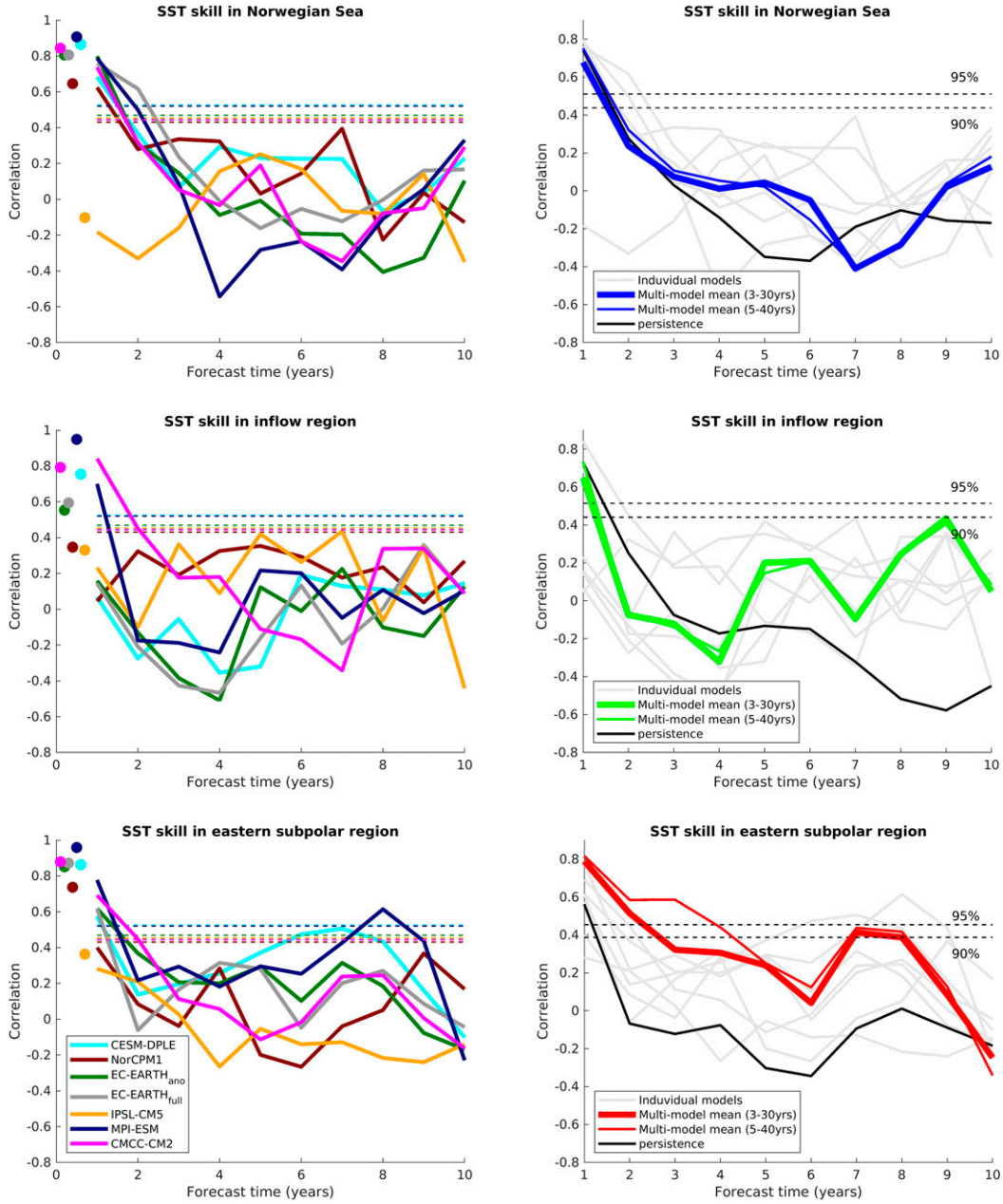

FIG. 6. Anomaly correlation coefficient of winter (January–April) SST for three different regions (see Fig. 5). (left) Correlation is between HadISST2 data and the ensemble mean of the individual prediction systems at different forecast times. At each forecast time, time series are 3–30-yr band-pass filtered prior to correlation. Correlation at zero lead time is shown by filled circles. (right) As in the left column at forecast time 1–10 years, but now also including the skill of the multimodel mean that has been correlated with HadISST2 data (thick line). The dashed lines show significant correlation at the 95% level by the standard two-sided Student's t test and taking into account autocorrelation (the 90% level is shown in addition for the multimodel mean). Thin colored lines represent the multimodel mean filtered with 5–40-yr band pass.

different resolution and models with higher resolution have more internal variability (e.g., Tang et al. 2019). Averaging the ensemble reduces this noise. Thus, when changing from single grid points to small regions, the results are overall the same for the CESM1-DPLE because it already has so many ensemble members.

In their observation-based study, Årthun et al. (2017) found evidence of predictability related to the poleward propagation of temperature anomalies, which they used in statistical prediction of Norwegian air temperature. It is important to note that the predicted signals estimated with dynamical prediction systems (i.e., those based on climate models, as in this study)

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

tend to be a lower bound of the true predictability that can be achieved in the real world. This is because these prediction systems are "imperfect" as they are affected by systematic model errors, observational uncertainties, and other limitations inherent to the initialization procedure. Large ensemble sizes (which are generally unaffordable for decadal climate prediction) are also warranted to achieve high predicted signals, as many members help improving their detectability over the climate noise. Another aspect potentially degrading the predicted signals in our analysis is the fact that the multimodel prediction ensemble mean used herein might not reflect a coherent physical evolution. This ensemble mean results from the averaging of members from different models, which might represent the propagation differently. We also note that for computing the skill or the spatiotemporal patterns for each lead time we combine predictions from different start dates, which, unlike the continuous experiments, might present discontinuities and/or inhomogeneities, also degrading the signals. All these limitations thus need to be kept in mind when interpreting our results from the decadal climate predictions.

3. Results

a. Skill in predicting winter SST anomalies along the Atlantic water pathway

We aim to capture skill in regions influenced by the advection of Atlantic-origin water. We therefore average winter SST within specific latitude bands and SST intervals along the Atlantic water pathway to the north, separating into three different regions (colored regions in Fig. 5). The SST intervals are based on the SST climatology (January–April seasonal mean) for each individual prediction system, and hence there are some minor regional differences across the prediction systems in the latitudinal bands (Fig. 1 in the online supplemental material). We define the three regions accordingly: 1) the eastern subpolar North Atlantic (hereafter only subpolar region), defined as SST $> 9°C$ and between 48° and 61°N; 2) the inflow region to the Nordic seas (hereafter simply called the inflow region), defined as within the regions of 61°–66°N, 20°W–10°E; and 3) the northern Norwegian Sea (hereafter simply Norwegian Sea), as SST $> 3°C$ and between 66° and 75°N. In this respect the inflow region is also covering large parts of the southern Norwegian Sea. A threshold of 3°C is chosen as the dividing line in the Nordic seas, as this represents the typical temperature of Atlantic Water in the eastern part of the Norwegian Sea (Eldevik et al. 2009). A threshold of 9°C is chosen as the dividing line in the southernmost region, as this captures the Atlantic Water in the eastern part of the subpolar North Atlantic.

The skill is calculated using the ensemble mean from the respective prediction systems and HadISST2 data. Time series have already been smoothed by bandpass filtering before calculation of skill, as described in section 2b. The results are synthesized in Fig. 6: In each row we show the results for one of the regions, and the left column shows all individual prediction systems. In the right column we show the skill of multimodel mean for each region.

In Fig. 6, we also show the SST skill at zero lead time (filled circles in left column). Information about how zero lead time is extracted from the prediction systems is given in Table 1. This varies to some extent from one system to another, but the zero-lead-time data in general reflect the initial state in predictions or in the data used to initialize the systems. As most of the prediction systems are initialized during October/November, we have compared the zero-lead-time data with HadISST2 data averaged over the months October, November, and December. Overall, we find high correlation at zero lead time, in particular for the eastern subpolar region and the Norwegian Sea. There is a larger spread in the correlations for the inflow region compared to the other two regions.

At short forecast times (1–2 years lead), we find skills similar to or above the one arising from persistence in the subpolar region and the Norwegian Sea. The persistence is here defined as the autocorrelation of HadISST2 data at time lags of 1–10 years (further discussed in the following subsection). Note that 1-yr lead means the first winter after the predictions are initialized. At longer forecast times, the skill is lower and not significant (except MPI-ESM in the subpolar region). In the first three forecast years in the Norwegian Sea, the individual predictions show a similar decrease in correlation (except IPSL-CM5), but at longer forecast times there is a large spread in the correlation from the different systems. These results are consistent with a previous study in the Nordic seas and Barents Sea; CMIP5 climate predictions show skill in predicting SST anomalies on the averaged forecast time 1–3 years (Langehaug et al. 2017, using linear detrending before assessing skill).

The skills of the prediction systems differ to a large extent for all forecast times in the inflow region, indicating that this region is difficult to predict. The inflow region covers partly the Greenland–Scotland Ridge, which has a complex bathymetry and is where the Atlantic Water interacts with the dense overflow water from the Nordic seas (e.g., Eldevik and Nilsen 2013). Only MPI-ESM and CMCC-CM2 have initial significant skill in the inflow region.

One prediction system shows re-emerging skill at longer forecast times; MPI-ESM in the subpolar region at forecast time of 8 years has significant skill. The other prediction systems also show an increase in skill at longer forecast times at a level mostly above the persistence, but not reaching significant values. The multimodel mean is well above the persistence, reaching significant values in the subpolar region and the inflow region (at the 90% level). Previous results have shown that the skill in the subpolar North Atlantic is not from external forcing (e.g., Matei et al. 2012; Msadek et al. 2014), which indicates that ocean dynamics contributes to the rise in skill. This underlines the potential role of ocean circulation in bringing predictability to the subpolar region and toward the Nordic seas.

To better understand the predictive skill in the different prediction systems, we next examine, in each of the prediction systems, the periodicity of the SST time series in the eastern

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

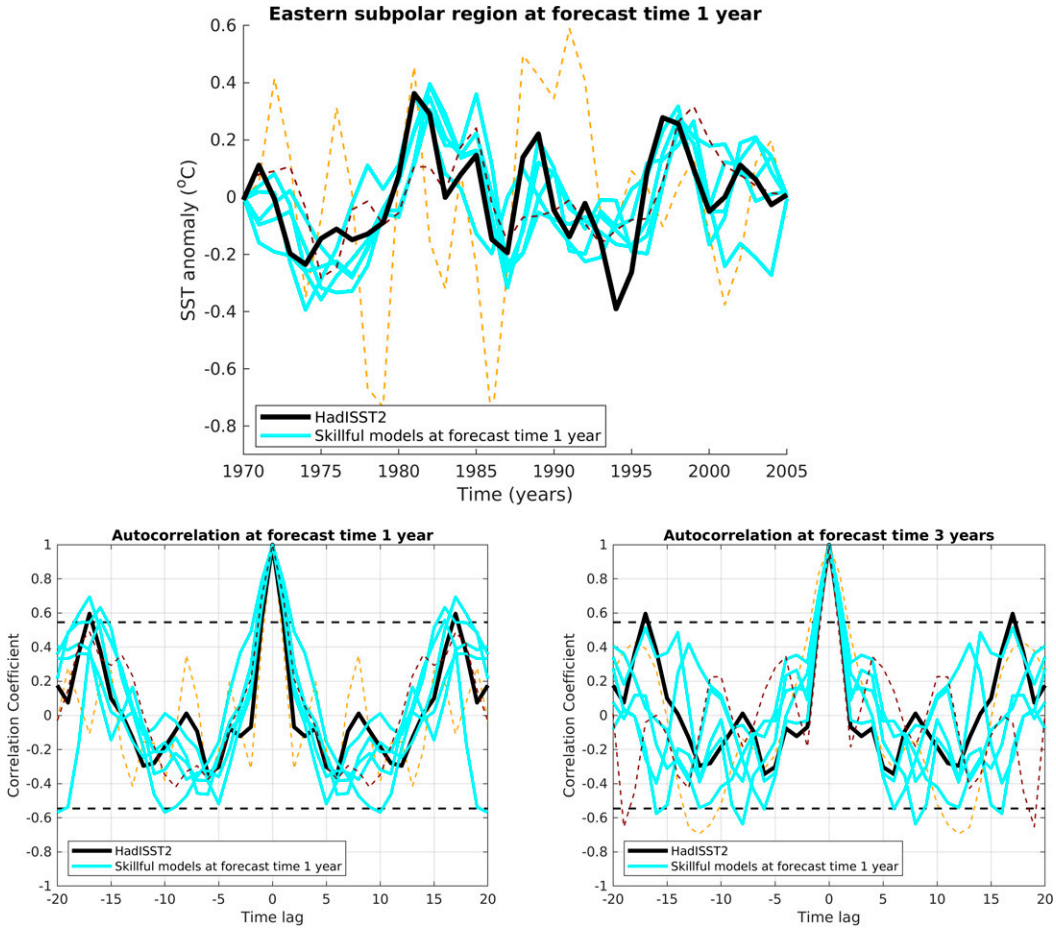

FIG. 7. (top) Time evolution of winter SST for the eastern subpolar region using HadISST2 (black line) and the ensemble mean of the individual prediction systems having significant skill at forecast time of 1 year (blue lines; see Fig. 6). The other prediction systems have not significant skill at lead year 1 (dashed lines; NorCPM1 and IPSL-CM5). All time series are 3–30-yr band-pass filtered. (bottom) The autocorrelation of the time series for 1970–2005 at forecast time of (left) 1 and (right) 3 years. Dashed horizontal lines indicate significance level.

subpolar region and poleward propagation of SST anomalies along the Atlantic water pathway.

b. Assessing periodicity of SST anomalies using autocorrelation

By using autocorrelation, we can identify whether a time series has a dominant period. It has earlier been shown that observation-based SST in the Norwegian Sea has a dominant spectral period at 14 years, which rises above the confidence interval of the red noise spectrum (Årthun et al. 2017). This dominant period has been related to poleward propagation of ocean anomalies originating in the North Atlantic Ocean. As such, a similar dominant period is expected to be found in the subpolar region studied herein.

Figure 7 is an attempt to look at the periodicity or frequency of SST anomalies in the subpolar region (i.e., how often do SST anomalies occur in this region?). The time series are too short (1970–2005) to adequately address frequency, but still we find a suggestive difference between forecast time of 1 year and forecast time of 3 years. We first show the SST time series from the subpolar region (Fig. 7, top panel). At a forecast time of 1 year, the prediction systems follow fairly well the observation-based SST in the subpolar region. The autocorrelation of these time series is shown in Fig. 7 (bottom-left panel), and the autocorrelation from the prediction systems shows behavior similar to the autocorrelation of HadISST2 data. However, at forecast time of 3 years, we find a larger spread in the autocorrelation from the prediction systems and they fail to reproduce the observation-based autocorrelation (Fig. 7, bottom-right panel), indicating a degradation of predictive skill from forecast time of 1 year to forecast time of 3 years. This is the case for most of the prediction systems. Although the prediction systems show a larger spread, we note however that they do in general show an increase in autocorrelation with increasing time lag (from a time lag of 5–10 years to a time lag of about 20 years).

In Fig. 7, we find one dominant peak in the autocorrelation of observation-based SST in the subpolar region, occurring at

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

long time lags (about 17 years), with a significant correlation (black curve in bottom panels). At forecast time of 1 year, this peak is fairly well reproduced by the prediction systems. However, on longer forecast times, this peak is not reproduced to the same extent. We already see this at forecast time of 2 years, but to a larger degree at forecast time of 3 years. These results illustrate how the predictions stop reproducing the correct range of spectral properties as assessed by the autocorrelation already 3 years into the prediction.

c. *Assessing the spatiotemporal SST pattern in the dynamical prediction systems*

In this section, we apply the method described in section 2, using cross-correlation between the northernmost station and all other stations. In HadISST2, we have noticed that the SST time series at St7 has a relatively high and positive correlation with the stations upstream and with an increasing time lag (Fig. 8, top panel, repeating the top-right panel of Fig. 4).

In general, we find that the representation of the cross-correlation differs largely among the prediction systems and that the cross-correlation pattern varies with forecast time for each prediction system (Figs. 2–8 in the online supplemental material). It is, thus, a challenge for the prediction systems to accurately reproduce the observed spatiotemporal SST pattern. As an example, we show the cross-correlation for each prediction system for the forecast range of 2–4 years (Fig. 8). We have chosen to show this forecast range because the prediction systems start to diverge from each other at this range, in terms of skill, and because of the dropping off of skill starting at this forecast range (Fig. 6). Each of the prediction systems represents differently the lagged relationships between the different stations, which tend to differ from the ones inferred from the observations. The major agreement between the prediction systems and observations concerns the stations located in the Nordic seas, while the major discrepancies are seen for the subpolar region, which is farthest from St7. We note that also the two observation-based datasets differ on the connection to the subpolar region (Fig. 4).

The prediction system that shows the most accurate representation of some qualitative aspects of the observed lagged correlation is CESM1-DPLE, as it is the only one showing a significant link between stations 1–6 and station 7, with the time lag for the maximum correlation increasing upstream. This prediction system has the largest number of ensemble members (Table 1). There are some differences in the specific time lags, with a longer time lag from St1 to St7 in this prediction system (positive correlation between St1 and St7 at a time lag of 9–10 years). We note that these correlations do not represent a physical propagation of anomalies, as in observation-based data or in historical runs where time series are continuous. The correlations from the prediction systems are a result of separate hindcasts put together for a given forecast time.

Likewise, the lagged correlations from NorCPM1 shows a short time lag compared to observation-based data. NorCPM1 shows positive correlations between St7 and all other stations, but mostly on shorter time lags than for HadISST2. These differences among the prediction systems with respect to the time lags could be related to different surface velocities in the respective climate models, as will be discussed later in section 4. Of the remaining prediction systems, the IPSL-CM5 and EC-EARTH ano models have positive significant correlations mainly in the Nordic seas (the latter extend to St3 just south of the Greenland–Scotland Ridge), while the other three prediction systems have positive significant correlations only within a limited region in the Nordic seas.

Furthermore, we investigate the spatiotemporal SST pattern at a subsurface depth of about 200 m, the typical depth of the core of the Atlantic water, in three of the prediction systems (Fig. 9 in the supplemental material). Intermodel differences discussed above for SST do exist at this depth with mostly similar characteristics to the surface. An exception occurs in IPSL-CM5, which shows a somewhat different cross-correlation pattern at subsurface compared to the surface, with shorter time lags for the maximum correlations. Both CESM1-DPLE and NorCPM1 show similar results at surface and subsurface. The similarity in the spatiotemporal SST pattern at both surface and subsurface for a given prediction system illustrates that the SST anomalies along the Atlantic water pathway are not only a surface signal, but also reflect what happens deeper in the water column (within the Atlantic water layer).

d. *Is the misrepresentation due to poor initialization or incorrect model physics?*

We have shown that it is a challenge for the prediction systems to realistically represent the cross-correlation based on HadISST2. The difference with the observation-based pattern could both be due to poor initialization and/or poor model physics. To better understand the causes for the misrepresentation we have assessed the cross-correlations at zero forecast time. In this way, we can test whether the misrepresentation is due to a poor synchronization of the models with observations at the time of initialization. Particularly poor agreement at lead time zero would indicate a poor initialization that likely contributes to the misrepresentation. A good agreement at lead time zero, however, does not necessarily imply that the initialization strategy is optimal for predictive purposes once the models run free, as subsequent initialization shocks and/or imbalances could quickly degrade the skill (e.g., Bilbao et al. 2021). Information on how model data are extracted at zero forecast time is given in Table 1. We find that the cross-correlation at zero lead time for the different prediction systems is fairly similar to the cross-correlation based on HadISST2 (Fig. 9). We note that although IPSL-CM5 has low SST skill at zero lead time, the spatiotemporal SST pattern is well reproduced at zero lead time. The latter tells that there is a link between the variability at the different stations. Overall, the results in Fig. 9 show that there is good agreement between observations and prediction systems at zero lead time in terms of their SST pattern.

To investigate the potential role of incorrect model physics in the misrepresentation of the spatiotemporal SST pattern,

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

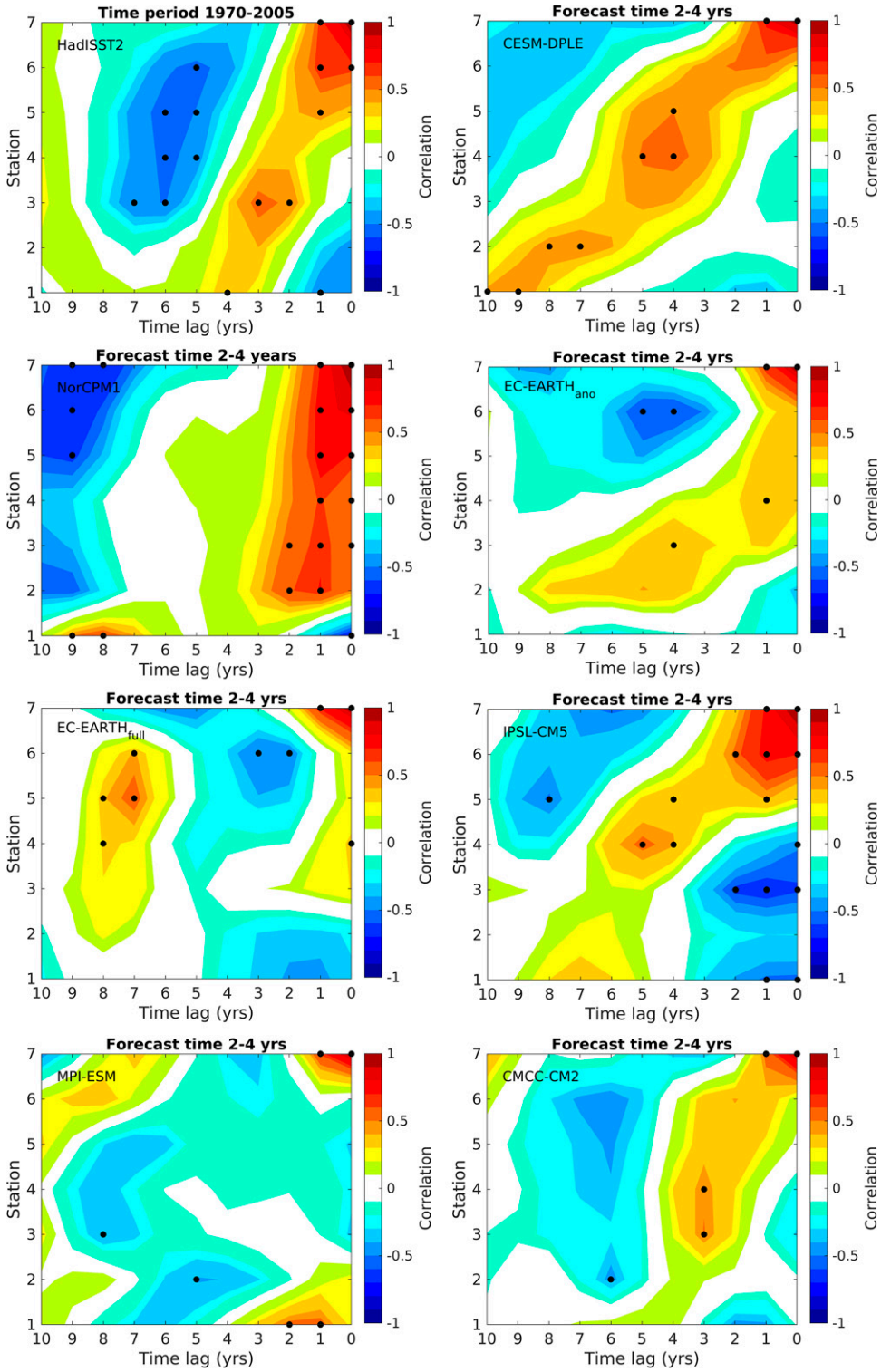

FIG. 8. Cross-correlation of winter SST between the northernmost station and all seven stations (shown in Fig. 1) at forecast time 2–4 years for the period 1970–2005. (top left) HadISST2 is shown for comparison with (remaining panels) the different prediction systems (names are given in upper-left corner of each panel). Significant correlations are marked by black circles, calculated at the 90% level by the standard two-sided Student's t test (taking into account autocorrelation). Time series are filtered by a 3–30-yr band-pass filter prior to correlation.

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

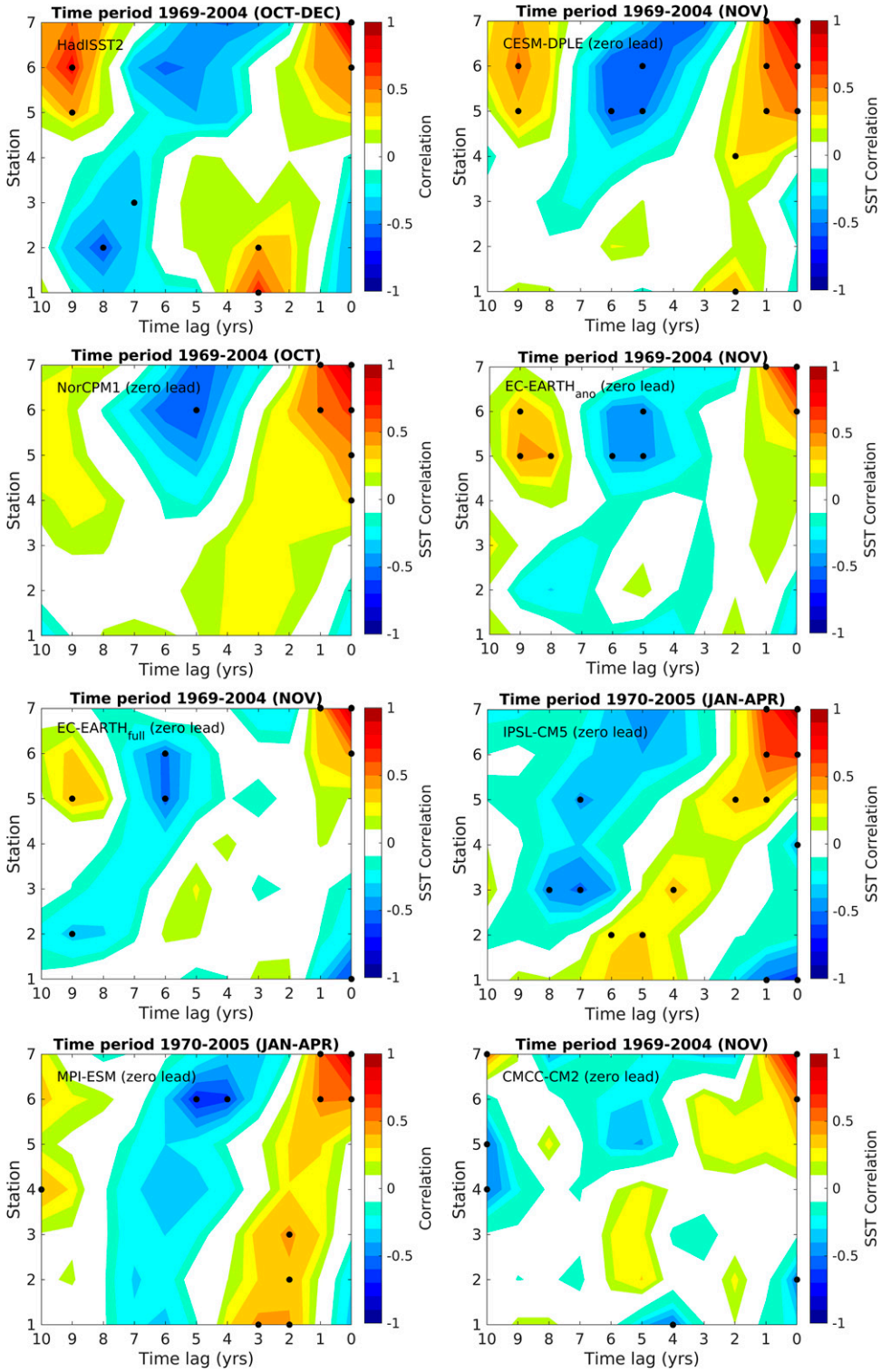

FIG. 9. As in Fig. 8, but using hindcasts at zero lead time and thus showing the period 1969–2004 for most prediction systems (Table 1). (top left) Note that HadISST2 cross-correlation uses the October–December mean for 1969–2004.

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

we have additionally computed the pattern in a set of historical (uninitialized) runs from the different climate models. These runs have continuous time series, which allows us to test to what extent the models realistically represent a physical propagation, without interferences from the drift dynamics present in the predictions. For each model we have three historical members, with the exception of CMCC, which only has one member available (Table 1). We have analyzed the pattern for each historical simulation individually, where we are only showing one of the three members for illustrative purposes (Fig. 10). Note that Fig. 10 is based on a single member, while Fig. 8 is based on a prediction ensemble. In a single historical member the effect of the propagation might be partly masked by atmospheric noise, whose effects can, at least partly, average out in the prediction ensemble. To quantify how well each member from each model and experiment represents this pattern, we calculate the pattern correlation between the observation-based and the modeled spatiotemporal SST patterns. This is done for the different individual historical members, as well as for the ensemble mean of the prediction systems at different forecast times (Fig. 11).

Initially (at zero forecast time), the prediction systems have relatively high skill in representing the spatiotemporal SST pattern (mostly above 0.7), but at higher forecast times the skill generally decreases, with differences across models. The skill of the historical members in representing the spatiotemporal SST pattern is also shown in Fig. 11 (each horizontal line represents one member). There is a large spread in the horizontal lines, showing that the cross-correlation is sensitive to which member is chosen. In other words, internal variability can notably affect the cross-correlation. To assess the skill of the historical members more robustly in reproducing the spatiotemporal SST pattern, we assess longer time periods for all members (Fig. 12, red curve). For each historical member, we produce the spatiotemporal pattern several times by using a running window with a length of 36 years (same time length as for HadISST2 data; 1970–2005). The length of the historical runs differs among the models (see caption of Fig. 12), and thus the number of repeating spatiotemporal SST patterns differs from one model to another. The red curve presents all correlations based on the historical members. In addition, we show the distribution when only looking at the period 1970–2005 (Fig. 12, red dashed curve). This is spikier as it contains only 16 correlations (compared to the red solid line that contains 1352 correlations when using the longer time periods). The red dashed line aligns well with the thick red line, suggesting that the distribution is not sensitive to the window considered. Both distributions show a large spread in the correlations and most of them are lower than the observation-based distribution (HadISST2, black curve). The observation-based distribution describes the sensitivity of the observed pattern to the temporal window, and thus to internal variability noise. To test the significance, we have built a set of synthetic time series with the same spectral properties of the observations to produce a random distribution of spatiotemporal SST patterns. For this, first, we produce 500 random white noise time series for each station (i.e., 7×500 time series with a temporal length of 36 years). Second, we use the first-order autoregressive coefficient of the observed HadISST2 time series to generate 7×500 red noise synthetic time series, which are then used to create 500 different spatiotemporal SST patterns. Finally, we compute the pattern correlation 500 times between the observation-based SST pattern and the synthetic SST patterns. In Fig. 12, we use the 95th percentile of the synthetic distribution to define our significance threshold (vertical blue dashed line).

We now look into the representation of the spatiotemporal SST pattern by the individual members of all the prediction systems, to better understand the ensemble spread. In this way, we can learn something about the reliability of the predictions, such as how the ensemble spread (i.e., in terms of how well the spatiotemporal observed SST pattern is represented) evolves with forecast time. In Fig. 12, we show the skill of individual members to represent the observed spatiotemporal SST pattern (gray curves). The dark gray curve represents all forecast times (1–10 years), whereas the light gray curve only represents forecast time = 1 year. We find that the average correlations for all forecast times are generally low (compared to HadISS2; black curve), with the distribution for forecast time = 1 year showing slightly higher values, which could suggest a small growth of uncertainty at the longer lead times. More interestingly, this analysis shows that already at short forecast times (1 year), the prediction systems show a relatively large ensemble spread, very similar to the one observed for the historical simulations when sampled over different temporal windows.

Overall, the results in Figs. 11 and 12 suggest that the misrepresentation of the SST pattern at longer forecast times is not due to too weak climate synchronization at initialization time. It could rather arise from initialization shocks/adjustments as well as from the development of biases due to incorrect model physics in representing the propagation of SST anomalies from the subpolar North Atlantic to the Fram Strait.

Although most members have low correlations, there are several members in the high end of the distribution with significant and relatively high correlations. In a follow-up study, it would be interesting to investigate these members more closely to determine which factors explain their improved representation of the SST pattern, and to assess whether if subsampled together they yield higher predictive skill in the North Atlantic and Arctic regions. Other studies have shown that a careful selection of the ensemble members based on their realism according to some observed metrics or properties can substantially enhance skill (e.g., Smith et al. 2020; Acosta Navarro et al. 2020).

4. Discussion

a. Skill and limits of current dynamical climate predictions in the northern North Atlantic–Arctic region

Dynamical climate prediction systems show promising results, demonstrating the capability of predicting changes

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

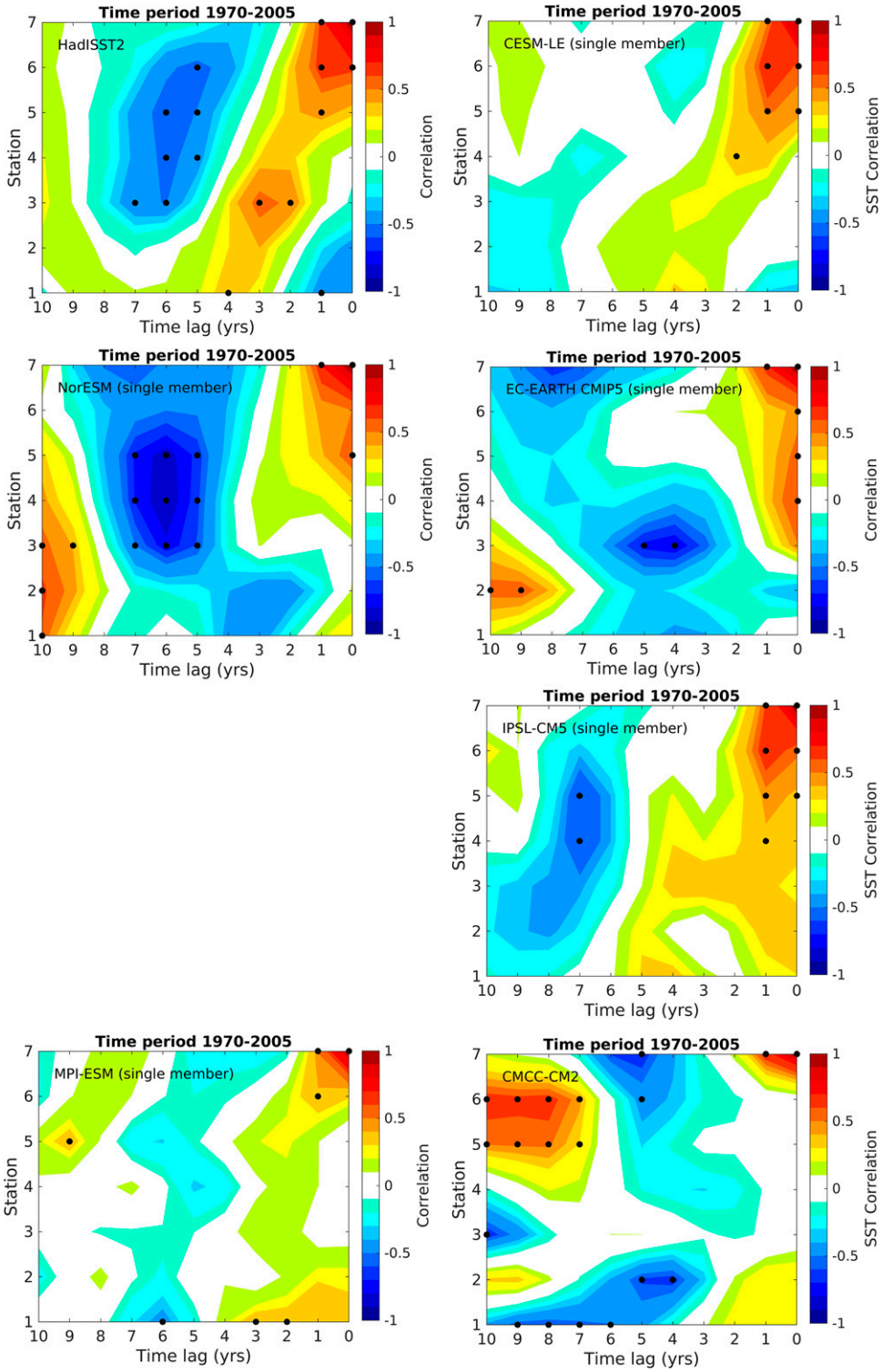

FIG. 10. As in Fig. 8, but using one historical member from the climate models (Table 1). The historical member with highest pattern correlation with HadISST2 is shown. EC-EARTH CMIP5 is used in both EC-EARTH ano and EC-EARTH full. Note that only one member is available from CMCC-CM2.

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

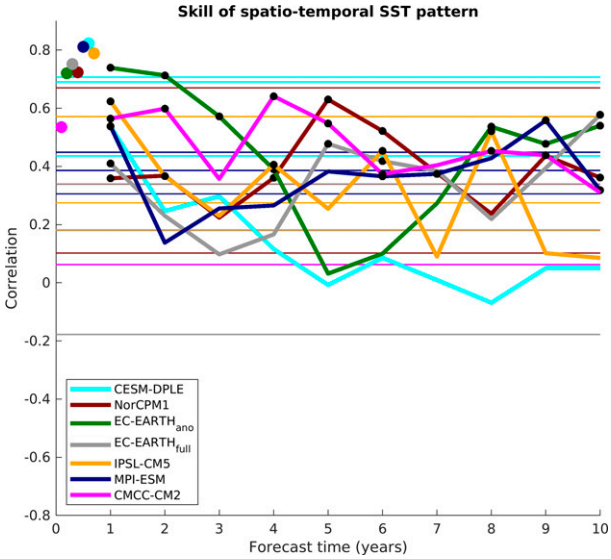

FIG. 11. Correlation between simulated and observation-based SST pattern (HadISST2). Simulated SST patterns are based on three different data sets: 1) historical runs (thin horizontal lines), 2) hindcasts at different forecast times (solid curves), and 3) hindcasts at zero lead time (filled circles). Significant correlations at the 90% level (black circles) by the standard two-sided Student's t test and taking into account autocorrelation. All correlations at zero lead time are significant. Note that HadISST2 cross-correlation uses the October–December mean (1969–2004) for most correlations at zero lead time, and otherwise the January–April mean (1970–2005). See Table 1 for more details.

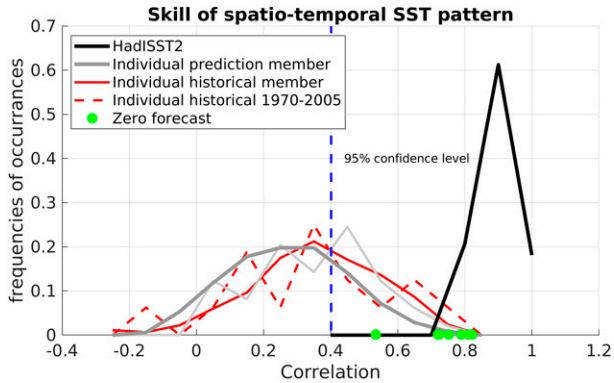

FIG. 12. Skill in representing the spatio-temporal SST pattern for individual members from all prediction systems for different forecast times (year 1 in light gray and all forecast times in dark gray). For each forecast time from each system, SST time series are constructed by choosing the same member for each time step. Skill of individual members from the historical runs from all models (red curve) is based on slightly different time periods with a running window of 36 years over these periods: MPI-ESM (1850–2005; 363 correlations), EC-EARTH (1850–2005; 363), IPSL-CM5 (1850–2005; 363), CESM-LE (1920–2005; 153), NorESM (1950–2014; 90), and CMCC-CM2 (1960–2014; 20). Skill for the historical runs only for the period 1970–2005 is shown as a red dashed line. Skill for the ensemble-mean predictions at zero forecast time is shown as green circles (shown as colored circles in Fig. 11). All model data are correlated with the ensemble mean of HadISST2 for 1970–2005. Skill of 10 individual members from HadISST2 over slightly different time periods (running window of 36 years over 1960–2010) is also shown (black curve). The significance threshold line, i.e., 95th percentile of the distribution when using white noise time series, is the vertical line.

in ocean surface temperature several years in advance in the subpolar North Atlantic (e.g., Robson et al. 2012; Yeager et al. 2012; Hazeleger et al. 2013a; Msadek et al. 2014). The source of predictive skill in this region is related to the northward advection of warm (and saline) subtropical water by ocean circulation (e.g., Matei et al. 2012; Yeager and Robson 2017).

Farther north, in the Norwegian Sea, dynamical climate predictions show capability of predicting ocean surface temperature only 1–2 years in advance (Langehaug et al. 2017, using CMIP5 models). This is also the case in the newer-generation climate predictions used in this study. For longer lead times (i.e., when predicting several years ahead), our results indicate that the skill in the eastern subpolar region and the Norwegian Sea is more limited and weaker than in the central part of the subpolar North Atlantic (e.g., Yeager 2020). We have found that only a few prediction systems show some skill when predicting ocean surface temperature several years ahead, and we have shown that the results are highly model dependent with large differences in the timing, region, and skill of anomaly propagation. As a result, the systems seem to be unable to capitalize on the potential predictability that has been demonstrated by observational data (Årthun et al. 2017).

We note that we have focused on interannual-to-decadal variability in the models. Other studies, focusing on more long-term changes, have found higher skill related to the poleward heat transport toward the Arctic (e.g., Yeager et al. 2015). Recent results with seasonal climate predictions in the Nordic seas also show promising results (Wang et al. 2019).

What about skill elsewhere in the Nordic seas and the Arctic Ocean? Johnson et al. (2018) demonstrate potential for decadal predictability of freshwater content in the Arctic Ocean related to long-term changes in sea level pressure. On the other hand, predictability related to the export of sea ice and liquid freshwater from the Arctic Ocean is limited (Schmith et al. 2018). Other studies using dynamical prediction systems (Germe et al. 2014; Kimmritz et al. 2019; Dai et al. 2020) show some skill in predicting sea ice extent in the Nordic seas (also for the pan-Arctic region) up to one year ahead. Higher predictability is shown for winter sea ice in the Barents Sea, as it is highly influenced by Atlantic water inflow (e.g., Årthun et al. 2017).

b. Process-oriented approach to understand why there is skill or lack thereof

We have found that climate models have difficulties in reproducing a realistic spatiotemporal SST pattern. This suggests that the mechanism that is identified to be a source of predictability along the Atlantic water pathway is not properly represented in the models. This feature seems to be translated into the prediction systems, which show low accuracy in predicting the spatiotemporal SST pattern, in particular at

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

longer forecast times. A common problem in most of the models and prediction systems is the linkage between anomalies in the eastern subpolar region and the Nordic seas. Several models and systems show lagged correlations between the Fram Strait and upstream until the Greenland–Scotland Ridge. However, south of the ridge, the lagged correlations are poor or do not exist. It thus appears difficult for the models to circulate ocean anomalies from the eastern subpolar region and across the Greenland–Scotland Ridge. We note that also the two observation-based datasets differ on the connection to the subpolar region, highlighting a region where the propagation seems more uncertain.

A natural follow-up question then is this: Why is it difficult for the models to realistically represent the spatiotemporal SST variability along the Atlantic water pathway? Although advection is proposed to be a key mechanism in poleward propagation of SST anomalies (Krahmann et al. 2001; Årthun et al. 2017)—supported by the propagation speed of SST anomalies matching that of radioactive tracers (Karcher et al. 2004; Gao et al. 2005)—there are several factors or mechanisms that can influence or modify the propagation. Such possible contributions could come through ocean feedback mechanisms (e.g., Huck et al. 1999; te Raa and Dijkstra 2002), exchanges between the subtropical and the subpolar North Atlantic (Lozier 2010), interactions between Atlantic Water and Arctic Water in the Norwegian Sea (Mork et al. 2019), or the non-Doppler effect related to Rossby waves (e.g., Liu 1999). Westward propagating Rossby waves are known to be important features in propagating information across the wide North Atlantic basin (Rossby et al. 1939; Killworth et al. 1997). These large-scale planetary waves, inducing anomalous currents, can lead to adjustments of the subpolar gyre circulation and Atlantic meridional overturning circulation, and thereby generate SST anomalies in the North Atlantic (e.g., Muir and Fedorov 2017; Årthun et al. 2021). However, the impact of Rossby waves in our study region, the Atlantic water pathway, appears to be relevant only on interannual time scales, contributing to the variable volume transport of the Norwegian Atlantic Current (Orvik and Skagseth 2003). A detailed investigation into the dominant mechanisms of propagation in each model is beyond the scope of this study. In the following, we have therefore chosen to concentrate on two aspects that we believe can provide first answers to our question above, regarding the representation of the poleward propagation of decadal-scale SST anomalies in climate models.

First, the propagation of the SST anomalies is dependent on the ocean currents carrying the signals downstream. Global climate models have long-standing issues in their representations of the Gulf Stream and its subsequent path as the North Atlantic Current (e.g., Langehaug et al. 2012). Furthermore, the Greenland–Scotland Ridge, separating the North Atlantic basin and the Nordic seas, is relatively shallow compared to the surrounding basins and the topography and circulation pattern around the ridge is complex. The Atlantic inflow across the ridge takes place through narrow channels, and the flow is characterized by high mesoscale activity, leading to high transport variability (Zhao et al. 2018). Higher-resolution climate models and prediction systems might

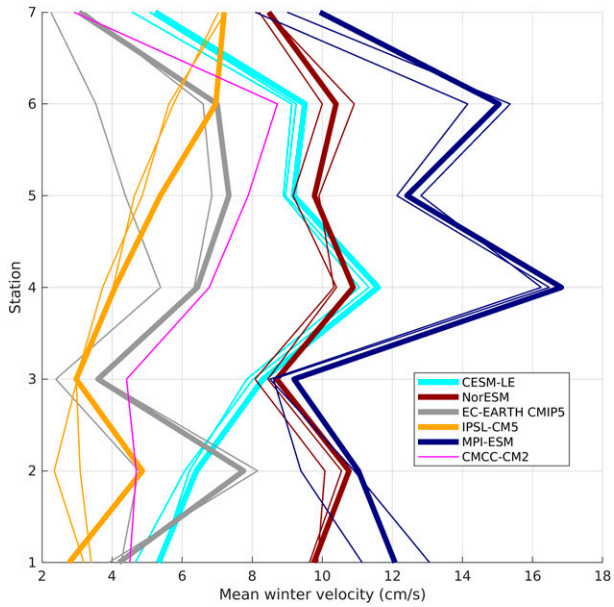

FIG. 13. Winter mean surface speed along the Atlantic water pathway (same stations as in Fig. 1) for the historical runs over the time period 1970–2005. Each curve represent one member from the historical run, and the thick line is the same member as displayed in Fig. 10.

therefore help to better represent processes related to Atlantic Water close to the Greenland–Scotland Ridge (e.g., Guo et al. 2016), and thus the circulation of anomalies along the Atlantic water pathway.

To assess intermodel differences in surface currents Fig. 13 shows the mean surface velocity along the Atlantic water pathway in the different models (the surface speed is averaged over the same 5° × 5° boxes as when analyzing SST anomalies). The models show large differences in surface velocities. However, a robust relationship between the simulated surface velocities and the time lags in the simulated spatio-temporal SST pattern is not found. One reason could be that there are additional factors that impact the propagation along the Atlantic water pathway, as described above. In addition, the models have difficulties in representing the pattern, and thus it is difficult to identify the time lags. Further investigations, for instance using a subset of the individual prediction members, would be very interesting but might be complex and require a dedicated study, which is therefore beyond the scope of this study. We here underline that there are large intermodel differences, and that the circulation of Atlantic Water appears challenging for the models to represent.

We have discussed the importance of having a realistic representation of the surface circulation. To further elucidate the picture, we here also briefly discuss the travel speed (or propagation speed) of temperature and salinity anomalies in the northern North Atlantic (e.g., Sundby and Drinkwater 2007; Årthun et al. 2017; Broomé and Nilsson 2018). The propagation speed is found to be much smaller than the actual current speed (e.g., the Norwegian Atlantic Current), where the

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

average propagation speed of SST anomalies along the Atlantic water pathway from the subpolar North Atlantic to the Fram Strait is found to be about 2 cm s^{-1} (Årthun et al. 2017). It has been shown that this difference in propagation speed and speed of the main current is a result of shear dispersion (Broomé and Nilsson 2018). According to their calculations that take eddy mixing into account, a current with a core speed of about 15 cm s^{-1} would result in a tracer advection speed of 2 cm s^{-1}, consistent with the average propagation speed proposed by Årthun et al. (2017). These results suggest that the propagation speed of the SST anomalies in the different models is dependent on both the speed of the main poleward current (i.e., the models' representation of the Norwegian Atlantic Current) and the parameterization of eddy mixing across the current. If we relate the averaged speed in Fig. 13 with the time lag of the cross-correlations for some of the prediction systems (Fig. 8), we find some relationships. For instance, NorESM and CESM-LE have similar velocity in the Nordic Seas, but largely differ in the subpolar North Atlantic with NorESM having the highest velocities. This seems to be consistent with the cross-correlation in their prediction systems, with a large time lag in CESM-DPLE and short time lag in NorCPM1. These two systems are the only ones with significant correlations between St7 and St2 in Fig. 8. Another example is NorESM with an overall higher velocity than IPSL-CM5. This seems to be consistent with the cross-correlation in their prediction systems; in the Nordic seas, there is a large time lag in IPSL-CM5 and a short time lag in NorCPM. However, as noted above, it is difficult to calculate a statistically robust relationship between surface velocity and time lags in the cross-correlation.

Second, the atmospheric influence on SST anomalies can be large. Atmospheric forcing has, for example, been shown to alter the along-path modification of SST anomalies by changing the current speed or by changing the air–sea heat fluxes (e.g., Furevik 2001). We emphasize that the SST propagation as demonstrated in Årthun et al. (2017) explains 55% of the variance of the bandpass filtered data. This indicates that although the propagation dominates a large portion of the SST variability along the Atlantic water pathway, other factors also influence the SST variability. Based on a heat budget analysis, Asbjørnsen et al. (2019) demonstrated that air–sea heat fluxes play an important role for Norwegian Sea heat content anomalies, acting to limit the predictability along the Atlantic water pathway in the Nordic seas. It is a major challenge for current state-of-the-art coupled global climate models to properly simulate ocean–atmosphere interactions (e.g., Feser et al. 2011), which could therefore be a potential source of intermodel differences in propagation speed.

c. Model differences and large spread in predictive skill

The global climate models have biases and differ substantially from each other in many aspects. This is shown both in this study (e.g., the characteristics of the Atlantic water pathway) and in other CMIP5 comparison studies investigating variability in sea ice extent and SST in the Nordic seas (Langehaug et al. 2017) and decadal variability in the North Atlantic (Ba et al. 2014; Menary et al. 2015). A study by Langehaug et al. (2012), which included both IPSL-CM4 and MPI-M ESM, showed that there are large model differences in the position of the North Atlantic Current and also in water mass transformation in the eastern subpolar region. A later study (Deshayes et al. 2014), also assessing IPSL-CM5A-LR and MPI-M ESM, documented that the models have clear differences in the extent of Atlantic Water both horizontally and with depth in the subpolar region. These model differences can contribute to the spread in skill that we see in this study.

The climate models and predictions systems used in this study differ in their spatial resolutions and their initialization approaches (Table 1). Most of the models have a similar resolution and use both surface and subsurface observations in the initialization of their ocean components, except for IPSL-CM5A-LR with its lower resolution (about 2°) and initialization of only SST. We have clearly seen that all of the models and systems differ largely, and it is rather challenging to identify specific settings or approaches that can lead to improved prediction skill for the North Atlantic–Arctic region. We note, however, that the prediction system (CESM1-DPLE) that most resembles the observed spatiotemporal SST pattern also has the highest number of ensemble members. This may suggest that a large ensemble is important to identify predictable signals as suggested in several studies (e.g., Scaife et al. 2019; Smith et al. 2019).

d. How to enhance climate prediction in the North Atlantic–Arctic region

There are several factors that are key to enhancing climate prediction, such as more ensemble members, higher model resolution, better understanding of the mechanisms taking place in our study region, more observations to feed into the initialization, and deeper investigations into the effects of different initialization methods. Furthermore, the recent study by Smith et al. (2020) shows the importance of postprocessing climate predictions to extract the predictable signal. In this study, we assessed how both models and prediction systems represent a spatiotemporal SST pattern identified in observations. This analysis helps us to better understand how the models and systems behave. More process-based or mechanistic ways than applied herein of assessing models and systems are important. Such approaches give us greater understanding of what is happening in the models and point to specific areas where models need to be improved. Improving the poleward propagation of anomalies could potentially enhance skill in the Atlantic water pathway toward the Arctic.

5. Conclusions

A key result from this study is that most of the climate models and dynamical decadal prediction systems have difficulties in representing the observed spatiotemporal SST variability along the Atlantic water pathway. In particular, the variability from the subpolar region, across the Greenland–

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

Scotland Ridge and into the Nordic seas, is challenging to simulate. This means that the mechanism that is identified to be a source of predictability along the Atlantic water pathway is not properly represented in the models, and hence limits their predictive capacity in this region on interannual to decadal time scales. In the eastern subpolar region, south of the Greenland–Scotland Ridge, the predictive skill is overall higher than farther north with significant skill at forecast times of 1–3 and 7–8 years. Farther north, the dynamical climate predictions show capability of predicting SST changes only at forecast time of 1 year (up to 2 years for some models).

A few prediction systems show re-emerging skill at longer forecast times. The multimodel mean is well above the persistence and reaching significant values in the eastern subpolar region, suggesting an important role of ocean circulation in bringing predictability to the subpolar region and toward the Arctic. If the predictive skills were high along the Atlantic water pathway for all forecast times (i.e., if the system was able to predict most variations in SST), we would expect that the system also reproduces the poleward propagation of anomalies for all forecast lead times. However, this relationship would not necessarily hold the other way around; a well-represented poleward propagation might enhance skill, but other model biases could still hamper predictive skill along the Atlantic water pathway.

Observational studies demonstrate predictability several years in advance along the Atlantic water pathway, thus suggesting a great potential for improvement of dynamical climate predictions. As this is currently a challenge for state-of-the-art dynamical prediction systems, one essential question for future research is how propagation of SST anomalies and their circulation with the ocean surface currents can be improved in dynamical prediction systems.

Acknowledgments. The research leading to these results has received funding from the Blue-Action Project (European Union's Horizon 2020 research and innovation program, Grant 727852), the Trond Mohn Foundation with the project Bjerknes Climate Prediction Unit (BCPU, Grant BFS20 18TMT01), the NordForsk under the Nordic Centre of Excellence (ARCPATH, 76654), and from the Bjerknes Centre with the project SKD MEDEVAC. The research leading to these results has also received funding from the German Federal Ministry of Education and Research (BMBF) through the JPI Climate/JPI Oceans NextG-Climate Science-ROADMAP (FKZ: 01LP2002A; DM and Norwegian Grant 316618/JPIC/JPIO-04; HRL and NK and ANR-19-JPOC-003; JM). PO was funded by the Spanish Ministry for the Economy, Industry and Competitiveness through the grant RYC-2017-22772. SY also receives financial support from the Danish National Center for Climate Research (NCKF). The National Center for Atmospheric Research is a major facility sponsored by the U.S. National Science Foundation (NSF) under Cooperative Agreement 1852977. EM is supported by the U.S. NSF Office of Polar Programs Grant 1737377. The prediction simulations using EC-EARTH anomaly initialization system were performed by SMHI on resources provided by the Swedish National Infrastructure for Computing (SNIC).

Data availability statement. For observation-based SST analyses the Hadley Centre Sea Ice and SST data set, both version 1.1 (HadISST1) and version 2 (HadISST2), are used. We thank the climate modeling groups listed in Table 1 for producing and making available their model output. All model data have been formatted by the respective modeling groups so that each hindcast for each variable is included in one single netCDF file. Details on these model data and how to request access are available from the modeling groups and Helene R. Langehaug at the Nansen Environmental and Remote Sensing Center.

REFERENCES

Acosta Navarro, J. C., and Coauthors, 2020: Link between autumnal Arctic sea ice and Northern Hemisphere winter forecast skill. *Geophys. Res. Lett.*, **47**, e2019GL086753, https://doi.org/10.1029/2019GL086753.

Årthun, M., and T. Eldevik, 2016: On anomalous ocean heat transport toward the Arctic and associated climate predictability. *J. Climate*, **29**, 689–704, https://doi.org/10.1175/JCLI-D-15-0448.1.

——, ——, E. Viste, H. Drange, T. Furevik, H. L. Johnson, and N. S. Keenlyside, 2017: Skillful prediction of northern climate provided by the ocean. *Nat. Commun.*, **8**, 15875, https://doi.org/10.1038/ncomms15875.

——, ——, and L. H. Smedsrud, 2019: The role of Atlantic heat transport in future Arctic winter sea ice loss. *J. Climate*, **32**, 3327–3341, https://doi.org/10.1175/JCLI-D-18-0750.1.

——, R. C. J. Wills, H. L. Johnson, L. Chafik, and H. R. Langehaug, 2021: Mechanisms of decadal North Atlantic climate variability and implications for the recent cold anomaly. *J. Climate*, **34**, 3421–3439, https://doi.org/10.1175/JCLI-D-20-0464.1.

Asbjørnsen, H., M. Årthun, Ø. Skagseth, and T. Eldevik, 2019: Mechanisms of ocean heat anomalies in the Norwegian Sea. *J. Geophys. Res. Oceans*, **124**, 2908–2923, https://doi.org/10.1029/2018JC014649.

Ba, J., and Coauthors, 2014: A multi-model comparison of Atlantic multidecadal variability. *Climate Dyn.*, **43**, 2333–2348, https://doi.org/10.1007/s00382-014-2056-1.

Bethke, I., and Coauthors, 2021: NorCPM1 and its contribution to CMIP6 DCPP. *Geosci. Model Dev.*, **14**, 7073–7116, https://doi.org/10.5194/gmd-14-7073-2021.

Bilbao, R., and Coauthors, 2021: Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth Syst. Dyn.*, **12**, 173–196, https://doi.org/10.5194/esd-12-173-2021.

Borchert, L. F., A. Düsterhus, S. Brune, W. A. Müller, and J. Baehr, 2019: Forecast-oriented assessment of decadal hindcast skill for North Atlantic SST. *Geophys. Res. Lett.*, **46**, 11 444–11 454, https://doi.org/10.1029/2019GL084758.

Broomé, S., and J. Nilsson, 2018: Shear dispersion and delayed propagation of temperature anomalies along the Norwegian Atlantic Slope Current. *Tellus*, **70A** (1), 1–13, https://doi.org/10.1080/16000870.2018.1453215.

Brune, S., A. Düsterhus, H. Pohlmann, W. A. Müller, and J. Baehr, 2018: Time dependency of the prediction skill for the

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

North Atlantic subpolar gyre in initialized decadal hindcasts. *Climate Dyn.*, **51**, 1947–1970, https://doi.org/10.1007/s00382-017-3991-4.

Chelton, D. B., 1983: Effects of sampling errors in statistical estimation. *Deep-Sea Res.*, **30A**, 1083–1103, https://doi.org/10.1016/0198-0149(83)90062-6.

Chepurin, G. A., and J. A. Carton, 2012: Subarctic and Arctic sea surface temperature and its relation to ocean heat content 1982–2010. *J. Geophys. Res.*, **117**, C06019, https://doi.org/10.1029/2011JC007770.

Dai, P., Y. Gao, F. Counillon, Y. Wang, M. Kimmritz, and H. R. Langehaug, 2020: Seasonal to decadal predictions of regional Arctic sea ice by assimilating sea surface temperature in the Norwegian Climate Prediction Model. *Climate Dyn.*, **54**, 3863–3878, https://doi.org/10.1007/s00382-020-05196-4.

Deshayes, J., R. Curry, and R. Msadek, 2014: CMIP5 model intercomparison of freshwater budget and circulation in the North Atlantic. *J. Climate*, **27**, 3298–3317, https://doi.org/10.1175/JCLI-D-12-00700.1.

Eldevik, T., and J. E. Ø. Nilsen, 2013: The Arctic–Atlantic thermohaline circulation. *J. Climate*, **26**, 8698–8705, https://doi.org/10.1175/JCLI-D-13-00305.1.

——, ——, D. Iovino, K. Anders Olsson, A. B. Sandø, and H. Drange, 2009: Observed sources and variability of Nordic seas overflow. *Nat. Geosci.*, **2**, 406–410, https://doi.org/10.1038/ngeo518.

Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, 2011: Regional climate models add value to global model data: A review and selected examples. *Bull. Amer. Meteor. Soc.*, **92**, 1181–1192, https://doi.org/10.1175/2011BAMS3061.1.

Furevik, T., 2000: On anomalous sea surface temperatures in the Nordic seas. *J. Climate*, **13**, 1044–1053, https://doi.org/10.1175/1520-0442(2000)013<1044:OASSTI>2.0.CO;2.

——, 2001: Annual and interannual variability of Atlantic Water temperatures in the Norwegian and Barents Seas: 1980–1996. *Deep-Sea Res. I*, **48**, 383–404, https://doi.org/10.1016/S0967-0637(00)00050-9.

Gao, Y., H. Drange, M. Bentsen, and O. M. Johannessen, 2005: Tracer-derived transit time of the waters in the eastern Nordic seas. *Tellus*, **57B**, 332–340, https://doi.org/10.3402/tellusb.v57i4.16553.

Germe, A., M. Chevallier, D. Salas y Mélia, E. Sanchez-Gomez, and C. Cassou, 2014: Interannual predictability of Arctic sea ice in a global climate model: Regional contrasts and temporal evolution. *Climate Dyn.*, **43**, 2519–2538, https://doi.org/10.1007/s00382-014-2071-2.

Giorgetta, M. A., and Coauthors, 2013: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *J. Adv. Model. Earth Syst.*, **5**, 572–597, https://doi.org/10.1002/jame.20038.

Guo, C., M. Ilicak, M. Bentsen, and I. Fer, 2016: Characteristics of the Nordic seas overflows in a set of Norwegian Earth System Model experiments. *Ocean Modell.*, **104**, 112–128, https://doi.org/10.1016/j.ocemod.2016.06.004.

Hazeleger, W., and Coauthors, 2013a: Predicting multiyear North Atlantic Ocean variability. *J. Geophys. Res. Oceans*, **118**, 1087–1098, https://doi.org/10.1002/jgrc.20117.

——, V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. J. Doblas-Reyes, K. Wyser, and M. Caian, 2013b: Multiyear climate predictions using two initialization strategies. *Geophys. Res. Lett.*, **40**, 1794–1798, https://doi.org/10.1002/grl.50355.

Hirahara, S., M. Alonso-Balmaseda, E. de Boisseson, and H. Hersbach, 2016: Sea surface temperature and sea ice concentration for ERA5. ERA Rep. Series No. 26, 25 pp., https://www.ecmwf.int/node/16555.

Holliday, N. P., and Coauthors, 2008: Reversal of the 1960s to 1990s freshening trend in the northeast North Atlantic and Nordic seas. *Geophys. Res. Lett.*, **35**, L03614, https://doi.org/10.1029/2007GL032675.

Huck, T., A. Colin de Verdière, and A. J. Weaver, 1999: Interdecadal variability of the thermohaline circulation in box-ocean models forced by fixed surface fluxes. *J. Phys. Oceanogr.*, **29**, 865–892, https://doi.org/10.1175/1520-0485(1999)029<0865:IVOTTC>2.0.CO;2.

Johnson, H. L., S. B. Cornish, Y. Kostov, E. Beer, and C. Lique, 2018: Arctic Ocean freshwater content and its decadal memory of sea-level pressure. *Geophys. Res. Lett.*, **45**, 4991–5001, https://doi.org/10.1029/2017GL076870.

Karcher, M. J., and Coauthors, 2004: The dispersion of 99Tc in the Nordic seas and the Arctic Ocean: A comparison of model results and observations. *J. Environ. Radioact.*, **74**, 185–198, https://doi.org/10.1016/j.jenvrad.2004.01.026.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1.

Killworth, P. D., D. B. Chelton, and R. A. de Szoeke, 1997: The speed of observed and theoretical long extratropical planetary waves. *J. Phys. Oceanogr.*, **27**, 1946–1966, https://doi.org/10.1175/1520-0485(1997)027<1946:TSOOAT>2.0.CO;2.

Kimmritz, M., F. Counillon, L. H. Smedsrud, I. Bethke, N. Keenlyside, F. Ogawa, and Y. Wang, 2019: Impact of ocean and sea ice initialisation on seasonal prediction skill in the Arctic. *J. Adv. Model. Earth Syst.*, **11**, 4147–4166, https://doi.org/10.1029/2019MS001825.

Krahmann, G., M. Visbeck, and G. Reverdin, 2001: Formation and propagation of temperature anomalies along the North Atlantic Current. *J. Phys. Oceanogr.*, **31**, 1287–1303, https://doi.org/10.1175/1520-0485(2001)031<1287:FAPOTA>2.0.CO;2.

Langehaug, H. R., P. B. Rhines, T. Eldevik, J. Mignot, and K. Lohmann, 2012: Water mass transformation and the North Atlantic Current in three multicentury climate model simulations. *J. Geophys. Res.*, **117**, C11001, https://doi.org/10.1029/2012JC008021.

——, D. Matei, T. Eldevik, K. Lohmann, and Y. Gao, 2017: On model differences and skill in predicting sea surface temperature in the Nordic and Barents Seas. *Climate Dyn.*, **48**, 913–933, https://doi.org/10.1007/s00382-016-3118-3.

——, A. B. Sandø, M. Årthun, and M. Ilıcak, 2019: Variability along the Atlantic water pathway in the forced Norwegian Earth System Model. *Climate Dyn.*, **52**, 1211–1230, https://doi.org/10.1007/s00382-018-4184-5.

Liu, Z., 1999: Forced planetary wave response in a thermocline gyre. *J. Phys. Oceanogr.*, **29**, 1036–1055, https://doi.org/10.1175/1520-0485(1999)029<1036:FPWRIA>2.0.CO;2.

Lozier, M. S., 2010: Deconstructing the conveyor belt. *Science*, **328**, 1507–1511, https://doi.org/10.1126/science.1189250.

Marshall, D. P., and H. L. Johnson, 2013: Propagation of meridional circulation anomalies along western and eastern boundaries. *J. Phys. Oceanogr.*, **43**, 2699–2717, https://doi.org/10.1175/JPO-D-13-0134.1.

Marshall, J., H. Johnson, and J. Goodman, 2001: A study of the interaction of the North Atlantic Oscillation with ocean circulation. *J. Climate*, **14**, 1399–1421, https://doi.org/10.1175/1520-0442(2001)014<1399:ASOTIO>2.0.CO;2.

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC

Matei, D., H. Pohlmann, J. Jungclaus, W. Müller, H. Haak, and J. Marotzke, 2012: Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *J. Climate*, **25**, 8502–8523, https://doi.org/10.1175/JCLI-D-11-00633.1.

Menary, M. B., D. L. R. Hodson, J. I. Robson, R. T. Sutton, R. A. Wood, and J. A. Hunt, 2015: Exploring the impact of CMIP5 model biases on the simulation of North Atlantic decadal variability. *Geophys. Res. Lett.*, **42**, 5926–5934, https://doi.org/10.1002/2015GL064360.

Mignot, J., J. García-Serrano, D. Swingedouw, A. Germe, S. Nguyen, P. Ortega, E. Guilyardi, and S. Ray, 2016: Decadal prediction skill in the ocean with surface nudging in the IPSL-CM5A-LR climate model. *Climate Dyn.*, **47**, 1225–1246, https://doi.org/10.1007/s00382-015-2898-1.

Mork, K. A., Ø. Skagseth, and H. Søiland, 2019: Recent warming and freshening of the Norwegian Sea observed by Argo data. *J. Climate*, **32**, 3695–3705, https://doi.org/10.1175/JCLI-D-18-0591.1.

Msadek, R., and Coauthors, 2014: Predicting a decadal shift in North Atlantic climate variability using the GFDL forecast system. *J. Climate*, **27**, 6472–6496, https://doi.org/10.1175/JCLI-D-13-00476.1.

Muir, L. C., and A. V. Fedorov, 2017: Evidence of the AMOC interdecadal mode related to westward propagation of temperature anomalies in CMIP5 models. *Climate Dyn.*, **48**, 1517–1535, https://doi.org/10.1007/s00382-016-3157-9.

Nigam, S., A. Ruiz-Barradas, and L. Chafik, 2018: Gulf Stream excursions and sectional detachments generate the decadal pulses in the Atlantic multidecadal oscillation. *J. Climate*, **31**, 2853–2870, https://doi.org/10.1175/JCLI-D-17-0010.1.

O'Mahony, M., 1986: *Sensory Evaluation of Food: Statistical Methods and Procedures*. Dekker, 487 pp.

Ortega, P., J. Mignot, D. Swingedouw, F. Sévellec, and E. Guilyardi, 2015: Reconciling two alternative mechanisms behind bi-decadal variability in the North Atlantic. *Prog. Oceanogr.*, **137**, 237–249, https://doi.org/10.1016/j.pocean.2015.06.009.

Orvik, K. A., and Ø. Skagseth, 2003: The impact of the wind stress curl in the North Atlantic on the Atlantic inflow to the Norwegian Sea toward the Arctic. *Geophys. Res. Lett.*, **30**, 1884, https://doi.org/10.1029/2003GL017932.

Persechino, A., J. Mignot, D. Swingedouw, S. Labetoulle, and E. Guilyardi, 2013: Decadal predictability of the Atlantic meridional overturning circulation and climate in the IPSL-CM5A-LR model. *Climate Dyn.*, **40**, 2359–2380, https://doi.org/10.1007/s00382-012-1466-1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 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, https://doi.org/10.1029/2002JD002670.

Robson, J. I., R. T. Sutton, and D. M. Smith, 2012: Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s. *Geophys. Res. Lett.*, **39**, L19713, https://doi.org/10.1029/2012GL053370.

Rossby, C. G., and Coauthors, 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Mar. Res.*, **2**, 38–55, https://doi.org/10.1357/002224039806649023.

Saravanan, R., and J. C. McWilliams, 1998: Advective ocean–atmosphere interaction: An analytical stochastic model with implications for decadal variability. *J. Climate*, **11**, 165–188, https://doi.org/10.1175/1520-0442(1998)011<0165:AOAIAA>2.0.CO;2.

Scaife, A. A., and Coauthors, 2019: Does increased atmospheric resolution improve seasonal climate predictions? *Atmos. Sci. Lett.*, **20**, e922, https://doi.org/10.1002/asl.922.

Schmith, T., S. M. Olsen, I. M. Ringgaard, and W. May, 2018: Limited predictability of extreme decadal changes in the Arctic Ocean freshwater content. *Climate Dyn.*, **51**, 3927–3942, https://doi.org/10.1007/s00382-018-4120-8.

Smith, D. M., and Coauthors, 2019: Robust skill of decadal climate predictions. *npj Climate Atmos. Sci.*, **2**, 13, https://doi.org/10.1038/s41612-019-0071-y.

——, and Coauthors, 2020: North Atlantic climate far more predictable than models imply. *Nature*, **583**, 796–800, https://doi.org/10.1038/s41586-020-2525-0.

Sundby, S., and K. Drinkwater, 2007: On the mechanisms behind salinity anomaly signals of the northern North Atlantic. *Prog. Oceanogr.*, **73**, 190–202, https://doi.org/10.1016/j.pocean.2007.02.002.

Swingedouw, D., J. Mignot, S. Labetoulle, E. Guilyardi, and G. Madec, 2013: Initialisation and predictability of the AMOC over the last 50 years in a climate model. *Climate Dyn.*, **40**, 2381–2399, https://doi.org/10.1007/s00382-012-1516-8.

Tang, S., H. von Storch, X. Chen, and M. Zhang, 2019: "Noise" in climatologically driven ocean models with different grid resolution. *Oceanologia*, **61**, 300–307, https://doi.org/10.1016/j.oceano.2019.01.001.

te Raa, L. A., and H. A. Dijkstra, 2002: Instability of the thermohaline ocean circulation on interdecadal timescales. *J. Phys. Oceanogr.*, **32**, 138–160, https://doi.org/10.1175/1520-0485(2002)032<0138:IOTTOC>2.0.CO;2.

Wang, Y., F. Counillon, N. Keenlyside, L. Svendsen, S. Gleixner, M. Kimmritz, P. Dai, and Y. Gao, 2019: Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF. *Climate Dyn.*, **53**, 5777–5797, https://doi.org/10.1007/s00382-019-04897-9.

Yashayaev, I., and D. Seidov, 2015: The role of the Atlantic Water in multidecadal ocean variability in the Nordic and Barents Seas. *Prog. Oceanogr.*, **132**, 68–127, https://doi.org/10.1016/j.pocean.2014.11.009.

Yeager, S. G., 2020: The abyssal origins of North Atlantic decadal predictability. *Climate Dyn.*, **55**, 2253–2271, https://doi.org/10.1007/s00382-020-05382-4.

——, and J. I. Robson, 2017: Recent progress in understanding and predicting Atlantic decadal climate variability. *Curr. Climate Change Rep.*, **3**, 112–127, https://doi.org/10.1007/s40641-017-0064-z.

——, A. Karspeck, G. Danabasoglu, J. Tribbia, and H. Teng, 2012: A decadal prediction case study: Late twentieth-century North Atlantic ocean heat content. *J. Climate*, **25**, 5173–5189, https://doi.org/10.1175/JCLI-D-11-00595.1.

——, A. R. Karspeck, and G. Danabasoglu, 2015: Predicted slowdown in the rate of Atlantic sea ice loss. *Geophys. Res. Lett.*, **42**, 10 704–10 713, https://doi.org/10.1002/2015GL065364.

——, and Coauthors, 2018: Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bull. Amer. Meteor. Soc.*, **99**, 1867–1886, https://doi.org/10.1175/BAMS-D-17-0098.1.

Zhao, J., A. Bower, J. Yang, X. Lin, and P. Hollidat, 2018: Meridional heat transport variability induced by mesoscale processes in the subpolar North Atlantic. *Nat. Commun.*, **9**, 1124, https://doi.org/10.1038/s41467-018-03134-x.

Unauthenticated | Downloaded 05/25/22 12:58 PM UTC