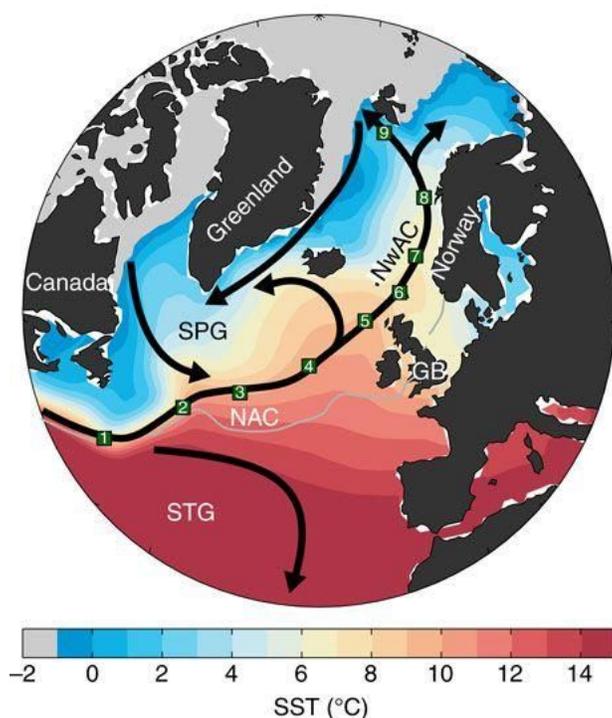# Benchmark performance of state-of- the-art prediction systems

Report describing the benchmark performance of state-of-the-art prediction systems on seasonal-to-decadal timescales, focusing on mechanisms and including forecast calibration techniques

Skillful prediction of northern climate provided by the ocean, Årthun, M. et al. Skillful prediction of northern climate provided by the ocean. Nat. Commun. 8, 15875 doi: 10.1038/ncomms15875 (2017).

About this document

Deliverable: D4.2 Benchmark performance of state-of- the-art prediction systems
Work package in charge: WP4 Enhancing the capacity of seasonal-to-decadal prediction in the Arctic and over the Northern Hemisphere.
Actual delivery date for this deliverable: 7 December 2019
Dissemination level: The general public (PU)

Lead authors
Max Planck Institute for Meteorology (MPI): Daniela Matei, Elisa Manzini, Katja Lohmann, Ying Dai, Johann Jungclaus.

Other contributing authors:
Netherlands eScience Center (NLeSC): Yang Liu, Laurens Bogaardt, Jisk Attema and Wilco Hazeleger
University of Bergen (UiB): Noel Keenlyside, Sebastien Barthelemy
University of Southampton (UoS): Sybren S. Drijfhout
National Oceanography Centre (NOC): Jennifer V. Mecking
Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC): Silvio Gualdi, Alessio Bellucci, Panos Athanasiadis
Danish Meteorological Institute (DMI): Shuting Yang, Torben Schmith, Tian Tian, Steffen M. Olsen
Centre National de la Recherche Scientifique (CNRS): Didier Swingedouw, Juliette Mignot, Marion Devilliers
Nansen Environmental and Remote Sensing Center (NERSC): Helene Langehaug, Yiguo Wang, Francois Counillon, Yongqi Gao
National Center for Atmospheric Research (NCAR): Elizabeth Maroon, Stephen Yeager, Gokhan Danabasoglou
Mercator: Gill Garric

We support Blue Growth!
Visit us on: www.blue-action.eu

Follow us on Twitter: @BG10Blueaction

Access our open access documents in Zenodo:
https://www.zenodo.org/communities/blue-actionh2020

Index

Summary for publication

The Blue-Action WP4 research within Task 4.1 has been mainly focusing on assessing in a mechanistic way the key processes responsible for the Arctic-lower latitude predictive linkages and their representation in the current state-of-art prediction systems.

Attribution of predictive skill of North Atlantic upper-ocean salt content.

The North Atlantic region is known to be a region with predictive skill at decadal scales in climate parameters such as sea surface temperature. Recently, decadal scale predictability of upper ocean salinity in the North Atlantic has also been demonstrated, a result of importance because it may lead to marine ecosystem prediction. Here, the origin of the upper ocean salinity predictability is investigated. By analysing hindcasts performed with the MPI-ESM decadal prediction system, it has been found that upper ocean salinity predictability can be attributed to the initialization of the circulation of the North Atlantic Ocean. Specifically, the upper ocean salinity predictability is related to the salt advection along the flow path of the ocean water from the subtropics.

Predictability of the North Atlantic Cold Blob

The subpolar North Atlantic sea surface temperature was near-record low during 2015, leading to a cold blob pattern in North Atlantic sea surface temperature anomalies for that year. During summer 2015, the cold blob has been linked to the European atmospheric heat wave of that summer. It is therefore of interest to determine the potential predictability of this extreme ocean event, given its societally relevant impacts related to its links to atmospheric heat waves. Analyses of five different prediction systems have shown that only two out of the five prediction systems predicted a weak, coherent cold blob in summer 2015, and none of the prediction systems captured the full magnitude of the observed cold blob within their ensemble. Although high predictive skill on seasonal-to-decadal timescales generally exists in the subpolar North Atlantic, the limited success obtained for the cold blob of summer 2015 points to the necessity of systematically assessing the prediction of extreme events to benchmark the performance of state-of-the- art prediction systems.

Propagation of Thermohaline Anomalies and their predictive potential

The heat transport of the Gulf Stream's extension toward the Arctic Ocean influences western European climate, Arctic sea ice conditions, and northern fisheries. Here, we address the question to what extent anomalies in sea surface temperatures along the Atlantic water pathway, which extend from the subpolar North Atlantic to the Nordic Seas, are predictable. The possibility of the existence of a predictive skill along the Atlantic water pathway is based on the observation that warm and cold anomalies propagate along the Atlantic water pathway with a certain time lag, and that this occurred repeatedly over the last 60 years. The analysis of the six decadal prediction systems available to the project has shown that most models have difficulties in representing this mechanism, but for lead times longer than 3 or 4 years some models represent the mechanism to some extent. Further analysis including the predictability of subsurface properties are necessary to test the robustness of the prediction results.

Predictability of the Rapid Warming of the North Pacific Ocean around 1990

The motivation to explore the predictability of the North Pacific Ocean region is the potential influence of North Pacific sea surface temperature on the Arctic stratosphere winter circulation, a known valuable source of predictability of tropospheric weather regimes. To this end, we address the strong and rapid warming of the North Pacific Ocean that occurred around 1990, with sea surface temperatures (SSTs) increasing by 2°C from 1988 to 1991. Simultaneously, the North Pacific Ocean underwent a quasi-decadal shift from a prolonged cold state in the 1980s to a warm state in the early-mid 1990s. The latter period has been characterized by a weakened Aleutian Low, a strengthened stratospheric polar vortex, and a lack of sudden stratospheric warming. Using the MPI-ESM decadal prediction system we demonstrate, for the first time, that the 1990 North Pacific rapid and strong warming in SSTs is predictable. The source of the predictive skill lays largely in the initialization with peak El Niño and the subsequent skillful prediction of La Niña, thereafter translated into skill for North Pacific SSTs *via* La Niña teleconnections. The quasi-decadal strengthening of the stratospheric polar vortex is also captured by the hindcasts, and potential for skillful predictions of the decreased frequency of SSWs is found.

Predictability of North Atlantic atmospheric anomalies

The multi-year predictive skill of North Atlantic atmospheric anomalies is investigated in hindcasts from the Community Earth System Model – Decadal Prediction Large-Ensemble (CESM-DPLE) simulations from NCAR. The atmospheric anomalies considered are (1) the number of winter days for which North Atlantic high-latitude blocking has been detected, and (2) the winter season index value of the Northern Atlantic Oscillation (NAO). The systematic analysis of prediction skill using anomalies averaged across all possible lead-year ranges has revealed significant skills over various lead-year ranges, for both atmospheric blocking and the NAO.

Trends in North Atlantic subpolar gyre temperatures

The variability of temperature over the North Atlantic subpolar gyre (SPG) is investigated in observations and model simulations, motivated by the extreme strong negative trend observed over 2004-2018. Large decadal variability, here measured by 15-year linear trends, is found in the Atlantic Ocean (SPG) sea surface temperature, in both models and observations. For the past and current climate, it is found that the extreme strong negative trend observed over 2004-2018 may occur as an extreme event. In the future, extreme cooling events of similar strength may still be possible, up to a global warming of about 6°C.

New prediction systems and analysis techniques

To advance prediction capabilities and analysis, a new prediction system with an improved inclusion of salinity in the initial conditions and based on an updated climate model has been completed, a method to systematically analyze the best skill scores for temperature, and a new debiasing technique have been developed, making use of the IPSL-CM decadal prediction system and by collecting different sources of climate information. It has been found that for European temperatures, the summer season is the most skillful season for decadal prediction, the best lead time is the average over 1-9 years, and that the best windows of opportunity for predictions is 1981-2000.

Forecast calibration techniques: Extended range Arctic sea forecast

Operational Arctic sea ice forecasts are of crucial importance to commercial and scientific activities in the Arctic region. Here we focus on the prediction of the sea ice in the Barents Sea at weekly to monthly time scales and propose a deep learning approach to calibrate forecasts of Arctic sea ice at weather time scales. With the input fields from atmospheric and oceanic reanalysis data sets, the deep learning technique is indeed able to learn the variability of the Arctic sea ice within historical records and effectively predict regional sea ice concentration patterns at weekly to monthly time scales. This method outperforms predictions with climatology and persistence at chosen time scales and is promising to act as a fast and cost-efficient operational sea ice forecast system in the future.

Work carried out

Attribution of predictive skill of North Atlantic upper-ocean salt content (MPI)

Predictive skill of North Atlantic upper-ocean salinity has, in contrast to upper-ocean temperature, so far not received much attention in the literature. However, given that upper-ocean salinity in the Northeastern Atlantic has been suggested to be an indicator and predictor of the abundance and distribution of marine ecosystem species (e.g., Miesner and Payne, 2018), the investigation of the predictive skill of North Atlantic upper-ocean salinity is highly desirable. Within FP7 project NACLIM (grant agreement n.308299) we have assessed the predictive skill of North Atlantic upper-ocean salt content using initialized decadal prediction experiments, in the following referred to as hindcasts, performed with six coupled climate models from the fifth phase of the Coupled Model Intercomparison Project (CMIP5): MPI-ESM-LR, EC-Earth, HadCM3, GFDL-CM2.1, CanCM4, MIROC5. Based on the multi-model ensemble mean, we demonstrate decadal-scale predictive skill of upper-ocean salt content in the entire subpolar North Atlantic and the eastern part of the Nordic Seas (upper panels in Figure 1). At forecast times from about five years onwards, the skill based on the initialized hindcasts is significantly above the skill of the persistence forecast.

An interesting question not addressed within NACLIM is which mechanism extends the predictive skill beyond the persistence forecast, and we shall attempt to answer it within BlueAction WP4, based on hindcasts of our own model - the Max Planck Institute for Meteorology Earth System Model (MPI-ESM). Like for all individual models, the predictive skill of upper-ocean salt content based on the MPI-ESM hindcasts (lower panels in Figure 1) is generally lower than the skill based on the multi-model ensemble mean. MPI-ESM shows the relatively high skill at longer forecast times in the subpolar North Atlantic, especially in the eastern part, where MPI-ESM performs even better than the multi-model ensemble mean. MPI-ESM, however, lacks any significant skill of upper-ocean salt content at longer forecast times in the eastern part of the Nordic Seas. This result is somewhat surprising, as Langehaug et al. (2017), based on the MPI-ESM model, demonstrate significant skill at longer forecast times of sea surface temperature in the eastern part of the Nordic Seas.

Following Matei et al. (2012), we correlate the salt content from the initialized hindcasts at the various forecast times with ocean circulation indices at the start of the hindcasts. Regarding the upper-ocean salt content in the eastern and western subpolar North Atlantic, a remarkable correspondence between the evolution of the predictive skill (lower panels in Figure 1) and the correlation with the initial strength of the North Atlantic subpolar gyre (upper panels in Figure 2) is found at forecast times from about five

years onwards. For the upper-ocean salt content in the eastern subpolar North Atlantic, there is also a close correspondence between the predictive skill and the correlation with the initial strength of the Atlantic meridional overturning circulation (lower panels in Figure 2). These results imply that the correct initialization of the North Atlantic Ocean circulation and thus of the salt advection along the flow path of the subtropical water can explain the predictive skill at forecast times from about five years onwards.
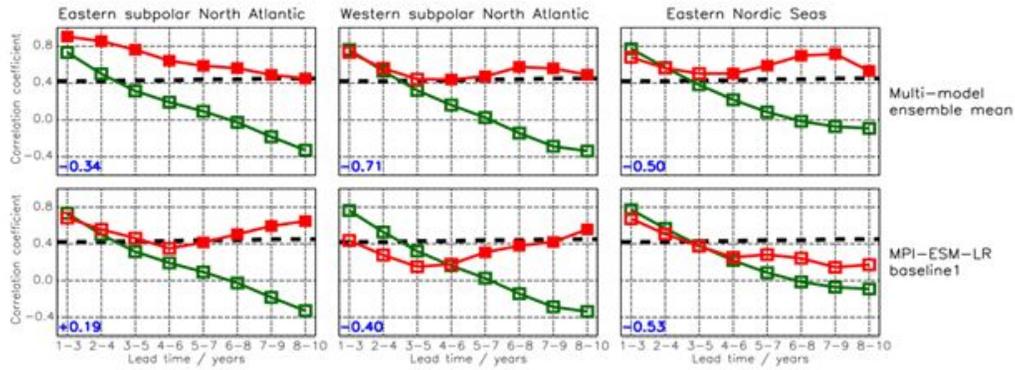

Figure 1: *Anomaly correlation coefficient between the upper-ocean (0-500 meter) salt content from the ISHII dataset and respectively from the initialized hindcasts (red) and from a damped persistence model based on the ISHII dataset (green) according to lead time. Filled red squares indicate that the skill of the hindcasts is significantly above the skill of the persistence forecast. The number in the lower left corner of the panels gives the anomaly correlation coefficient between the upper-ocean (0-500 meter) salt content from the ISHII dataset and from the historical simulations (for the period covered by the hindcasts). The black dashed line gives the 95% significance level according to a one-sided t-test. Upper panels from FP7 project NACLIM (grant agreement n.308299).*
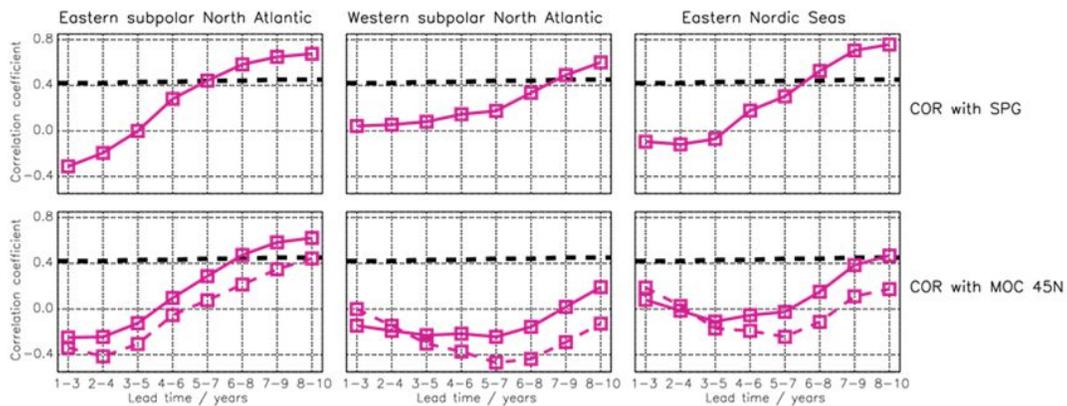
Figure 2: *Correlation between the strength of the North Atlantic subpolar gyre (upper panels) or the Atlantic meridional overturning circulation (AMOC, lower panels) at the start of the MPI-ESM hindcasts and the upper-ocean salt content in the MPI-ESM hindcasts. For the correlation with the AMOC strength, solid (dashed) magenta lines indicate that the AMOC strength has been linearly detrended (has not been detrended) prior to analysis. The black dashed line gives the 95% significance level.*

Predictability of the North Atlantic Cold Blob (NCAR, UiB, NERSC, DMI, CNRS, MPI, UoS, NOC; External Partners: University of Reading, UK Met Office)

The 2015 subpolar cold SST anomaly in the North Atlantic is examined as a case study to test seasonal-to-decadal prediction systems for a particularly challenging and societally relevant event. Previous studies indicate that the 2015 North Atlantic cold SST anomaly contributed to the summer 2015 European heat wave (Duchez et al. 2016, Mecking et al. in press). Earlier analysis of the Community Earth System Model Decadal Prediction Large Ensemble (CESM-DPLE) shows that subpolar SST are well-predicted all the way to the end of the 10-year hindcasts (Yeager et al. 2018). This 2015 cold event, however, is an exception: it was not encompassed within the full spread of the CESM-DPLE's 40 ensemble members. The cold anomaly was not captured even in the November 2014 hindcast set, which was initialized with a cold anomaly only eight months prior to the peak of the cold blob. Instead, the cold anomaly evolved toward an ensemble mean state that was anomalously warm during summer 2015. A heat budget analysis of the CESM forced ocean and sea ice simulation (FOSI) used to initialized the CESM-DPLE indicates that there is too little winter cooling due to surface heat fluxes and ocean advection in the CESM-DPLE. This lack of cooling is in turn connected to anomalously weak surface winds, especially in January and March of 2015, which were instead months of strong positive North Atlantic Oscillation (NAO) conditions. Further analysis of the CESM-DPLE November 2014 hindcasts indeed indicated that none of the ensemble members had NAO conditions that were as positive as those observed.

Based on these results from the CESM-DPLE, a multi-model study was initiated to determine if the inability to forecast this extreme cold event was limited to the CESM-DPLE and to test other decadal prediction systems on this particularly challenging event. Key scientific questions to examine include how well the cold blob was captured by each system with only 6-9 months of lead time, if a predictive success (or failure) with this event was associated with anomalous subpolar cooling due to surface heat fluxes, and how well each system predicted the positive NAO conditions of 2015. A successful prediction would be judged by whether the observed quantity (e.g., subpolar SST, NAO, etc.) was encompassed within the ensemble spread and whether any signal of the observed quantity was present in the ensemble mean. If models were successful for the shortest lead time of 6-9 months, then hindcasts initialized from earlier dates would next be examined.

Model output (sea surface temperature, sea level pressure, sea ice concentration, surface heat flux) were provided by all contributing partners and collected by NCAR. NCAR led the analysis of five different prediction systems: CESM-DPLE, DePreSys3, EC-Earth3, IPSL-CM5A-LR, NorCPM1 (Swingedouw et al. 2013, Dunstone et al. 2016, Counillon et al. 2016, Wang et al. 2017, Yeager et al. 2018). Anomalies for all systems were calculated using a standard drift-correction method (Goddard et al. 2013) with a 1983-2014 climatology. This approach is necessary because some of the prediction systems use full-field initialization approaches that often causes spurious drift. Even though some systems were anomaly-initialized and had little drift, a priority was to analyze all systems as consistently as possible. The evolution of the ensemble mean and spread of SST, surface heat fluxes, and atmospheric circulation were examined. The main results are summarized here.

Mean subpolar North Atlantic SST anomalies were calculated for all hindcast systems in the box (45°W-20°W, 45°N-65°N) and are shown in Figure 3. The cold blob peaked in magnitude during summer 2015, and a successful forecast is indicated if the observed anomalies (blue line) is encompassed within the spread of ensemble members (grey lines). All systems were initialized in Fall 2014 or Winter 2015 from a cold state (with the exception of one of the two IPSL-CM5A-LR ensemble sets). By summer 2015, however, none of the ensembles contain a member that has subpolar SST as cold as that observed. Three of the models (DePreSys3, IPSL-CM5A-LR, NorCPM1) do contain members with negative anomalies during summer 2015, and produce a weak cold blob in the ensemble mean. Of these three, a single, coherent cold blob in the observed location is seen in two of the systems (Figure 4). Ensemble mean anomalies in the other systems (CESM-DPLE, EC-Earth3) are instead net positive by summer 2015; these hindcast simulations were unable to sustain the initial cold anomalies.
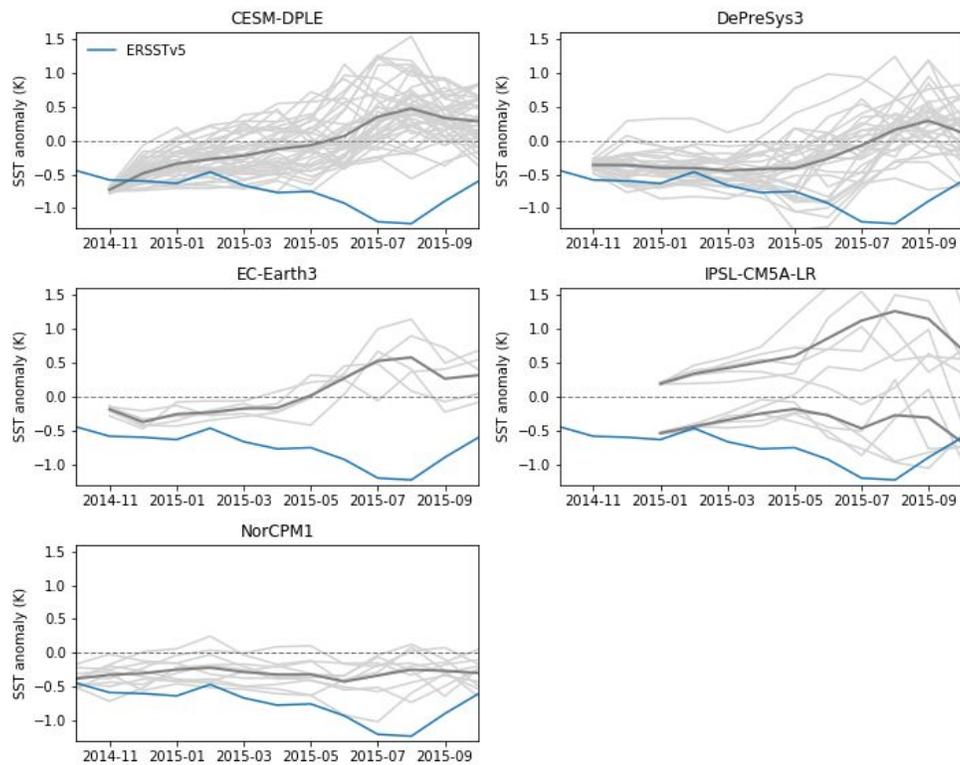

Figure 3: *Evolution of subpolar North Atlantic SST predictions during 2015 from five different climate prediction systems. The displayed predictions are derived from the hindcast initialized in late 2014 or early 2015, depending on the model. Dark grey lines indicate the ensemble mean, while light grey lines indicate individual ensemble members. The IPSL-CM5A-LR simulations include two ensemble sets for this hindcast, differing only in their initial conditions. Blue line shows the observed value and is derived from the Extended Reconstructed SST version 5 (ERSSTv5 – Huang et al. 2017).*
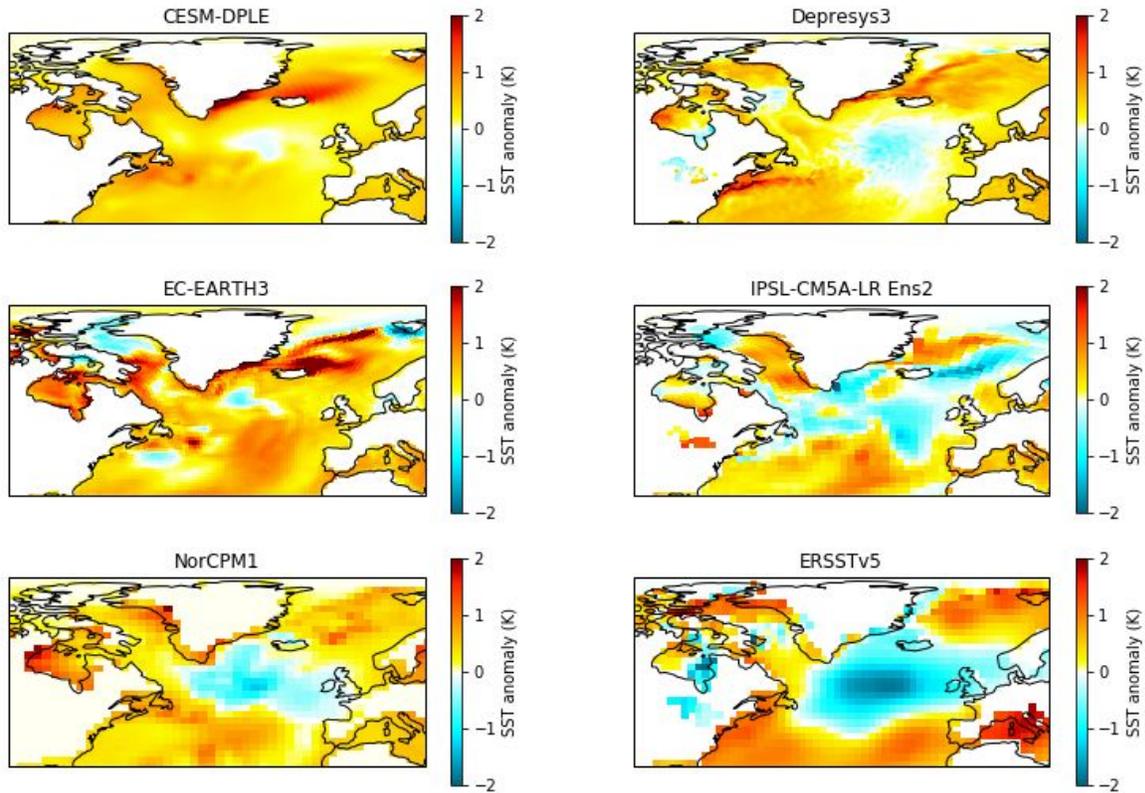
Figure 4: Summer (June-July-August) 2015 ensemble mean predicted and observed (bottom-right panel) subpolar North Atlantic SST anomalies.

Examination of reanalysis surface heat fluxes and a CESM hindcast heat budget indicate that strong cooling occurred through the winter and spring of 2015 and contributed to the deepening of the cold blob. In January and March 2015, this cooling was especially pronounced and was associated with strong North Atlantic Oscillation (NAO) positive conditions. Surface heat fluxes during the winter in the five prediction systems were overly positive compared to reanalysis (Figure 5). The two systems that produced weak, coherent cold anomalies (NorCPM1 and Depresys) are weakly negative during February or March 2015. During May-August 2015, all systems produced positive surface heat flux anomalies that acted to warm the ocean, while reanalysis indicates negative surface heat flux anomalies. These positive summer anomalies also likely contributed to the lack of a strong predicted cold blob.
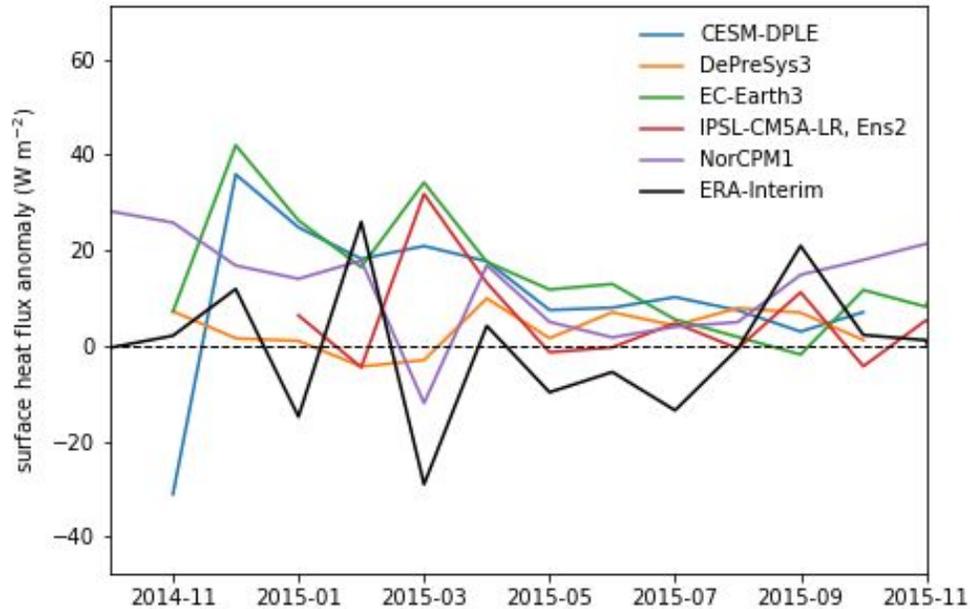
Figure 5: Subpolar North Atlantic net surface heat flux in prediction systems and ERA-Interim reanalysis (Dee et al. 2011) during 2014-2015. The ensemble mean surface heat flux is shown for each model.

The lack of negative surface heat fluxes in the prediction systems can be associated with too weak turbulent heat fluxes due to weak surface winds. The question is therefore if the prediction system captured the observed notable positive NAO conditions occurred in January and March 2015, because positive NAO conditions are known to produce strong surface winds in the subpolar North Atlantic. A NAO index was calculated in observations and prediction systems. In observations, the NAO was defined as the first principal component of the monthly mean sea level pressure field. To calculate NAO in each of the prediction systems, the monthly mean sea level pressure anomalies from each prediction system were interpolated to the observed grid and then projected on the normalized, observed empirical orthogonal functions. The first principal component is interpreted to be the NAO and so it indicates the strength of the observed NAO weighing pattern (Figure 6). Observed NAO was greater than one for most of winter and spring 2015. None of the systems produced an ensemble member with an NAO as persistently positive as observed. The ensemble mean NAO in DePreSys3 was positive in April and May 2015, which may have allowed greater cooling during these months instead of March; however, the timing of negative surface heat fluxes in Figure 5 in February/March does not coincide with strong positive NAO, which instead occurred in April/May. NorCPM1, which also produced a weak cold blob, has weakly positive ensemble mean NAO conditions in December-March, but surface heat fluxes were negative during only March. As such, there is weak support for the hypothesis that models with more positive NAO conditions also had negative surface heat flux anomalies that could sustain or amplify a modelled cold blob. Future examination of this case study will be on the differences in upper ocean initialization and evolution.
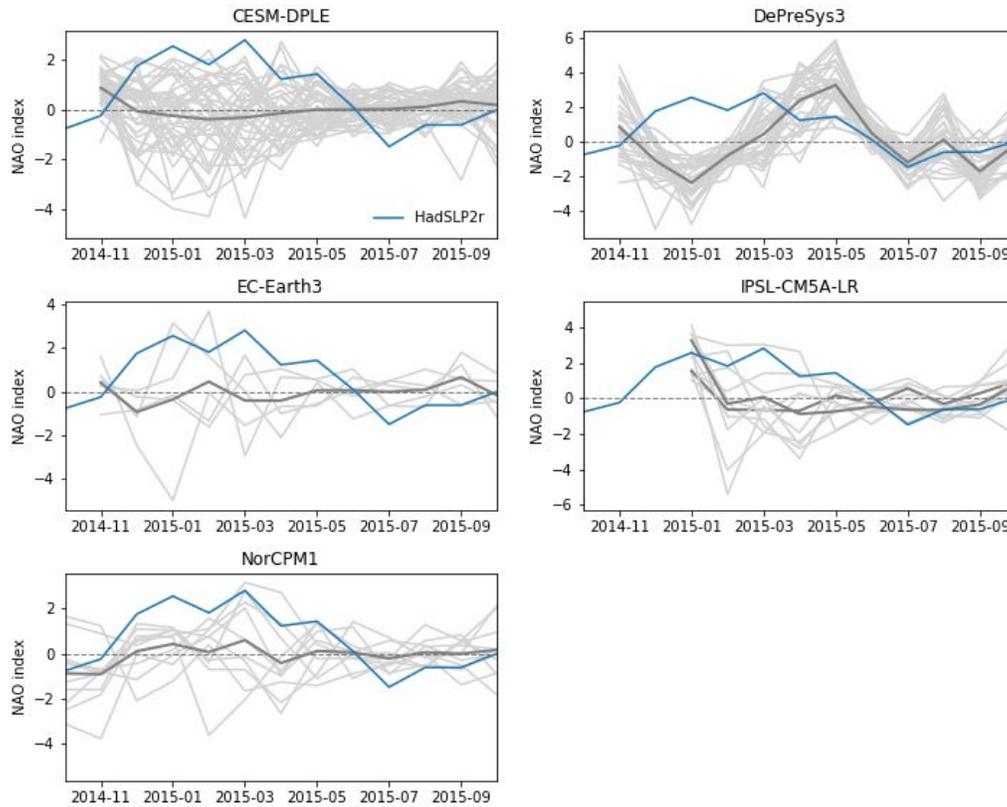
Figure 6*: Monthly mean NAO in climate prediction systems and observations (HadSLP2r, Allan and Ansell, 2006)). NAO for individual ensemble members are plotted in light grey, ensemble means in dark grey, and for observations in blue. Two ensembles with differing initial conditions are presented for IPSL-CM5A-LR.*

Propagation of Thermohaline Anomalies and their predictive potential (NCAR, UiB, DMI, CNRS, MPI)

We hypothesize that predictive skill or lack thereof along the Atlantic water pathway in the subpolar North Atlantic and the Nordic Seas is related to the northward propagation of thermohaline anomalies. To test this, we use the six decadal predictions systems available within the project. The decadal hindcasts start every year, except one system that start every second year, and runs for 10 years. We assess the predictive skill of about 40 (20 if every second year) hindcasts that start in the time period 1960-2001. For each hindcast, several members have been run, and herein the main focus is on the ensemble mean. See Table 1 for more details on each of the prediction systems. Since we are focusing on interannual-to-decadal variability, we apply a third-order 3-30 year band-pass Butterworth filter to all time series, prior to the calculation of the correlation. We have tested the robustness of using this filter by also applying a 5-40 year band-pass filter. This gives overall same results in the predictive skill along the Atlantic water pathway.

Observations show that there exists a lagged correlation for SST anomalies along the Atlantic water pathway, suggesting that the anomalies are advected northward with time (Århun et al., 2017; Langehaug et al., 2018, Figure 8 upper panels). The observational analysis has been performed by using SST time series from stations along the Atlantic water pathway (see magenta stations on map in Figure 7). Here we confront each of the decadal prediction system with the propagation mechanism

identified from observations. We use cross-correlation or lagged correlation to assess how SST anomalies along the Atlantic water pathway co-vary.

We calculated skill of winter SST along the Atlantic water pathway by averaging SST within specific latitude bands and SST intervals (Figure 7). Initially, there is relatively high skill in the eastern subpolar region. The skill drops at lead times 2-4 years and then increases on longer lead times (8-10 years; 5-7 years for CESM-DPLE), becoming significant for some models (Figure 7).

As described above, observation-based data (HadISST1) show that SST at the northernmost station has a relatively high and positive correlation with all other stations upstream, but the correlation weakens, and the time lag increases as we move upstream (upper panel, Figure 8). This pattern is relatively similar for the two mean lead times shown in Figure 8, although somewhat variable for the correlation with stations 1 and 2 (located in the subpolar region). Applying the same analysis on all models, we find in general that the cross-correlation patterns differs with lead times and across models (Figures 8-9). On medium lead times (2-4 years), most models do not represent the cross-correlations as seen in HadISST1, although two of the models appear to capture the pattern but with weak correlations. On the other hand, on longer lead times (8-10 years; 5-7 years for CESM-DPLE), the cross-correlation patterns resemble to a larger degree what is seen in HadISST. However, the time lag for each of the stations vary largely among the models. This could imply that the advection speed is different in the models. In general, there is an increase in skill at longer lead times when the propagation mechanism is better represented compared to shorter lead times. The skill is not very high (and for some models not significant).

Model	Ocean Resolution	Ensemble size	Start year of each prediction	Full-field (F) or Anomaly (A)	Start date	Length (years)	Initialization strategy	Num. of hind-casts	Start years used
NorCPM1 (norcpm-cmip6_hindcast)	320 x 384 51+2 layers (isopycnal)	10	Yearly (1960-2018)	A	Nov 1st	10	initialized with EnKF anomaly assimilation of SST and TS-profiles	42	1960-2001
CESM-DPLE	320 x 384 60 layers	40	Yearly (1954–2015)	F	Nov 1st	10	initialized with CESM1.1 forced with historical atm state and flux fields	42	1960-2001
IPSL-CM5A-LR (DEC3)	182 x 149 31 layers	3	Yearly (1961–2013)	A	Jan 1st	10	surface nudging of SST (ERSST)	41	1961-2001
EC-Earth (v2.3) SMHI/DMI	320 x 160 42 layers	10	Yearly (1960–2005)	A	Nov 1st	10	initialized with ocean reanalysis (and atm/ice?)	41	1961-2001
EC-Earth (v2.3) BSC	362 x 292 42 layers	5	Every second year (1960–2010)	F	Nov 1st	10	initialized with ERA-40 and ERA interim atm/land reanalyses and NEMOVAR-S4 ocean reanalysis	21	1960-2000
MPI CMIP5 MPI-ESM-LR	256 x 220 40 layers	10	Yearly (1961–2011)	A	Jan 1st	10	3D T&S from ORAs4 ocean reanalysis; full-field initialization for atm. from ERA40/ERAInterim	41	1961-2001

Table 1: *Model and data assimilation specifications of the six dynamical prediction systems used in this study.*
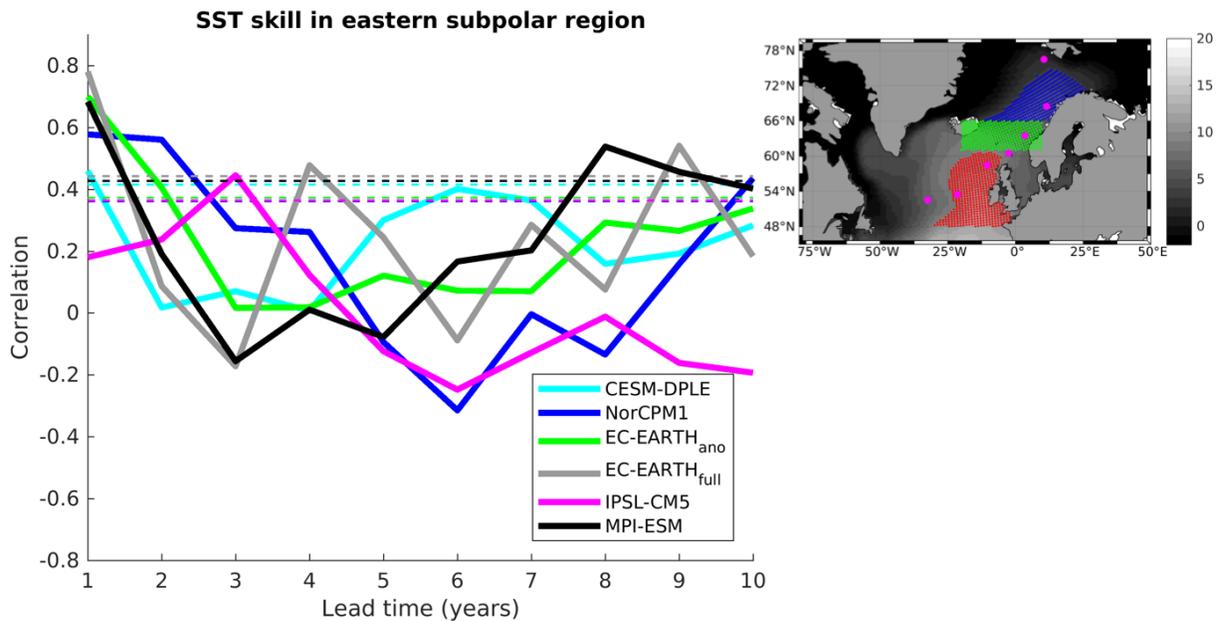
Figure 7: Anomaly correlation coefficient of winter (Jan-Apr) SST for the eastern subpolar North Atlantic (solid lines). The correlation is between HadISST1 data and the ensemble mean of the hindcasts at different lead times. At each lead time, time series are band pass filtered prior to correlation. Significant correlations at the 90% level (dashed lines) by the standard two-sided Student's t-test (taking into account autocorrelation). The eastern subpolar North Atlantic is defined as the domain with SST larger than 9° C (red region on map) and within the latitude and longitude bands 48-61° N and 40° W-40° E, respectively.
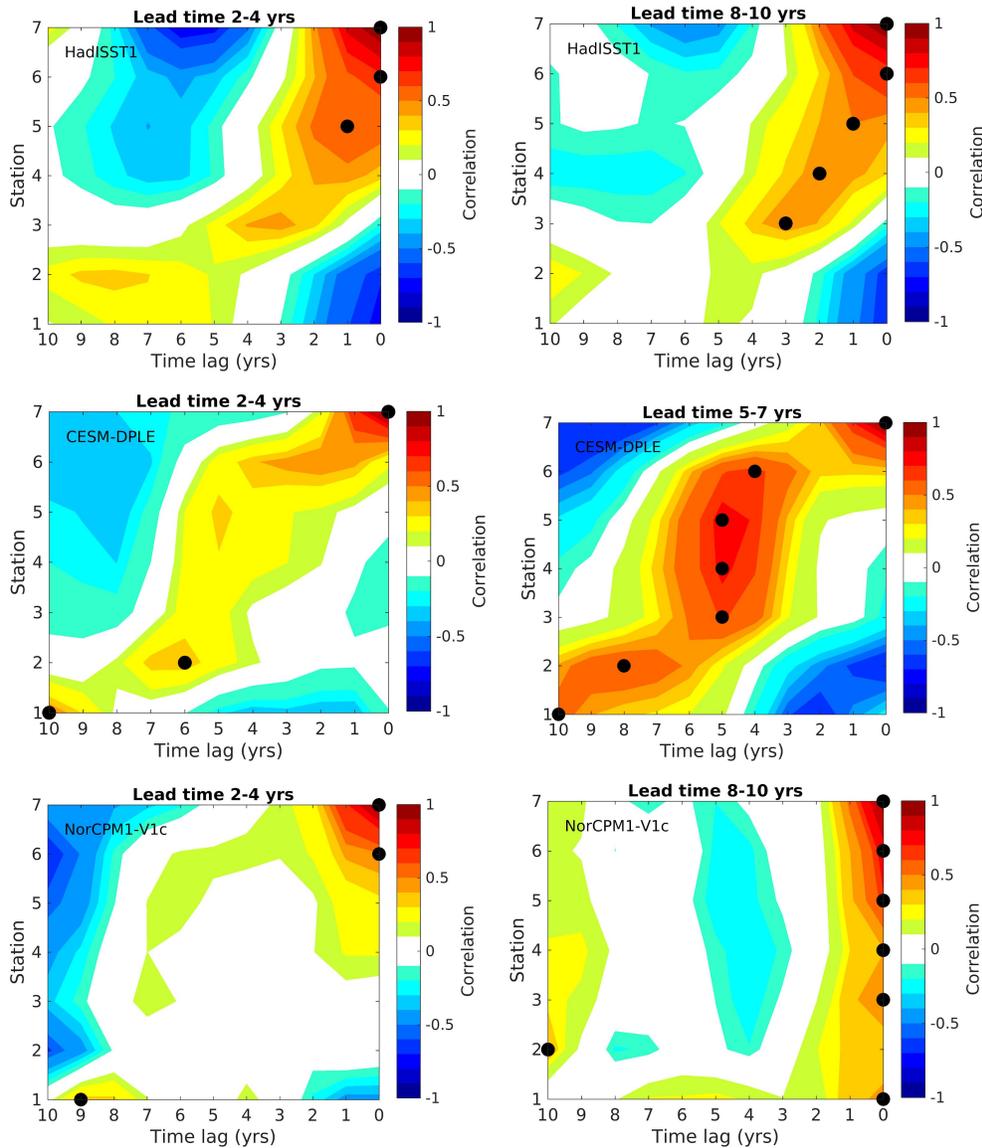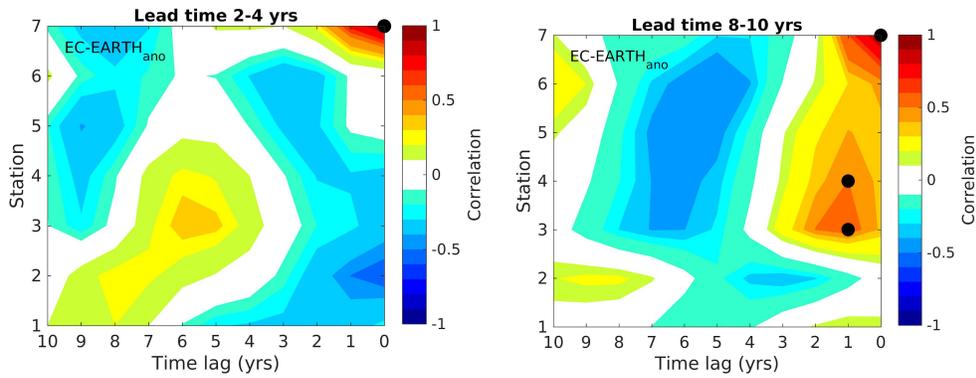
Figure 8: Cross-correlation of winter SST between the northernmost station and all seven stations (Figure 7, magenta circles). Black circles mark the maximum significant positive correlation for each station. Time series are band pass filtered prior to correlation. Significant correlations at the 90% level by the standard two-sided Student's t-test (taking into account autocorrelation). Shown are the observations (upper panels) and three decadal prediction systems.
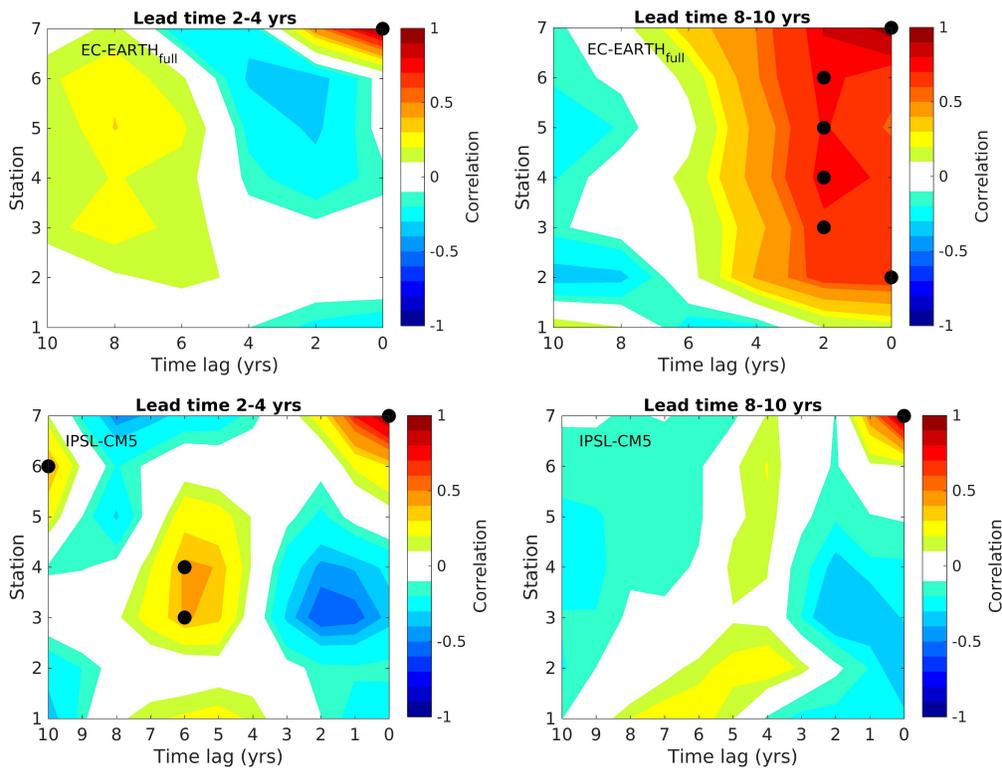
Figure 9: Same as for Fig. 8, but for the other three decadal prediction systems.

Predictability of the Rapid Warming of the North Pacific Ocean around 1990 (MPI, NCAR)

A strong and rapid warming of the North Pacific Ocean has been observed around 1990, with the SST increasing by 2°C from 1988 to 1991. Simultaneously, the North Pacific Ocean underwent a quasi-decadal shift from a prolonged cold state in the 1980s to a warm state in the early-mid 1990s. The latter period has been characterized by a w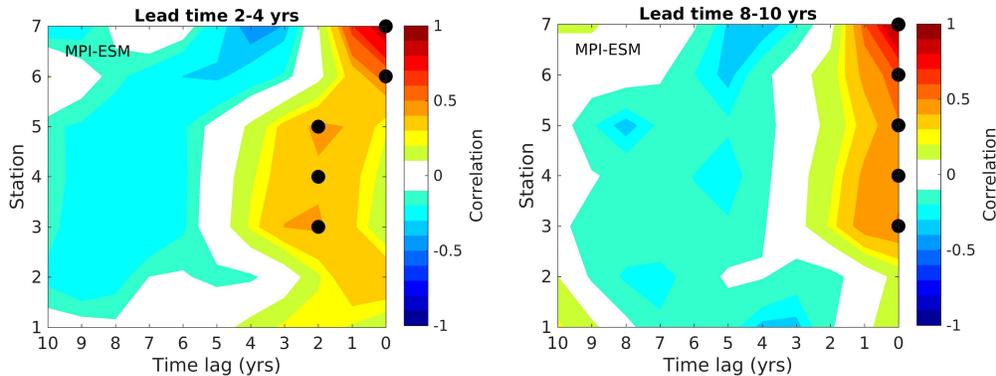eakened Aleutian Low, a strengthened stratospheric polar vortex. The special interest in North Pacific SSTs is motivated by their potential influence on the Arctic stratosphere winter climate (Hurwitz et al. 2012), a known valuable source of predictability of tropospheric weather regimes (Baldwin and Dunkerton 2001). Here we analyze hindcasts and historical experiments from the MPI-ESM decadal prediction systems to address the following questions: Are the strong and rapid increase in the North Pacific SSTs and the following sustained warm anomaly in SST predictable? To what extent are the associated weakened Aleutian Low and the strengthened stratospheric polar vortex predictable? Can we identify the source of predictive skill?

Prediction skill for the rapid warming is demonstrated by the evolution of the SSTs, for the ensemble hindcasts initialized at the end of 1987 (Figure 10, right panels). The sharp SST increase and the following persistence in positive SST anomalies during the 1990s are simulated by the initialized hindcasts, for each of the model resolution considered. On the contrary, the rapid warming and the persistent anomalies are not captured by any of the uninitialized historical experiments (Figure 10, left panels). Furthermore, the 1991-1997 ensemble mean of the pooled anomalies from the three MPI-ESM initialized hindcasts show a weakened Aleutian Low (Figure 11, left), westward of the dateline, the location where a trough in the stationary waves is located (Garfinkel et al. 2010). Therefore, the predicted geopotential height positive anomaly favors a reduction of the strength of the stationary waves and consequently a deeper and stronger stratospheric vortex (Fig. 11, right). As a first attempt to evaluate the predictability of changes in the frequency of occurrence of sudden stratospheric warming (SSWs) events, we analyzed daily occurrence of extremes in geopotential height anomalies, at lead year one for the winter seasons from 1990/91 to 1996/97. We find that averaged over the seven winters, the lead year 1 prediction provide a 24% reduction of positive geopotential height extremes (a proxy for SSWs) in the early-mid 1990s, wherein a single member produces a 57% reduction that matches with the reduction in positive geopotential height extremes from the reanalysis for the considered period.
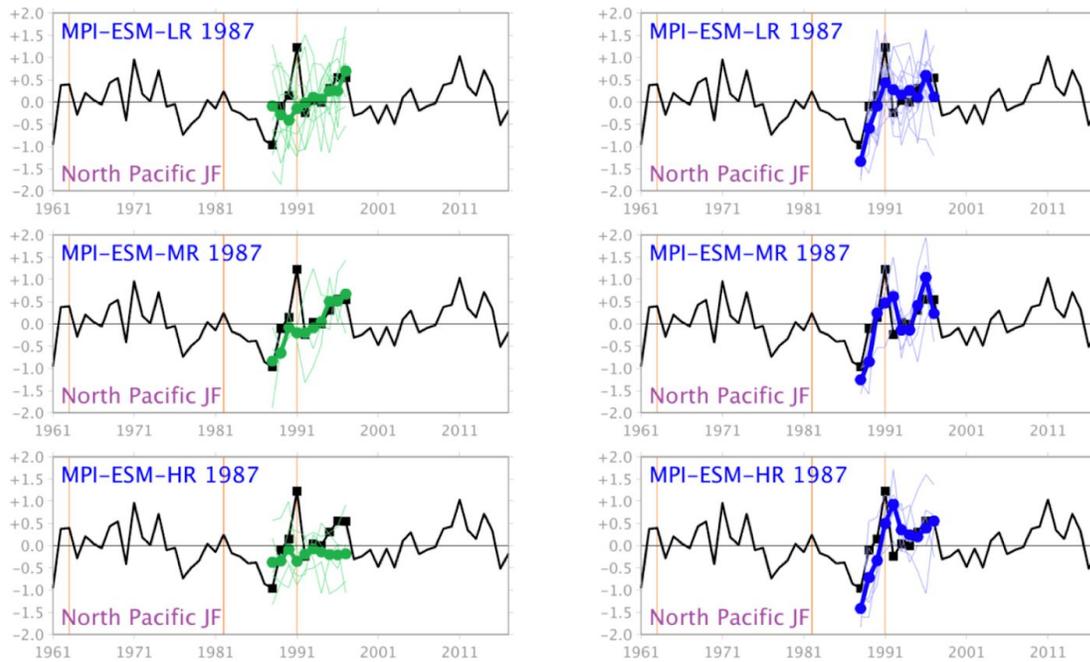
Figure 10: *North Pacific SST anomalies averaged over January-February. Anomalies form the HadISST1 reanalysis are shown in all panels (black curves). The ensemble mean anomaly of the historical experiment are shown at left (thick green), and the ensemble mean anomaly of the hindcasts initialized at the end of 1987 are shown at right (thick blue), for MPI-ESM-LR (top), MPI-ESM-MR (middle) and MPI-ESM-HR (bottom). Single realisations are depicted by the thin colored curves.The yellow vertical lines indicate volcanic eruptions: Mt. Agung (March 1963), El Chichón (April 1982) and Mt. Pinatubo (June 1991).*
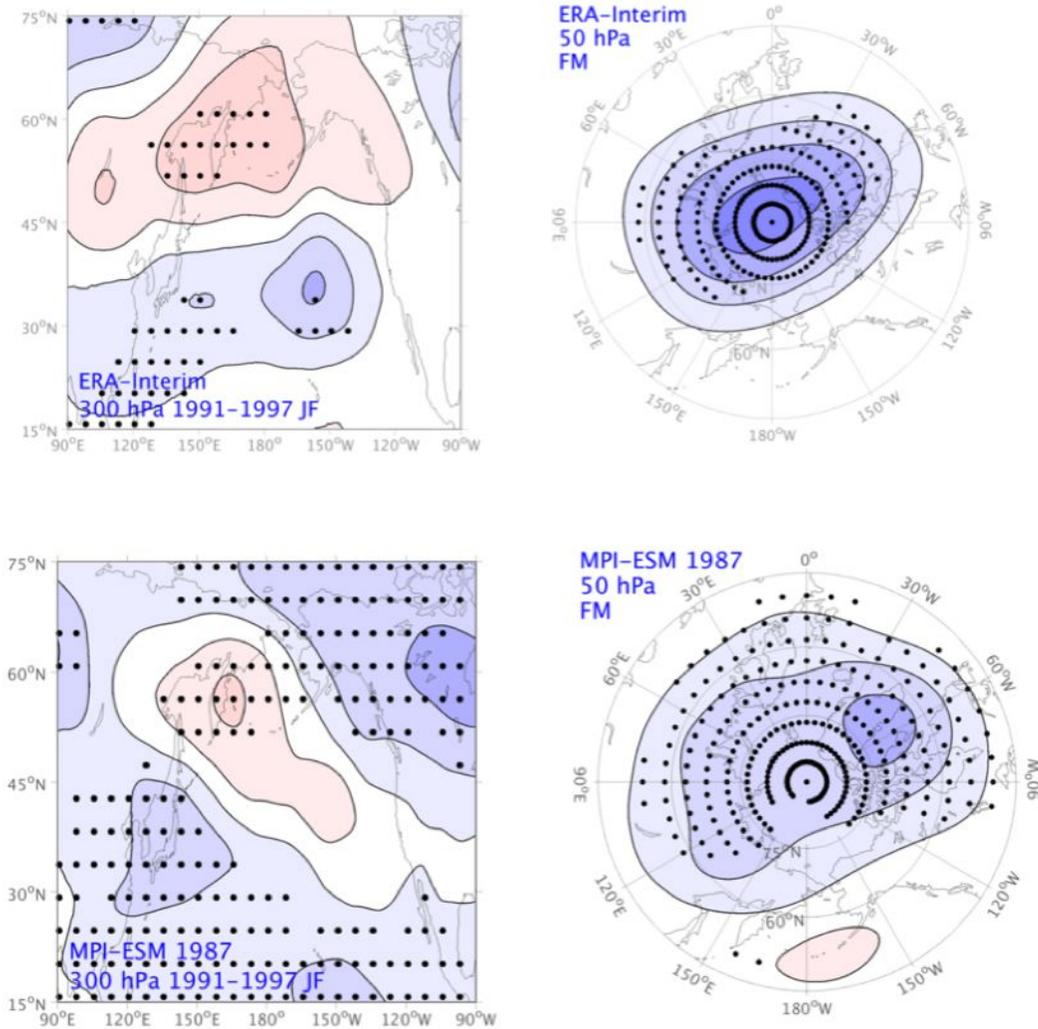
Figure 11: (top) Geopotential height anomalies from ERA-interim reanalysis averaged over 1991-1997, January-February, 300 hPa (left), and February-March, 50 hPa (right). In the left panel, contours start from ±10 m with an interval of 20 m. In the right panel, contours start from ±35 m with an interval of 70 m. Zero lines are omitted. Stippling denotes anomalies that are statistically significant at the p < 0.10 level based on a two-tailed Monte-Carlo test. (bottom) same as (top) but for the ensemble mean of all 20 members of the hindcasts. The simulated geopotential height anomalies are scaled to have the same standard deviation as those in the ERA-interim reanalysis.

The initial conditions associated with the skillful prediction are characterized by a deepened Aleutian Low and a peak El Niño (Figure 12). Initialized with peak El Niño at the end of 1987, the hindcast captures the 1988 January-February (JF) El Niño (0.86 K), negative SST anomalies over the North Pacific (-1.34 K), and negative geopotential height anomalies on the Northern side of the Aleutian low (Figure 12 upper right), as compared with the respective anomalies from reanalysis and observation (Figure 12 upper left). Thereafter, the hindcast predicts the subsequent La Niña in 1989 (-1.08 K) with its teleconnection, the negative PNA (Figure 12, second to top). However, the simulated La Niña takes

longer to decay and the La Niña conditions persist during 1990 (-0.81 K). In 1991, the North Pacific SSTs become above-normal (0.4 K), while La Niña decays, and the Equatorial Pacific is close to ENSO-neutral conditions. Overall, the ensemble hindcasts initialized at the end of 1987 reproduce the observed features reasonably well, including the ENSO phase transition from 1988 to 1989, the negative Pacific North Atlantic pattern from 1989 onward.
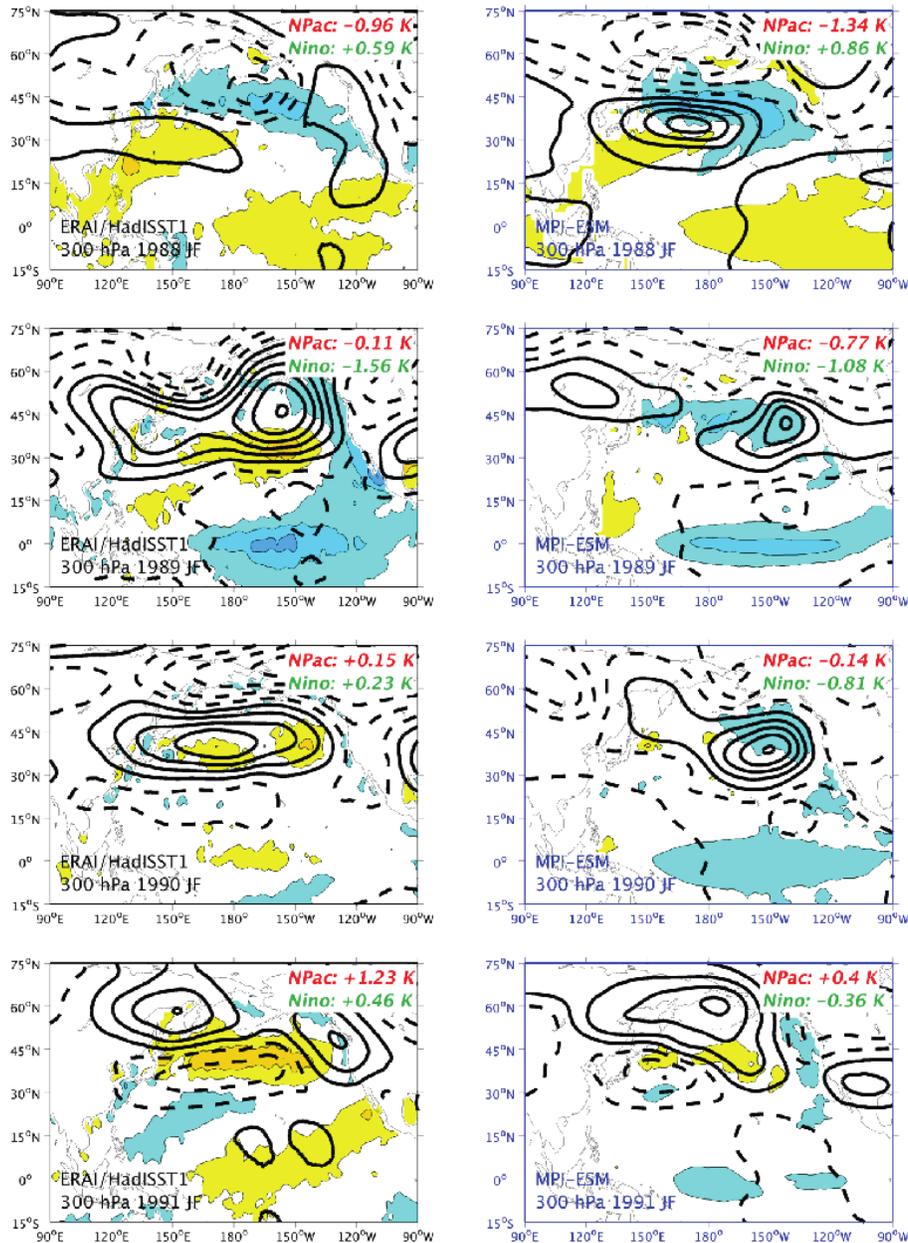

Figure 12: SST anomalies (shading) and geopotential height anomalies at 300 hPa (contours) averaged during winter months (January-February) from 1988 to 1991 for (left panels) the HadISST1 and ERA-interim reanalysis and (right panels) the ensemble mean of all 20 members of hindcasts initialized at the end of 1987. For SST anomalies,

contours start from ±0.4 K with an interval of 0.8 K. For geopotential height anomalies, contours start from ±20 m with an interval of 40 m. Zero lines are omitted. The red and green text at the upper-right corner for each panel denotes the area-averaged SST anomalies over the North Pacific region (40°-50°N, 160°-200°E) and the Niño-3.4 region (5°S-5°N, 120°-170°W). The simulated geopotential height anomalies are scaled to have the same standard deviation as those in the ERA-interim reanalysis.

Predictability of North Atlantic atmospheric anomalies (*CMCC, NCAR, WHOI)*
As continuous efforts on the development of decadal predictions are paying off, skillfully predicting climate anomalies on a multi-annual timescale is now becoming an operational service for many institutions worldwide. At this timescale, predictability originates mostly from the initialization of the ocean, yet the respective atmospheric circulation response (specifically over the midlatitude North Atlantic) had been elusive. Recently, with the availability of large-ensemble simulations, the atmospheric signal has become detectable and shown to exhibit statistically significant skill. This work shows results from the analysis of the Community Earth System Model – Decadal Prediction Large-Ensemble (CESM-DPLE) simulations from NCAR.

Considering the Euro-Atlantic domain, the predictable atmospheric anomalies represent a forced response to oceanic low-frequency variability that strongly resembles the Atlantic Multidecadal Variability (AMV). The latter is correctly reproduced in the CESM-DPLE decadal hindcasts, possibly because of a realistic ocean initialization and ocean dynamics. The occurrence of atmospheric blocking in certain areas of the Euro-Atlantic domain determines the concurrent circulation regime and the phase of known teleconnections, such as the North Atlantic Oscillation (NAO), consequently affecting the storm tracks and the frequency and intensity of extreme weather events. Therefore, skillfully predicting the decadal fluctuations of blocking frequency and the NAO is of importance for deducing conditions favouring or disfavouring societally relevant impacts from extreme weather events.

Figure 13a documents the predictive skill of the 40-member ensemble mean for the number of blocking days in winter occurring in high-latitude North Atlantic, referred to as high-latitude blocking (HLB). The anomaly correlation coefficient (ACC) is shown for averages taken over all possible lead-year ranges, determined by the lead-year (ordinate) and the end hindcast year available of that lead-year (abscissa). Evidently, for HLB blocking the skill is statistically significant over various lead-year ranges, reaching as high as 0.65 for LY[1–8], namely lead-year 1, averaging over 8 years. Clearly, averaging over more years reveal predictability also at higher leads, namely up to leads year 5 and 6. It is noted that the predictive skill for HLB is largely unaffected by linear detrending of the timeseries (not shown). Figure 13c shows the ACC for the NAO index. As for HLB blocking, the skill is statistically significant over various lead-year ranges, reaching as high as 0.63 for LY[2–8] and 0.58 for LY[1–8], which is comparable to that for HLB. The NAO index is defined here after zonal averaging of mean sea level pressure (MSLP) at 35°N and 65°N, between 80°W and 30°E. This definition (Jianping et al., 2003) accounts for the zonal migration of the NAO centers of action. The traditional definition yields a comparable ACC matrix with correlation differences that do not exceed 0.09. Linear detrending of both the observed and the model NAO time series reduces the ACC values (not shown) indicating that MSLP over the North Atlantic exhibits trends associated with external forcing.

The right panels of Figure 13 show how the ACC increases with the ensemble size. Each line in these plots corresponds to a lead-year range (cell) in the respective ACC matrix. The skill increases almost monotonically with the ensemble size, and this strengthens our confidence that the detected skill

relates to a real predictable signal. Furthermore, the skill has clearly not saturated at N=40 (even more so for HLB compared to the NAO), pointing to the potential benefit of further increasing the ensemble size. For both HLB and the NAO, Figure 13 shows also (dashed-dotted lines) the skill of the sub-ensemble mean against a single member of the ensemble (averaged across all possible member permutations). The fact that the CESM-DPLE ensemble mean has more skill in predicting the observed anomalies than predicting any member of the ensemble itself (treated as observations) is a common feature across decadal and seasonal prediction systems alike, particularly referring to the mid-latitude North Atlantic. In terms of physical mechanisms, in the associated study (Athanasiadis et al., NPJ Climate and Atmospheric Science; in review) this predictability is shown to originate from the realistic initialization of the North Atlantic ocean determining the evolution of SST anomalies over the Subpolar Gyre and associated changes in the GSE SST front impacting on the atmospheric circulation.
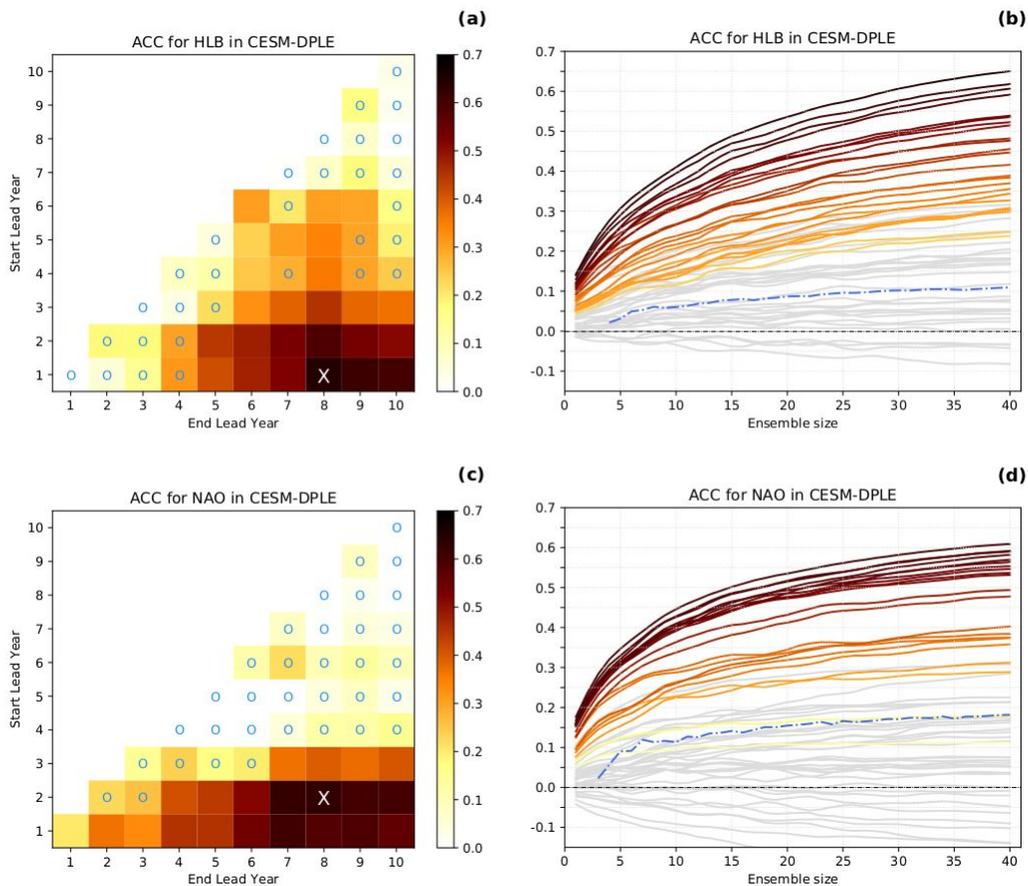
Figure 13: *The predictive skill for the CESM-DPLE ensemble-mean measured by the Anomaly Correlation Coefficient (ACC) for High Latitude Blocking (HLB) in **(a)** and the North Atlantic Oscillation (NAO) in **(c)**. Each cell below the diagonal corresponds to a different lead-year range, over which the ACC is calculated, defined by the lead-year ("Start Lead Year", ordinate) and the hindcast year available for that lead-year ("End Lead Year", abscissa). The markers (o) indicate not statistically significant correlations. In (a) and (c), an X marker indicates the lead-year range with the highest ACC (0.65 for HLB and 0.63 for NAO). In **(b)** and **(d)**, the respective skill is computed as a function of the ensemble size (averaged for all possible member combinations). Each line corresponds to a different lead-year range. Lines in colour correspond to statistically significant correlations for the full ensemble (N=40)*

following the same colour code as in (a) and (b). The dashed-dotted lines show the skill of the sub-ensemble mean against a single member of the ensemble (averaged for all possible combinations).

Trends in North Atlantic ocean subpolar gyre temperatures (DMI)

Large decadal variability is found in the Atlantic ocean subpolar gyre (SPG) temperature in reanalysis data sets and models, both in the unforced (piControl) and the climate change (hist+rcp8.5) experiments. Figure 14 illustrates the characteristics of the 15-year linear trend in the SPG surface temperature from the ocean reanalysis EN4, and from IPSL-CM6A-LR model for the unforced piControl experiment and from the EC-Earth model for the hist+rcp8. experiments. A period of 15 year is used here to calculate the decadal trend of the SPG surface temperature. However the choice of the length does not change the characteristics of the decadal trend. For instance, a linear trend of 12 years or 20 years shows similar statistics as the 15-year trend. It can be seen that warming and cooling trends of decadal timescales comparable with the observed amplitudes occur frequently and alternatively, in both unforced and climate change experiments, up to the historical period. The extreme strong negative trend observed over 2004-2018 (indicated in the red line) may occur as an extreme event in the piControl and the historical period when the anthropogenic forcing is not very strong (Figure 14b, c). But as the climate warms, such a strong negative trend becomes very rare (Figure 11d). It is interesting to note that alternative appearance of warming and cooling trends seems to have a decadal frequency of around 10-30 years.
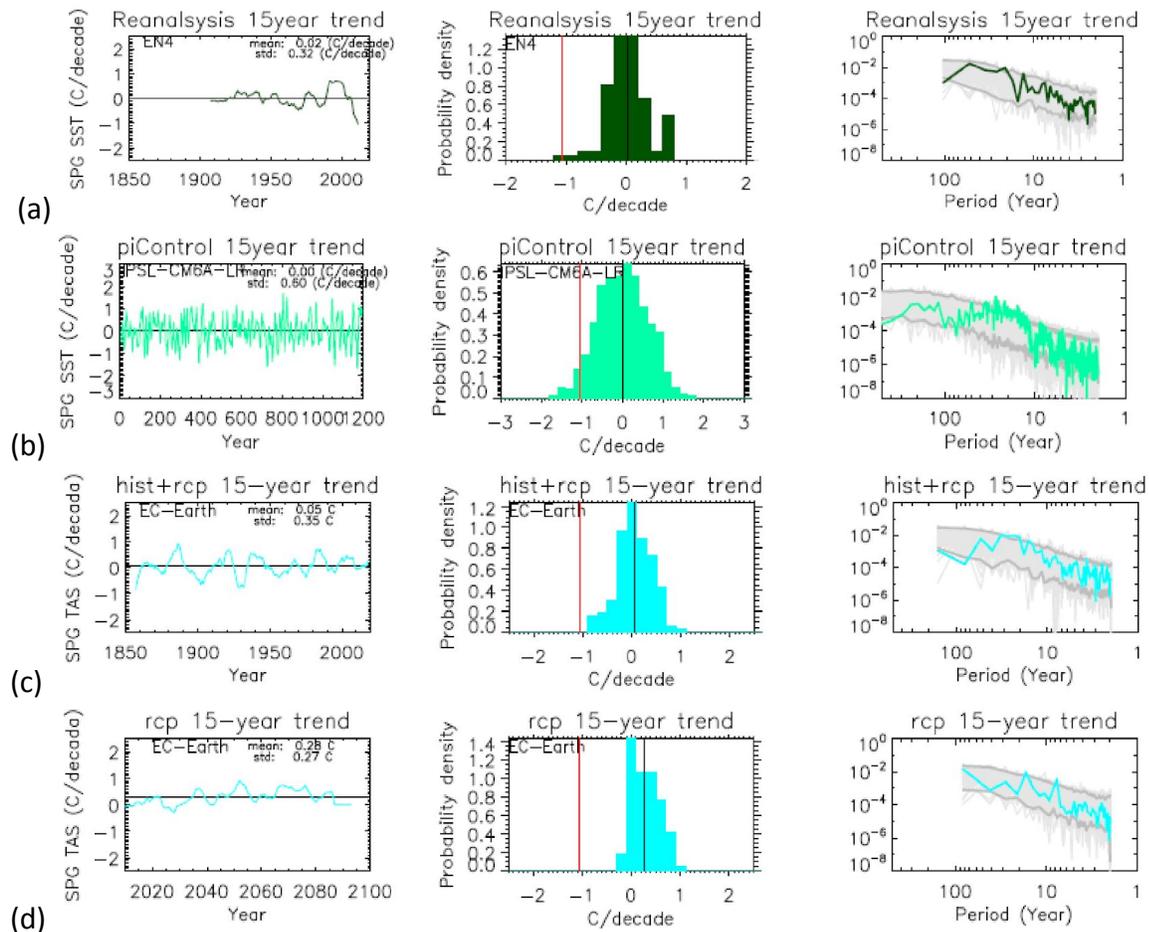
Figure 14: Time series of the 15-year trend of the SPG surface temperature (left panels, unit: °C/decade), the corresponding histogram (middle panels, unit: %) and power spectrum (right panels) for (a) the ocean analysis EN4 dataset; (b) the CMIP6 piControl by IPSL-CM6A-LR model; (c) the CMIP5 historical experiment and (d) RCP8.5 scenario experiment by the EC-Earth model. The vertical red line on the histograms (middle panels) indicates the observed 15-year trend of SPG SST for the period 2004-2018, which is the strongest negative trend observed for the region since 1900 in the EN4 data record. The shading and grey lines on the power spectrum (right panels) indicate the significance between 0.05 and 0.95 levels calculated based on surrogate data.

Thereafter, we have extended the analysis to all available CMIP5 models (Figure 15). The SPG temperature will continue to increase as global warming continues. However, most CMIP5 models simulate that the warming in the SPG region will be less than the increase of global mean temperature under the RCP8.5 scenario, implying that the region may continue to be a "warming hole". Across all models, the decadal trends become less variable and move towards more positive trend in the future. But extreme cooling events of similar strength as the recent observed one may still be possible, up to a warming of about 6°C.
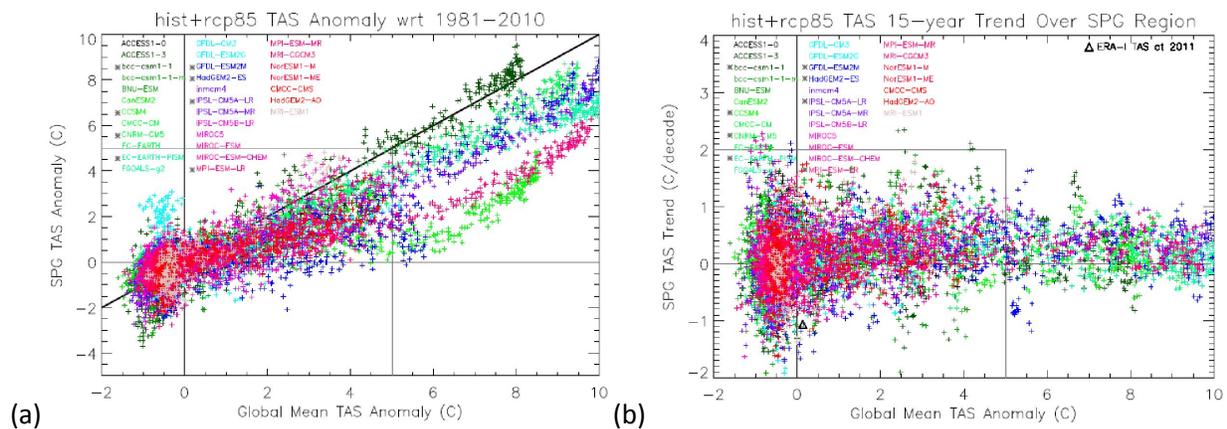

Figure 15: Scatter plots of (a) the annual mean SPG surface temperature changes (unit: °C) and (b) the corresponding 15-year trends (unit: °C/decade) against the global mean temperature changes as simulated by 31 CMIP5 models. All changes are respect to the time average of 1981 to 2010. Each dot represents an annual mean value.

New prediction systems and analysis techniques (CNRS/IPSL-CNRS/EPOC)
A new prediction system, with an improved inclusion of salinity in the initial conditions and the use of the new IPSL-CM6 model has been developed by CNRS/IPSL-EPOC. Results will be disseminated soon through CMIP6 database (DCCP-A). To further analyzed the prediction skill over Europe from prediction systems, a method has been developed to systematically analyze the best skill scores for temperature. This is obtained by systematically varying the seasons of average, the lead time periods and the time windows of the analysis to potentially identify some useful windows of opportunity. This leads to a three-dimension field of skill score with the different combinations possible (Figure 16). The systematic analysis of skill applied to IPSL-CM5 decadal prediction system highlights that for Europe, the summer season is the most skillful season for decadal prediction. The best lead time is 1-9 years. It also showed that the best windows of opportunity for predictions is 1981-2000 (1-6 years for May-September when computed over the whole period 1961-2013, cf. Figure 16). This period is including the abrupt warming

in subpolar gyre SST in 1995, which may explain the increase in skill due to the relatively correct prediction of this event.
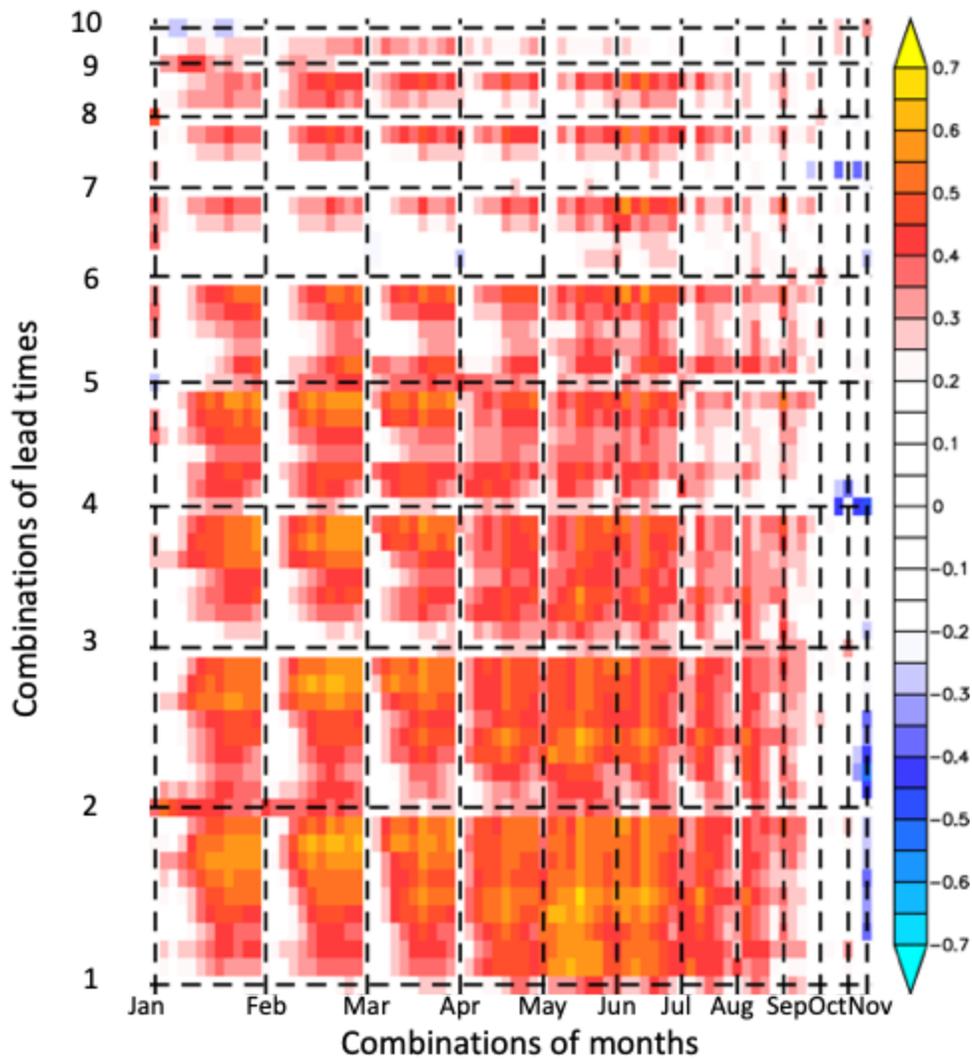
Figure 16: *Matrix of the ACC skill scores for 2-meter temperature averaged over Europe simulated by the IPSL-CM5 decadal prediction system, computed using time series including one start date every year over the period 1961-2013. The x-axis is showing the different possible combinations of months and the y-axis the different possible combinations of lead time averaging.*

Two different sources of climate information at the daily time scale have been used for establishing a debiasing technique: A high resolution regional reanalysis from EU WATCH project for Europe that provides a kind of statistical downscaling and the global-scale 20CR reanalysis that provides a longer time frame. This debiasing technique is based on the quantile-quantile approach and uses the cumulative distribution function - transform (CDF-t). This technique aims at debiasing the climate predictions by using climate reconstruction and has been implemented in collaboration with the private company The Climate Data Factory (at no cost). Some preliminary results are presented in Figure 17

showing the ACC skill scores for lead time 1-3 years. It shows that the debiasing is providing some additional significant skills in Central Europe. Some further improvements and refinement of the statistical methods are under discussion that should hopefully further enhance skill. This bias-corrected and statistically downscaled prediction system using WATCH (10 km resolution) are also in the process of being used to assess the potential for any predictability in the phenological stages of grape vine by comparing the prediction of the phenological stages with a few observations available within France from the CLIMOWINE database.
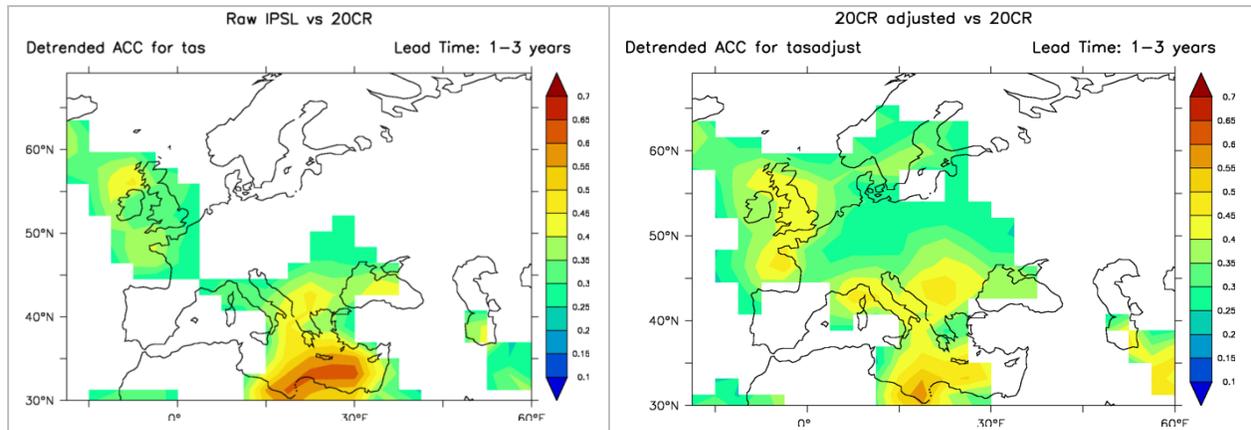

Figure 17: *Anomalies correlation (ACC) scores for 2-meter temperature in the atmosphere for lead-time 1-3 years for the IPSL-CM5A decadal prediction system as compared to 20CR reanalysis. On the left is the ACC using the raw data of the decadal prediction system (starting dates from 1960 to 2015) and on the left is bias-corrected data using the CDF-t method.*

Forecast calibration techniques: Extended range Arctic sea forecast (NLeSC)

Operational Arctic sea ice forecasts are of crucial importance to commercial and scientific activities in the Arctic region. Currently, numerical climate models, including General Circulation Models (GCMs) and regional climate models, are widely used to generate the Arctic sea ice predictions at weather time-scales (Metzger et al. 2014, Hebert et al. 2015, Smith et al. 2016). However, these numerical climate models require near real-time input of weather conditions to assure the quality of the predictions and these are hard to obtain and the simulations are computationally expensive. In this study, we propose a deep learning approach to forecasts of Arctic sea ice at weather time scales.

We focus on the prediction of the sea ice in the Barents Sea at weekly to monthly time scales. To work with such spatial-temporal sequence problems, Convolutional Long-Short Term Memory Networks (ConvLSTM) are useful (Xingjian et al. 2015). ConvLSTM are LSTM (Long-Short Term Memory) networks with convolutional cells embedded in the LSTM cells. This approach is unsupervised learning and it can make use of enormous amounts of historical records of weather and climate. The structure of ConvLSTM is illustrated in Figure 18. Previous studies have shown that ConvLSTM is suitable for many weather prediction problems (Xingjian et al. 2015; Kim et al. 2017 and 2019). More information about ConvLSTM is provided by Xingjian et al. (2015).
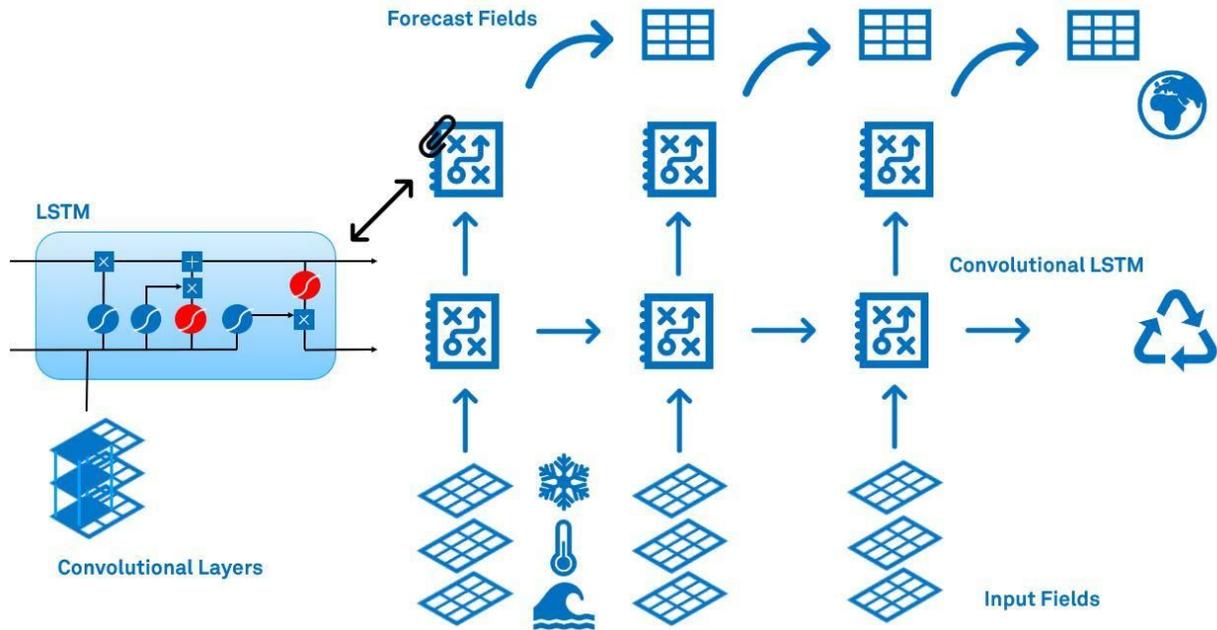
Figure 18: *Structure of the convolutional long-short term memory neural networks.*

We use input fields from atmospheric (ERA-Interim) and oceanic (ORAS4) reanalysis data sets to examine the capability of ConvLSTM to sea ice prediction. The sea ice fields are also taken from ERA-Interim, with a resolution of 0.75 x 0.75 degree (about 30km in the Barents Sea). We performed potential lead time dependent sea ice predictions with multiple fields from 1979 to 2016, including sea ice concentration (SIC), ocean heat content (OHC) , 2-meter temperature (T2M), sea level pressure (SLP), geopotential height at 500hPa (Z500) and 850 hPa (Z850), net surface turbulent and radiation fluxes (SFlux), and 10-meter meridional (V10m) and zonal (U10m) winds, which are averaged to weekly means. Different combinations of input fields are tested to evaluate the contribution from different predictors. The ConvLSTM network is trained with data from 1979 – 2008 (30 years). Each field contains 24 x 56 points and the training takes about 8-12 hours with GTX-1060 GPU card, depending on the size of the network and the input fields. Data from 2009 to 2012 is taken for cross-validation and predictions are made from 2013 to 2016. We predict sea ice 6 weeks ahead. The results are evaluated against persistence and climatology.

An overview of the predictions with different input fields with ConvLSTM is shown in Figure 19. The root mean square error (RMSE) of lead time dependent predictions are illustrated together with those from climatology and persistence. The predicting errors increase with the prediction time step. It is found that predictions made by ConvLSTM with SIC and one extra variable are generally better than those by persistence and climatology, except for ConvLSTM predictions with SIC and T2M or Z850. For the first week, the performances are very similar among different methods, including persistence. Starting from week 3, predictions with SIC and one extra variable become much better than persistence and it is prevalent for the following weeks. This indicates that some of the chosen variables can provide predictability to sea ice prediction at weekly to sub-monthly time scales. By adding more data to ConvLSTM network, the quality of prediction can be improved significantly (see prediction with SIC, OHC, T2M, SLP, Z500, Z850, SFlux and UV10m with ConvLSTM in Figure 19). It should be noted that

predictions with SIC and Z850 or T2M are worse than those with only SIC using ConvLSTM. This means the extra input data adds "noise" to the network and somewhat reduces the performance of ConvLSTM.
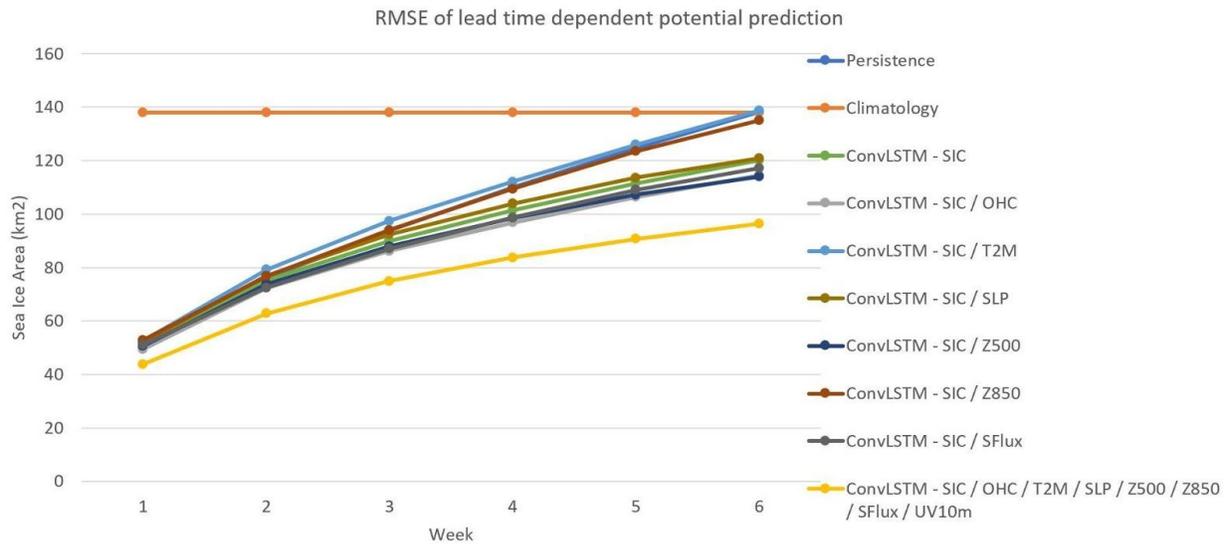
Figure 19: RMSE of potential lead time prediction of sea ice area in the Barents Sea with different combination of input fields. The results are compared with those from persistence and climatology.

The results with different combinations of input fields can also be treated as sensitivity tests on the predictors since deep neural networks are able to account for the non-linear relations between variables. Given the RMSE from predictions made by ConvLSTM with SIC and one extra field in Figure 19, it can be observed that SFlux, OHC and Z500 add more skills to the predictions, while SLP, Z850, and T2M cannot contribute to sea ice predictions at the chosen time scales.

Further evaluation of the predictions with ConvLSTM in each month is provided in Figure 20. It includes mean square error (MAE) of 1st week predictions with SIC and OHC using ConvLSTM in each month. ConvLSTM performs well in summer and the transition times between summer and winter. However, in winter the skills of ConvLSTM drop down, especially for the regions near the Eurasian continent. Qualitatively, the results with ConvLSTM compare well with the predictions made by the state-of-the-art operational models (Metzger et al. 2014, Hebert et al. 2015, Smith et al. 2016).
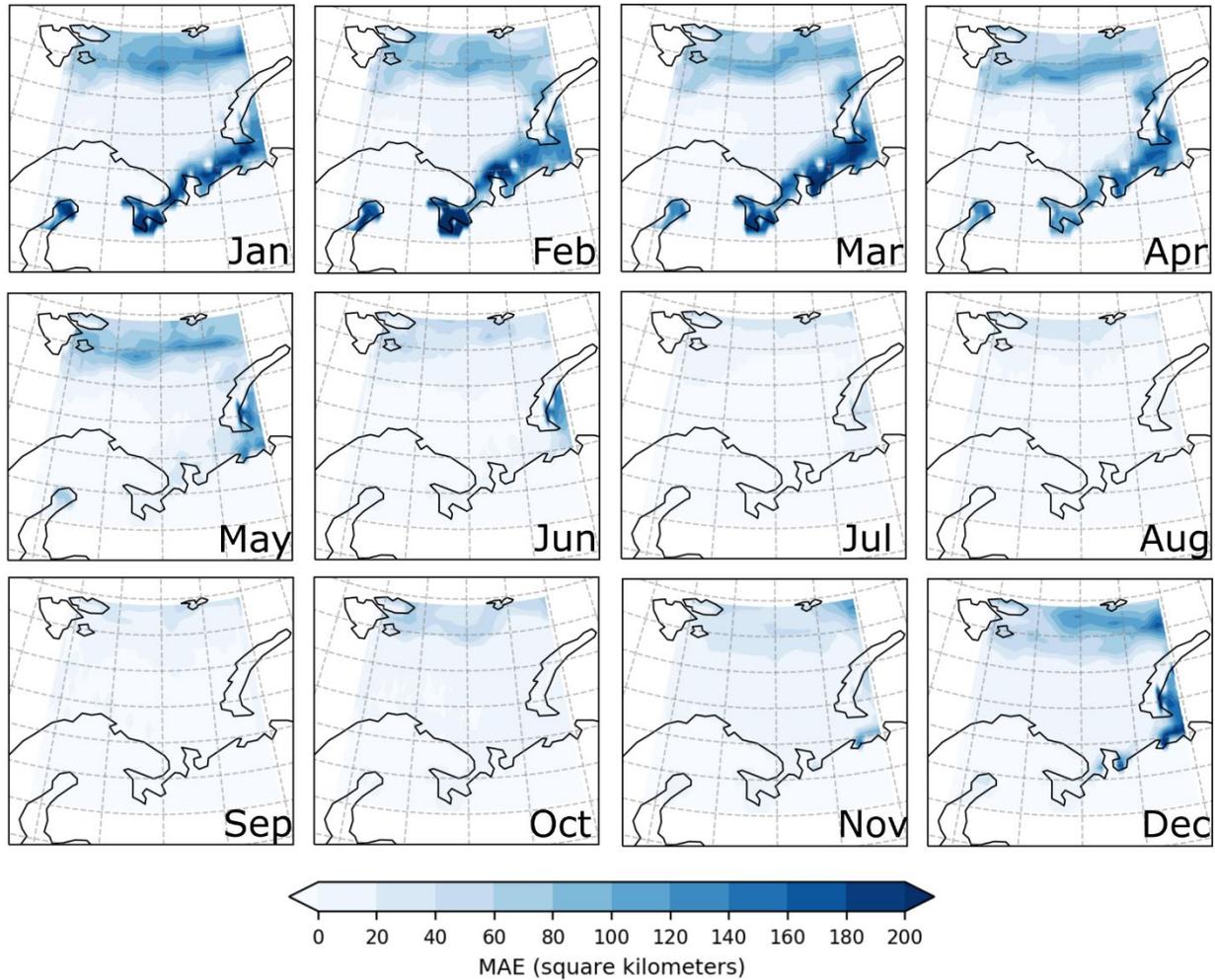
Figure 20: *Mean square error (MAE) of 1st week predictions of sea ice area, with SIC and OHC predictors and using ConvLSTM for each calendar month.*

Main results achieved

Attribution of predictive skill of North Atlantic upper-ocean salt content

Recently, decadal scale predictability of upper ocean salinity in the North Atlantic has been found, providing further evidence to the predictive skill of properties of the North Atlantic Ocean. Upper ocean salinity predictability is of importance because it may lead to marine ecosystem prediction. Here, the examination of the origin of the upper ocean salinity predictability has demonstrated that the predictive skill of upper ocean salinity in the North Atlantic resides in the ocean circulation. Specifically, the predictability at forecast times from about five years onwards of the upper ocean salinity can be attributed to the initial strength of the North Atlantic subpolar gyre, for the entire subpolar North

Atlantic and the Eastern Nordic Seas. In the eastern subpolar North Atlantic, predictive skill at 7-10 years may also arise from the initial strength of the Atlantic meridional overturning circulation.

Predictability of the North Atlantic Cold Blob

High predictive skill on seasonal-to-decadal timescales generally exists in the subpolar North Atlantic sea surface temperature. A question is however, if predictive skill remains high also for extreme ocean events. Here we have used the near-record low subpolar North Atlantic sea surface temperature observed in 2015 (the "cold blob") as a case study to test the ability of the project seasonal-to-decadal prediction systems to indeed predict cold subpolar extremes. Model output was gathered from the project five prediction systems. We have found that for a nine-month forecast, only two of the prediction systems predicted a weak, coherent cold blob, and none of the prediction systems captured the full magnitude of the observed cold blob within their ensemble. Preliminary results of the analysis of the multi-model predictions indicate that this prediction failure is possibly linked not only to a deficiency in the atmospheric circulation's evolution, namely, too weak surface winds associated with the failure to predict the phase and strength of the North Atlantic Oscillation, but also to upper ocean initialization. Further examination of this case could point the way towards improvements in the initialization and model structure of the next generation of decadal prediction systems.

Propagation of Thermohaline Anomalies and their predictive potential

Observations have shown that there exists a lagged correlation for SST anomalies along the Atlantic water pathway, where water flows poleward, from the subpolar North Atlantic to the Nordic Seas. Specifically, warm and cold anomalies propagate along the Atlantic water pathway, with a certain time lag, and this occurred repeatedly over the last 60 years. The northward advections with time of SST anomalies may therefore provide for a mechanism for the predictability of the state of the Atlantic Ocean along the Atlantic water pathway. To this end, we assess to what extent the six decadal prediction systems available with the project can predict the winter SST along the Atlantic water pathway in the time period 1960-2010. We find that most models have difficulties in representing this mechanism, but for long lead times (after a few years) some models represent the mechanism to some extent. The reason for this might be that predictions on long lead times allow the time for correctly initialized SST anomalies in the subtropics to be advected northwards. The propagation mechanism appears to be a source for skill, but with large model differences in location and timing of the SST anomalies. To test the robustness of the mechanism, we plan to perform the same analysis for subsurface properties in the core of the Atlantic Ocean water.

Predictability of the Rapid Warming of the North Pacific Ocean around 1990

The North Pacific Ocean underwent a quasi-decadal shift from a prolonged cold state in the 1980s to a warm state in the early-mid 1990s. This shift was initiated by a strong and rapid warming of the North Pacific Ocean around 1990, with SSTs increasing by $2^{\circ}C$ from 1988 to 1991. In the decade thereafter, sustained SST warm anomalies persisted for several years and atmospheric anomalies were characterized by a weakened Aleutian Low, a strengthened stratospheric polar vortex, and a lack of sudden stratospheric warming. Using the MPI-ESM decadal prediction system we demonstrate predictive skill of the rapid warming of 1991 and the following ocean and atmospheric anomalies. The source of prediction skill lays in the initialization with peak El Niño and the subsequent predictions of La

Niña and its associated teleconnections. The SST development over North Pacific may be indeed be understood as resulting from the impact of the evolving atmospheric teleconnection based on the effects of ocean-atmosphere interactions: the negative Pacific North American anomaly can induce surface easterly wind anomalies, through diminishing the climatological mid-latitude westerlies, lessening heat fluxes at the ocean surface and weakening vertical mixing in the upper ocean, resulting in warmer SSTs over North Pacific Ocean.

Predictability of North Atlantic atmospheric anomalies

The number of atmospheric blocking winter day and the phase of the North Atlantic Oscillation are dominant features of the mid latitude atmospheric circulation, affecting the storm tracks and the frequency and intensity of extreme weather events. Here the 40-member hindcasts from the Community Earth System Model – Decadal Prediction Large-Ensemble (CESM-DPLE) simulations from NCAR are analysed for exploring the predictive skill of these atmospheric anomalies. The systematic analysis of prediction skill using anomalies averaged across all possible combinations of lead-year has revealed significant skills over various lead-year ranges, for both atmospheric blocking and the NAO. Specifically, high correlations are found when averaging over more years. For instance, the highest correlation for the number of winter blocking days is found at lead year 1 averaging over 8 years, and significant correlations are also found at higher leads, namely up to leads year 5 and 6. The NAO shows multi-year significant correlation skill up to lead year 3.

Trends in North Atlantic subpolar gyre temperatures

To understand what may drive abrupt North Atlantic subpolar gyre SPG changes, such as the extreme strong negative 2004-2018 trend in SPG surface temperatures, ocean reanalysis and model outputs are analyzed and intercompared. Specifically, the ocean reanalysis EN4 datasets, preindustrial control, historical and future scenario experiments are considered. The characteristics of the temperature changes from these reanalyses and the model experiments, including the decadal trends, their anomaly distributions and power spectra are assessed. Large decadal variability is found for both the reanalysis data sets. Comparable variations are also found from the preindustrial control and the historical experiments. However, into the future, extreme strong negative trends comparable to the 2004-2018 trend are found only up to a global warming of about 6°C.

New prediction systems and analysis techniques

A new prediction system, a method to systematically analyze the best skill scores for temperature, and a new debiasing technique have been developed, making use of the IPSL-CM decadal prediction systems and by collecting different sources of climate information. Specifically, the systematic analysis of the best skill scores for temperature has shown that the summer season is the most skillful season for decadal prediction, the best lead time is 1-9 years, and that the best windows of opportunity for predictions is 1981-2000. The debiasing technique is providing some additional significant skills for 2-m temperature skill for lead-time 1-3 years in Central Europe.

Forecast calibration techniques: Extended range Arctic sea forecast

To advance the operational prediction of the sea ice in the Barents Sea at weather to monthly time scales, a deep learning approach to forecast Arctic sea ice at weather time scales is proposed. The deep learning method makes use of Convolutional Long-Short Term Memory Networks (ConvLSTM). With the input fields from atmospheric and oceanic reanalysis data sets (training period: 1979 – 2008), it is found that ConvLSTM is able to learn the variability of the Arctic sea ice within historical records and effectively predict (from 2013 to 2016) regional sea ice concentration patterns at weekly to monthly time scales. Based on the known sources of predictability, sensitivity tests with different climate fields were also performed by making predictions with sea ice concentrations and other predictors using ConvLSTM. The influences of the different predictors on the quality of predictions have been evaluated with the help of ConvLSTM. This method outperforms predictions with climatology and persistence at chosen time scales and is promising to act as a fast and cost-efficient operational sea ice forecast system in the future.

Progress beyond the state of the art

WP4 has been significantly advancing the climate predictability research in several areas:

- In this work, we move beyond standard practices for statistical estimating skill and apply "skill attribution" to demonstrate predictive skill by investigating reproduction of underlying mechanisms in dynamical predictions. This novel process-oriented skill assessment has been combined with the more traditional skill assessments to assess predictability in North-Atlantic and Arctic sectors.
- To gain new insights, we have analysed state-of-the-art decadal climate prediction systems that are based on a mix of post CMIP5 and CMIP6 Earth System Model versions, as well an ensemble of modern reanalysis.
- We have performed the first multi-prediction system analysis to show predictability of the recent record-low North Atlantic "Cold Blob" to show importance of both atmospheric conditions and ocean state. The "Cold Blob" case study is a challenging test of seasonal-to-decadal prediction systems given the extreme nature of this anomaly. Continued analysis of this case will provide direction for improvements to next-generation climate prediction systems.
- We have also demonstrated for the first time multi-year prediction skill for North Atlantic blocking. This is a great step forward in long-term predictions of extreme weather statistics.
- We have moved towards a seamless evaluation of predictive skill by investigating how predictable the frequency of Sudden Stratospheric Warming events (a typical subseasonal-to-seasonal process) is in a large ensemble of single model initialised decadal predictions.
- Innovative forecast calibration techniques involving deep learning approaches have been developed for enhancing the current Arctic sea ice predictive capabilities at sub-seasonal time scales.
- The systematic analysis of skill on the three dimensions months, lead time and windows of time has been applied for the first time to a decadal prediction system and is providing a new methodology that improves our knowledge on the best period of prediction.
- The debiasing technique based on quantile-quantile is showing some improvements for the predictive skill over Europe. Furthermore, it provides a kind of downscaling approach that could

be useful for impacts studies and to produce data that can be potentially disseminated to the business sector thanks to the collaboration with the private company The Climate Data Factory.

Impact

The research in Task 4.1 contributing to D4.2 has made contributions to several expected impacts of the Blue-Action project.

We have improved the capacity to predict the weather and climate of the Northern Hemisphere, and make it possible to better forecast of extreme weather phenomena

- By identifying specific factors (inaccuracy in initial conditions, systematic model biases) that might limit the predictability of high-impact climate extreme events, such as the North Atlantic Cold Blob during summer 2015, in current generation of prediction systems.
- By demonstrating decadal predictive skill of impact-relevant climate quantities when employing large ensemble of initialised predictions
- By developing innovative forecast calibration techniques that could lead to an enhancement in the predictive skill over the Arctic.

Improve the capacity of climate models to represent Arctic warming and its impact on regional and global atmospheric and oceanic circulation

By performing a mechanistic assessment of key processes responsible for the Arctic-lower latitude predictive linkages and their representation in the current state-of-art prediction systems with a particular focus on the propagation of thermohaline anomalies along the North Atlantic current towards the Nordic Seas and Arctic.

Contribute to better servicing the economic sectors that rely on improved forecasting capacity

The demonstrated decadal-scale predictive skill of North Atlantic upper-ocean salt content is of interest to fisheries and marine conservation as upper-ocean salinity in the north-eastern North Atlantic has been suggested to be an indicator of abundance and distribution of marine ecosystem species.

Improve the professional skills and competences for those working and being trained to work within this subject area

WP4 early career scientists have attended specialised training events (e.g. YOPP-APEC Polar Prediction School 2018; MPG Professional Communication in Science) that helped improved both their professional skills and communication competence.

Impact on the business sector

The predictive capabilities of the benchmark predictions have been successfully integrated in the first pilot ecosystem predictions developed within WP5. This expands the climate services sector to fishing industry.

Lessons learned and Links built

We have shown that the well-documented decadal predictability of North Atlantic sea surface temperature leads to a significant predictable component in the midlatitude tropospheric circulation (e.g. assessed through blocking). Thus through ocean-atmosphere interaction skill may extend beyond near-surface atmospheric variables, such as T2M and MSLP shown previously [e.g. Bellucci et al., 2014; Smith et al., 2019], However, further model developments to increase the unrealistically low signal-to-noise ratio is needed to achieve more skillful decadal climate predictions and services. Next generation climate models with increased ocean (eddy-resolving) and atmospheric resolution developed within the H2020 PRIMAVERA and CMIP6 HighResMIP projects promise improved predictability by more realistically representing coupled processes. Pilot high resolution decadal predictions going beyond the CMIP6 DCPP setups are currently being tested by several partners and will be reported within the D4.5 deliverable. The power of large ensemble of multi-model predictions for achieving significant multi-year to decadal predictability of several impact relevant oceanic and atmospheric processes has also been demonstrated. In addition, we have shown the need for mechanistic assessments of prediction skill. We have also shown that deep-learning techniques can be useful for sub-seasonal predictions; these could prove useful for longer term predictions or as forecast calibration technique.

Contribution to the top level objectives of Blue-Action

This deliverable contributes to the achievement of all the objectives and specific goals indicated in the Description of the Action, part B, Section 1.1

Objective 1 Improving long range forecast skill for hazardous weather and climate events

The study on the Predictability of the Rapid Warming of the North Pacific Ocean around 1990 and its associated atmospheric impacts demonstrated potential for skillful predictions of the decreased frequency of sudden stratospheric warming events in the 1990s. The onset of major SSWs has significant follow-on effects on surface weather and climate in the Northern Hemisphere extratropical regions.

Objective 2 Enhancing the predictive capacity beyond seasons in the Arctic and the Northern Hemisphere

WP4 multiyear-to-decadal predictability studies had a strong focus on the attribution of drivers responsible for achieving predictability/limiting the predictive skill of impact-relevant climate quantities over the Arctic and Northern Hemisphere. This involved comparing processing and mechanisms responsible for predictability in both the observations and state-of-art climate prediction systems. By employing very large ensemble of initialised predictions we have for the first time demonstrated significant decadal predictive skill of key features of atmospheric variability over the North Atlantic region. Innovative forecast calibration techniques involving deep learning approaches have been developed for enhancing the current Arctic sea ice predictive capabilities at sub-seasonal time scales.

Objective 5 Optimizing observational systems for predictions

The coordinated multi-model analyses of ensembles of initialised decadal hindcasts have also evaluated the impact of observational data (component - ocean, atmosphere, sea ice, and spatial coverage - surface, 3D) used to initialise the predictions, particularly with respect to a more accurate representation of initial conditions over the North Atlantic-Arctic Sector.

Objective 6 Reducing and evaluating the uncertainty in prediction systems

Several oceanic and atmospheric reanalyses have been employed to estimate the current predictive skill potential and limitations over the North Atlantic-Arctic Sector. Both multi-model and single model large ensemble of non- and initialised climate predictions have been used to better estimate and reduce the uncertainties in predictions.

Objective 8 Transferring knowledge to a wide range of interested key stakeholders

WP4 leading scientists have taken part in the Blue-Action engagement event in Edinburgh to discuss how the ocean predictions can be used to respond to the climate emergency in Scotland.

References (Bibliography)

Allan, R. and T. Ansell (2006) A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004. *J. Climate,* **19**, 5816–5842, https://doi.org/10.1175/JCLI3937.1

Baldwin, M., and T. Dunkerton (2001) Stratospheric harbingers of anomalous weather regimes. Science, 294, 581-584.

Counillon, F., N. Keenlyside, I. Bethke, Y. Wang, S. Billeau, M. Lin Shen & M. Bentsen (2016) Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model, Tellus A: Dynamic Meteorology and Oceanography, 68:1, DOI: 10.3402/tellusa.v68.32437

Dee, D. P. and Coauthors (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q.J.R. Meteorol. Soc., 137: 553-597. doi:10.1002/qj.828.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M. and Knight, J. (2016) Skilful predictions of the winter North Atlantic oscillation one year ahead. Nature Geoscience, 9, 809–814.

Duchez, A., Frajka-Williams, E., Josey, S.A., Evans, D.G., Grist, J.P., Marsh, R., McCarthy, G.D., Sinha, B., Berry, D.I. and Hirschi, J.J. (2016) Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave. Environm. Res. Lett., 11, 074004.

Garfinkel, C.I., D.L. Hartmann (2010), and F. Sassi, Tropospheric Precursors of Anomalous Northern Hemisphere Stratospheric Polar Vortices, J Climate 23, 3282-3299, DOI: 10.1175/2010JCLI3010.1

Goddard, L., and Coauthors (2013) A verification framework for interannual-to-decadal predictions experiments. Climate Dynamics, 40, 245–272, https://doi.org/10.1007/s00382-012-1481-2.

Hebert, D. A., R. A. Allard, E. J. 171 Metzger, P. G. Posey, R. H. Preller, A. J. Wallcraft, M. W.Phelps, and O. M. Smedstad (2015) Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the us navy's arctic cap nowcast/forecast system. Journal of Geophysical Research: Oceans, 120 (12), 8327–8345.

Huang, B. and Coauthors (2017) NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5. NOAA National Centers for Environmental Information. doi:10.7289/V5T72FNM.

Hurwitz, M. M., P. A. Newman, and C. I. Garfinkel, 2012: On the influence of North Pacific sea surface temperature on the Arctic winter climate. J. Geophys. Res., 117, D19110.

Jianping L., Wang J.X.L. (2003) A New North Atlantic Oscillation Index and Its Variability. Advances in Atmospheric Sciences,20, 661–676.

Kim, S., S. Hong, M. Joh, and S.-K. Song (2017) Deeprain: ConvLSTM network for precipitation prediction using multichannel radar data. arXiv preprint arXiv:1711.02316.

Kim, S., H. Kim, J. Lee, S. Yoon, S. E. Kahou, K. Kashinath, and M. Prabhat (2019) Deep hurricane-tracker: Tracking and forecasting extreme climate events. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 1761–1769.

Langehaug H R, Matei D, Eldevik T, Lohmann K, Gao YG (2017) On model differences and skill in predicting sea surface temperature in the Nordic and Barents Seas. Climate Dynamics, 48, 913-933

Matei D, Pohlmann H, Jungclaus JH, Müller WA, Haak H, Marotzke J (2012) Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. Journal of Climate, 25, 8502–8523

Mecking, J.V., S.S. Drijfhout, J.J.-M. Hirschi, and A. Blaker (2019) Ocean and atmosphere influence on the 2015 European Heatwave. Env. Res. Letts., https://doi.org/ 10.1088/1748-9326/ab4d33.

Metzger, E. J., and Coauthors (2014) US navy operational global ocean and arctic ice prediction systems. Oceanography, 27 (3), 32–43.

Miesner AK, Payne MR (2018) Oceanographic variability shapes the spawning distribution of blue whiting (Micromesistius poutassou). Fisheries Oceanography, minor revision stage.

Smith, G. C., and Coauthors (2016) Sea ice forecast verification in the canadian global ice ocean prediction system. Quarterly Journal of the Royal Meteorological Society, 142 (695), 659–671.

Swingedouw, D., Mignot, J., Labetoulle, S. et al. (2013) Initialisation and predictability of the AMOC over the last 50 years in a climate model. *Climate Dynamics* 40, 2381-2399. DOI: 10.1007/s00382-012-1516-8.

Wang, Y., F. Counillon, I. Bethke, N. Keenlyside, M. Bocquet, M-l. Shen (2017) Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. Ocean Modelling 114, 33-44

Xingjian, S., Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 802–810.

Yeager, S.G., G. Danabasoglu, N.A. Rosenbloom, W. Strand, S.C. Bates, G.A. Meehl, A.R. Karspeck, K. Lindsay, M.C. Long, H. Teng, and N.S. Lovenduski (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

Dissemination and exploitation of Blue-Action results

Dissemination activities

Type of dissemination activity	Name of the scientist (institution), title of the presentation, event	Place and date of the event	Type of Audience	Estimated number of persons reached	Link to Zenodo upload
Participation to a workshop	Daniela Matei (MPI-M), Title: Decadal-scale predictive skill of North Atlantic upper-ocean salt content and its attribution to the initialization of the North Atlantic Ocean circulation	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	60	https://zenodo.org/record/3268766#.XcrGeoVelds
Participation to conference	Ying Dai (MPI-M):Predictability of the Rapid Warming of the North Pacific Ocean around 1990: Potential Attribution and Impacts, at the 27th IUGG General Assembly	Montreal (CAN), 8-18 July 2019,	Scientific Community (higher education, Research)	5000	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a workshop	Helene R. Langehaug (NERSC), 2-slides on "Predictive skill along the extension of the North Atlantic Current in the Norwegian Climate Prediction	Barcelona (ES), 3 July 2018	Scientific Community (higher education, Research)	30	https://zenodo.org/record/1434130#.W6i-Ki2B1TY

Participation to a workshop	Helene R. Langehaug (NERSC), Assessing poleward propagation of temperature anomalies in decadal hindcast experiments at the Joint workshop in Bergen, organised by the Bjerknes Climate Prediction Unit and the EU Climate Modelling Cluster	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	50-60	https://doi.org/10.5281/zenodo.3251800
Communication campaign	Helene R. Langehaug (NERSC), The push for predicting the future climate, Podcast	Bergen (NO), 16 August 2019	General Public	Downloaded 134 times	https://bjerknessenteret.podbean.com/e/the-push-for-predicting-the-future-climate/
Participation to a conference	Shuting Yang (DMI), title: The recent abrupt cooling over Subpolar North Atlantic: Exploring the variability of the North Atlantic (EGU 2019)	Vienna (AT), 8-9, April 2019	Scientific Community (higher education, Research)	80	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a workshop	Shuting Yang (DMI): On the climate variability and the recent abrupt cooling over Subpolar North Atlantic	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	60	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a conference	Shuting Yang (DMI): On the decadal variability in the Subpolar North Atlantic and its recent abrupt cooling trend (General Assembly of European Meteorological Society, 2019)	Copenhagen (DK), 9-13 September 2019	Scientific Community (higher education, Research)	80	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a conference	Tian Tian (DMI): Sensitivity study on decadal prediction skill to Arctic sea ice initialization in an Earth system model with a multi-category sea ice module (EGU 2019)	Vienna (AT), 8-9, April 2019	Scientific Community (higher education, Research)	80	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a workshop	Tian Tian (DMI): The role of Arctic sea ice initialisation in decadal climate prediction: linking the Arctic sea ice loss and the mid-latitude climates	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	60	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a conference	Tian Tian (DMI): Refinement methods of Arctic sea-ice initialization for improving the decadal prediction skill in the Arctic (General Assembly of	Copenhagen (DK), 9-13 September 2019	Scientific Community (higher education, Research)	80	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20

	European Meteorological Society, 2019)				
Participation to a workshop	Sébastien Barthélémy (UiB) Toward background error covariance hybridization for climate prediction	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	60	https://zenodo.org/record/3241214
Participation to a workshop	Sébastien Barthélémy (UiB) Toward background error covariance hybridization for climate prediction	Voss (NO), 3-5 June 2019	Scientific Community (higher education, Research)	50	https://zenodo.org/record/3241214
Participation to a workshop	Elizabeth Maroon (NCAR), Title: Was the 2015 subpolar North Atlantic cold blob predictable?, 2019 CESM Summer Workshop	Boulder, CO (USA), 17-19 June 2019	Scientific Community (higher education, Research)	100	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a workshop	Panos Athanasiadis (CMCC), Title: Decadal Predictability of North Atlantic Blocking and the NAO.	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	60	10.5281/zenodo.3556769
Participation to a conference	Panos Athanasiadis (CMCC), Title: Skillful decadal predictions of North Atlantic blocking.Participation to international conference (EGU)	Vienna (AU), 8 – 10 April 2019	Scientific Community (higher education, Research)	90	https://zenodo.org/communities/blue-actionh2020/?page=1&size=20
Participation to a workshop	Jennifer Mecking (UoS): Atmosphere versus Ocean in 2015 Cold Blob	Bergen (NO), 5 – 7 June 2019	Scientific Community (higher education, Research)	60	10.5281/zenodo.3260905

Peer reviewed articles

Title	Authors	Publication	DOI	Is Blue-Action correctly acknowledged?	Open Access granted?	If in Green OA, provide the link where this publication can be found
Ocean and atmosphere influence on the 2015 European heatwave	J.V. Mecking, S.S. Drijfhout, J.J-M. Hirschi, A.T. Blaker	Environmental Research Letters	10.1088/1748-9326/ab4d33	Yes	Yes	https://iopscience.iop.org/article/10.1088/1748-9326/ab4d33/meta

Other publications
In revision:
- Decadal-scale predictive skill of North Atlantic upper-ocean salt content and its attribution to the initialization of the North Atlantic Ocean circulation. Lohmann K, Matei D, Bersch M, Jungclaus JH, Pohlmann H, Kröger J, Modali K, and Müller W, Journal of Climate
- Decadal predictability of North Atlantic blocking and the NAO, P. J. Athanasiadis, S. Yeager, Y.-O. Kwon, A. Bellucci, D. W. Smith and S. Tibaldi, NPJ Climate and Atmospheric Science

In preparation
- A multimodel prediction comparison of the 2015 subpolar cold blob.. Maroon, S. Yeager, G. Danabasoglu, N. Rosenbloom, I. Bethke, F. Counillon, N. Dunstone, N. Keenlyside, J. Mignot, P.-A. Monerie, J. Robson, D. Swingedouw, T. Tian, Y. Wang, S. Yang
- Extended range Arctic sea ice forecast with Convolutional Long-Short Term Memory Networks. Y. Liu, L. Bogaardt, J. Attema, W. HazelegerMonthly Weather Review
- Attribution of predictive skill along the Atlantic water pathway. Langehaug et al.
- Background error covariance matrix hybridization for the quasi-geostrophic model. S. Barthélémy, F. Counillon, N. Keenlyside
- Impact of constraining sea ice volume in anomaly initialization in winter forecasts errors with the EC-Earth V3.3 Climate Prediction System. T. Tian, P. Karami, F. Massonnet, S. Yang, T. Koenigk, and T. Kruschke, Geoscientific Model Development Discussion
- A realistic Greenland and surroundings melting in a coupled climate model. M. Devilliers, D. Swingedouw, J. Mignot, J. Deshayes, G. Garric, M. Ayache, Climate Dynamics
- Was the 2015 subpolar cold blob predictable in the CESM Decadal Prediction Large Ensemble. E. Maroon, S. Yeager, G. Danabasoglu, N. Rosenbloom, Journal of Climate
- A multimodel prediction comparison of the 2015 subpolar cold blob. E. Maroon, S. Yeager, G. Danabasoglu, N. Rosenbloom, I. Bethke, F. Counillon, N. Dunstone, N. Keenlyside, J. Mignot, P.-A. Monerie, J. Robson, D. Swingedouw, T. Tian, Y. Wang, S. Yang

To be submitted
- Predictability of the Rapid Warming of the North Pacific Ocean around 1990: Potential Attribution and Impacts, Y. Dai, E. Manzini, H. Pohlman, W. Müller, D. Matei, GRL, to be submitted by 31 Dec. 2019

Uptake by the targeted audiences
As indicated in the Description of the Action, the audience for this deliverable is *the* general public (PU) is and is made available to the world via [CORDIS.](#)

This is how we are going to ensure the uptake of the deliverables by the targeted audiences:
- Participation to scientific conferences (EGU, AGU) and ad-hoc workshops with the scientific community organised by Blue-Action and other projects;
- Publication of peer-reviewed articles in open access.