Impact of Initialization Methods on the Predictive skill in NorCPM - An Arctic-Atlantic Case Study

Leilane Passos (leilanepassos@gmail.com)
University of Bergen: Universitetet i Bergen https://orcid.org/0000-0001-7277-2534
Helene Langehaug
Nansen Environmental and Remote Sensing Center
Marius Årthun
University of Bergen: Universitetet i Bergen
Tor Eldevik
University of Bergen: Universitetet i Bergen
Ingo Bethke
University of Bergen: Universitetet i Bergen
Madlen Kimmritz
Alfred Wegener Institute for Marine and Polar Research: Alfred-Wegener-Institut Helmholtz-Zentrum fur Polar- und Meeresforschung

Research Article

Keywords: Decadal prediction , Subpolar North Atlantic , Norwegian Sea, NorCPM

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-766415/v1

Climate Dynamics manuscript No.
(will be inserted by the editor)

Impact of initialization methods on the predictive skill in NorCPM - an Arctic-Atlantic case study

L. Passos · H. R. Langehaug · M. Årthun · T. Eldevik · I. Bethke · M. Kimmritz

Received: date / Accepted: date

Abstract The skilful prediction of climatic conditions on a forecast horizon of months to decades into the future remains a main scientific challenge of large societal benefit. Here we assess the hindcast skill of the Norwegian Climate Prediction Model (NorCPM) for sea surface temperature (SST) and sea surface salinity (SSS) in the Arctic-Atlantic region focusing on the impact of different initialization methods. We find the skill to be distinctly larger for the Subpolar North

Leilane Passos

University of Bergen and Bjerknes Centre for Climate Research

Nansen Environmental and Remote Sensing Center

E-mail: leilanepassos@gmail.com

H. R. Langehaug

Nansen Environmental and Remote Sensing Center and Bjerknes Centre for Climate Research

M. Årthun, T. Eldevik, and I. Bethke

University of Bergen and Bjerknes Centre for Climate Research

M. Kimmritz

Alfred Wegener Institute for Polar and Marine Research

Atlantic than for the Norwegian Sea, and generally for all lead years analyzed. For the Subpolar North Atlantic, there is furthermore consistent benefit in increasing the amount of data assimilated, and also in updating the sea ice based on SST with strongly coupled data assimilation. The predictive skill is furthermore significant for at least two model versions up to 8-10 lead years with the exception for SSS at the longer lead years. For the Norwegian Sea, significant predictive skill is more rare; there is relatively higher skill with respect to SSS than for SST. A systematic benefit from more complex data assimilation approach can not be identified for this region. Somewhat surprisingly, skill deteriorates quite consistently for both the Subpolar North Atlantic and the Norwegian Sea when going from CMIP5 to corresponding CMIP6 versions. We find this to relate to change in the regional performance of the underlying physical model that dominates the benefit from initialization.

Keywords Decadal prediction · Subpolar North Atlantic · Norwegian Sea · NorCPM

1 Introduction

Skillful decadal climate predictions are beneficial to society by potentially providing information to stakeholders, contributing to political and economical decision-making, and by guiding the planning of climate adaptation measures (Vera et al., 2010; Kushnir et al., 2019). Dynamical predictions are achieved by initialization of climate models with the observed climate state through data assimilation (DA) or alternative synchronization methods. The initialization can reduce the forecast error by considering the internal variability and the mean forced response in the

climate system (Meehl et al., 2009; Yeager et al., 2012), thus achieving skillful predictions regionally up to a decade (Smith et al., 2019).

Decadal climate predictions are still in their early stages of development, the first initialized coupled ocean-atmospheric efforts started in the beginning of the 2000s (e.g. Smith et al., 2007; Keenlyside et al., 2008). Since then, many procedures and techniques have been developed to deal with inherent initialization problems such as initial "shocks", model drift, and a growing variety of assimilated data. Developments range from the primary initialization of the surface ocean component to additional initialization of the subsurface ocean and other components of the climate system (Morioka et al., 2018); from updates of single model components (Weakly-Coupled DA) to cross-component updates during DA in particular model components (which we denote here as Strongly-Coupled DA) (Penny et al., 2019) are current aspects under development and investigation.

Comparisons between multiple DA approaches and techniques using the same model can help to understand how different initialization procedures affect the skill of certain physical processes (Polkova et al., 2019b), as for example the Atlantic Meridional Overtuning Circulation. To this end, the Norwegian Climate Prediction Model (NorCPM), developed by the Bjerknes Center for Climate Research, is a valuable tool. NorCPM consists of a fully-coupled Earth-system model performing ensemble-based sequential DA, based in the Norwegian Earth System model (NorESM). The Norwegian initiative has been through many buildup stages, evolving from surface ocean initialization (Counillon et al., 2016) to assimilation of subsurface observations (Wang et al., 2017) and initializing sea ice component (Kimmritz et al., 2018). A comparison of these stages and their effects on skill, focusing on a specific region or physical process, can contribute to the NorCPM

development by identifying strengths and deficiencies of the different initialization techniques.

On interannual-to-decadal timescales, the Arctic-Atlantic region (Figure 1a) is characterized by poleward propagation of thermohaline anomalies from the Subpolar North Atlantic (SPNA) towards the Arctic through the Norwegian Sea (NS) (Eldevik et al., 2009; Årthun and Eldevik, 2016). The poleward propagation of these anomalies is considered to be a source of climate predictability (e.g. Latif and Keenlyside, 2011), influencing climate variability over Scandinavia and the state of sea ice in the Arctic (Årthun et al., 2017). Based on dynamical climate predictions, the SPNA is one of the most predictable areas in the world (Yeager et al., 2012; Buckley et al., 2019). Some models have demonstrated significant skill up to 10 years in advance for sea surface temperature and upper ocean heat content (van Oldenborgh et al., 2012; Matei et al., 2012; Yeager et al., 2012; Buckley et al., 2019; Smith et al., 2020).

Despite the NS receiving thermohaline anomalies from the SPNA (Årthun and Eldevik, 2016), Langehaug et al. (2017) found skill of only up to 1-3 years in the NS, which is less than what is found for the SPNA region (Matei et al., 2012). Assessing similarities and differences in predictive skill between these two regions for different initialization techniques will improve our understanding of which approaches that are most beneficial in enhancing predictive skill and identify sources of uncertainty in the NS.

The present work focuses on investigating the sensitivity of decadal predictive skill of NorCPM to different initialization techniques, and assessing which one that leads to higher predictive skill for the SPNA and NS. The variables analyzed are those constrained by the DA: sea surface temperature (SST) and sea surface

salinity (SSS). These surface variables are important for marine ecosystems, Arctic sea ice, and atmospheric climate (Årthun et al., 2012, 2018b,a). Here, skill is defined as the ability of the prediction system (NorCPM) to reproduce the same variability of these quantities as in the reanalysis used to initialize NorCPM. In addition, SST variability in NorCPM is compared to that from HadSST2. Five versions of NorCPM are systematically analyzed (Table 1), to primarily investigate how different representations of the ocean initial state and external forcings can affect the predictive skill in the SPNA and the NS for different lead years.

2 Data and Methods

2.1 Norwegian Climate Prediction Model (NorCPM)

NorCPM (Counillon et al., 2014, 2016) is based on the Norwegian Earth System Model (NorESM) (Bentsen et al., 2013) adding the Ensemble Kalman Filter (EnKF) as the DA method (Evensen, 2003)). The fully coupled NorCPM consists of MICOM (ocean), CAM4 (atmosphere), CICE4 (sea ice), CLM4 (land model), and the coupler CPL7; its structure is in general based on the Community Climate System Model version 1.0.3 (CESM1) (Vertenstein et al., 2012). The ocean/sea-ice and atmospheric components have horizontal resolution of 1x1 and 1.9x2.5, respectively.

The DA in NorCPM is based on anomaly initialization. In this process, monthly anomalies are assimilated, generating a reanalysis field that is used to initialize the decadal hindcast. In anomaly DA, the choice of the climatology reference period is relevant to calculate the mean and the subsequent anomalies. In this study there are two reference periods (Table 1). The DA is applied in the ocean component

using sea surface temperature data from HadSST2 (Rayner et al., 2006) and, additionally, subsurface hydrographic profiles from EN4.2.1 (Good et al., 2013) dependent on the particular version. Each reanalysis has 30 ensemble members.

The predictions are initialized from the reanalysis and run freely for 10 years, where each hindcast has between 5 to 20 members depending on the version. In this study, we used results from five different versions: V5, V5w, and V5s, based on NorESM1-ME (Bentsen et al., 2013; Tjiputra et al., 2013) defined here as NorCPM-CMIP5; V6w, and V6s based on NorCPM1 (Bethke et al., 2021) defined here as NorCPM-CMIP6. Details of each version including their respective reference period for calculating anomalies are described in Table 1.

The version V5 assimilates SST through weakly-coupled data assimilation method (WCDA) in the ocean component (described further bellow). Version V5w uses the same assimilation method, but in addition to SST also assimilates hydrographic profiles of temperature and salinity. The implementation of the T-S profiles DA is described in Wang et al. (2017). These authors report an improvement of the system accuracy, generating a reanalyses field suitable to be used at seasonal-to-decadal predictions.

The version V5s (Bethke et al., 2021) assimilates SST and hydrographic profiles, updating the sea ice state through strongly-coupled data assimilation (SCDA) with updates and post-processing of the sea ice state following Kimmritz et al. (2018). The transition from WCDA to SCDA is the pathway defined for major operational forecast centers (Penny et al., 2017). Both DA approaches have two stages; the analysis stage where the DA is performed, and the forecast stage when the components of the system interact through the coupler. In NorCPM, the WCDA method only updates the other components during the forecast stage.

The information exchange between ocean and the other components (sea-ice, land, atmosphere) is achieved dynamically through cross-component fluxes during the forecast stage. By contrast, in the NorCPM's SCDA implementation, the information exchange between ocean and sea ice is additionally realized by using the cross-domain error covariance, allowing ocean observations to instantaneously impact the sea ice state variables during the analysis stage (there is no update for atmosphere and land in the analysis step). Therefore, the information exchange between ocean and sea ice is achieved statistically during the analysis stage, but also dynamically during the forecast stage. Additionally to the SCDA, the version V5s also has the observation error variance of the T-S profiles (EN4.2.1) inflated in the areas with sea ice concentration higher than 50% (Bethke et al., 2021). This additional inflation was applied to deal with the sparsity of TS profiles underneath the sea ice that makes the observations there unreliable.

The implementations V6w and V6s (Bethke et al., 2021) have the same approach as V5w and V5s, but use external forcings as prescribed by the Coupled Model Intercomparison Project Phase 6 (CMIP6), while CMIP5 forcing is used for V5, V5w and V5s. Besides that, the version NorCPM-CMIP6 has re-tuning, and minor code modifications unrelated to forcing upgrades (like bug fixes). The detailed description of the code changes and their effects is in Bethke et al. (2021). The main difference between CMIP5 and CMIP6 protocols is the set of climate forcings applied such as greenhouse gases (Meinshausen et al., 2017), ozone concentrations, atmospheric aerosols (tropospheric and volcanic) (Thomason et al., 2018), and solar forcing (Matthes et al., 2016). The CMIP6 protocol is more precise than the CMIP5 protocol in the way these external forcings should be implemented, however different models may require adaptations in the forcing implementation

due to their specific features and limitations (Lurton et al., 2020; Sellar et al., 2020). In NorCPM, the re-tuning due to the CMIP6 forcings's implementation included an increase of the condensation threshold for low clouds and a decrease of the snow albedo over sea ice adjusting parameters that affect snow metamorphosis (Bethke et al., 2021). Among the updates required to the CMIP6 protocol, it was verified that the land surface types and transient land-use caused an unrealistic land-cryosphere cooling trend over the historical period in NorCPM (Bethke et al., 2021).

V5, V5w and V5s were initialized each second year, while V6w and V6s were initialized every year. All simulations are initialized on November 1st. Figure 1b and Figure 1c show an example of the SST winter time series of the reanalysis (black curves) and the hindcast (red curves) in the SPNA and NS of the version V5s. Both areas show a positive SST trend between 1985 and 2010 for the reanalysis and also for the hindcast (although smaller trends in the hindcast compared to the reanalysis in the NS). The simulation period of the reanalysis and hindcast for each version is shown in Figure 2. In order to compare all versions, the only overlapping period is 1983-1999, comprising 9 initializations. To compare all hindcast versions with the same number of ensemble members, we use 5 members, which is the total ensemble number of V5w. For those versions with higher number of ensemble members we randomly selected five members. Additionally, in the implementations with yearly initialization frequency (V6w and V6s), only every second start date was used. In this way, all versions were systematically compared with the same ensemble size and initialization frequency.

2.2 Hindcast skill and uncertainty

In this study, the predictive skill is quantified using the anomaly correlation coefficient (ACC) and root mean square error (RMSE), which are usual ways to analyze decadal climate prediction skill according to Goddard et al. (2013). The skill assessment of the hindcast for SST and SSS for each version is done against the ensemble mean of the respective NorCPM reanalysis (30 members) and HadSST2. The skill is measured using the ACC as a function of the lead year. The first lead year here is the next year after the start date, and so on. SST and SSS are averaged (without weighted) in time every 3 lead years (1-3, 2-4, 4-6, 5-7, 6-8, 7-9, and 8-10), and then averaged in space (not detrended) for the SPNA and the NS regions, as defined in Figure 1a. After that, the ACC is calculated according to Equation 1.

$$ACC_{LY} = \frac{\sum_{ini=1}^{9}(hind_{ini} - \overline{hind_{LY}})(rean_{ini} - \overline{rean_{LY}})}{\sqrt{\sum_{ini=1}^{9}(hind_{ini} - \overline{hind_{LY}})^2}\sqrt{\sum_{ini=1}^{9}(rean_{ini} - \overline{rean_{LY}})^2}} \quad (1)$$

Where LY is the lead year and ini denote a certain start date for the relevant lead year. $hind$ is the ensemble mean of the hindcast of a start date, and $\overline{hind}$ means the mean of all start dates for the relevant lead year. $rean$ is the ensemble mean of the respective reanalysis of a start date, and $\overline{rean}$ means the mean of all start dates for the correspondent lead year. The ACCs are subject to sampling uncertainty due to the small number of initializations and limited ensemble size. To account for this uncertainty we computed the 25-75% bootstrap confidence interval and we consider two ACCs statistically separated if their confidence intervals do not overlap. The significance level is calculated by the standard two-sided Students t-test (O'Mahony, 1986) at 90% due also to the short compar-

ison period. The predictive skill is calculated for the winter period, defined here as January-April average. This study focus on winter because of the persistence of SST anomalies in this season, when the atmosphere-ocean coupling is most vigorous and creates a sea surface temperature anomaly that reaches the base of the deep winter mixed layer and reemerges in the following winter, causing the natural decadal climate variability in the North Atlantic (Alexander and Deser, 1995; Watanabe and Kimoto, 2000).

3 Predictive skill in different versions of NorCPM

The predictive skill of SST and SSS in different versions of NorCPM in the SPNA (black lines) and in the NS (red lines) is shown in Figure 3. The skill is higher in the SPNA than in the NS in most versions and lead years. In the SPNA, the version V5w and V5s are the only ones with significant skill at all lead years for SST (Figure 3 upper panel). For SSS, only V5s has values higher than or close the significance level (Figure 3 lower panel). Skill differences between implementations and versions are smaller at shorter lead years, and become more pronounced at medium and longer lead years.

In the NS, there is no single implementation that performs higher than the significance level at all lead years neither for SST nor for SSS (red lines in Figure 3 upper and lower panels). For SST, at shorter lead years, V5 is the only version with skill higher than the significance level, while at medium lead years V5w is the one with highest skill. At longer lead years, V5s and V6w have the highest skill in the NS (Figure 3 upper panel). For SSS, the versions with the highest skill are similar to the ones for SST (Figure 3 lower panel), although the differences

between versions at shorter lead years are smaller for SSS than for SST. We note that V6s presents a high anti-correlation at longer lead years for both SST (-0.8) and SSS (-0.75).

The predictive skill in the CMIP6 versions, V6w and V6s, is lower than the respective CMIP5 versions, V5w and V5s. The differences between them are more pronounced at medium and longer lead years. In the next subsections, we assess in more detail the effects of the different initialization techniques on the predictive skill in the SPNA and NS.

3.1 Skill effects of assimilating subsurface data using different climatology reference periods

The assimilation of subsurface data in the ocean is important for a realistic representation of mass transport, mixed-layer depth and eddy kinetic energy; however, the subsurface ocean has only been adequately observed in the last decades, and differences in the frequency and quality of observed temperature and salinity data can result in spurious signals in the assimilated field (Yang et al., 2017). In order to deal with this problem, one approach is for the models to calculate ocean transport processes by only initializing SST. In this way, the subsurface field is initialized indirectly (Keenlyside et al., 2008); similar approach used in version V5. However, recent studies have shown a skill improvement in the North Atlantic when assimilating subsurface data using, for example, lagged-initialization methods (Tatebe et al., 2012; Kröger et al., 2018). To evaluate the effects of subsurface initialization on the predictive skill in NorCPM we compare versions V5 and V5w in the SPNA and in the NS. Nevertheless, differences between these versions are not limited to

the type of data assimilated, since they use different climatology reference periods (Table 1). We will briefly come back to this in the discussion.

The inclusion of subsurface data increases the predictive skill and decreases the error of the hindcast for SST and SSS in the SPNA. In this area, the differences are more pronounced at medium and longer lead years (Figures 4a, 4c and Supplementary Figures 9a, 9c). Unlike the SPNA, SST skill in the NS is higher when using only surface data assimilation at shorter lead years (Figure 4b). At medium lead years, the addition of subsurface data does not generate statistically significant differences between the two versions. For SSS in the NS, adding subsurface data assimilation improves the predictive skill at medium lead years (Figure 4d). The RMSE for SST in the NS is slightly higher at shorter lead years in V5w, after that the error is statistically equal for both versions (Supplementary Figure 9b). In the NS, the RMSE for SSS is the same at short lead years between versions and slightly higher in V5w at medium and longer lead years (Supplementary Figure 9d). We note that using HadSST2 data instead of the reanalysis, we find similar results for both regions (Figures 4a, 4b and Supplementary Figures 9a, 9b).

3.2 Skill effects of weakly vs strongly-coupled DA and ensemble inflation

Sea ice is an important component of the climate system, since it helps regulate the heat transfer between the ocean and the atmosphere with a global effect on climate scales, impacting the slow-evolving thermohaline circulation (Holland et al., 2001; Liu et al., 2019). This component was thus chosen to initiate the implementation of SCDA in NorCPM (see Kimmritz et al., 2018, for DA of sea ice concentration updating the sea ice and the ocean state). In the strongly-coupled method the

observations are used to update other model components through cross-covariance error, being an alternative to improve consistency between the analyzed state of each component and eliminate initial shocks (Penny et al., 2017). In this approach, the main challenge is to deal with the differences in spatial and temporal scales between ocean, atmosphere and sea ice. The scales of ocean and sea ice are more alike than those of ocean/sea ice and atmosphere, making the jointly update of ocean-sea ice a natural "starter" for SCDA. Idealized and non-idealized studies have demonstrated advances in the strongly-coupled approach between ocean and atmosphere (Lu et al., 2015a,b; Sluka et al., 2016) and between ocean and sea ice (Kimmritz et al., 2018). Considering that the SCDA approach has recently been developed, it is important to understand whether it has a positive impact on the predictive skill, or whether it transfers biases from one component to another (Penny et al., 2017).

The evaluation of the predictive skills of V5w (WCDA) and V5s (SCDA+inflation) is shown in Figure 5. In addition to the jointly update of the ocean and sea ice during analysis stage (SCDA), an error inflation is applied as described in Section 2. The effect of these implementations on the predictive skill differs depending on the area. In the SPNA, V5s shows slightly higher predictive skill and lower RMSE than V5w for both SST and SSS (Figures 5a, 5c and Supplementary Figures 10a, 10c) at all lead times. In the NS, WCDA (V5w) has higher skill than V5s at shorter and medium lead times for SST (Figure 5b) and at medium lead times for SSS (Figure 5d).The RMSE of SST and SSS is statistically similar for both versions in the NS (Supplementary Figures 10b and 10d). Using HadSST2 data instead of the reanalysis, we find similar results for both regions (Figures 5a, 5b and Supplementary Figures 10a, 10b).

The difference between V6w and V6s is similar to the described above for V5w and V5s. Also in NorCPM-CMIP6 versions, the SCDA with additional inflation has a positive impact over the skill in the SPNA for almost all lead years; and for the NS, it maintains the skill at first lead years leading to a significant anti-correlation at longer lead years (Supplementary Figures 11 and 12). An overall assessment of the spatial peculiarities between V5w and V5s are detailed in Section 4.

3.3 Skill differences between NorCPM-CMIP5 and NorCPM-CMIP6

The NorCPM-CMIP6 version includes in the underlying physical model the CMIP6 forcings, re-tuning, and minor code modifications Bethke et al. (2021). To compare the predictive skill between NorCPM-CMIP5 and NorCPM-CMIP6, we here compare V5w (CMIP5) and V6w (CMIP6), since both have the same initialization approach (Table 1). For NorCPM-CMIP6 we use the same number of initializations as NorCPM-CMIP5 (every second year), as described in Section 2.

In the SPNA, the NorCPM-CMIP5 version has SST skill higher than the significance level for all lead years. The NorCPM-CMIP6 version has the same SST skill/RMSE as NorCPM-CMIP5 at shorter lead years (up to 3-5 years), but lower than the significance level after 5-7 years (Figure 6a) in addition to a higher RMSE (Supplementary Figure 13a). SSS has statistically the same predictive skill in both versions up to 2-4 years. After 3-5 lead years, the skill of SSS in NorCPM-CMIP6 is lower than in NorCPM-CMIP5 (Figure 6c) while the RMSE is statistically the same for both versions (Supplementary Figure 13c). In the NS, the NorCPM-CMIP5 version has higher SST skill than the NorCPM-CMIP6 version at shorter lead years (up to 4-6 years); after that the skill is statistically equal, and then lower

at 8-10 lead years (Figure 6b). At all lead years the RMSE is statistically similar for both versions (Supplementary Figure 13b). For SSS in the NS, the predictive skill of NorCPM-CMIP5 version is higher than NorCPM-CMIP6 at almost all lead years with the biggest differences between versions at medium lead years (Figure 6d), it is also when the RMSE for V6w is higher than in V5w (Supplementary Figure 13d). Using HadSST2 data instead of the reanalysis, we find similar results for both regions (Figures 6a, 6b and Supplementary Figures 13a, 13b).

4 Overall predictive skill between the Subpolar North Atlantic and the Norwegian Sea

According to the analysis made in Section 3, the version V5s (CMIP5, SCDA+error inflation) has the overall highest skill in the SPNA with significant values for almost all lead years for SST and SSS (Figure 3). In the NS, the version V5w (CMIP5, WCDA) is the only version with significant values at medium lead years for SST and SSS (Figure 3). Considering this, the comparison between the SPNA and the NS in this section will be evaluated based on V5s and V5w. The spatial evolution of skill with lead year for SST and SSS for both versions is shown in Figure 7 and Figure 8, respectively.

In the SPNA, the version V5w has a large area with skill higher than 0.6 at 1-3 lead years. Near Newfoundland Basin is the only area where the correlation at this lead year is null or negative (Figure 7a). This pattern remains up to 7-9 lead years (Figure 7k), with the exception of another area with low skill that forms at lead year 2-4 near the Rockall Plateau (Figure 7b). This is an area where branches of NAC merge and flow northward towards the NS (Daniault et al.,

2016); this low skill area remains up to 5-7 lead years (Figure 7i). At lead years 8-10, significant predictive skill in the SPNA is mainly localized in the Labrador Sea and south of Greenland (Figure 7l). Comparing version V5w and V5s at 1-3 lead years, V5s shows higher skill in most of the SPNA with predictive skill higher than 0.8 (Figure 7e). At 1-3 lead years there is a lower skill area near the Newfoundland Basin, and at 2-4 lead years there is also a lower skill area near the Rockall Plateau (Figures 7e and f). However, in V5s these two areas with poor skill are smaller compared to that in V5w. This pattern remains the same up to 6-8 lead years (Figure 7n). After that, the two areas with poor skill grow and merge, and at 8-10 lead years significant predictive skill in the SPNA is localized in the eastern part, near to Iceland, and in the Labrador Sea (Figure 7p). Most parts of the Labrador Sea still have values higher than 0.8, which is not seen in V5w at 8-10 lead years(Figure 7l). In both versions, NorCPM struggles to represent SST variability near Newfoundland Basin and Rockall Plateau.

In the NS, the version V5w has predictive skill of 0.6-0.8 at 1-3 lead years in most of the area from the Greenland-Scotland Ridge to the Knipovich Ridge (Figure 7a). The skill remains in the area, in a narrow region close to Norway, up to 7-9 lead years (Figure 7k). At 4-6 lead years, skill is also seen in the Barents Sea (Figure 7d). At 5-7 lead years, an anti-correlation area forms within the Norwegian and Lofoten basins (Figure 7i). This feature grows at the subsequent lead years becoming significant from 6-8 lead years (Figure 7j). At lead years 8-10, the anti-correlation area expands and there is no significant skill in the NS (Figure 7l). In the V5s case, the significant predictive skill in the NS at 1-3 lead years is only seen in a narrow region close to southern Norway (Figure 7e). At 3-5 lead years, the area with skill extends to the northern Norway (Figure 7g), and to the Barents

Sea at 4-6 lead years (Figure 7h). However, in this version the anti-correlation forms early, at 2-4 lead years, in addition to being in a larger area comprising the Norwegian and Lofoten basins (Figure 7f). The anti-correlation is significant from 3-5 to 6-8 lead years (Figures 7g, 7h, 7m, and 7n). After that, the anti-correlation is not significant and the area decreases in size (Figures 7o and 7p). Unlike V5w, in V5s at 8-10 lead years there is a large area in the NS, extending to the SPNA, with predictive skill of 0.6-0.8 (Figure 7p). These are the only lead years where V5s performs better than V5w in the NS.

In view of the above comparison, the SCDA and the error inflation (version V5s) have a positive impact on the predictive skill in the SPNA, especially in the Labrador Sea. However, this version has a negative effect on the predictive skill in the NS, specifically in the Norwegian and Lofoten basins, where a strong anti-correlation area is formed and develops early in the hindcast. Furthermore, the skill in the Barents Sea also degrades in this version.

5 Discussion

In this study, we have assessed the sensitivity of decadal predictive skill in the Subpolar North Atlantic (SPNA) and in the Norwegian Sea (NS) to the use of different types of data assimilation techniques in NorCPM. The SPNA is one of the areas with highest predictive skill at decadal time scales in climate prediction systems (Yeager and Robson, 2017). It is also an important source region of predictability to the NS (Årthun et al., 2017; Langehaug et al., 2017), and the correct representation of mechanisms underlying decadal variability is thus important for climate predictability in the Arctic-Atlantic region.

The predictability in the SPNA and in the NS has been associated with specific physical mechanisms that act on different time scales. The Atlantic Meridional Overturning Circulation (AMOC) is pointed out as a key source of decadal to multidecadal climate predictability in the SPNA. However, recently the Labrador Sea Water thickness anomalies have been indicated as precursor of upper ocean heat content predictability in this area, as well as an important driver influencing the southward ocean circulation (Ortega et al., 2020; Yeager, 2020). The Labrador Sea Water is formed by convective processes driven by high oceanic heat losses during winter. The intense heat loss is associated with sea ice formation (Yeager, 2020) and reduce the density stratification in the Labradror Sea. Dense Labrador Sea Water is carried southward as part of the deep branch of AMOC. Thus, the representation of sea ice plays an important role in the predictive skill of the SPNA. On a shorter time scale, SST variability with a period of 13-18 years has been found to dominate in the North Atlantic, and is suggested to contribute to recent cold anomaly in the SPNA (Årthun et al., 2021).

The pronounced decadal variability in the SPNA is linked to northward propagation of thermohaline anomalies and largely contributes to predictability in the NS (Årthun et al., 2017). The Greenland-Scotland Ridge (GSR) is where the warm and salty waters of the North Atlantic Current (NAC) enters the NS (Figure 1). The representation of the flow across the GSR is thus important since it is the gateway of thermohaline anomalies coming from the SPNA. Heuzé and Årthun (2019) suggest, based on analysis of 23 CMIP5 models, that the model resolution and bathymetry in the GSR is a key factor for the oceanic heat transport from the North Atlantic to the Arctic. In addition, analysis from observations and ocean state estimate show that unrealistic eddy fluxes, and air-sea heat fluxes in the

Norwegian Sea can limit the predictability carried by the poleward heat anomalies (Chafik et al., 2015; Asbjørnsen et al., 2019). The air-sea fluxes in this area are considerable; the Lofoten basin alone is responsible for 1/3 of the heat loss in the Nordic Seas, despite only covering 1/5 of its total area (Richards and Straneo, 2015).

The NorCPM study herein shows that the increase of complexity in the initialization method in the CMIP5 versions (V5 to V5s) results in a general skill increase for the SPNA, whereas the same is not achieved in the NS. Independent of the initialization approach used, there is a large skill difference between the SPNA and NS, suggesting that there are difficulties in representing the physical processes in the NS or predictability might be more limited in this region compared to the SPNA. In NorCPM, one of the physical processes taking place in the SPNA, AMOC, is in relatively good agreement with observations (V5; Counillon et al., 2016). On the other hand, in the NS the propagation of SST anomalies is not properly represented by the model (V6w; Langehaug et al., 2021), and the surface currents are rather broader in the SPNA and NS (NorESM1-M; the underlying model of NorCPM; Langehaug et al., 2018). A similar result is found in the system MPI-ESM-LR; the use of oceanic EnKF improves the predictive skill in the SPNA, especially in the Labrador and Irminger Seas, while decreases the skill in the NS compared to the historical run (Brune and Baehr, 2020). The spatial ACC maps in Figure 7 show significant SST skill in a narrow region close to Norway, but in the frontal region where Atlantic Water meets Arctic Water the skill is largely reduced. The ACC maps showed that SST skill decreases in the Norwegian and Lofoten basins. These areas are dominated by intense eddy activity and surface heat fluxes and the non-significant SST skill can indicate the struggling of the pre-

diction system to properly represent these processes. This will be further discussed below, when comparing the versions V5w and V5s.

The addition of subsurface data with the use of a different climatology reference period in the initialization method showed a general positive impact on the predictive skill in both regions. In the SPNA, this initialization method generated a considerable improvement of the skill in the surface ocean that lasted at medium and longer lead years, while for the NS the gain of skill was only related to SSS. Considering both ACC maps (Figures 7 and 8) and RMSE maps (Supplementary Figures 14, 15, and 16), there is low or no skill in similar regions for SST and SSS: in the Newfoundland Basin, and in the Norwegian and Lofoten basins. The reason for gain in skill only for SSS in the NS could be due to the fact that SST is more influenced by local surface forcing than the SSS in this area (Asbjørnsen et al., 2019; Furevik et al., 2002). In addition to the inclusion of subsurface data, the different climatology reference periods used might also have an influence on the skill. The sparsity and quality of the observed data before the 1980s can increase the uncertainties in the mean climate calculation, which in turn can affect the uncertainties of anomalies used in the data assimilation. Unfortunately, with the available implementations it was not possible to separate the effects of subsurface data assimilation and the different climatology reference periods.

Furthermore, we have looked at the impact of a jointly update of the ocean and sea ice (SCDA) and additional error inflation (comparison of versions V5w and V5s). When the sea-ice is corrected by the covariation with the ocean temperature, the surface ocean skill is slightly higher in the SPNA, while in the NS such increase is not verified. The reanalysis of V5s is the only one with strongly reduced bias in Arctic sea ice thickness, which grow back over lead years causing a strong drift in

the hindcast (Bethke et al., 2021). The main differences between V5w (WCDA) and V5s (SCDA) in the NS are: significant SST skill in the V5s is only found in the southwest area at 1-3 lead years compared to a larger area in V5w that extends towards the Knipowich Ridge (7a); a significant anti-correlation area over the Norwegian and Lofoten basins appears earlier in V5s compared to V5w, at 2-4 lead years in the former (7g). The anti-correlation area is also seen in the NorCPM-CMIP6 versions, being stronger in V6s than in V6w (Supplementary Figures 11 and 12), and it is also seen in the historical run of the respective model version (Figure not shown). The anti-correlation area is also associated with high SST RMSE (Supplementary Figures 14 and 15) extending from the Jan Mayen to the Mohn-Knipovich Ridges, the occurrence area of the Arctic Front. The Arctic Front is localized near the sea ice edge and is characterized by the interaction of the colder and fresher Arctic Water with the warmer and more saline Atlantic Water (Swift and Aagard, 1981). Along the front, observations show active air-sea interaction (Raj et al., 2019); while high-resolution numerical simulations verified cross-ridge exchange on Mohn-Knipovich Ridge leading to a cooling and freshening of the Norwegian Atlantic Current (Ypma et al., 2020). It is uncertain, however, how the Arctic Front is represented in NorCPM, and is beyond the scope of this study. However, the results herein shows that this region and associated processes are difficult to represent in a coarse climate model.

In the SPNA the SCDA version, V5s, increased the skill and decreased the RMSE for both SST and SSS. The SST RMSE reduction is seen in the Labrador Sea and over the MAR (Supplementary Figures 14e-h and 14m-o), while the SSS RMSE decrease happens in the Labrador Sea and on the Reykjanes Ridge, at 2-4 and 3-5 lead years (Supplementary Figures 16f and 16g, respectively). De-

spite the improvement in V5s version, in both V5s and V5w the model struggles to represent SSS over the MAR. Results from another dynamical prediction system, CESM-DPLE, show that the MAR plays an important role in the SPNA predictability, inducing a strong coupling between the AMOC and the SPNA circulation through the propagation of deep water mass anomalies (Yeager, 2020). However, this mechanism has not been assessed in the present study.

Initialized decadal hindcasts from CMIP6 improved the representation of SST in the SPNA compared to CMIP5 according to Borchert et al. (2021). This study analyzed 6(7) prediction systems from CMIP5 (CMIP6) and showed that 88% of the SPNA SST variance at lead years 5-7 is explained by CMIP6 hindcasts, while CMIP5 explains only 42%. Borchert et al. (2021) attribute this difference to a good representation of the SPNA SST in CMIP6 historical simulations due to an increase of ensemble size, as well as to a high predictive skill in CMIP6 after the end of the CMIP5 period. In our study, ensemble size, number of initializations, and period analyzed were the same for all versions. In NorCPM-CMIP6 the drop of skill is related to a wrong surface land temperature trend, in particular over the Canadian Arctic, that affects the atmospheric state and circulation over the North Atlantic and also the Nordic Seas (Bethke et al., 2021). Thus, the comparison here between versions NorCPM-CMIP5 and NorCPM-CMIP6 includes changes in the code and in the forcings, which makes it difficult to evaluate their isolated effects on the predictive skill.

Intercomparisons between prediction systems or versions can be an important approach to better understand how well some physical processes are represented in the model and might help to identify weaknesses and indicate potential directions for further development. To investigate the predictive skill in the Nordic Seas,

Langehaug et al. (2017) analyzed the results of three prediction systems. The authors suggested a relationship between resolution and SST skill in the analyzed systems, in addition to highlighting that unrealistic sea ice cover could be a relevant bias in the Nordic Seas. Three initialization approaches were tested by Polkova et al. (2014), and the best results for sea surface height in the North Atlantic were verified in the full state initialization with heat and freshwater flux corrections. On the other hand, Polkova et al. (2019a) found improvements in some parts of the North Atlantic for surface temperatures and upper ocean heat content with a combination of ensemble Kalman filter and filtered anomaly initialization. Based on this, it is not possible to define an optimal technique for all areas and variables even using the same prediction system (Höschel et al., 2019), which is analogue to the results of this work.

6 Summary and Conclusions

The goal of this study was to assess and quantify the impact of different initialization strategies on the development of the Norwegian Climate Prediction Model (NorCPM). The investigation was focused on the Subpolar North Atlantic (SPNA) and the Norwegian Sea (NS). The predictive skill was tested against NorCPM's own reanalysis, and for SST also against HadSST2.

The comparison among versions showed that the choice of Data Assimilation (DA) method appears to have largest impact on medium to long lead years. In the SPNA, increasing initialization complexity resulted in a general skill increase within NorCPM-CMIP5 versions, however, the same was not found for the NS. The additional assimilation of subsurface data in addition to a different climatology

reference period gave a general skill improvement in both areas (Section 3.1). An improvement was found for both SST and SSS in the SPNA, while in the NS the skill for SST was maintained at medium and longer lead years and improved for SSS between 2-4 and 5-7 lead years.

The joint update of the ocean and sea ice (Strongly Coupled DA) with additional inflation error increased surface ocean skill in the SPNA and decreased it in the NS. This version presented an area with highly negative ACC over the Norwegian and Lofoten basins associated with high RMSE between the Jan Mayen and Mohn-Knipovich Ridges, showing that this version struggles in particular to represent ocean processes in these areas in the NS. The comparison between NorCPM-CMIP5 and NorCPM-CMIP6 showed a reduced ocean skill in the latter due to erroneous land surface conditions that affect the atmospheric state over the North Atlantic and Nordic Seas.

In general, the comparison of skill in this study shows that the NorCPM-CMIP5 version with the Strongly Coupled DA has the highest skill in the SPNA and without it (Weakly Coupled DA) for the NS. Furthermore, the results from this study show that despite the NS being directly influenced by thermohaline anomalies coming from the SPNA, the circulation of the Atlantic Water in the Norwegian and Lofoten basins and its interaction with colder and fresher waters over the Jan Mayen and Mohn-Knipovich Ridges are aspects that need to be improved in NorCPM. A better representation of the complex bathymetry close to the Greenland-Scotland Ridge and in the NS along with an eddy permitting grid might be a way to improve NorCPM results in the NS.

The NS is the path by which thermohaline anomalies coming from the North Atlantic influence the Arctic sea ice cover (Onarheim et al., 2015), cod stock

(Årthun et al., 2018a) and the surface air temperature over Scandinavia (Årthun et al., 2017). The identification of the most suitable DA approach for the Arctic-Atlantic region contributes to the development of NorCPM. Identifying the main challenges to predictive skill in the NS furthermore contributes to a better representation of key processes not just in NorCPM, but also in other prediction systems. The results show that areas with processes as intense surface heat fluxes and eddy activity within the Norwegian-Lofoten Basins and in the area of the Arctic Front, might be key areas to improve NS skill. The investigation of the NS skill in other prediction systems might help to guide development for improvement in NorCPM.

Data availability

The observation-based SST used HadISST2 is available at https://www.metoffice.gov.uk/hadobs/hadsst2/. NorCPM-CMIP5 reanalysis and hindcast and NorCPM-CMIP6 reanalysis are available upon request to the modeling group. The CMIP6-DCPP (V6w and V6s) output can be acessed at https://doi.org/10.22033/ESGF/CMIP6.10844

Author contributions

LP coordinate the writing of this article. LP, HL and Å worked in the conceptualization of the article. LP, HL and IB contributed processing the results. TE, IB and MK contributed with the results evaluation. All co-authors contributed to the writing of the article.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

Computing and storage resources have been provided by UNINETT Sigma2 (nn9039k,ns9039k).

Financial support

LP, HL, MÅ, TE, IB and MK received funding from the Trond Mohn Foundation through the Bjerknes Climate Prediction Unit (BFS2018TMT01). HRL, MÅ, and TE has also received funding from the Blue-Action Project (European Unions Horizon 2020 research and innovation program, Grant 727852).

References

Alexander MA, Deser C (1995) A Mechanism for the Recurrence of Wintertime Midlatitude SST Anomalies. Journal of Physical Oceanography 25:122–137

Årthun M, Eldevik T (2016) On anomalous ocean heat transport toward the Arctic and associated climate predictability. Journal of Climate 29(2):689–704, DOI 10.1175/JCLI-D-15-0448.1

Årthun M, Eldevik T, Smedsrud LH, Skagseth, Ingvaldsen RB (2012) Quantifying the influence of atlantic heat on barents sea ice variability and retreat. Journal of Climate 25(13):4736–4743, DOI 10.1175/JCLI-D-11-00466.1

Årthun M, Eldevik T, Viste E, Drange H, Furevik T, Johnson HL, Keenlyside NS (2017) Skillful prediction of northern climate provided by the ocean. Nature Communications 8(May), DOI 10.1038/ncomms15875

Årthun M, Bogstad B, Daewel U, Keenlyside NS, Sandø AB, Schrum C, Ottersen G (2018a) Climate based multi-year predictions of the Barents Sea cod stock. PLoS ONE 13(10):1–13, DOI 10.1371/journal.pone.0206319

Årthun M, Kolstad EW, Eldevik T, Keenlyside NS (2018b) Time Scales and Sources of European Temperature Variability. Geophysical Research Letters 45(8):3597–3604, DOI 10.1002/2018GL077401

Årthun M, Wills RC, Johnson HL, Chafik L, Langehaug HR (2021) Mechanisms of decadal North Atlantic climate variability and implications for the recent cold anomaly. Journal of Climate 34(9):3421–3439, DOI 10.1175/JCLI-D-20-0464.1

Asbjørnsen H, Årthun M, Skagseth , Eldevik T (2019) Mechanisms of ocean heat anomalies in the Norwegian Sea. Journal of Geophysical Research: Oceans p 2018JC014649, DOI 10.1029/2018JC014649, URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2018JC014649

Bentsen M, Bethke I, Debernard JB, Iversen T, Kirkevåg A, Seland , Drange H, Roelandt C, Seierstad IA, Hoose C, Kristjánsson JE (2013) The Norwegian Earth System Model, NorESM1-M Part 1: Description and basic evaluation of the physical climate. Geoscientific Model Development 6(3):687–720, DOI 10.5194/gmd-6-687-2013

Bethke I, Wang Y, Counillon F, Keenlyside N, Kimmritz M, Fransner F, Samuelsen A, Langehaug H, Svendsen L, Chiu PG, Leilane Goncalves dos Passos MB, Guo C, Tjiputra J, Kirkevåg A, Olivier D, Seland , Vågene JS, Fan Y, Lawrence P (2021) NorCPM1 and its contribution to CMIP6 DCPP

628 Borchert LF, Menary MB, Swingedouw D, Sgubin G, Hermanson L, Mignot J
629 (2021) Improved Decadal Predictions of North Atlantic Subpolar Gyre SST in
630 CMIP6. Geophysical Research Letters 48(3):1–10, DOI 10.1029/2020GL091307
631 Brune S, Baehr J (2020) Preserving the coupled atmosphereocean feedback in
632 initializations of decadal climate predictions. Wiley Interdisciplinary Reviews:
633 Climate Change 11(3):1–19, DOI 10.1002/wcc.637
634 Buckley MW, DelSole T, Susan Lozier M, Li L (2019) Predictability of North At-
635 lantic Sea surface temperature and upper-ocean heat content. Journal of Climate
636 32(10):3005–3023, DOI 10.1175/JCLI-D-18-0509.1
637 Chafik L, Nilsson J, Skagseth , Lundberg P (2015) On the flow of Atlantic water
638 and temperature anomalies in the Nordic Seas toward the Arctic Ocean. JOurnal
639 of Geophysical Research: Oceans pp 7897–7918, DOI 10.1002/2014JC010472.
640 Received
641 Counillon F, Bethke I, Keenlyside N, Bentsen M, Bertino I, Zheng F (2014)
642 Seasonal-to-decadal predictions with the ensemble kalman filter and the Nor-
643 wegian earth System Model: A twin experiment. Tellus, Series A: Dynamic Me-
644 teorology and Oceanography 66(1), DOI 10.3402/tellusa.v66.21074
645 Counillon F, Keenlyside N, Bethke I, Wang Y, Billeau S, Shen ML, Bentsen M
646 (2016) Flow-dependent assimilation of sea surface temperature in isopycnal co-
647 ordinates with the Norwegian Climate Prediction Model. Tellus, Series A: Dy-
648 namic Meteorology and Oceanography 68(1), DOI 10.3402/tellusa.v68.32437
649 Daniault N, Mercier H, Lherminier P, Sarafanov A, Falina A, Zunino P, Pérez
650 FF, Ríos AF, Ferron B, Huck T, Thierry V, Gladyshev S (2016) The north-
651 ern North Atlantic Ocean mean circulation in the early 21st century. Progress
652 in Oceanography 146(June):142–158, DOI 10.1016/j.pocean.2016.06.007, URL

http://dx.doi.org/10.1016/j.pocean.2016.06.007

Eldevik T, Nilsen JE, Iovino D, Anders Olsson K, Sandø AB, Drange H (2009) Observed sources and variability of Nordic seasoverflow. Nature Geoscience 2(6):406–410, DOI 10.1038/ngeo518, URL http://dx.doi.org/10.1038/ngeo518

Evensen G (2003) The Ensemble Kalman Filter: Theoretical formulation and practical implementation. Ocean Dynamics 53(4):343–367, DOI 10.1007/s10236-003-0036-9

Furevik T, Bentsen M, Drange H, Johannessen JA, Korablev A (2002) Temporal and spatial variability of the sea surface salinity in the Nordic Seas. Journal of Geophysical Research: Oceans 107(12):1–16, DOI 10.1029/2001jc001118

Goddard L, Kumar A, Solomon A, Smith D, Boer G, Gonzalez P, Kharin V, Merryfield W, Deser C, Mason SJ, Kirtman BP, Msadek R, Sutton R, Hawkins E, Fricker T, Hegerl G, Ferro CA, Stephenson DB, Meehl GA, Stockdale T, Burgman R, Greene AM, Kushnir Y, Newman M, Carton J, Fukumori I, Delworth T (2013) A verification framework for interannual-to-decadal predictions experiments. Climate Dynamics 40(1-2):245–272, DOI 10.1007/s00382-012-1481-2

Good SA, Martin MJ, Rayner NA (2013) EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. Journal of Geophysical Research: Oceans 118(12):6704–6716, DOI 10.1002/2013JC009067

Heuzé C, Årthun M (2019) The Atlantic inflow across the Greenland-Scotland ridge in global climate models (CMIP5). Elementa 7(1), DOI 10.1525/elementa.354

678 Holland MM, Bitz CM, Eby M, Weaver AJ (2001) The role of ice-ocean interactions
679 in the variability of the North Atlantic thermohaline circulation. Journal of Cli-
680 mate 14(5):656–675, DOI 10.1175/1520-0442(2001)014⟨0656:TROIOI⟩2.0.CO;2
681 Höschel I, Illing S, Grieger J, Ulbrich U, Cubasch U (2019) On skillful decadal pre-
682 dictions of the subpolar North Atlantic. Meteorologische Zeitschrift 28(5):417–
683 428, DOI 10.1127/metz/2019/0957
684 Keenlyside NS, Latif M, Jungclaus J, Kornblueh L, Roeckner E (2008) Ad-
685 vancing decadal-scale climate prediction in the North Atlantic sector. Nature
686 453(7191):84–88, DOI 10.1038/nature06921
687 Kimmritz M, Counillon F, Bitz CM, Massonnet F, Bethke I, Gao Y (2018) Op-
688 timising assimilation of sea ice concentration in an Earth system model with
689 a multicategory sea ice model. Tellus, Series A: Dynamic Meteorology and
690 Oceanography 70(1):1–23, DOI 10.1080/16000870.2018.1435945, URL https:
691 //doi.org/10.1080/16000870.2018.1435945
692 Kröger J, Pohlmann H, Sienz F, Marotzke J, Baehr J, Köhl A, Modali K, Polkova
693 I, Stammer D, Vamborg FS, Müller WA (2018) Full-field initialized decadal
694 predictions with the MPI earth system model: an initial shock in the North At-
695 lantic. Climate Dynamics 51(7-8):2593–2608, DOI 10.1007/s00382-017-4030-1,
696 URL http://dx.doi.org/10.1007/s00382-017-4030-1
697 Kushnir Y, Scaife AA, Arritt R, Balsamo G, Boer G, Doblas-Reyes F, Hawkins
698 E, Kimoto M, Kolli RK, Kumar A, Matei D, Matthes K, Müller WA, OKane
699 T, Perlwitz J, Power S, Raphael M, Shimpo A, Smith D, Tuma M, Wu B
700 (2019) Towards operational predictions of the near-term climate. Nature Climate
701 Change 9(2):94–101, DOI 10.1038/s41558-018-0359-7

Langehaug HR, Matei D, Eldevik T, Lohmann K, Gao Y (2017) On model differences and skill in predicting sea surface temperature in the Nordic and Barents Seas. Climate Dynamics 48(3-4):913–933, DOI 10.1007/s00382-016-3118-3

Langehaug HR, Sandø AB, Årthun M, Ilcak M (2018) Variability along the Atlantic water pathway in the forced Norwegian Earth System Model. Climate Dynamics 52(1-2):1211–1230, DOI 10.1007/s00382-018-4184-5, URL http://dx.doi.org/10.1007/s00382-018-4184-5

Langehaug HR, Counillon F, Ortega P, Matei D, Keenlyside N, Maroon E (2021) Propagation of Thermohaline Anomalies and their predictive potential in the Northern North Atlantic (Accepeted). Journal of Climate pp 1–22

Latif M, Keenlyside NS (2011) A perspective on decadal climate variability and predictability. Deep-Sea Research Part II: Topical Studies in Oceanography 58(17-18):1880–1894, DOI 10.1016/j.dsr2.2010.10.066, URL http://dx.doi.org/10.1016/j.dsr2.2010.10.066

Liu W, Fedorov A, Sévellec F (2019) The mechanisms of the Atlantic meridional overturning circulation slowdown induced by Arctic sea ice decline. Journal of Climate 32(4):977–996, DOI 10.1175/JCLI-D-18-0231.1

Lu F, Liu Z, Zhang S, Liu Y (2015a) Strongly coupled data assimilation using leading averaged coupled covariance (LACC). Part I: Simple model study. Monthly Weather Review 143(9):3823–3837, DOI 10.1175/MWR-D-14-00322.1

Lu F, Liu Z, Zhang S, Liu Y, Jacob R (2015b) Strongly coupled data assimilation using leading averaged coupled covariance (LACC). Part II: CGCM experiments. Monthly Weather Review 143(11):4645–4659, DOI 10.1175/MWR-D-15-0088.1

Lurton T, Balkanski Y, Bastrikov V, Bekki S, Bopp L, Braconnot P, Brockmann P, Cadule P, Contoux C, Cozic A, Cugnet D, Dufresne JL, Éthé C, Foujols

MA, Ghattas J, Hauglustaine D, Hu RM, Kageyama M, Khodri M, Lebas N, Levavasseur G, Marchand M, Ottlé C, Peylin P, Sima A, Szopa S, Thiéblemont R, Vuichard N, Boucher O (2020) Implementation of the CMIP6 Forcing Data in the IPSL-CM6A-LR Model. Journal of Advances in Modeling Earth Systems 12(4):1–22, DOI 10.1029/2019MS001940

Matei D, Pohlmann H, Jungclaus J, Müller W, Haak H, Marotzke J (2012) Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. Journal of Climate 25(24):8502–8523, DOI 10.1175/JCLI-D-11-00633.1

Matthes K, Funke B, Anderson ME, Barnard L, Beer J, Charbonneau P, Clilverd MA, Dudok de Wit T, Haberreiter M, Hendry A, Jackman CH, Kretschmar M, Kruschke T, Kunze M, Langematz U, Marsh DR, Maycock A, Misios S, Rodger CJ, Scaife AA, Seppälä A, Shangguan M, Sinnhuber M, Tourpali K, Usoskin I, van de Kamp M, Verronen PT, Versick S (2016) Solar Forcing for CMIP6 (v3.1). Geoscientific Model Development Discussions 6(June):1–82, DOI 10.5194/gmd-2016-91

Meehl GA, Goddard L, Murphy J, Stouffer RJ, Boer GJ, Danabasoglu G, Dixon K, Giorgetta MA, Greene AM, Hawkins E, Hegerl G, Karoly D, Keenlyside N, Kimoto M, Kirtman B, Navarra A, Pulwarty R, Smith D, Stammer D, Stockdale T (2009) Decadal Prediction Can It Be Skillful? Bulletin of the American Meteorological Society 90(10):1467–1486, DOI 10.1175/2009BAMS2778.I, URL http://journals.ametsoc.org/doi/10.1175/2009BAMS2778.1%0Ahttps://journals.ametsoc.org/doi/pdf/10.1175/2009BAMS2778.1

Meinshausen M, Vogel E, Nauels A, Lorbacher K, Meinshausen N, Etheridge DM, Fraser PJ, Montzka SA, Rayner PJ, Trudinger CM, Krummel PB, Beyerle U,

752 Canadell JG, Daniel JS, Enting IG, Law RM, Lunder CR, O'Doherty S, Prinn
753 RG, Reimann S, Rubino M, Velders GJ, Vollmer MK, Wang RH, Weiss R (2017)
754 Historical greenhouse gas concentrations for climate modelling (CMIP6). Geo-
755 scientific Model Development 10(5):2057–2116, DOI 10.5194/gmd-10-2057-2017
756 Morioka Y, Doi T, Storto A, Masina S, Behera SK (2018) Role of subsurface
757 ocean in decadal climate predictability over the South Atlantic. Scientific Re-
758 ports 8(1):8523, DOI 10.1038/s41598-018-26899-z, URL http://www.nature.
759 com/articles/s41598-018-26899-z
760 van Oldenborgh GJ, Doblas-Reyes FJ, Wouters B, Hazeleger W (2012) Decadal
761 prediction skill in a multi-model ensemble. Climate Dynamics 38(7-8):1263–
762 1280, DOI 10.1007/s00382-012-1313-4
763 O'Mahony M (1986) Sensory Evaluation of Food: Statistical Methods and Pro-
764 cedures. Series: Food science and technology, Marcel Dekker, Inc., New York,
765 DOI 10.1177/001088048402400413
766 Onarheim IH, Eldevik T, Årthun M, Ingvaldsen RB, Smedsrud LH (2015) Skillful
767 prediction of Barents Sea ice cover. Geophysical Research Letters 42(13):5364–
768 5371, DOI 10.1002/2015GL064359
769 Ortega P, Robson J, Menary M, Sutton R, Blaker A, Germe A, Hirschi J, Sinha B,
770 Hermanson L, Yeager S (2020) Labrador Sea sub-surface density as a precursor
771 of multi-decadal variability in the North Atlantic: a multi-model study. Earth
772 System Dynamics Discussions pp 1–25, DOI 10.5194/esd-2020-83
773 Penny SG, Akella S, Alves O, Bishop C, Buehner M, Chevallier M, Counillon
774 F (2017) Coupled data assimilation for integrated earth system analysis and
775 prediction. Bulletin of the American Meteorological Society 98(7):ES169–ES172,
776 DOI 10.1175/BAMS-D-17-0036.1

777 Penny SG, Bach E, Bhargava K, Chang CC, Da C, Sun L, Yoshida T (2019)
778 Strongly Coupled Data Assimilation in Multiscale Media: Experiments Using
779 a Quasi-Geostrophic Coupled Model. Journal of Advances in Modeling Earth
780 Systems 11(6):1803–1829, DOI 10.1029/2019MS001652
781 Polkova I, Köhl A, Stammer D (2014) Impact of initialization procedures on the
782 predictive skill of a coupled ocean-atmosphere model. Climate Dynamics 42(11-
783 12):3151–3169, DOI 10.1007/s00382-013-1969-4
784 Polkova I, Brune S, Kadow C, Romanova V, Gollan G, Baehr J, Glowienka-Hense
785 R, Greatbatch RJ, Hense A, Illing S, Köhl A, Kröger J, Müller WA, Pankatz
786 K, Stammer D (2019a) Initialization and Ensemble Generation for Decadal Cli-
787 mate Predictions: A Comparison of Different Methods. Journal of Advances in
788 Modeling Earth Systems 11(1):149–172, DOI 10.1029/2018MS001439
789 Polkova I, Köhl A, Stammer D (2019b) Climate-mode initialization for
790 decadal climate predictions. Climate Dynamics 53(11):7097–7111, DOI 10.1007/
791 s00382-019-04975-y, URL https://doi.org/10.1007/s00382-019-04975-y
792 Raj RP, Chatterjee S, Bertino L, Turiel A, Portebella M (2019) The Arctic Front
793 and its variability in the Norwegian Sea. The Arctic Front and its variability in
794 the Norwegian Sea pp 1–22, DOI 10.5194/os-2018-159
795 Rayner NA, Brohan P, Parker DE, Folland CK, Kennedy JJ, Vanicek M, Ansell
796 TJ, Tett SF (2006) Improved analyses of changes and uncertainties in sea surface
797 temperature measured in Situ since the mid-nineteenth century: The HadSST2
798 dataset. Journal of Climate 19(3):446–469, DOI 10.1175/JCLI3637.1
799 Richards CG, Straneo F (2015) Observations of water mass transformation and
800 eddies in the Lofoten basin of the Nordic seas. Journal of Physical Oceanography
801 45(6):1735–1756, DOI 10.1175/JPO-D-14-0238.1

802 Robson JI, Sutton RT, Smith DM (2012) Initialized decadal predictions of the
803 rapid warming of the North Atlantic Ocean in the mid 1990s. Geophysical Re-
804 search Letters 39(19):1–6, DOI 10.1029/2012GL053370
805 Sellar AA, Walton J, Jones CG, Wood R, Abraham NL, Andrejczuk M, Andrews
806 MB, Andrews T, Archibald AT, de Mora L, Dyson H, Elkington M, Ellis R,
807 Florek P, Good P, Gohar L, Haddad S, Hardiman SC, Hogan E, Iwi A, Jones
808 CD, Johnson B, Kelley DI, Kettleborough J, Knight JR, Köhler MO, Kuhlbrodt
809 T, Liddicoat S, Linova-Pavlova I, Mizielinski MS, Morgenstern O, Mulcahy J,
810 Neininger E, O'Connor FM, Petrie R, Ridley J, Rioual JC, Roberts M, Robert-
811 son E, Rumbold S, Seddon J, Shepherd H, Shim S, Stephens A, Teixiera JC,
812 Tang Y, Williams J, Wiltshire A, Griffiths PT (2020) Implementation of U.K.
813 Earth System Models for CMIP6. Journal of Advances in Modeling Earth Sys-
814 tems 12(4):1–27, DOI 10.1029/2019MS001946
815 Sluka TC, Penny SG, Kalnay E, Miyoshi T (2016) Assimilating atmospheric ob-
816 servations into the ocean using strongly coupled ensemble data assimilation.
817 Geophysical Research Letters 43(2):752–759, DOI 10.1002/2015GL067238
818 Smith DM, Cusack S, Colman AW, Folland CK, Harris GR, Murphy JM (2007)
819 Improved surface temperature prediction for the coming decade from a global
820 climate model. Science 317(5839):796–799, DOI 10.1126/science.1139540
821 Smith DM, Eade R, Scaife AA, Caron LP, Danabasoglu G, DelSole TM, Delworth
822 T, Doblas-Reyes FJ, Dunstone NJ, Hermanson L, Kharin V, Kimoto M, Mer-
823 ryfield WJ, Mochizuki T, Müller WA, Pohlmann H, Yeager S, Yang X (2019)
824 Robust skill of decadal climate predictions. npj Climate and Atmospheric Sci-
825 ence 2(1):1–10, DOI 10.1038/s41612-019-0071-y

826 Smith DM, Scaife AA, Eade R, Athanasiadis P, Bellucci A, Bethke I, Bilbao R,
827 Borchert LF, Caron L, Counillon F, Danabasoglu G, Delworth T, Dunstone NJ,
828 Flavoni S, Hermanson L, Keenlyside N, Kharin V, Kimoto M, Merryfield WJ,
829 Mignot J, Mochizuki T, Modali K, Monerie P, Müller WA, Nicolí D, Ortega P,
830 Pankatz K, Pohlmann H, Robson J, Ruggieri P, Swingedouw D, Wang Y, Wild S,
831 Yeager S, Yang X, Zhang L (2020) North Atlantic climate far more predictable
832 than models imply. Nature 583(July), DOI 10.1038/s41586-020-2525-0, URL
833 http://dx.doi.org/10.1038/s41586-020-2525-0
834 Swift JH, Aagard K (1981) Seasonal transitions and water mass formation in the
835 Iceland and Greenland seas * of the seas between Greenland and Scandinavia
836 is that of HELLAND- HANSEN and NANSEN (1909) (see Fig . 1 for geo-
837 graphical names). However , they lacked great submarine ri. Deep-sea research
838 28A(10):1107–1129
839 Tatebe H, Ishii M, Mochizuki T, Chikamoto Y, Sakamoto TT, Komuro Y, Mori
840 M, Yasunaka S, Watanabe M, Ogochi K, Suzuki T, Nishimura T, Kimoto M
841 (2012) The initialization of the MIROC climate models with hydrographic data
842 assimilation for decadal prediction. Journal of the Meteorological Society of
843 Japan 90(A):275–294, DOI 10.2151/jmsj.2012-A14
844 Thomason LW, Ernest N, Millán L, Rieger L, Bourassa A, Vernier JP, Man-
845 ney G, Luo B, Arfeuille F, Peter T (2018) A global space-based stratospheric
846 aerosol climatology: 1979-2016. Earth System Science Data 10(1):469–492, DOI
847 10.5194/essd-10-469-2018
848 Tjiputra JF, Roelandt C, Bentsen M, Lawrence DM, Lorentzen T, Schwinger
849 J, Seland, Heinze C (2013) Evaluation of the carbon cycle components in the
850 Norwegian Earth System Model (NorESM). Geoscientific Model Development

851 6(2):301–325, DOI 10.5194/gmd-6-301-2013

852 Vera C, Barange M, Dube OP, Goddard L, Griggs D, Kobysheva N, Odada
853 E, Parey S, Polovina J, Poveda G, Seguin B, Trenberth K (2010) Needs as-
854 sessment for climate information on decadal timescales and longer. Procedia
855 Environmental Sciences 1(1):275–286, DOI 10.1016/j.proenv.2010.09.017, URL
856 http://dx.doi.org/10.1016/j.proenv.2010.09.017

857 Vertenstein M, Craig T, Middleton A, Feddema D, Fischer C (2012)
858 CESM1.0.4 Users Guide. Tech. rep., National Center for Atmospheric Research
859 (NCAR), URL http://www.cesm.ucar.edu/models/cesm1.0/cesm/cesm_doc_
860 1_0_4/book1.html

861 Wang Y, Counillon F, Bethke I, Keenlyside N, Bocquet M, Shen Ml (2017) Op-
862 timising assimilation of hydrographic profiles into isopycnal ocean models with
863 ensemble data assimilation. Ocean Modelling 114:33–44, DOI 10.1016/j.ocemod.
864 2017.04.007, URL http://dx.doi.org/10.1016/j.ocemod.2017.04.007

865 Watanabe M, Kimoto M (2000) On the persistence of decadal SST anomalies
866 in the North Atlantic. Journal of Climate 13(16):3017–3028, DOI 10.1175/
867 1520-0442(2000)013⟨3017:OTPODS⟩2.0.CO;2

868 Yang C, Masina S, Storto A (2017) Historical ocean reanalyses (19002010) using
869 different data assimilation strategies. Quarterly Journal of the Royal Meteoro-
870 logical Society 143(702):479–493, DOI 10.1002/qj.2936

871 Yeager S (2020) The abyssal origins of North Atlantic decadal predictability. Cli-
872 mate Dynamics DOI 10.1007/s00382-020-05382-4, URL https://doi.org/10.
873 1007/s00382-020-05382-4

874 Yeager S, Karspeck A, Danabasoglu G, Tribbia J, Teng H (2012) A Decadal Predic-
875 tion Case Study: Late Twentieth-Century North Atlantic Ocean Heat Content.

876 Journal of Climate 25:5173–5189, DOI 10.1175/JCLI-D-11-00595.1

877 Yeager SG, Robson JI (2017) Recent Progress in Understanding and Predicting
878 Atlantic Decadal Climate Variability. Current Climate Change Reports pp 112–
879 127, DOI 10.1007/s40641-017-0064-z

880 Ypma SL, Georgiou S, Dugstad JS, Pietrzak JD, Katsman CA (2020) Pathways
881 and Water Mass Transformation Along and Across the Mohn-Knipovich Ridge
882 in the Nordic Seas. Journal of Geophysical Research: Oceans 125(9), DOI 10.
883 1029/2020JC016075

Tables and Figures

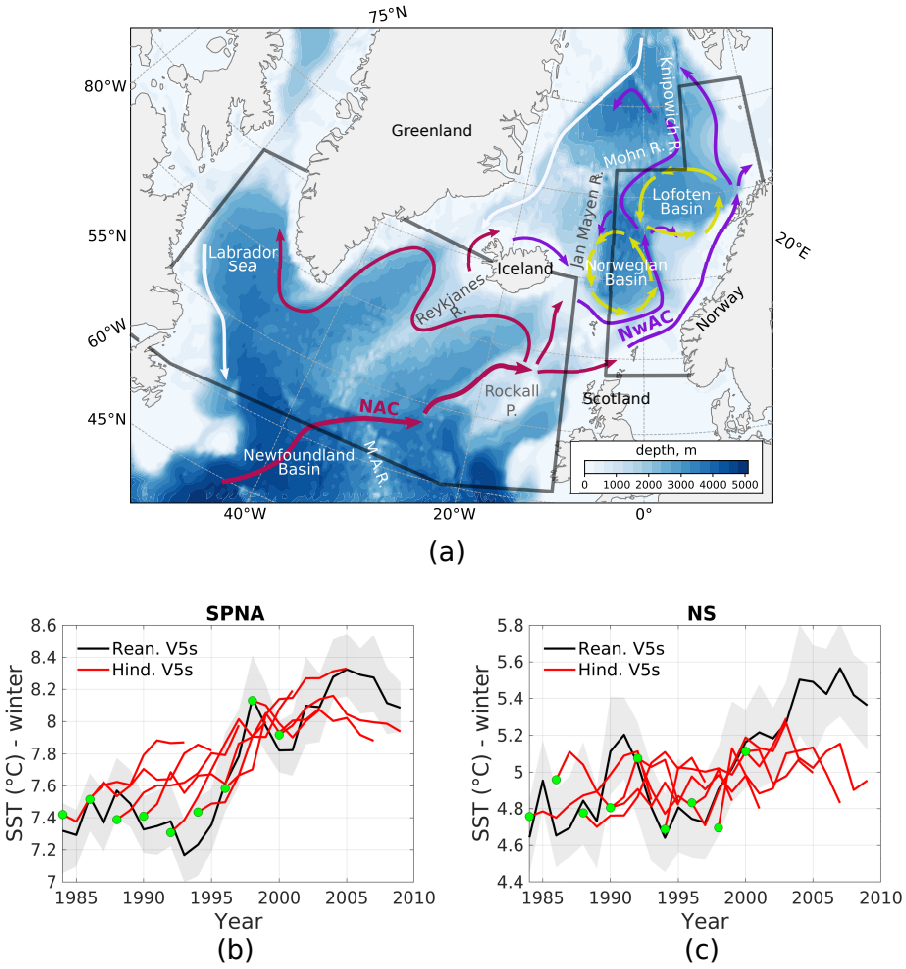

Fig. 1 a) Schematic diagram highlighting the SPNA area (60W-10W, 50N-65N), according to Robson et al. (2012), and the NS, according to Asbjørnsen et al. (2019), in transparent black. The main currents analyzed in this study are the North Atlantic Current (NAC; pink) and the Norwegian Atlantic Current (NwAC; purple) with its recirculation (yellow). The white arrows indicate cold and fresh surface water from the Arctic. Ridges and Plateau are indicated with R. and P. respectively. b) Hindcast winter SST in the SPNA in version V5s (red). c) Hindcast winter SST in the NS in version V5s (red). The black curves show the respective reanalysis used to initialize the model. Grey shading represents the spread of the ensemble members of the reanalysis. The starting time of each hindcast is indicated by a green circle.

Table 1 NorCPM versions characteristics. The number of members for those versions with more than 5 is in brackets.

Version	Assimilated Data	CMIP Forcing	Assimilation Method	Ensemble Size	Climatology Reference Period	Initialization Frequency
V5 (NorESM1-ME)	SST (HadSST2)	CMIP5	WCDA	5(20)	1950-2010	Every 2 years
V5w (NorESM1-ME)	SST (HadSST2) Hydrographic data (EN4)	CMIP5	WCDA	5	1980-2010	Every 2 years
V5s (NorESM1-ME)	SST (HadSST2) Hydrographic data (EN4)	CMIP5	SCDA	5(10)	1980-2010	Every 2 years
V6w (NorCPM1)	SST (HadSST2) Hydrographic data (EN4)	CMIP6	WCDA	5(10)	1980-2010	Every 1 years
V6s (NorCPM1)	SST (HadSST2) Hydrographic data (EN4)	CMIP6	SCDA	5(10)	1950-2010	Every 1 years

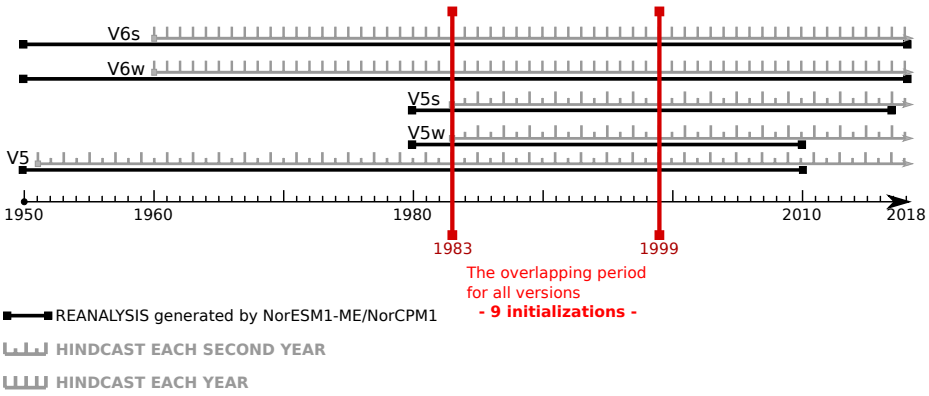

Fig. 2 Experiment time period for each version of NorCPM. The black line indicate the period of the reanalysis used and the respective hindcast (grey line). The vertical size of the grey lines indicate the initialization frequency of each experiment, which varies between every year (V6w and V6s) and every second year (V5, V5w, and V5s). The red lines indicate the overlapping period used for comparison.

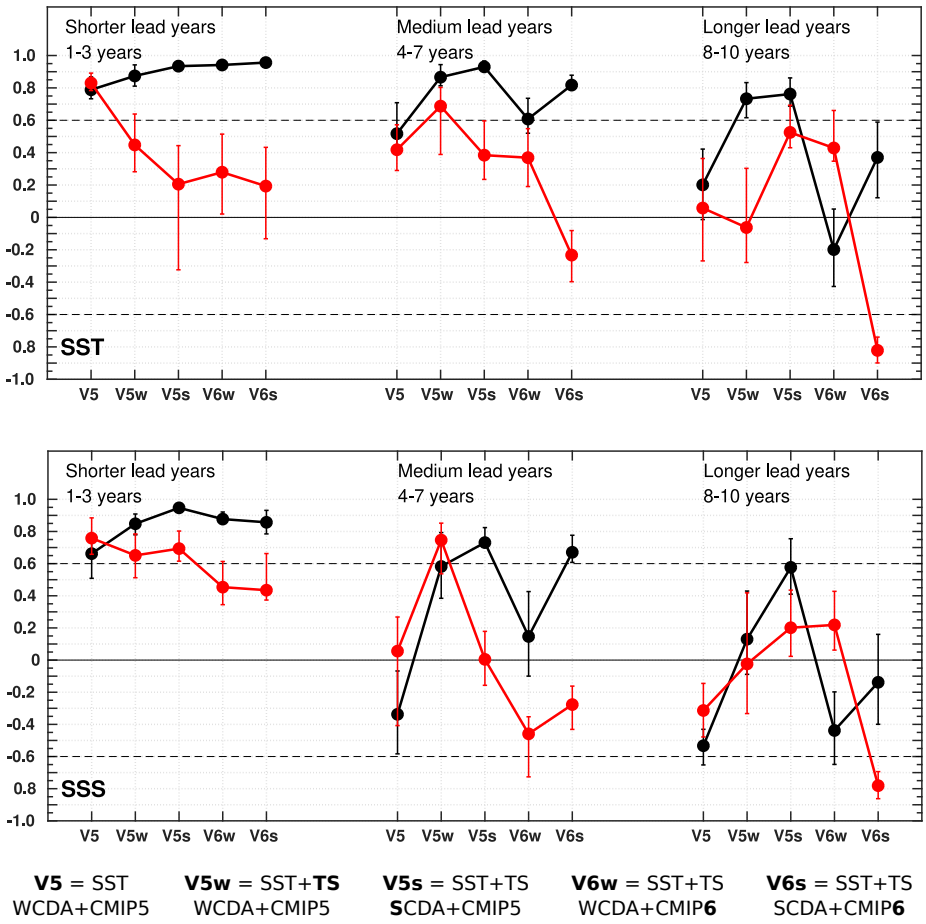

Fig. 3 Anomaly correlation coefficient for winter (Jan-Apr) SST (upper panel) and SSS (lower panel) at shorter, medium and longer lead years in different versions of NorCPM in the SPNA (black) and in the NS (red). The vertical bars indicate the 25 and 75 percentiles and the dashed grey lines show the 90% significance level.

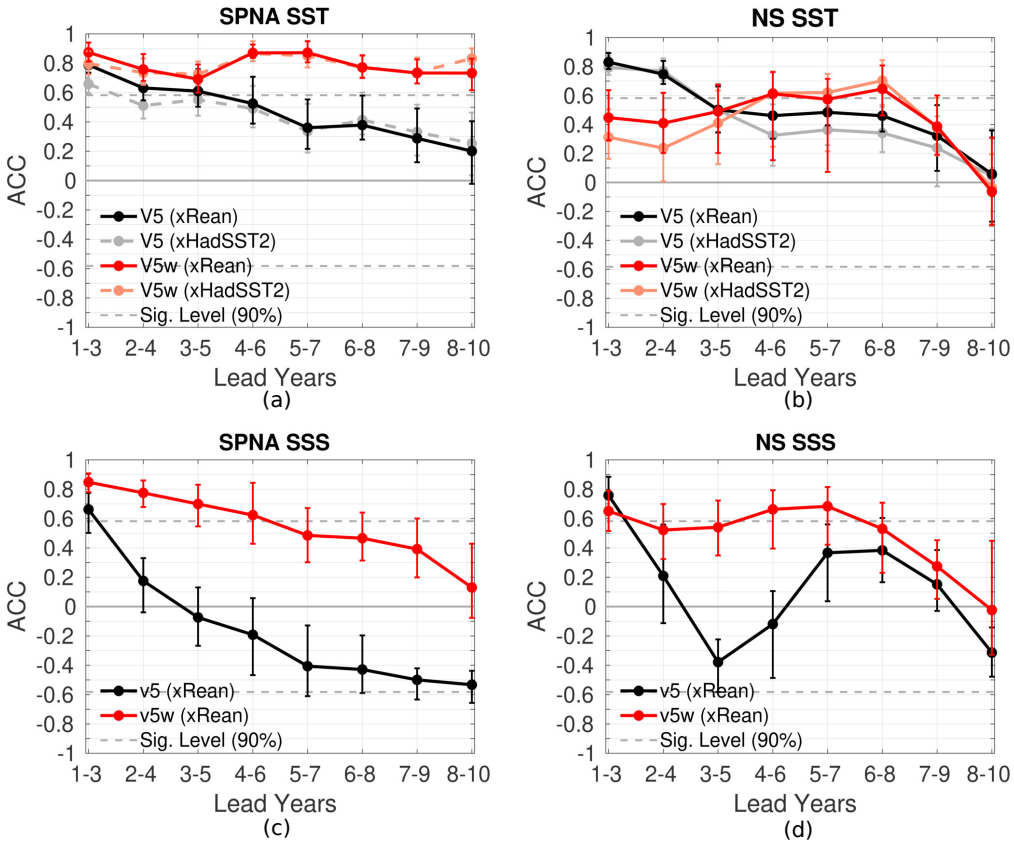

Fig. 4 Anomaly correlation coefficient for winter (Jan-Apr) between NorCPM hindcasts V5 and V5w and the respective reanalysis and HadSST2 for SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles and the dashed grey lines show the 90% significance level.

.

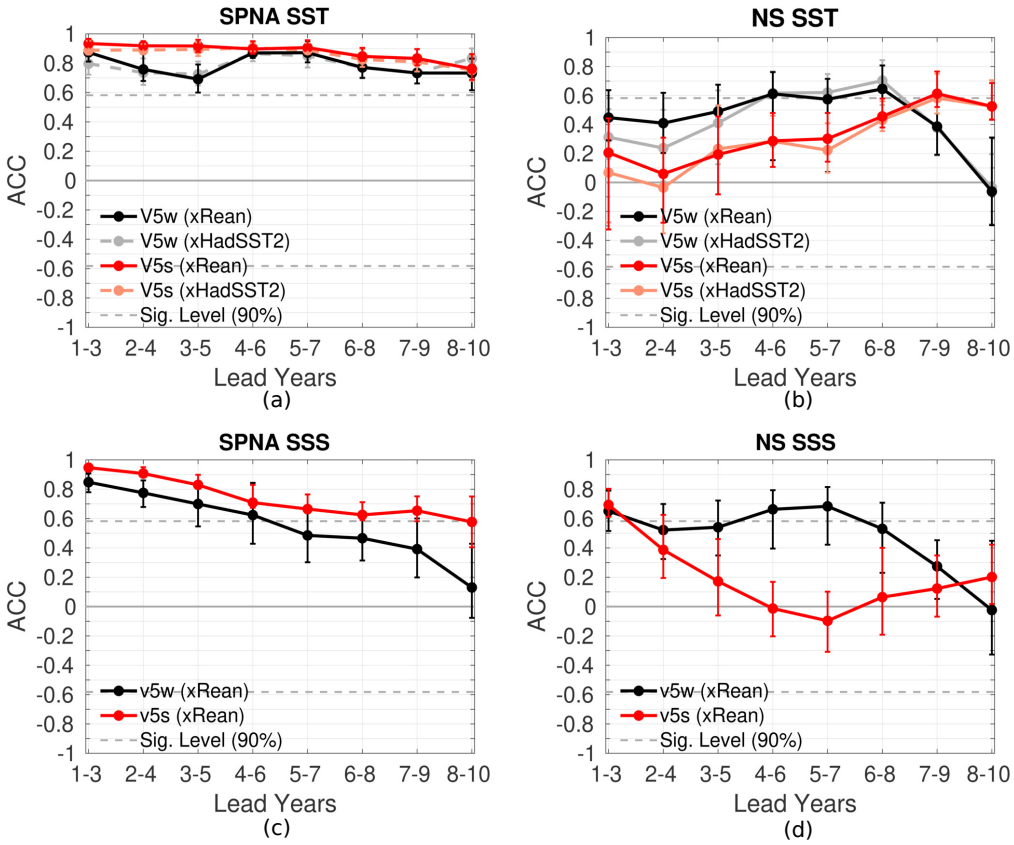

Fig. 5 Anomaly correlation coefficient between NorCPM hindcasts V5w and V5s and the respective reanalysis and HadSST2 for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles and the dashed grey lines show the 90% significance level.

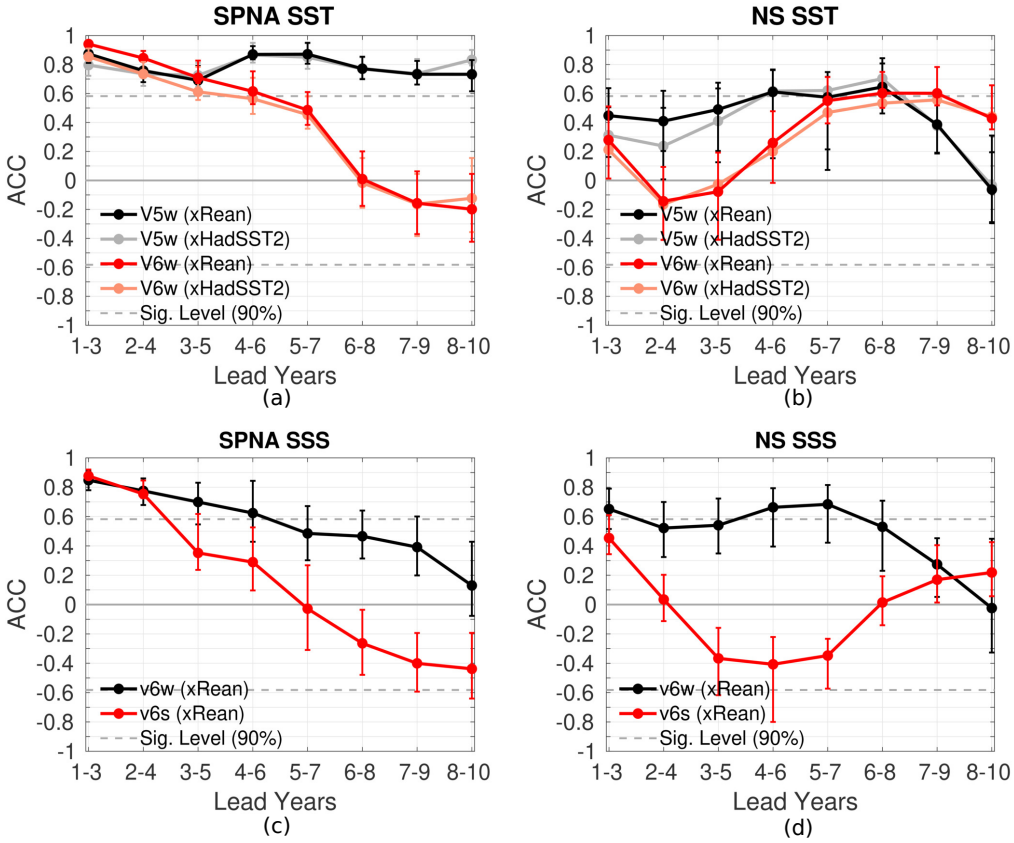

Fig. 6 Anomaly correlation coefficient between NorCPM hindcasts V5w and V6w and the respective reanalysis and HadSST2 for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles and the dashed grey lines show the 90% significance level.

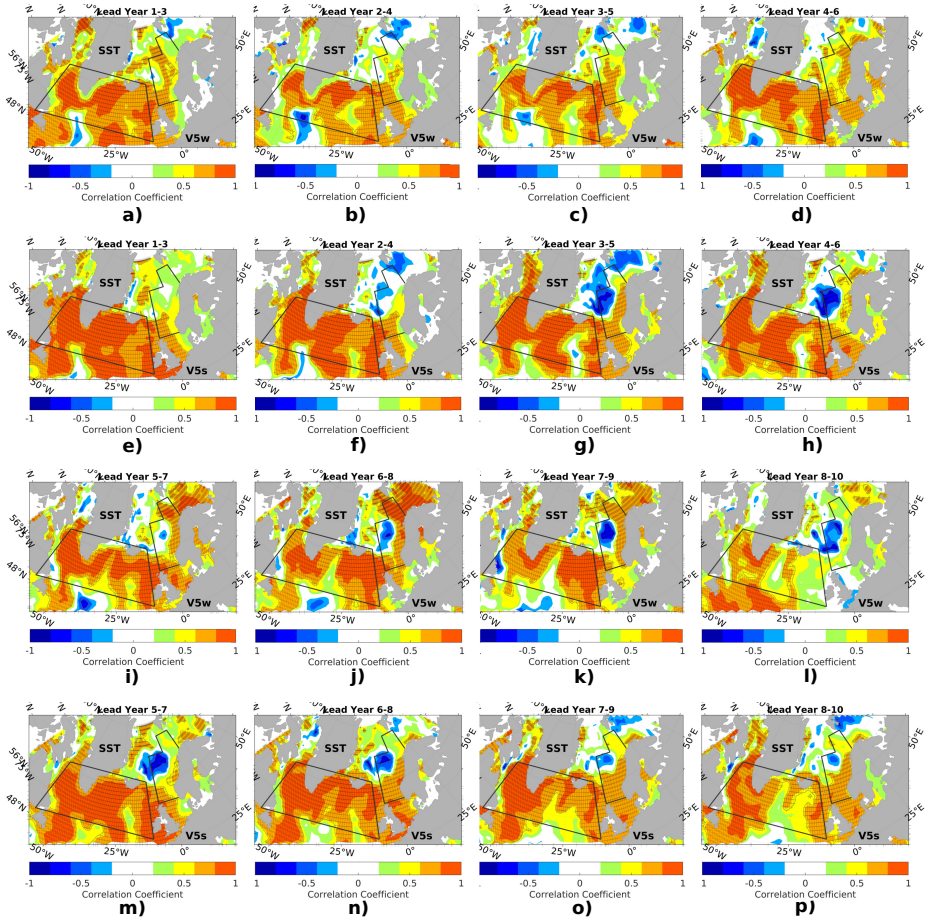

Fig. 7 Spatial anomaly correlation coefficient of SST in each grid point for V5w: 1-3 years (a), 2-4 years (b), 3-5 years (c), 4-6 years (d), 5-7 years (i), 6-8 years (j), 7-9 years (k), 8-10 years (l). For V5s: 1-3 years (e), 2-4 years (f), 3-5 years (g) and 4-6 years (h), 5-7 years (m), 6-8 years (n), 7-9 years (o), 8-10 years (p). The SPNA and NS regions are shown by the black boxes. Significant values are indicated by the hatched regions.

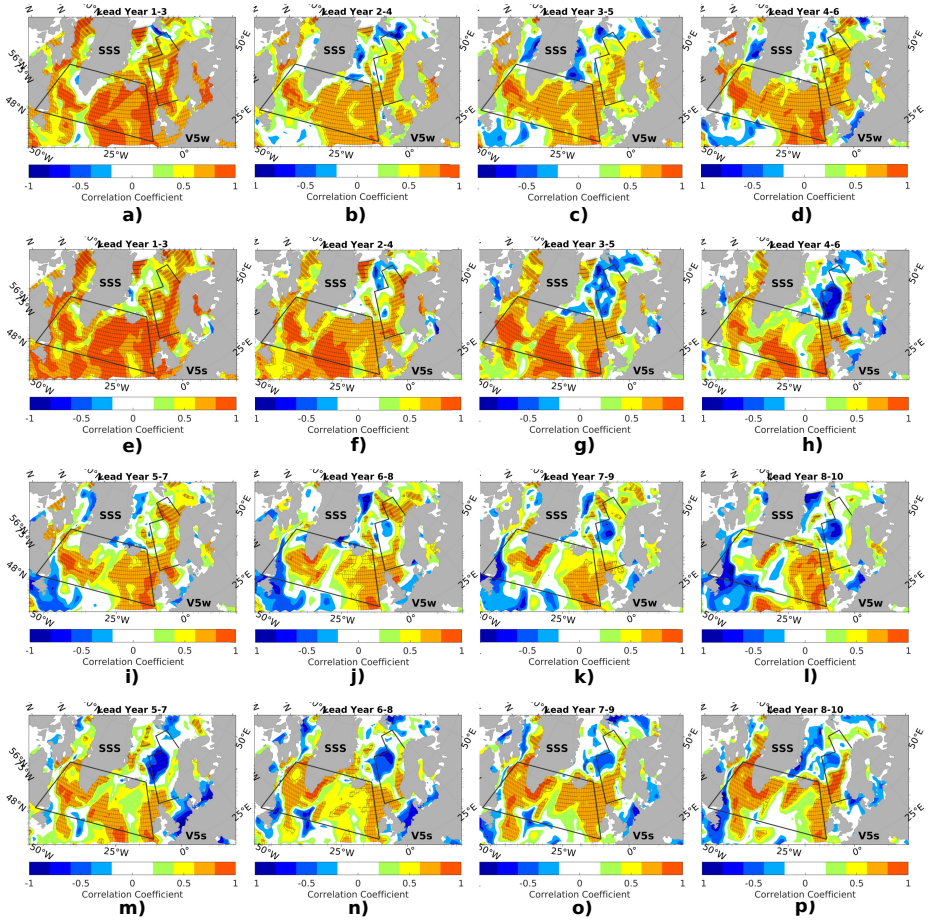

Fig. 8 Spatial anomaly correlation coefficient of SSS in each grid point for V5w: 1-3 years (a), 2-4 years (b), 3-5 years (c), 4-6 years (d), 5-7 years (i), 6-8 years (j), 7-9 years (k), 8-10 years (l). For V5s: 1-3 years (e), 2-4 years (f), 3-5 years (g) and 4-6 years (h), 5-7 years (m), 6-8 years (n), 7-9 years (o), 8-10 years (p). The SPNA and NS regions are shown by the black boxes. Significant values are indicated by the hatched regions.

7 Supplementary Figures

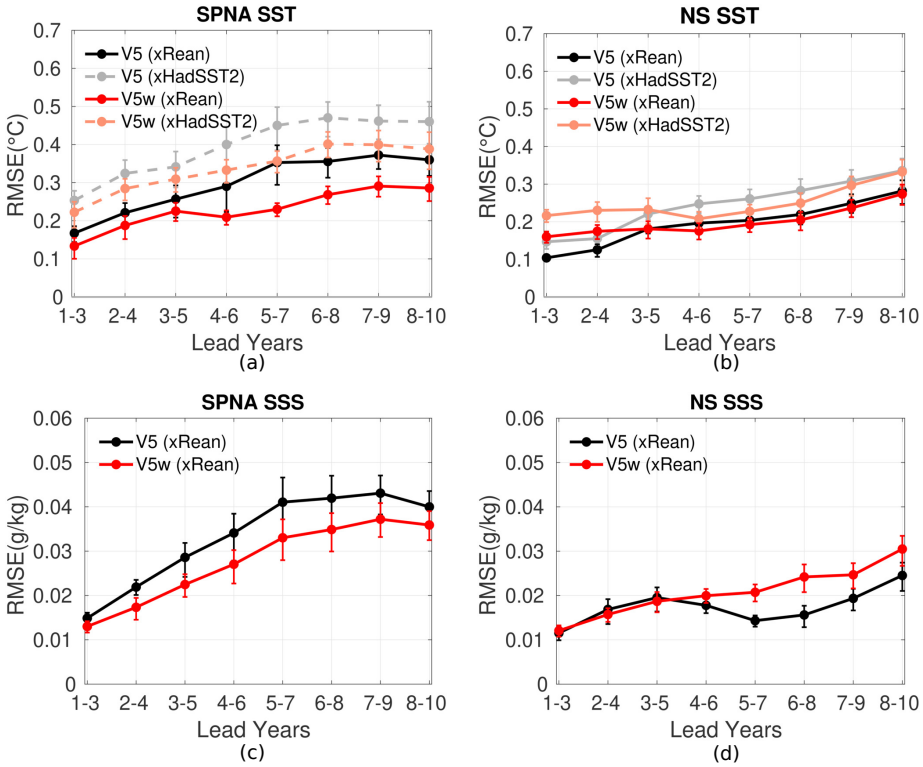

Fig. 9 RMSE between NorCPM hindcasts V5 and V5w and the respective reanalysis and HadSST2 for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles.

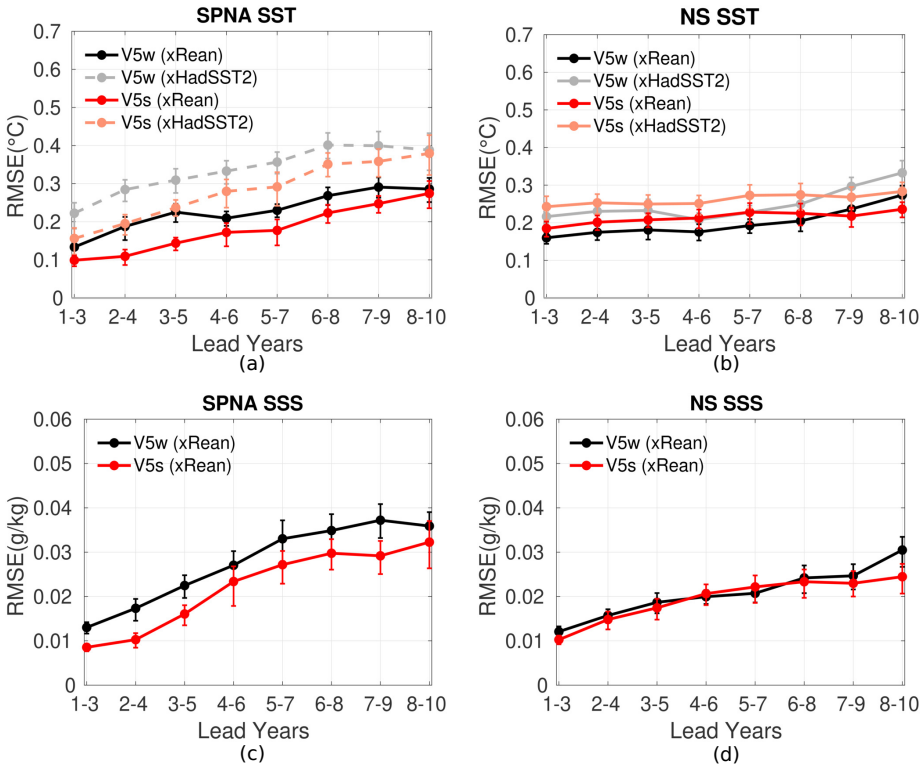

Fig. 10 RMSE between NorCPM hindcasts V5w and V5s and the respective reanalysis and HadSST2 for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles.

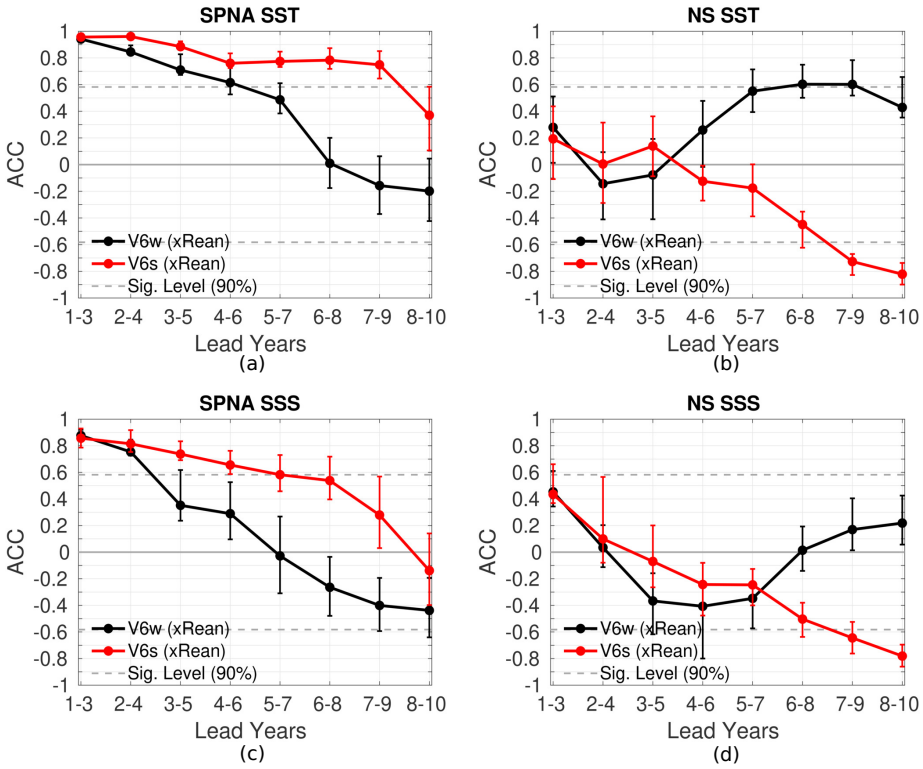

Fig. 11 Anomaly correlation coefficient between NorCPM hindcasts V6w and V6s and the respective reanalysis for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles and the dashed grey lines show the 90% significance level.

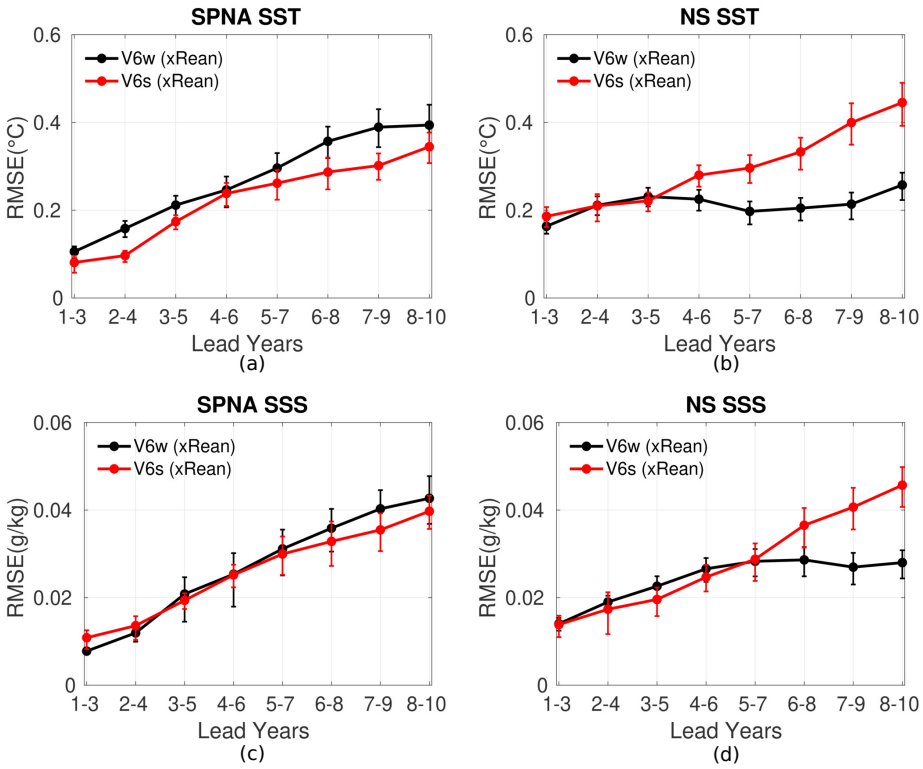

Fig. 12 RMSE between NorCPM hindcasts V5w and V6w and the respective reanalysis for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles.

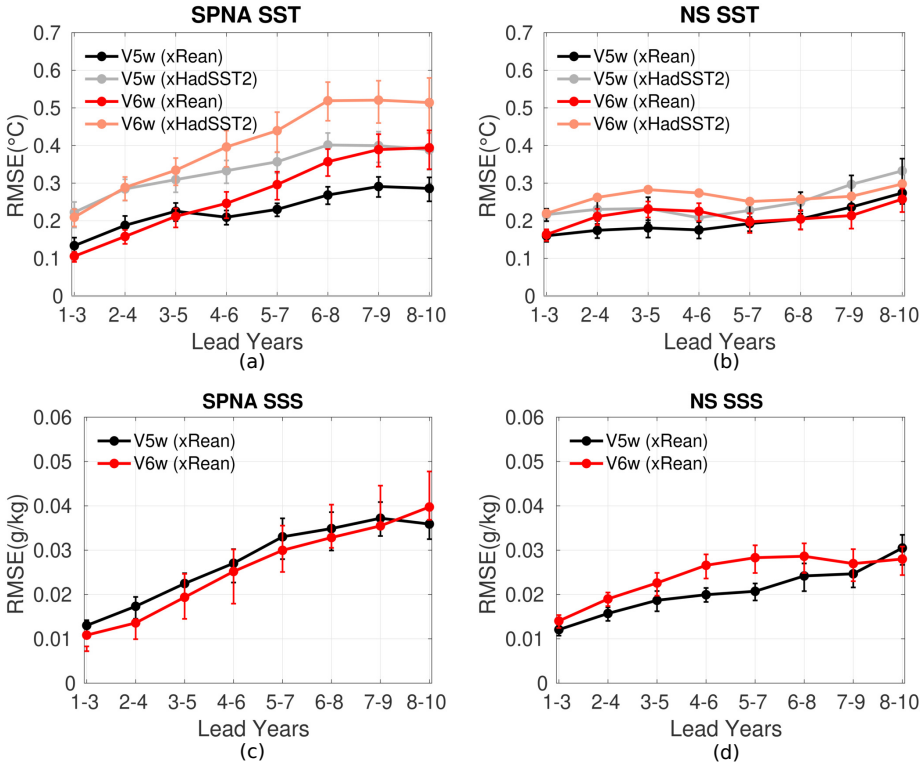

Fig. 13 RMSE between NorCPM hindcasts V5w and V6s and the respective reanalysis and HadSST2 for winter (Jan-Apr) SST from SPNA (a) and NS (b) and SSS from SPNA (c) and NS (d). The vertical bars indicate the 25 and 75 percentiles.

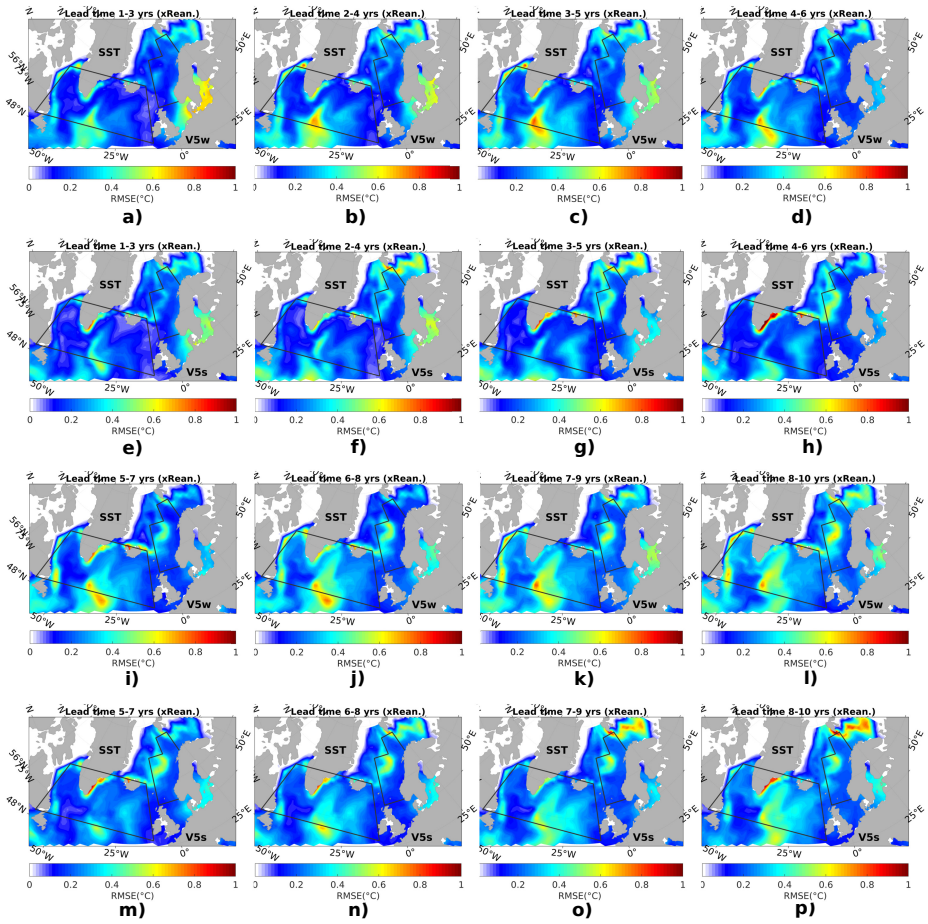

Fig. 14 Spatial RMSE of SST in relation to the respective reanalysis in each grid point for V5w: 1-3 years (a), 2-4 years (b), 3-5 years (c), 4-6 years (d), 5-7 years (i), 6-8 years (j), 7-9 years (k), 8-10 years (l). For V5s: 1-3 years (e), 2-4 years (f), 3-5 years (g) and 4-6 years (h), 5-7 years (m), 6-8 years (n), 7-9 years (o), 8-10 years (p). The SPNA and NS are defined as area averaged according Figure 1.

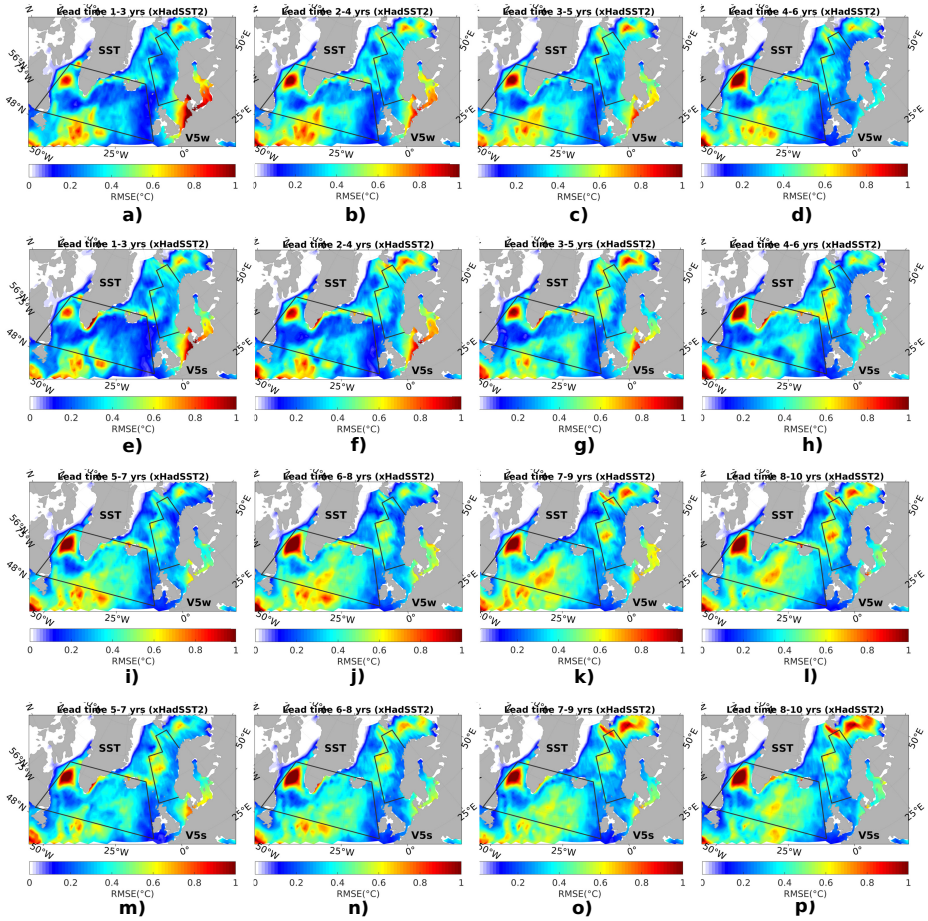

Fig. 15 Spatial RMSE of SST in relation to HadSST2 in each grid point for V5w: 1-3 years (a), 2-4 years (b), 3-5 years (c), 4-6 years (d), 5-7 years (i), 6-8 years (j), 7-9 years (k), 8-10 years (l). For V5s: 1-3 years (e), 2-4 years (f), 3-5 years (g) and 4-6 years (h), 5-7 years (m), 6-8 years (n), 7-9 years (o), 8-10 years (p). The SPNA and NS are defined as area averaged according Figure 1.

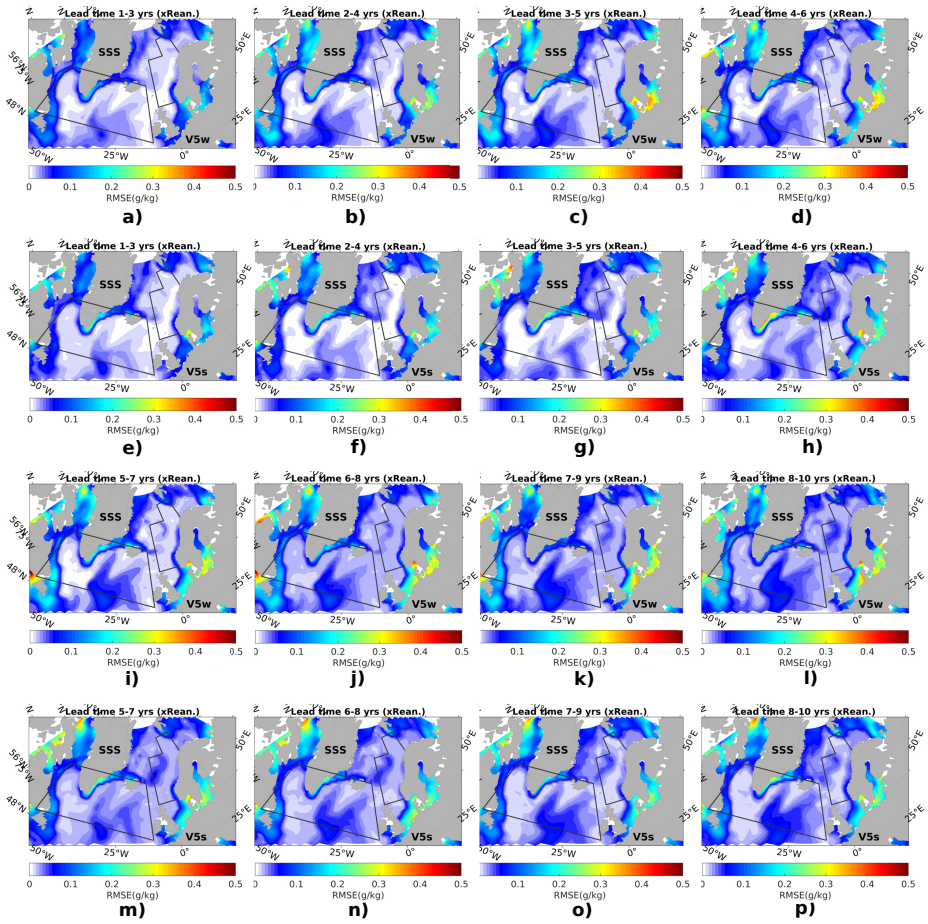

Fig. 16 Spatial RMSE of SSS in relation to the respective reanalysis in each grid point for V5w: 1-3 years (a), 2-4 years (b), 3-5 years (c), 4-6 years (d), 5-7 years (i), 6-8 years (j), 7-9 years (k), 8-10 years (l). For V5s: 1-3 years (e), 2-4 years (f), 3-5 years (g) and 4-6 years (h), 5-7 years (m), 6-8 years (n), 7-9 years (o), 8-10 years (p). The SPNA and NS are defined as area averaged according Figure 1.