



**HORIZON 2020
THEME SC5-2017**



European Climate Prediction system

(GRANT AGREEMENT 776613)

European Climate Prediction system (EUCP)

Deliverable 1.3

“Recommendations for the development of a new
generation of climate forecast systems.”

Deliverable Title		
Brief Description		
WP number	WP1	
Lead Beneficiary	CMCC	
Contributors	<p><i>(BSC): Francisco Doblas-Reyes, Markus Donat; Roberto Bilbao, Aude Carreric, Juan Acosta Navarro, Pablo Ortega, Yohan Ruprich-Robert (CMCC): Panos Athanasiadis; (CNRS-IPSL): Didier Swingedouw; (UKMO): Leon Hermanson, Doug Smith, Adam Scaife, Anca Brookshaw; (SMHI): Tim Kruschke, Shiyu Wang, Frederik Schenk; (UOXF): Daniel J. Befort, Christopher H. O'Reilly, Antje Weisheimer.</i></p>	
Creation Date	02/08/2021	
Version Number	v3	
Version Date	15/12/2021	
Deliverable Due Date	31/01/2022	
Actual Delivery Date	21/01/2022	
Nature of the Deliverable	R	<i>R – Report</i> <i>P - Prototype</i> <i>D - Demonstrator</i> <i>O - Other</i>
Dissemination Level/ Audience	x	<i>PU - Public</i> <i>PP - Restricted to other programme participants, including the Commission services</i>

		<i>RE - Restricted to a group specified by the consortium, including the Commission services</i>	
		<i>CO - Confidential, only for members of the consortium, including the Commission services</i>	
Version	Date	Modified by	Comments
V0	02/08/2021	Panos Athanasiadis	Created to collect contributions
V1	30/09/2021	Partners	Aiming at a full draft
V2	15/12/2021	Panos Athanasiadis	Complete for internal review
V3	15/02/2022	Carol McSweeney, Jason Lowe	Coordination Team review

Table of Contents

1. Executive Summary	5
2. Project Objectives	6
3. Detailed Report	7
3.1 Benefits of model initialization for Arctic sea-ice predictions (BSC)	
3.2 Improving large-ensemble multi-model predictions on mid-latitudes by exploiting process-driven metrics describing key teleconnection mechanisms (BSC)	
3.3 Towards a high-resolution prediction system (BSC)	
3.4 Impact of Volcanic Eruptions in Decadal Predictions (BSC)	
3.5 Increasing the ensemble size is key for achieving higher skill in decadal predictions of North Atlantic blocking and the NAO (CMCC)	
3.6 Increased model resolution contributes to mitigating long-standing SST and blocking biases in the Euro-Atlantic sector (CMCC)	
3.7 Atmospheric model resolution determines the response to Gulf Stream SST front variability (CMCC)	
3.8 Investigation of the impact of Greenland ice sheet melt on decadal predictability (CNRS-IPSL)	
3.9 Priorities for development of improved climate forecast systems (Met Office)	
3.10 Benefits from higher resolution and coupled assimilation approaches (SMHI)	
3.11 Representation of model uncertainty in multiannual predictions (UOXF)	
3.12 To what extent predicting the North Atlantic variability could improve our ability to predict climate over the globe? (BSC, CMCC, CNRS-IPSL, Met Office)	
4. Lessons learnt	41
5. Links built	42
6. Acronyms	42
7. References	42

1. Executive Summary

This deliverable report gives an overview of WP1 activities aiming at providing recommendations for the development of a new generation of climate forecast systems following evidence from recent research conducted by EUCP partners contributing to the Task 1.3. A number of key messages emerging from these activities are outlined below, while all the associated details (including original results, most still unpublished) are included in Section 3.

- It has been shown that realistically initializing sea-ice in decadal predictions reduces errors in the representation of sea-ice extent (**S3.1**), which controls high-latitude air–sea fluxes and impacts the atmospheric circulation.
- Sub-sampling techniques can improve skill (**S3.2**) and thus such methods should be further explored.
- Increasing model resolution improves the representation of deep convection in the Labrador Sea (**S3.3**) with important benefits for the AMOC, model biases and the associated drift in decadal predictions. Such improvements can also be important for seasonal predictability.
- It is shown that including volcanic aerosol forcing in decadal predictions has a significant impact on the forecasts (**S3.4**), thus pointing to the need of re-running such forecasts after major future volcanic eruptions.
- Evidence is presented, corroborating similar findings in recent literature, that increasing the ensemble size of decadal predictions monotonically increases the predictive skill for the NAO and North Atlantic blocking (**S3.5**). This is the case also in seasonal forecasts (e.g., Scaife et al., 2014; Athanasiadis et al., 2017). Thus, it is recommended for future decadal prediction systems to increase their ensemble sizes so as to achieve larger multi-model ensembles than currently available.
- Increasing the horizontal resolution in climate models, particularly towards an eddy-permitting ocean resolution (0.25°), is shown to mitigate important long-standing SST biases in the extratropical North Atlantic with beneficial impacts also on European blocking and the North Atlantic jet (**S3.6**).
- Using a multi-model set of atmosphere-only historical simulations it is demonstrated that only high-resolution ($<50\text{km}$) models are able to realistically simulate the atmospheric circulation response to extratropical SST forcing (**S3.7**). This has important implications for predictability in the Euro-Atlantic sector, indicating that future operational climate prediction systems should have similar, or higher resolutions.
- Greenland ice-sheet melting is shown to be important for SST errors at the beginning of the North Atlantic current and the broader subpolar gyre. Thus, to reduce the respective SST biases better representation of continental ice-sheet melting is needed (**S3.8**).
- A summary of recommendations is provided following recent published and unpublished research pointing to the need for: (i) larger ensemble sizes, (ii) higher oceanic resolution, towards 0.25° (iii) better understanding of model deficiencies to better account for and ultimately resolve such deficiencies, including model biases and the signal-to-noise problem (**S3.9**).
- Significant improvements in skill are demonstrated for higher-resolution decadal hindcasts, especially for the first forecast year in the Pacific basin (**S3.10**). This points to a potential for better predictions in the extended seasonal range (up to 12 months) subject to increases in model

resolution, as mentioned above. Also, advances in assimilation techniques are expected to benefit climate predictions and allow more robust forecast verification.

- The use of stochastic parameterization schemes is shown to benefit climate predictions in the tropics from the seasonal range up to 2 years, effectively accounting for model uncertainty as in multi-model ensembles (S3.11). Consequently, the implementation of such schemes (stochastic physics) is recommended for future climate prediction systems.
- Finally, recent research based on coordinated multi-model experiments (DCPP-C) also points to the need to reducing model biases so as to better simulate the climatic impacts of the AMV (S3.12).

2. Project Objectives

WITH THIS DELIVERABLE, EUCP HAS CONTRIBUTED TO THE ACHIEVEMENT OF THE FOLLOWING OBJECTIVES (DESCRIPTION OF ACTION, SECTION 1.1):

No.	Objective	Yes	No
1	Develop an ensemble climate prediction system based on high-resolution climate models for the European region for the near-term (1–40 years), including improved methods used to characterise uncertainty in climate predictions, regional downscaling, and evaluation against observations [WP1, WP2, WP3, WP5].	X	
2	Use the climate prediction system to produce consistent, authoritative and actionable climate information [WP1, WP2, WP3, WP4, WP5, WP6]. This information will be co-designed with users to constitute a robust foundation for Europe-wide climate service activities to support regional, national and local climate-related risk assessments and climate change adaptation programmes.		X
3	Demonstrate the value of this climate prediction system through high impact extreme weather events in the near past and near future (1-40 years) drawing on convection permitting regional climate models translated into risk information for, and more importantly with, targeted end users [WP3, WP4, WP5, WP6].		X
4	Develop, and publish, methodologies, good practice and guidance for producing and using EUCP’s authoritative climate predictions for 1–40 year timescales [WP1, WP2, WP3, WP4, WP5, WP6].	X	

3. Detailed Report

Introduction

The following subsections (3.1–3.12) provide scientific evidence/background for a series of recommendations towards the development of a new generation of climate forecast systems. Such recommendations include: increasing ensemble sizes and model resolution (both in the ocean and the atmosphere), better understanding of model deficiencies in the current forecast systems, improving the representation of physical processes (such as continental ice-sheet melting), being able to produce quickly the volcanic forcings corresponding to major future eruptions so as incorporate such forcings in climate predictions, implementing better parameterizations of subgrid processes as well as new parameterizations accounting for model uncertainty (such as stochastic physics schemes), and continuing working on reducing model biases, which have multiple implications for the quality of climate predictions. Admittedly, these recommendations spring from the work of the individual partners and thus do not represent the result of an exhaustive and orchestrated multi-model analysis of current forecast systems. Nevertheless, they point to specific improvements that the new generation of climate prediction systems can benefit from.

3.1 — Benefits of model initialization for Arctic sea-ice predictions (BSC)

Methods: Table 3.1 shows the CMIP6 model experiments used at BSC for the analysis in Sections 3.1 and 3.2. In Section 3.1, the EC-Earth model results (initialization approach i1) were excluded due to a well-known problem in the decadal predictions in which the Labrador convection collapses a few years after initialization with large negative impacts on Arctic sea ice (Bilbao et al., 2021). In Section 3.2, only one of the versions of NorCPM1 was used due to unavailability of the necessary data.

Table 3.1.1: Description of the CMIP6 experiments used in the decadal prediction analysis.

Model	DCPP members	Historical members	Start dates available
IPSL-CM6A-LR	10	10	1961-2017
MPI-ESM1-2-HR	5	5	1960-2019
NORCPM1 (two initializations)	10/10	10/10	1960-2018
CANESM5	10	10	1961-2017
MIROC	10	10	1960-2018
EC-Earth (initialization i1)	10	10	1960-2018

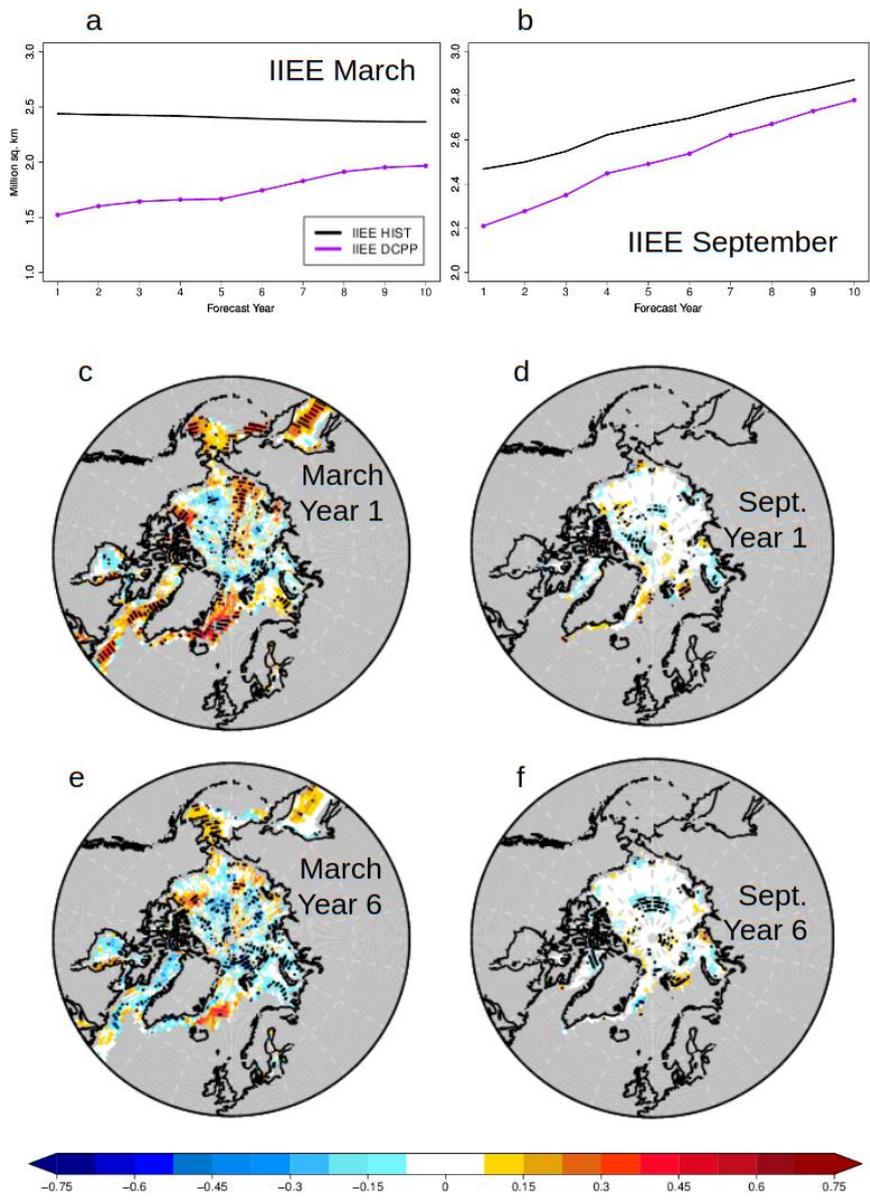


Figure 3.1.1: Ensemble mean, pan-Arctic integrated ice edge error (IIEE, Goessling et al., 2016) as a function of forecast year for 6-model ensembles of CMIP6 in DCPP (55 members) and Historical experiments (55 members) in a) March and b) September. The IIEE is computed from the SIC absolute values. Difference in anomaly correlation of sea-ice concentration between DCPP and Historical 6-model ensemble means for c) March monthly-mean of the 1st forecast year, d) September monthly-mean of the 1st forecast year, e) March monthly-mean of the 6th forecast year and d) September monthly-mean of the 6th forecast year. Dots indicate statistical significance at the 95% confidence level. Observational reference: HadISSTv1.1.

At BSC, we have investigated the added value of model initialization for Arctic sea-ice decadal predictions. For this purpose, two ensembles (6 models - 55 members each, Table 3.1.1) of multi-model simulations are used: DCPP, which is initialized and driven with historical changes in external forcings, and Historical, which is only driven by external forcing changes. The comparison between them allows

for the isolation and quantification of the benefits of model initialization on predictive skill. Both ensembles show a very similar level of Arctic sea-ice extent decrease in March and September in recent decades (not shown), in line with observations, but underestimate the magnitude of the trend, whose representation is not improved with initialization. However, a more appropriate verification metric of sea-ice skill, the integrated ice edge error (IIEE; Goesling et al. 2016), that integrates absolute errors in the position of the simulated sea-ice edge, shows advantages for both, March and September sea-ice up to forecast year 10 in DCPD with respect to Historical (Fig. 3.1.1 a–b). Not surprisingly, the largest differences between both ensembles are seen in forecast year 1 and decrease with time, a difference that is statistically significant in all forecast years, for both March and September sea-ice. Spatial maps of skill differences of sea-ice concentration (SIC) anomalies between DCPD and Historical also show superior performance of DCPD in the 1st March (Fig. 3.1.1.c) in the Atlantic and Pacific sectors of the Arctic, with little effect in the central Arctic. The 6th forecast March shows a much more limited beneficial impact of initialization (Fig. 3.1.1.e), visible only in the Pacific sector. In September, the impact of initialization is much smaller in general, both in forecast years 1 and 6 (Fig. 3.1.1.d,f), despite having a significantly lower IIEE in DCPD than Historical.

3.2 — Improving large-ensemble multi-model predictions on mid-latitudes by exploiting process-driven metrics describing key teleconnection mechanisms (BSC)

Previous analyses of seasonal re-forecasts revealed that one of the possible reasons behind the limited winter skill over the Eurasian continent is an overall underestimation of an observed teleconnection between the Barents-Kara sea-ice and atmospheric circulation over Northern Eurasia (Acosta Navarro et al., 2020). Based on that work, we subsampled the members from a 6-model DCPD ensemble in which the teleconnection is strongest and closest to the observed values (Fig. 3.2.1.a). In this case, instead of limiting the analysis to the first winter (DJF) after initialization, DJF means of years 2–10 were used together with the sea-ice extent in the Barents-Kara region the previous Novembers (i.e. from forecasts years 1–9), to compute the teleconnection strength at decadal timescales. It has been found that skill for winter sea level pressure (Fig. 3.2.1.b) in the Eurasian region and also for surface variables, such as air temperature (Fig. 3.2.1.c) and precipitation (not shown), can be substantially and significantly improved, when this methodology is applied. In fact, these results show stronger positive impacts on the skill for forecast years 2–10 than for the seasonal forecast analysed previously. These results are in line with other recent studies highlighting the potential of physical-constraints in the analysis of large prediction ensembles to circumvent some of the systematic model problems, including the signal-to-noise paradox (Dobrynin et al., 2018; Smith et al., 2020; Donegan et al., 2020).

Process-based constrained multi-model decadal predictions

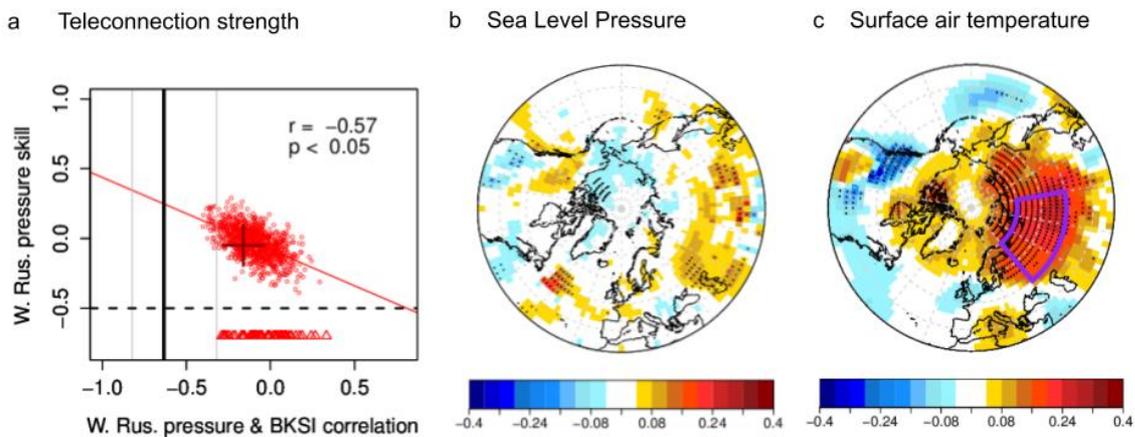


Figure 3.2.1: a) Strength of the Barents/Kara Sea Ice-Western Russia teleconnection in winter as defined in Acosta Navarro et al. (2020) and its link with the skill for Western Russia’s atmospheric circulation (purple box in c) in the multi-model ensemble members selections (dots). Each one of the 1000 dots represents the average of 11 members randomly chosen. The strength of the link in observations is represented by the thick vertical black line (thin lines indicate its uncertainty), and for the individual forecast members with triangles. b) Increment in ACC skill for winter sea level pressure when using the subset of ensemble members with a more realistic representation of the teleconnection, as compared to the full multi-model ensemble. ACC scores are computed after linearly detrending the anomalies. Stippling represents significant values at the 95% confidence level. c) The same as for b), but for surface air temperature. Decadal predictions include 55 members from 6 different DCPD systems, and their skill is computed for forecast years 2–10. Observational references for evaluation are HadISSTv1.1 (a) and JRA55 (b–c).

3.3 — Towards a high-resolution prediction system (BSC)

Among the various techniques being considered to improve the predictive skill of decadal prediction systems, increasing the horizontal resolution of climate models is a promising avenue. Through the use of an eddy-permitting configuration of a global coupled climate model we expect to improve the representation of key teleconnection mechanisms, by enabling previously unresolved interactions of ocean eddies with the atmosphere (Mahajan et al., 2018), which in turn could increase the predictive skill of the systems, in particular over land.

The BSC is currently working on developing a decadal prediction system at eddy-permitting horizontal ocean resolution. We use the latest version of the coupled model EC-Earth3.3 (Döscher et al., 2021) in its high-resolution (HR) configuration (Haarsma et al., 2020). As described in the EUCP MS3, an important effort has been dedicated to the tuning of this coupled configuration, to reduce the model biases and improve process-representation in the North Atlantic, a key region for decadal prediction skill. One particularly interesting feature of this HR model version is the improvement in the simulated variability of the deep convection in the Labrador Sea and the Atlantic Meridional Overturning Circulation (AMOC) compared to the standard resolution (SR) version, which occurred too intermittently in the latter due to an overly strong local density stratification. The biases in the Labrador Sea caused a problem in the decadal prediction system based on EC-Earth3.3-SR, for which an

initialization shock occurred leading to a collapse of the Labrador Sea convection (Bilbao et al., 2021), inducing a quick degradation of the predictive skill in the Subpolar North Atlantic. Some short preliminary tests performed in prediction mode show that this problem is not present in the high-resolution configuration with EC-Earth3.3-HR, for which the Labrador Sea convection remains active and stable, positively impacting the strength of the AMOC (see Figure 3.3.1).

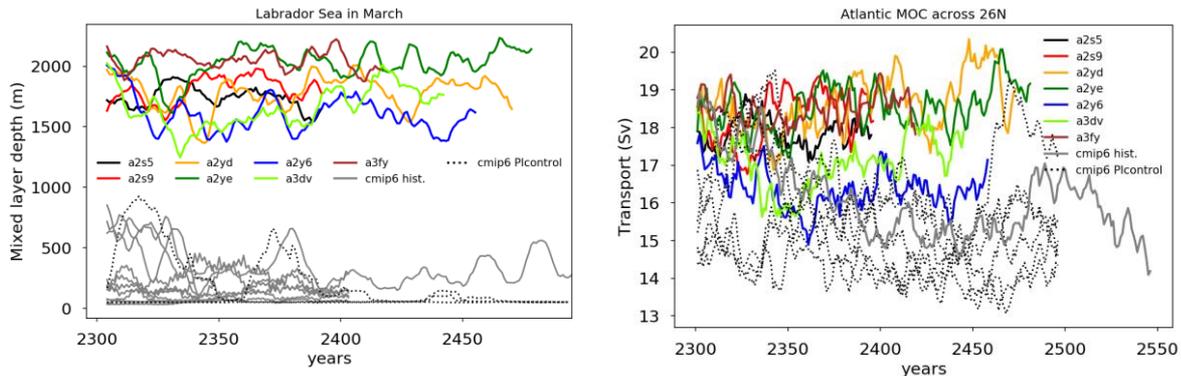


Figure 3.3.1: (left) Time series of the mixed layer depth in the Labrador Sea in March for the different high-resolution experiments with a fixed forcing of the year 1980 (coloured lines) compared to the standard-resolution experiments with a fixed forcing of the pre-industrial period (black dashed lines) and with the historical forcing (1850–2014) (grey lines). The mixed layer depth is a proxy of the deep ocean convection. (right) Timeseries of the AMOC strength at 26°N for the same experiments (only one standard resolution experiment with historical forcing - grey line). The decrease in the formation of deep-water masses directly impacts the strength of the AMOC. The y-axis on the figures correspond to simulation years, not to calendar years.

Additional improvements of the future decadal prediction system can be achieved by enhancing the initialization procedure. At the BSC, we have developed a new refined strategy to improve the in-house reconstructions used to initialize both our standard and high-resolution simulations. This protocol is developed to circumvent a non-stationary bias present in ORAS5 (Tietsche et al., 2020), the currently operational ocean reanalysis of the ECMWF, which has also been produced at an eddy-permitting resolution. This is indeed the product assimilated in the ocean in-house reconstructions used to initialize all our ongoing predictions. The best performing approach resulted from combining the assimilation of two datasets, sea temperature and salinity at the surface from the ORAS5 ocean reanalysis and 3D ocean temperature and salinity below the mixed layer from the EN4 ocean reanalysis.

In a subsequent step we produced a retrospective seasonal forecast system with the tuned version of the coupled model, whose oceanic and sea-ice initial conditions come from the new in-house HR reconstruction, which follows the best strategy previously identified at standard resolution. The atmospheric initial conditions of the seasonal forecast were taken from the ERA5 reanalysis and interpolated to the same grid of the HR version of IFS, the atmospheric component of EC-Earth. The ensemble of initial conditions has been generated by introducing random perturbations in the temperature fields of the interpolated atmospheric initial conditions.

The retrospective forecasts produced so far include 14 ensemble members, each of them initialised on the 1st of May and with a forecast range of 4 months to cover the whole summer. The different forecasts have been initialised each year from 1990 to 2015.

A preliminary analysis of the seasonal forecast has been conducted, comparing the results with a previous set of HR seasonal forecasts using a pre-CMIP6 version of EC-Earth3, produced in the framework of the APPLICATE project. That previous HR version of the model showed substantial degradations of skill in the Arctic when compared to the equivalent SR version, affecting the northern hemisphere winter extratropical forecast skill. The reduced skill was linked to a suboptimal tuning in the Arctic region. The sea-ice cover prediction has been improved in the new CMIP6-HR version developed in EUCP, as illustrated in Figure 3.3.2, which compares the skill to predict the summer Arctic sea-ice in the two HR systems.

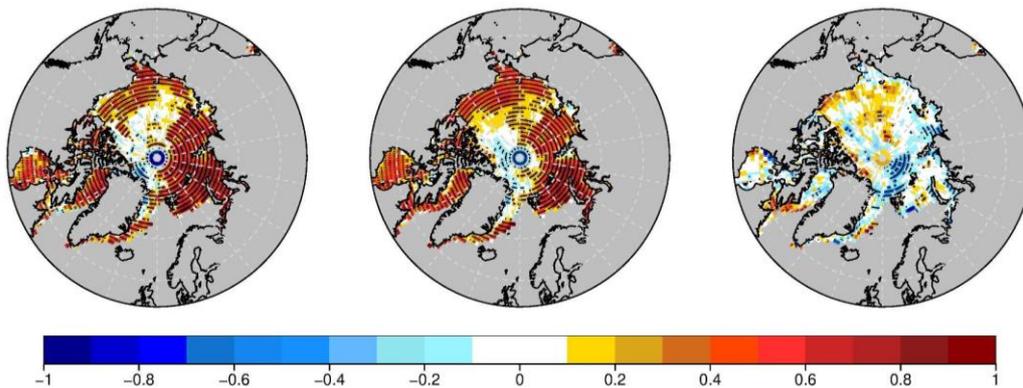


Figure 3.3.2: (left) JJA sea-ice concentration anomaly correlation coefficient between 10-member ensemble mean from EC-Earth3.2 and CERSAT for the period 1993–2015. (centre) Same as (left) but for the new tuned EC-Earth3.3 configuration. (right) Added value of the new tuned version of the model shown as the difference in correlation coefficient between EC-Earth3.2-HR and EC-Earth3.3-HR, both against CERSAT.

Given the promising results of this tuned version of the model and this new initialization protocol, the final step will be to produce a decadal prediction system. Given the high computational needs for running the standard decadal prediction system recommended in Boer et al. (2016), which are unaffordable at these resolutions, we will perform a scientifically reasonable reduced setup (see Table 3.3.1). The experiment will be a set of retrospective forecasts with 5 ensemble members, each of them initialised once every 3 years over the 1960–2015 period, and with a forecast range of 3 years. This reduced decadal prediction system will allow to investigate in particular the added-value of horizontal resolution for the representation and predictability of dynamically driven oceanic processes, and in particular those that have been related to the ocean circulation, like the cold blob or the rapid warming of the subpolar gyre in the mid-90s.

Table 3.3.1: Comparison of the BSC set-up of the current seasonal forecast and the future reduced decadal forecast, with the resulting number of simulated years. Given the length of each member, the production of a decadal prediction system at high resolution is quickly highly demanding in terms of computer resources.

	Number of members	Number of forecasted months	Number of start dates	Simulated years
Produced seasonal forecast system	14	4	26	121.3
Reduced decadal forecast system	5	36	19 (every 3 years over the 1960–2015 period)	285

3.4 — Impact of volcanic eruptions in decadal predictions (BSC)

Explosive volcanic eruptions have climate impacts on seasonal-to-decadal timescales with high predictive potential (e.g. Menegoz et al., 2018; Hermanson et al., 2020). In recent decades, three major volcanic eruptions have occurred and are included in decadal prediction hindcasts: Mount Agung (Mar–Sept 1963), El Chichón (Feb–Mar 1982) and Mount Pinatubo (Jun 1991). This coordinated multi-model analysis seeks to determine the climate signals that were caused by those volcanic eruptions.

The Decadal Climate Prediction Project component C (DCPP-C, Boer et al., 2016) includes a protocol to investigate the impact of volcanic eruptions on decadal climate variability and prediction, which consists in repeating the hindcasts (DCPP-A) initialised before the large volcanic eruptions (s1962, s1981 and s1990) and excluding the volcanic aerosol forcing associated with these eruptions (beyond the background climatological forcing). The impact of the volcanic eruptions is therefore determined by subtracting the hindcasts with and without the volcanic aerosols (DCPP-A - DCPP-C). Here we show results from six state-of-the-art decadal prediction systems from CMIP6: CanESM5, CESM1, EC-Earth3, HadGEM3, IPSL-CM6 and CMCC. Each prediction system has 10 ensemble members of 10 forecast years with and without the volcanic forcing for the three start-dates closest to the respective volcanic eruptions, making a total of 180 members.

Following the volcanic eruptions, the global mean surface temperature drops in response to the negative radiative forcing and recovers after approximately 5–7 years (Fig. 3.4.1). As expected, the surface temperature response varies with the magnitude of the eruption. The response among models is comparable and consistent with previous studies (e.g. Hermanson et al., 2020).

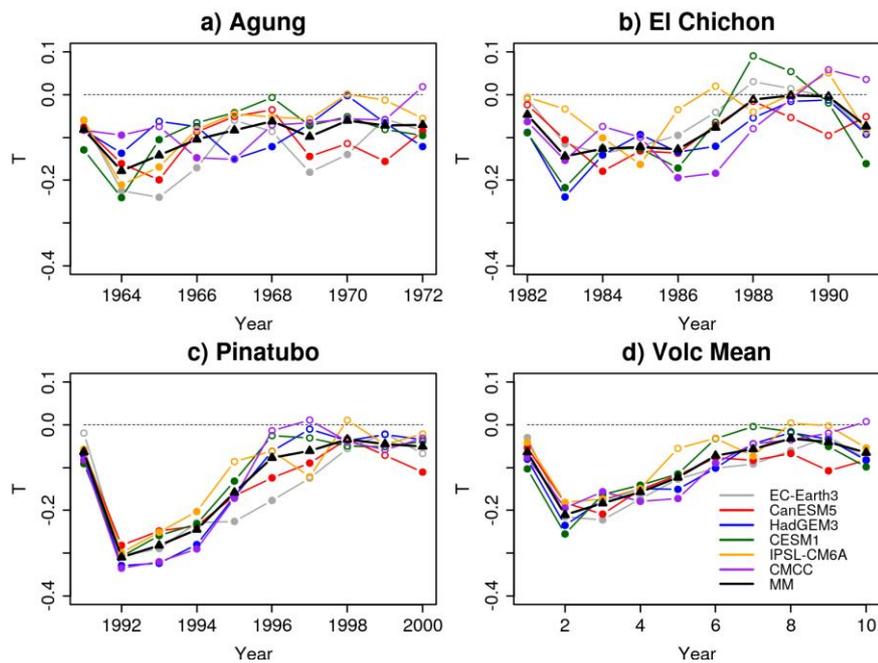


Figure 3.4.1: Global annual mean surface temperature ($^{\circ}\text{C}$) response (dcpA–dcpC) to the eruption of a) Mount Agung (1963), b) El Chichón (1982), c) Mount Pinatubo (1991) and d) the mean of the three eruptions. Filled dots indicate statistically significant differences between the hindcasts with and without the volcanic forcing.

The surface temperature response in the first year following the volcanic eruptions shows a distinct pattern, characterised by widespread cooling, which is largest in the Tropics, and a warming in the Eurasian Arctic sector (Fig. 3.4.2). This pattern is consistent across all the models and for the individual volcanic eruptions. The warming in north Eurasia has been linked to changes in the atmospheric circulation in response to volcanic forcing, in particular to the North Atlantic Oscillation (e.g. Swingedouw et al., 2017). Here we find that the North Eurasian warming occurs throughout the first year, regardless of the season, and the sea level pressure response, although weak, is consistent with a strengthening of the westerly winds over North Europe which might explain the warming. This mechanism is currently under investigation.

For forecast years 2–5, the cooling is worldwide, and the Arctic shows largest anomalies. At later forecast times (years 6–9) the temperature anomalies decrease, consistent with the recovery following the eruption, and differences among the models and eruptions emerge.

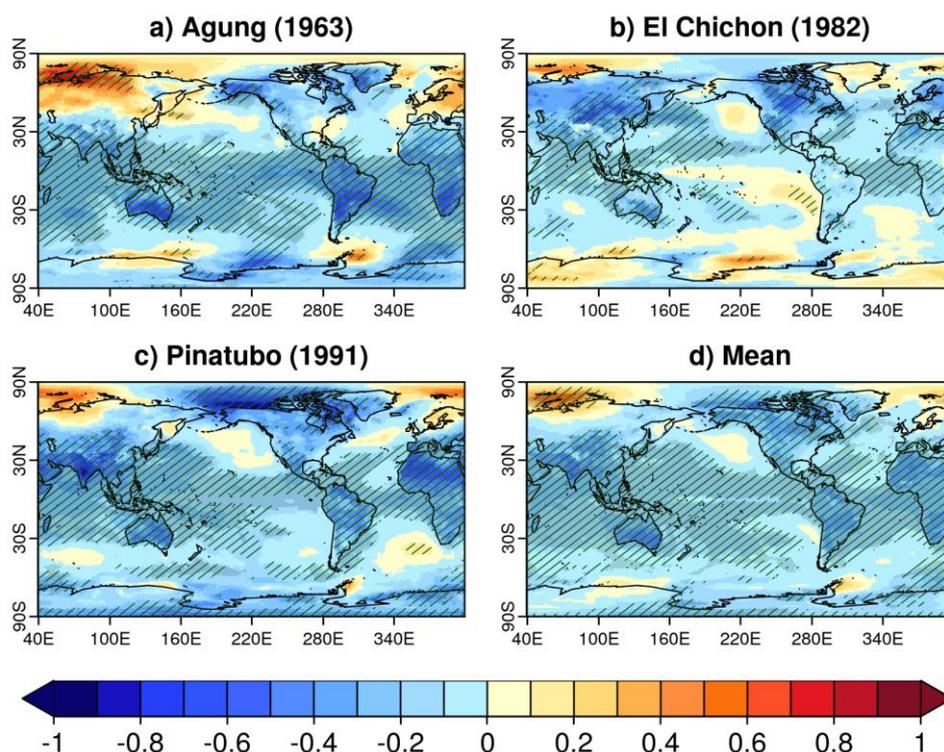


Figure 3.4.2: Multi-model mean surface temperature ($^{\circ}\text{C}$) response ($\text{dcpA}-\text{dcpC}$) the first year (June–May) following the eruptions of a) Mount Agung (1963), b) El Chichón (1982), c) Mount Pinatubo (1991) and d) the mean of the three volcanoes. Hatching indicates statistically significant anomalies and shading indicates model sign consistency.

In the ocean, volcanic impacts persist on longer timescales, and in particular the North Atlantic is a region where recent volcanic eruptions have been shown to induce changes on decadal timescales (e.g. Swingedouw et al., 2017). Preliminary results show an increase in the mixed layer depth in the Labrador Sea the first 3 winters following the eruption and an increase in the Atlantic Meridional Ocean Circulation (AMOC) strength in years 2 to 9.

These results highlight the strong effects of large volcanic eruptions on climate, ranging from months to decades, which can be important at the local scale, and thus substantially degrade the skill of the predictions when the latter are made without including the forcing related to such major volcanic eruptions when these occur. Hence, it is important to develop methods to quickly generate the volcanic forcing associated with such eruptions, so that new predictions can be produced soon after their occurrence.

3.5 — Increasing the ensemble size is key for achieving higher skill in decadal predictions of North Atlantic blocking and the NAO (CMCC)

Atmospheric predictability has generally been found to be quite limited on multi-annual timescales. New decadal prediction experiments from NCAR are here shown to exhibit remarkable skill in reproducing the observed multi-annual variations of wintertime blocking frequency over the North

Atlantic and of the North Atlantic Oscillation (NAO) itself. This is partly due to the large ensemble size that allows the predictable component of the atmospheric variability to emerge from the background chaotic component. The occurrence of blocking in certain areas of the Euro-Atlantic domain determines the concurrent circulation regime and the phase of known teleconnections, such as the NAO, consequently affecting the storm-track and the frequency and intensity of extreme weather events. Therefore, skilfully predicting the decadal fluctuations of blocking frequency and the NAO may be used in statistical predictions of near-term climate anomalies, and it provides a strong indication that impactful climate anomalies may also be predictable with improved dynamical models.

We studied the decadal predictability of the NAO and that of blocking frequency anomalies in winter utilizing the Community Earth System Model (CESM) Decadal Prediction Large Ensemble (DPLE) simulations, documented by Yeager et al. (2018). These are decadal hindcast ensemble simulations consisting of 40 members, initialized every November from 1954 to 2015 and run for 10 years. Daily mean geopotential height fields at 500 hPa (Z500) and monthly mean sea-level pressure (MSLP) data are used from these simulations, while the respective data from NCAR/NCEP Reanalysis are used for verification. The detection of atmospheric blocking was performed following the two-dimensional extension [Scherrer et al., 2006] of the original method introduced by Tibaldi and Molteni (1976). Precisely, after determining, for each ensemble member, the number of days of instantaneous blocking in each winter season (December to March, DJFM), a 5-day threshold for persistence was applied to determine the days belonging to prolonged blocking episodes, hereafter referred to as blocking days. It is noted that prior to the blocking detection mean bias correction has been applied to the daily Z500 fields by subtracting a lead-year dependent daily climatology (to account also for model drift) and then adding the resulting anomalies to the respective observed climatology. All Z500 fields were interpolated onto a coarser regular grid ($2.5^\circ \times 2.5^\circ$), and blocking detection was performed in the latitudinal zone 30° – 75° N.

An area of high blocking frequency is identified for this analysis (70° W– 10° E, 57.5° – 75° N) relating to North Atlantic high-latitude blocking (HLB). The rationale for choosing this area is that HLB tends to be accompanied by a southerly displaced eddy-driven jet (negative NAO), as shown in previous studies. Conversely, the absence of blocking over the North Atlantic tends to coincide with a positive NAO regime and an eddy-driven jet close to its climatological position. Hence, the area relating to HLB is chosen due to its direct link to the NAO, which is the leading mode of variability in the North Atlantic sector and the most documented variability pattern representing the atmospheric response to extratropical North Atlantic SST anomalies.

The Anomaly Correlation Coefficient (ACC) is used to evaluate skill. For assessing the statistical significance of correlation coefficient values, accounting for autocorrelation, the effective degrees of freedom were calculated following Bretherton et al. (1999). A one-sided Student's t-test against the null hypothesis of non-positive correlation has been applied at the 0.05 significance level.

Figure 3.5.1.a documents the predictive skill of the 40-member ensemble mean for the number of blocking days in winter belonging to blocking episodes occurring anywhere in the above-defined HLB area. Specifically, for HLB it is required that the centre of the detected blocking falls within the selected area. The ACC is shown for all possible lead-year ranges, determined by the start lead-year (ordinate) and the end lead-year (abscissa). For example, for the initialization year 1990 the lead-year range LY[3–8] represents the average of the DJFM anomalies falling between December 1992 and March 1998. The

blue markers (open circles) in this figure indicate that the respective correlation coefficients were found to be not statistically significant. Hereafter, this type of plot is referred to as ACC matrix. Evidently, for HLB the skill is statistically significant over various lead-year ranges, reaching as high as 0.65 for LY[1–8] (indicated by the “X” marker). Figure 3.5.1.c shows the ACC matrix 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 hereafter zonally averaging the mean sea level pressure (MSLP) at 35°N and 65°N, between 80°W and 30°E. This definition accounts for the zonal migration of the NAO centres of action (Jianping et al., 2003). The traditional definition (Wallace and Gutzler, 1981) yields a comparable ACC matrix with correlation differences that do not exceed 0.09.

In Fig. 3.5.1 panels *b* and *d* show how the ACC increases with ensemble size. Each line in these plots corresponds to a lead-year range (individual cell) in the respective ACC matrix. The skill increases almost monotonically with the ensemble size, which 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, Fig. 3.5.1 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 any single member of the ensemble is a common feature across decadal and seasonal prediction systems alike (Smith et al., 2019), particularly referring to the mid-latitude North Atlantic.

Given the lower signal-to-noise ratio characterizing climate forecasts as compared to the real world, ensemble averaging is essential for extracting the predictable signal from the bulk of chaotic variability, particularly for the atmospheric circulation over the North Atlantic. Anomalies that survive ensemble averaging do so thanks to being similar across a large proportion of the ensemble, and therefore are indicative of a common predictable component. Instead, anomalies that vanish through averaging do so because they are different from member to member, as chaotic noise would be. A large ensemble size makes the isolation of the signal (through averaging) more effective, and it is this well-documented effect that makes CESM-DPLE particularly powerful for assessing atmospheric decadal predictability at mid-latitudes.

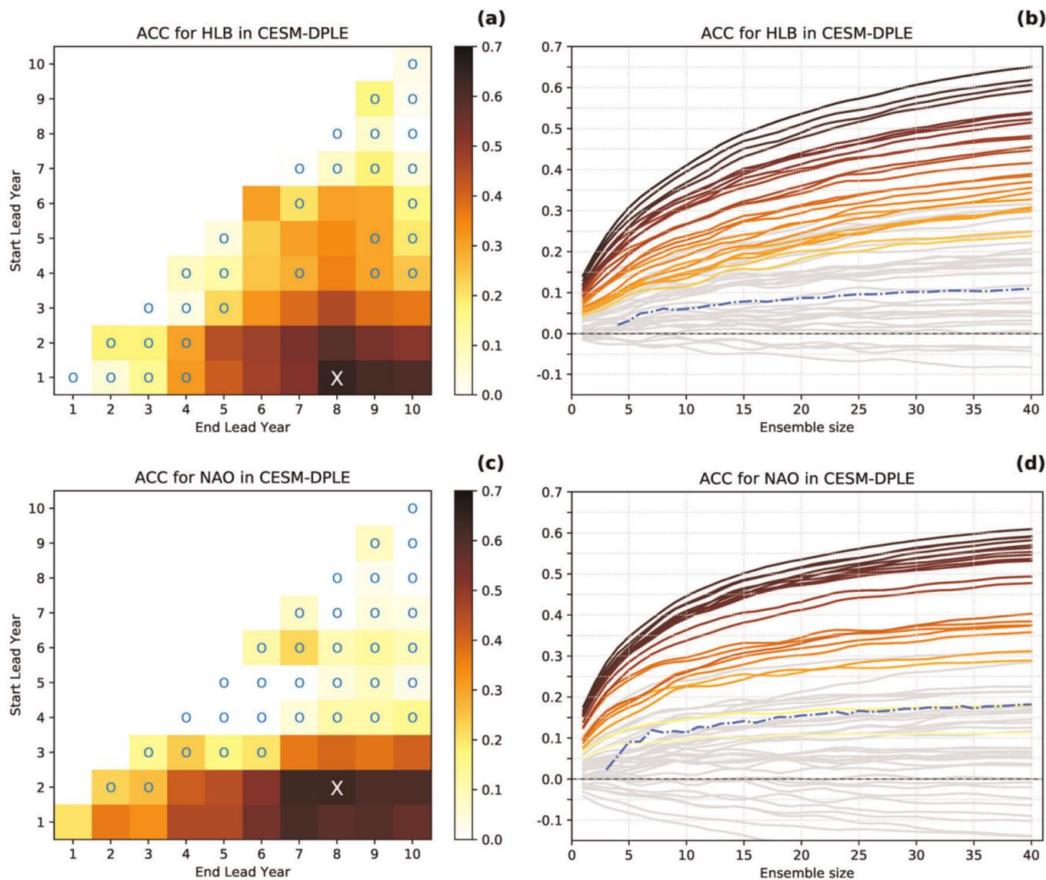


Figure 3.5.1: Predictive skill for high-latitude blocking and the NAO. The predictive skill for the CESM-DPLE ensemble-mean measured by the anomaly correlation coefficient (ACC) for high-latitude blocking (HLB) in **a** and the North Atlantic Oscillation (NAO) in **c**. Each cell below the diagonal corresponds to a different lead-year range defined by the start lead-year (ordinate) and the end lead-year (abscissa). The cyan 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). Reproduced from Athanasiadis et al. (2020).

3.6 — Increased model resolution contributes to mitigating long-standing SST and blocking biases in the Euro-Atlantic sector (CMCC)

Starting to resolve the oceanic mesoscale in climate models is a step change in model fidelity affecting time-mean biases and the representation of processes linked to predictability. This study examines how certain obstinate biases in the midlatitude North Atlantic respond to increasing resolution (from 1° to 0.25° in the ocean) and how such biases in sea surface temperature (SST) affect the atmosphere. Using a multi-model ensemble of historical climate simulations run at different horizontal resolutions, it is shown that a severe cold SST bias in the central North Atlantic, common to many ocean models, is

significantly reduced with increasing resolution. The associated bias in the time-mean meridional SST gradient relates to a positive bias in low-level atmospheric baroclinicity that subsequently causes significant atmospheric circulation biases downstream. These SST biases cause severe diabatic heating biases in the interior of the atmosphere that also affect blocking and the jet. Further increases in oceanic and atmospheric resolution are expected to lead to additional improvements in the representation of Euro-Atlantic climate.

This multi-model study is based on the historical coupled simulations (1950–2014) performed following the HighResMIP protocol (Haarsma et al., 2016). In this protocol, these simulations are referred to as “hist-1950”. The respective models are documented as follows: CMCC-CM2 (Cherchi et al. 2019), CNRM-CM6 (Voldoire et al. 2019), EC-Earth3P (Haarsma et al., 2020), ECMWF-IFS (Roberts et al. 2018), HadGEM3-GC31 (Roberts et al. 2019), MPI-ESM1-2 (Gutjahr et al. 2019), AWI-CM-1.0 (Sein et al. 2017). Each of these models was run in its coupled configuration, provided with observed greenhouse-gas and aerosol concentrations, in at least two different resolutions, with most of the models increasing simultaneously both the atmospheric and the oceanic resolution, while two models only increased the atmospheric resolution (CMCC-CM2 and MPI-ESM-1.2), and two other models (ECMWF-IFS and HadGEM3-GC31) also run configurations at intermediate resolutions aiming to assess separately the impact of increasing the oceanic and the atmospheric resolution. Table 3.6.1 shows the different configurations and the respective nominal resolutions in midlatitudes. Two paired sub-ensembles have been introduced grouping all available model configurations into two classes: “LR” includes the coarsest configuration of each model, while “HR” includes all the remaining configurations with increased resolution either in atmosphere, or both in ocean and atmosphere.

This study focuses on the boreal winter season (DJF) when air–sea contrasts and the oceanic forcing on the atmosphere are stronger. Daily data were used for geopotential height at 500 hPa to assess blocking. The respective daily mean data from the ERA5 (Hersbach et al., 2020) reanalysis were used as an observational reference. Monthly mean data from HadISST2 (Kennedy et al., 2017) were used to compute SST model biases and observational climatologies.

Table 3.6.1: HighResMIP models and simulations. Columns detail the model name, the atmosphere grid spacing at 50°N, nominal ocean grid spacing, and the respective ensemble.

No.	Model	Atm. grid (km)	Ocean grid (km)	LR / HR
1	AWI-CM-1-1-LR	129	50	LR
2	AWI-CM-1-1-HR	67	25	HR
3	CMCC-CM2-HR4	64	25	LR
4	CMCC-CM2-VHR4	18	25	HR
5	CNRM-CM6-1	142	100	LR
6	CNRM-CM6-1-HR	50	25	HR
7	EC-Earth3P	71	100	LR
8	EC-Earth3P-HR	36	25	HR
9	ECMWF-IFS-LR	50	100	LR
10	ECMWF-IFS-MR	50	25	HR
11	ECMWF-IFS-HR	25	25	HR
12	HadGEM3-GC31-LL	135	100	LR
13	HadGEM3-GC31-MM	60	25	HR
14	HadGEM3-GC31-HM	25	25	HR
15	HadGEM3-GC31-HH	25	8	HR
16	MPI-ESM1-2-HR	67	40	LR
17	MPI-ESM1-2-XR	34	40	HR

Most of the examined HighResMIP models were found to exhibit a reduction in wintertime SST bias in the central North Atlantic with increasing resolution. This reduction is shown for the respective multi-model ensemble means (“LR” and “HR”) in Fig. 3.6.1.a. Even though the SST bias reduction pattern is quite different from model to model, the most significant improvements occurred in the models that increased not only the atmospheric but also the oceanic resolution (Table 3.6.1). The SST bias changes seen in the central North Atlantic generally correspond to a reduction in a cold SST bias that most LR models have. In fact, similar cold biases have been a long-standing issue in past-generation climate models. There is evidence that these cold biases are endogenous to the respective oceanic model components and are reduced with increasing resolution, likely thanks to better resolving the oceanic mesoscale eddies.

On the other hand, an improvement with increasing resolution was found for wintertime blocking frequency near the storm-track exit (Fig. 3.6.1.b). This improvement is evident also in Schiemann et al. (2020) who examined the same multi-model ensemble. Blocking in this broad area is referred to as European blocking, it relates to anticyclonic Rossby wave breaking and is dynamically consistent with a poleward deviation of the eddy-driven jet and the storm-track.

A possible causal relationship between the reduction in the cold SST biases and the improvements in European blocking frequency involves the modification of near-surface atmospheric baroclinicity by the underlying SST gradients. Such an influence has been clearly demonstrated in previous studies (Nakamura et al., 2008; Nakamura and Yamane, 2009; Hotta and Nakamura, 2011). Athanasiadis et al. (in review) provide an elaborate diagnostic analysis connecting these two aspects and demonstrating the causal link between improvements in blocking and SST biases. Thus, it is shown that increasing model resolution, primarily in the ocean (from 1° to 0.25°), leads to a significant reduction in a typical

wintertime cold SST bias at the beginning of the North Atlantic Current, which is common among most of the examined models. This improves model biases in the adjacent meridional SST gradient, which in turn, is shown to affect low-level baroclinicity and the maximum Eady Growth Rates at the central North Atlantic, directly upstream to the storm-track exit. The reduction of this cold bias ultimately leads (for each model and for the multi-model mean) to a significant reduction of the respective deficit in European blocking frequency and northern-jet occurrence, which is mediated via an associated reduction in eddy–mean flow interaction biases.

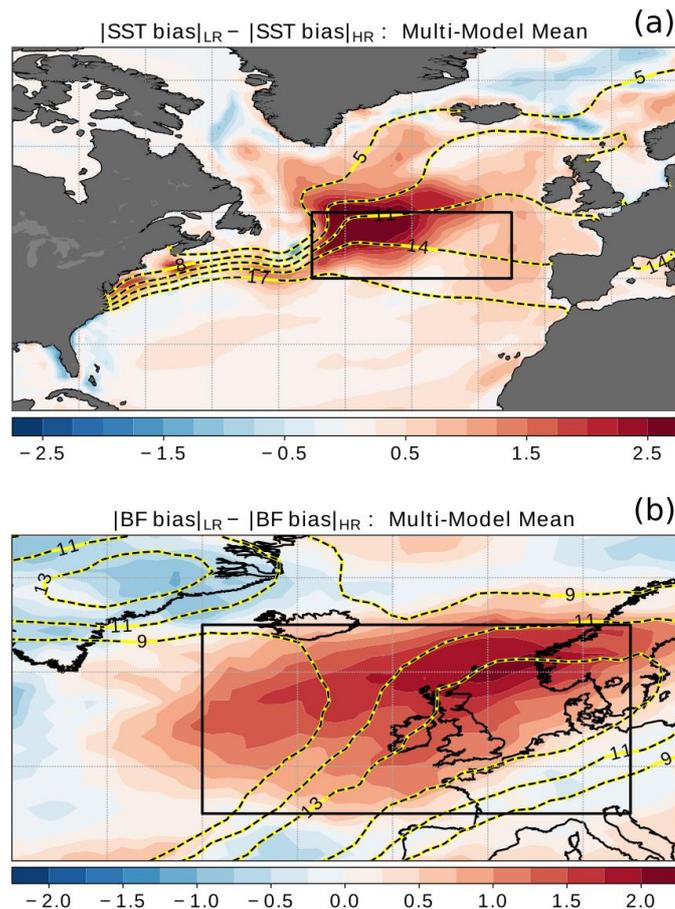


Figure 3.6.1: Differences in wintertime absolute bias between “LR” and “HR” multi-model means (Table 3.5.1). Upper panel: SST biases (shading, in K) with the HadISST2 1950–2014 climatology in contours (°C). Lower panel: blocking frequency biases (shading, in % of blocked days) with the ERA-JOINT climatological blocking frequency in contours (% of blocked days). Details on the blocking detection method are provided in Athanasiadis et al. (in review).

There is a clear indication that increasingly resolving the oceanic mesoscale (with model resolutions comparable to 0.25°) immediately benefits ocean representation. Further increases in oceanic resolution are expected to lead to additional improvements to both model components. Even if the oceanic resolution was found to be key, referring specifically to the above-discussed improvements in European blocking and the eddy-driven jet, the importance of atmospheric resolution should not be undervalued. The presented analysis focused on coupled historical simulations; nevertheless, assessing also the respective multi-model atmosphere-only forced historical HighResMIP simulations (highresSST-present) showed that increasing the atmospheric resolution alone also brought noticeable improvements

in the representation of North Atlantic eddy-driven jet. This is in agreement with numerous recent studies highlighting the importance of high atmospheric resolution (e.g., Willison et al., 2013; Smirnov et al., 2015; Parfitt et al., 2016; Schiemann et al., 2017, 2020).

3.7 — Atmospheric model resolution determines the response to Gulf Stream SST front variability (CMCC)

Part of atmospheric variability in the extra-tropics is linked to the variability of western boundary current SST fronts, with the latter acting as a source of atmospheric predictability, especially on interannual and decadal time scales (Joyce et al. 2019; Athanasiadis et al. 2020; Kohyama et al., 2021). However, general circulation models do not accurately reproduce the atmospheric response to prescribed SST front variability as estimated from observations. This aspect has been assessed in single-model studies using idealized forcings and finding the atmospheric resolution to be a limiting factor (e.g. Smirnov et al., 2015). Yet, previous analyses are based on idealised experimental frameworks forcing the atmosphere with fixed and unrealistic SST anomalies.

In this work we analyse the atmospheric response to Gulf Stream SST front (GSF) shifts in a multi-model ensemble of historical atmosphere-only simulations forced with observed SSTs (1950–2014). In agreement with previous studies, the atmospheric response is found to be resolution dependent. The large-scale response includes a meridional shift in the North Atlantic eddy-driven jet and storm track that is homo-directional to the SST front displacement. This atmospheric response, which is stronger but similar in the observations, is not realistically reproduced by the low-resolution (100 km) models. Our results suggest that interannual to decadal predictability may be higher than what models currently indicate.

The atmospheric response has been investigated in the context of the High-Resolution Model Intercomparison Project (HighResMIP), by analysing six historical simulations performed with three AGCMs, each run with two different horizontal resolution set-up. The AGCMs have been forced with the same observed SSTs. Understanding the impact of horizontal resolution on air-sea interaction can shed light on the effective role of the extratropical ocean on atmospheric variability, with important implications for climate predictions and climate change studies. HighResMIP provides an ideal framework for a multi-model analysis of the impact of increasing model resolution on the atmospheric response to oceanic forcing.

Hereafter we will refer to model configurations with a nominal resolution coarser than 50 km as R100 models, and those with a nominal resolution finer than or equal to 50 km as R50+ models. Each model has been forced with the HadISST2 sea-ice concentration and SST dataset, provided at daily frequency in the period 1950–2014 on a 0.25° grid. For each model a multi-member ensemble of simulations has been used, and the results in the following sections refer to the respective ensemble means. This specific experimental design and the use of multi-member ensembles allows a more robust identification of the atmospheric response forced by the observed oceanic variability as the ensemble averaging aids the forced response to emerge from the chaotic atmospheric variability, which is particularly high at midlatitudes. Table 3.7.1 provides details on the model configurations considered in this study. Additional details about the experimental set-up can be found in Haarsma et al. (2016) (*highresSST-present* experiments). The model results have been compared to ERA5 reanalysis here used as a surrogate of observations.

Table 3.7.1: HighResMIP models. Columns detail the institution name, the model name, the nominal resolution and number of members used for analysis

Institution	Model	Nominal Resolution (km)	Members
EC-Earth-Consortium	EC-Earth3P	100	3
	EC-Earth3P-HR	50	3
MOHC	HadGEM3-GC31-MM	100	3
	HadGEM3-GC31-HM	50	3
ECMWF	ECMWF-IFS-LR	50	8
	ECMWF-IFS-HR	25	6

Previous studies have shown that the north–south shift of GSF represents the leading variability mode of SST variability in the GSF area on interannual and longer timescales (e.g. Joyce et al., 2009). The winter season has been selected because this is the time of the year characterised by the most intense heat exchanges between ocean and atmosphere, resulting in a stronger impact of the ocean variability on the atmosphere. The GSF has been defined as the line of maximum SST gradient. The SST gradient vector magnitude has been calculated for the winter-mean SST fields. The latitude of the GSF has been averaged in the 50°–68°W longitudinal range, where the GS is more zonally oriented.

In Figure 3.7.1 (upper panel) the GSF latitude time series in the period 1950–2014 is shown. The years corresponding to the “North” (“South”) phase are highlighted with red (blue) stars. The index features an interannual variability component associated with SST front latitudinal shift of about 0.2°–0.5°. A lower frequency, decadal-scale component is also evident associated with weaker amplitude meridional displacements. The North Atlantic SST variability described above has been previously interpreted as the oceanic response to North Atlantic Oscillation (NAO) forcing via surface heat fluxes and Ekman currents, explaining an important portion of the extratropical SST variability on seasonal and interannual timescale. However, it has also been shown that close to the GSF the SST variability is primarily driven by oceanic processes, such as oceanic heat transport and diffusion, with the former admittedly remaining subject to the atmospheric forcing. Therefore, the SST anomalies described here are the surface oceanic signature of both historical atmospheric forcing and intrinsic oceanic variability. Some authors have argued that meridional shifts of the GSF may be even key in explaining decadal NAO variability and predictability (Joyce et al., 2019; Athanasiadis et al., 2020).

This zonally averaged latitude of the GSF has been used to define the “North” and “South” phases of the front shift via the respective upper and lower tercile categories. Figure 3.7.1 shows the SST composites obtained by averaging all years over the upper and lower terciles, i.e. the “North” and “South” phases of the GSF, consisting of 22 and 21 years, respectively. The resulting SST pattern shows a tripolar structure extending to the entire North Atlantic in both GSF phases but of opposite sign. The “North” (“South”) phase is associated dominantly with positive (negative) anomalies along the climatological position of the GSF in winter, as well as in the midlatitude North Atlantic to the south of the GSF and the North Atlantic current, and negative (positive) anomalies further north in the subpolar gyre region and south of 30°N. The anomalies are strongest close to the GSF with values exceeding 1 K (absolute departures). In the remaining part of the basin, the SST anomalies are much weaker and lower than 0.4 K.

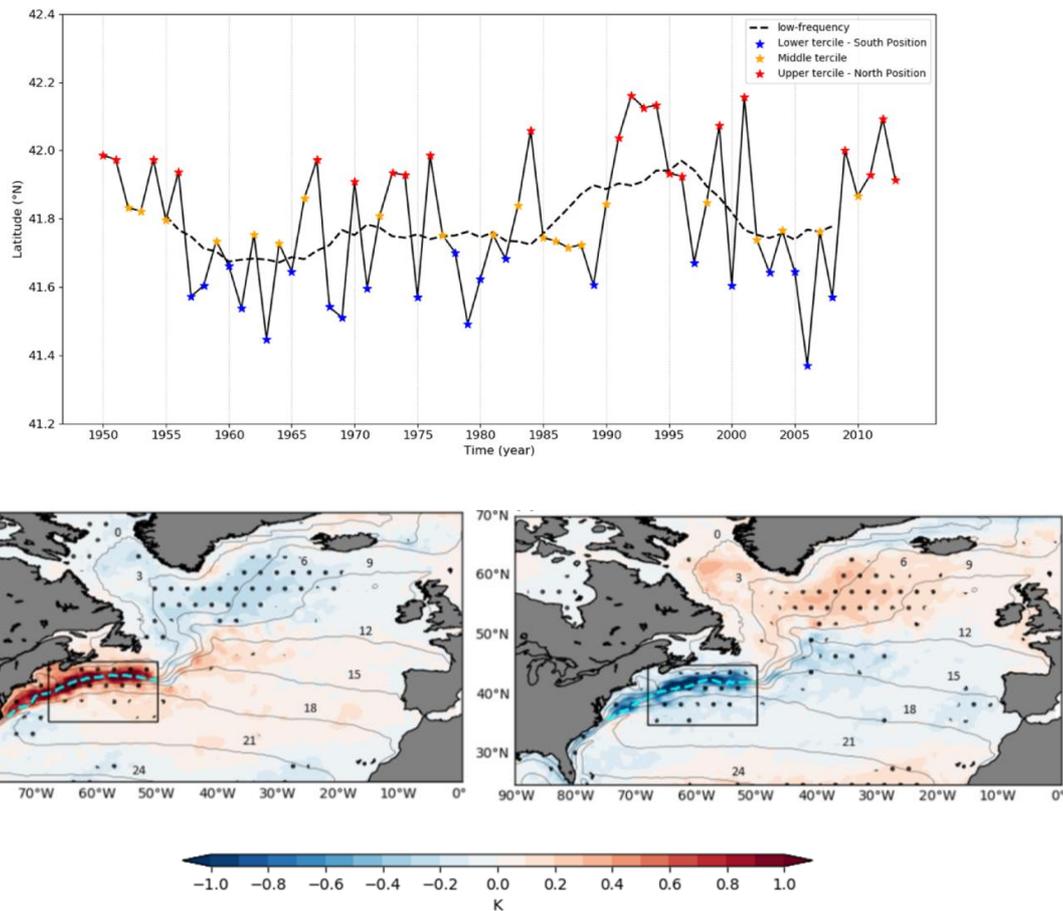


Figure 3.7.1: (Top) Winter-mean latitude of the GSF averaged in the range 50°–68°W. Red, yellow and blue stars represent years in which the GSF latitude falls, respectively, in the upper, middle and lower tercile categories (i.e. North, Middle and South position). The dashed line is the 10-year running mean applied to GSF latitude timeseries. (Bottom) SST (K; colour shaded) anomalies associated to “North” (left) and “South” (right) phases of the GSF in winter (DJF) in HadISST2 dataset. The “North” (“South”) phase has been defined as the upper (lower) tercile of GS SST front mean latitude. The respective climatological position of the SST front is indicated by the cyan dashed line. The black contours indicate winter SST climatology. Black dots denote anomalies that were found to be statistically significant at the 90% confidence level.

Famoos Paolini et al. (in review) have conducted an elaborate diagnostic analysis examining aspects of the local- and large-scale atmospheric response to this GSF variability, including a near-surface heat budget analysis and standard eddy-mean flow interaction diagnostics to determine the character of this response and the underlying physical processes, as well as its dependence on model resolution. Presenting the entire study here is beyond the scope of this Deliverable, but the key finding relevant to EUCP is that the R100 (low-resolution) models cannot simulate a realistic response to GSF variability, in contrast to the R50+ (high-resolution) models. This is demonstrated in Fig. 3.7.2 showing the respective zonal wind anomalies at 850 hPa (composite differences between the “North” and the “South” phases of the interannual GSF variability). Only the R50+ models exhibit a response that resembles the observed poleward shift of the eddy-driven jet, which is homo-directional to the GSF shift. In contrast, R100 models do not seem to agree between themselves and with the observations. In the manuscript it is shown that these discrepancies are directly linked to how models respond locally to the anomalous

diabatic heating associated with the GSF shifts (effectively “replacing” cold water with much warmer water in a narrow strip along the climatological position of the front, as seen in Fig. 3.7.1). Low resolution (R100) models balance the anomalous diabatic heating near the surface and in the lower-troposphere interior via cold meridional advection by the mean flow and consequently fail to reproduce the large-scale baroclinicity anomalies that are found in ERA5 and in the high-resolution (R50+) models at the poleward side of the GSF, linking the changes in storm track activity to those in the eddy-driven jet. This is considered an important limitation in view of the models’ representation of processes linking to seasonal and decadal predictability.

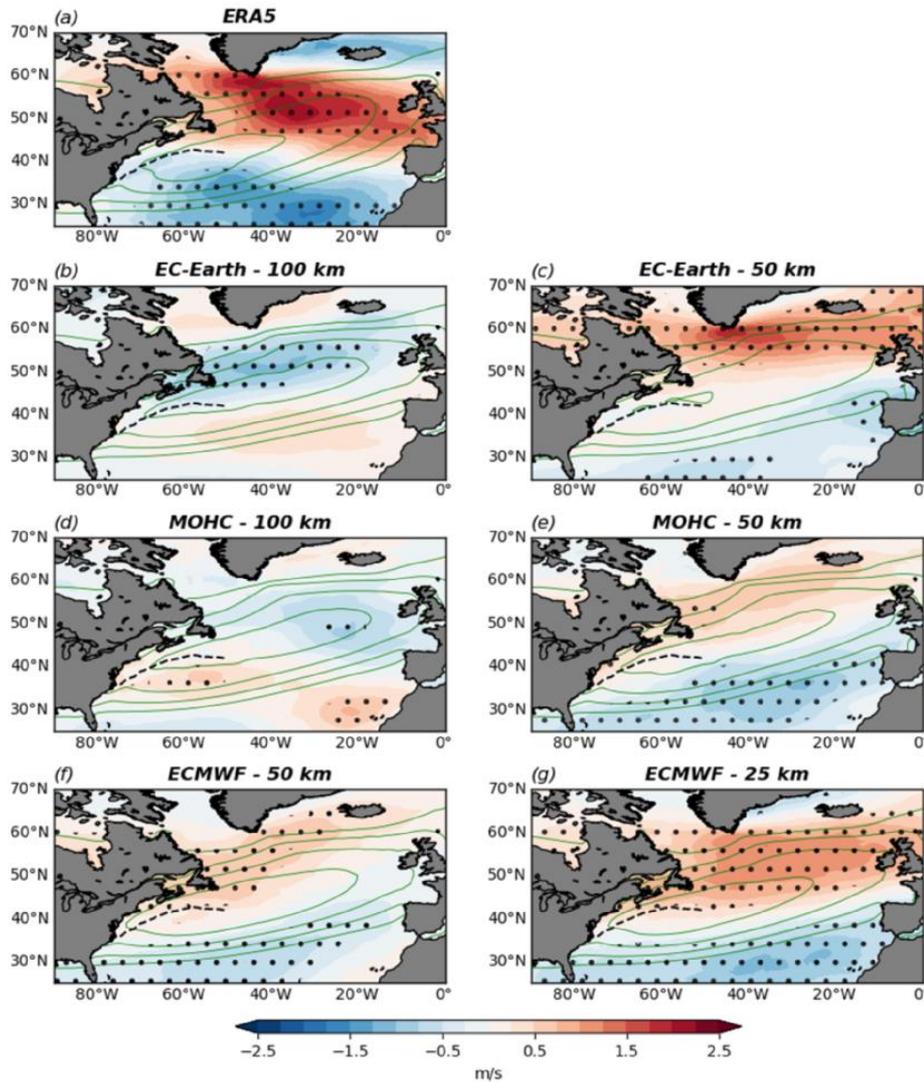


Figure 3.7.2: 850 hPa zonal wind (m s^{-1} ; colour shaded) response to the GSF shift in winter (DJF). (a) ERA5. (b, c) EC-Earth. (d, e) MOHC. (f, g) ECMWF. Beside institution name, the model nominal resolution in km is reported. Black dots denote anomalies that were found to be statistically significant at the 90% confidence level. Green contours indicate the winter climatology of zonal wind at 850 hPa every 2 m s^{-1} from 4 m s^{-1} . The winter climatological position of the GSF is indicated by the black dashed line.

3.8 — Investigation of the impact of Greenland ice sheet melt on decadal predictability (CNRS-IPSL)

CNRS-IPSL has performed new hindcasts using the IPSL-CM6A-LR model based on a new initialisation scheme as compared to those published within DCP-A. We have tried to better integrate the information of SSS through an amplification of the observed SSS variability based on those of the model in the nudging procedure. From this new nudged simulation, we have launched a few new hindcasts before the 1995 abrupt events, for start-dates going from 1988 to 1994 (i.e. 7 start-dates, each with 10 members and 10-year long), that we use as a case study of performance. These hindcast simulations are used as *reference* hindcasts. Indeed, to evaluate the role of Greenland ice sheet (GrIS) melting, we have included the estimate of this melting from Devilliers et al. (2021) in a new nudged simulation where the melting was added and also in hindcast simulations, named *melting* hindcasts. The target was then to compare those melting hindcasts simulations with the reference ones.

The two sets of hindcasts using the IPSL-CM6A-LR model with 7 start-dates are analyzed in terms of RMSE. Fig. 3.8.1 is showing the evolution of SST in the Labrador Sea region for both sets. It is worth noting that the 1995 abrupt warming is slightly better captured when the melting is included, notably for what concerns the amplitude of the warming, since the *melting* hindcasts do show a larger warming around 1995, as this was the case in historical simulations (not shown here). Furthermore, to better quantify the improvements, we define the mean RMSE of the ensemble mean of each start-date over 10 years, for each grid point. Figure 3.8.2 shows the results and highlights some improvements mainly in the west subpolar gyre, while elsewhere in the North Atlantic the differences of RMSE are very small. This shows that including the GrIS melting in the initial conditions and along the hindcasts appears to improve the prediction around the 1995 abrupt warming event in the eastern subpolar gyre. Nevertheless, the start-dates are too few to allow a robust statistical assessment of these improvements. The mechanisms that might explain these improvements are still under investigation. Also, a shortcoming of the approach is that it includes an observed melting, that is not known per se in forecast mode, as it is also the case for external forcing. Thus, those sensitivity simulations should be taken as a first proof of concept to evaluate the potential impacts of Greenland ice sheet melting, but in real forecast mode, forecasts of Greenland ice sheet melting will have to be produced by the model itself rather than being taken from observations.

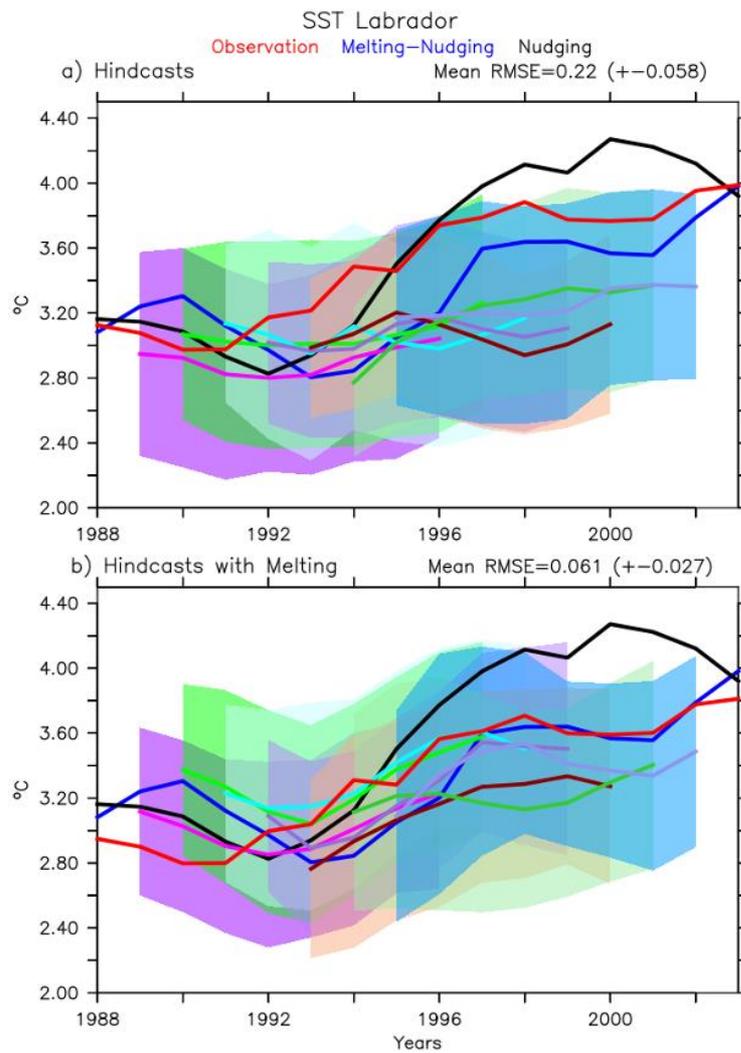


Figure 3.8.1: Evolution of the SST in the Labrador Sea (40–60°W, 45–65°N) in the HadISST observation in red, in the nudged simulation including (in blue) or not (in black) the Greenland ice sheet melting for a) the hindcasts with the melting and b) the hindcasts including the melting. The thick lines represent the ensemble mean of the different starting dates, while the overlap represents one standard deviation of the spread. A 3-year running mean has been applied to all time series.

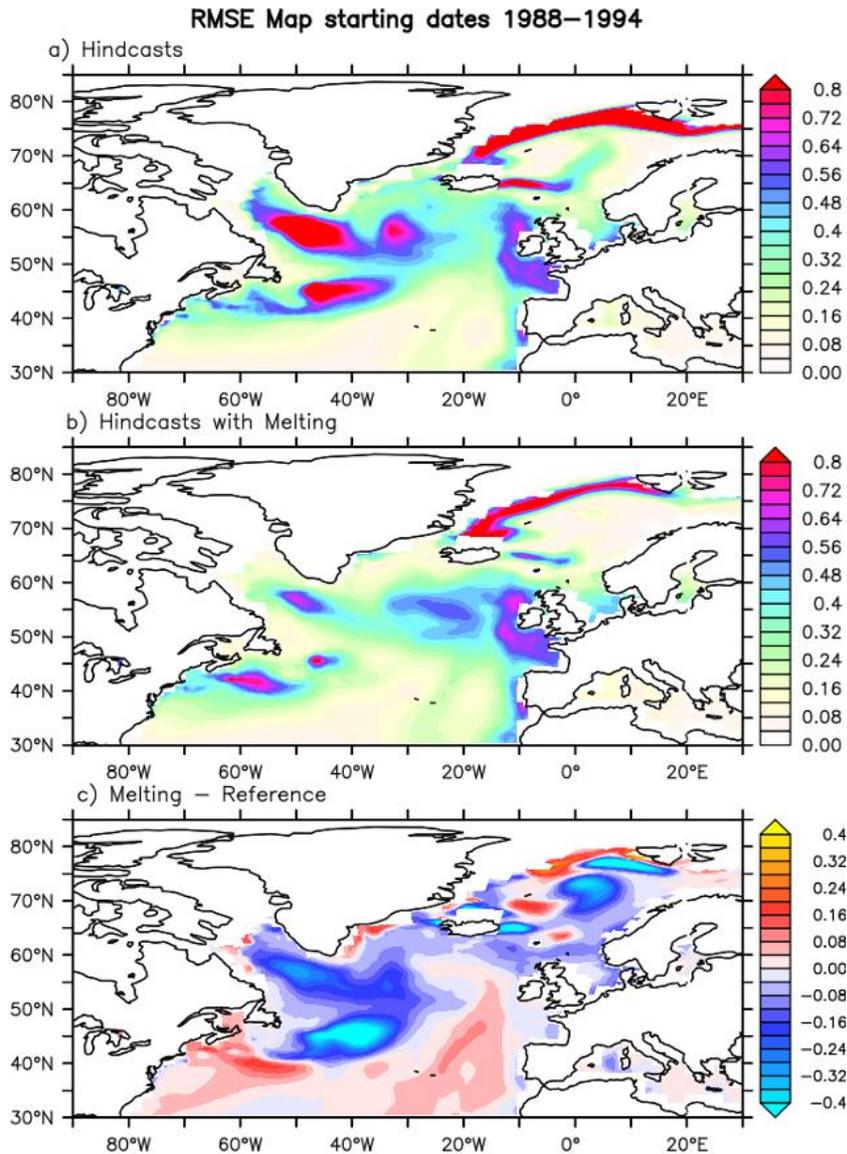


Figure 3.8.2: Root Mean Square Error (RMSE) for the SST averaged for the 7 starting dates of the ensemble mean of the hindcasts over 10 years. a) for the hindcasts that do not include the melting and b) for the hindcasts including it. c) Difference of RMSE between the Melting and Reference hindcasts.

3.9 — Priorities for the development of improved climate forecast systems (Met Office)

The priorities for the development of improved climate forecast systems include:

Implementation of large ensembles

With the current generation of models, many of which are affected by anomalously weak signal-to-noise ratios in the mid-latitudes, large ensembles are necessary to extract the predictable information for downstream applications and users (Smith et al., 2019, 2020). Increasing atmospheric resolution even

by a factor of two from current operational resolutions does not appear to have a significant impact on forecast skill and it is only at much higher (currently unaffordable) levels that impacts appear to emerge (Scaife et al, 2019). Given that increasing ensemble size is much cheaper, it should be given the highest priority to enable skilful predictions.

The use of large multi-model combinations is also attractive in this respect, but current decadal predictions systems typically use only 10 members each, and so we recommend that all systems should increase ensemble size as a priority. Large ensembles also benefit hindcasts, allowing not just improved accuracy in the estimation of the skill of predictions, but also in the derivation of corrections that can be applied to post-processing signal-to-noise deficiencies in outputs.

Resolution

There are clear benefits to using higher ocean resolution (at least 0.25 degrees). At present, some systems remain at 1 degree and we recommend this resolution upgrade for all participating systems.

Improved understanding

Improving understanding of the way the Earth System responds to external forcing in models is key to gaining confidence in the respective climate predictions, as well as diagnosing model deficiencies for targeted development. It would also allow the community to increase the information provided from the multi-model in the regular Climate Updates issued by the WMO Lead Centre for Annual-to-Decadal Climate Prediction.

In the longer term, addressing the modelling deficiencies which cause the signal-to-noise paradox is crucial. In such investigation priority should, arguably, be given to the analysis of the ocean–atmosphere interaction (Osso et al., 2020; Zhang et al., 2021; Czaja et al., 2019) and atmospheric transient-eddy feedbacks in coupled models (Scaife et al., 2019; Smith et al., accepted).

In light of the current literature, which we feel is still inconclusive on the benefits of coupled data assimilation or initial perturbation methods, we feel these areas offer less promise and should be lower priority.

There remains a great need for co-development of forecast products and services with users, particularly to ensure the best use of forecast information and the best post-processing and metrics for predictions and to avoid naïve use of site-specific forecast information from individual ensemble members.

Better use of observations

Observations and observation-based products with global coverage are necessary both for initialization and verification of model predictions; they also need to cover multiple decades in the past. More research into methodology, testing and comparison of ocean (re)analyses is needed to improve the ability to reconstruct the 3-D temperature and salinity ocean state during pre-Argo times (i.e. before ~2004). Robust methods are also needed for the initialization of sea-ice thickness, in forecasts and hindcasts, and for land surface fields, all of which are often inconsistently analysed over the historical periods, due for example to limited length observational records of these quantities.

3.10 — Benefits from higher resolution and coupled assimilation approaches (SMHI)

When trying to improve existing (decadal) climate prediction systems, there are in general three aspects to be optimized. First, the climate model in use may be improved, that means its physical core equations and parametrizations and/or its resolution. Second, the underlying initialization procedure and assimilation approach, that means the way how to ingest the information about the real world's climate evolution may be optimized. Third, the output of climate predictions may be modified by improved post-processing (e.g. bias- and drift-correction) and calibration procedures.

While all three components are important, arguably the largest potential to improve towards a new generation of climate forecast systems is in the former two aspects. The SMHI-contribution to EUCP task 1.3 focuses on analyzing potential benefits for decadal prediction by increasing the (horizontal) resolution of its forecast system and a coupled assimilation approach to (hopefully) derive improved initial conditions for decadal climate predictions.

Benefits from higher resolution modelling for prediction skill

As part of its contribution to CMIP6-DCPP and the Nordic Council of Ministers project ARCPATH (2016–2020) SMHI built a new decadal climate prediction system in collaboration with the Danish Meteorological Institute (DMI; see Tian et al., 2021, for reference). The (atmosphere-ocean) general circulation model (AOGCM) in use is EC-Earth (v3.3.1.1; Döscher et al., 2021), incorporating model components for the atmosphere (IFS c36r4), the ocean (NEMO3.6), and sea-ice (LIM3).

The “observational” data used for initializing ocean and sea-ice fields is taken from ECMWF’s ORAS5 ocean reanalysis. For all fields and systems an anomaly initialization approach is used, that means the observational anomalies (compared to the climatological period 1979–2014) are added to the model’s climatology. This approach usually prevents substantial initialization shocks and subsequent model drifts. Atmospheric fields have been initialized from ECMWF reanalysis data.

The quasi-operational decadal prediction system that was set up together with DMI consists of 15 ensemble members in total, initialized annually on 1 November throughout the period 1960–2020. This decadal prediction system is participating in the international exchange of annual-to-decadal climate predictions led and coordinated by the UK Met Office (<https://hadleyserver.metoffice.gov.uk/wmolc/>) and contributing to CMIP6-DCPP. This quasi-operational system makes use of the standard resolution of EC-Earth3 which means approx. 80 km in the atmosphere (TL255) with 91 levels in the vertical and a nominal resolution of 1° in the ocean with 75 levels.

As part of SMHI’s contribution to task 1.3 of EUCP, we set up an analogous prediction system employing EC-Earth3 in high resolution, that is approximately 40 km in the atmosphere (TL511) and a nominal resolution of 0.25° in the ocean. The number and distribution of the levels in the vertical is the same as for the standard resolution. It should be noted that the model version and the tuning configuration in particular that was used for this effort was an intermediate version in the course of the joint BSC/SMHI tuning of EC-Earth3-HR (see WP1 MS3 as well as the respective description in Section 3.2; BSC contribution). This is not identical to the final EC-Earth3-HR, which has not yet been released. Still, it was found that this version of the high-resolution configuration contained a number of improvements compared to the standard resolution version potentially beneficial for climate prediction

(especially when applying an anomaly-initialization approach as done by SMHI/DMI). Further details regarding the specific model configuration and some climatological analyses are found in the EUCP milestone report MS3.

The initialization approach taken for this high-resolution prediction system is identical to the quasi-operational system. Potential improvements in forecast skill are therefore only attributable to effects of the horizontal resolution, hence a better representation of predictable processes in the coupled climate system and/or a “better match” of anomalies derived from reanalyses (produced with respective models) to the high-resolution configuration of EC-Earth3. 10-member ensembles of climate predictions were initialized on each 1 November 1990–2004 and integrated for 5 years and 2 months. The limited number of start-dates (only 15 compared to 71 with the standard resolution system) and shorter forecast integration (62 months instead of 122) was chosen in the light of the substantial computational costs (a factor of ~20 higher than for the standard resolution) for this model configuration.

In the following, we compare the decadal prediction skill of HR and SR based on the initializations performed for both configurations (1990–2004) in terms of the ensemble means using the Anomaly Correlation Coefficient (ACC) relative to ERA5-reanalysis data. This deterministic metric can assess the accuracy between the forecast and reference data. In the left column of Fig. 3.10.1, the ACC for the first forecast year is shown for T255 (top), T511 (centre) and the differences between the two (bottom), that is the gain from resolution. For the T255 predictions, the ACC is high and statistically significant at 5% significance level over the central Eurasian continent, Northern Africa and South America. Higher and significant ACC is also found over the North Atlantic, the tropical Indian Ocean and Pacific Ocean. Low ACC is mainly found over parts of the Northern Hemisphere and the southern polar region. In the Northern Hemisphere, negative and non-significant ACC is found over two distinct regions: Canada and the northern Siberian region adjacent to the northern polar. Compared to T255, T511 has a similar ACC spatial pattern as T255, but the ACC is much higher in most regions, particularly in the tropical oceans and northern North America. However, reduced ACC is also found in some sub-regions, especially over the central Eurasian continent and the North Atlantic Ocean south of Greenland. The right column of Fig. 3.10.1 shows analogous results for longer lead times, that is the average over lead years 2–5. It can be seen that the prediction skill has substantially increased globally both for T255 and T511 except over the eastern Pacific Ocean. In the T255 predictions a stronger negative ACC is found over the central East Pacific Ocean; the T511 predictions show a similar negative correlation spatial pattern, but the magnitude is much smaller.

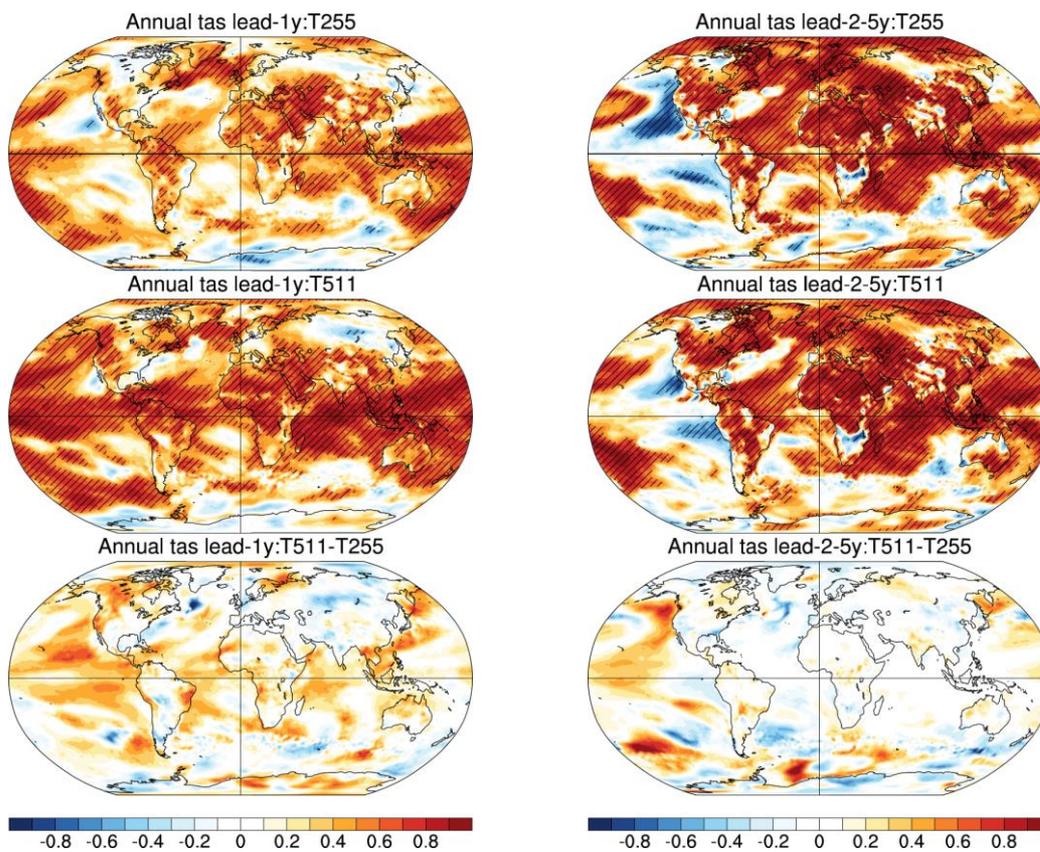


Figure 3.10.1: ACC for the annual mean 2m-temperature calculated over the start-dates 1990–2004 for EC-Earth3 standard resolution (T255) (top), high resolution (T511) (middle) and the differences between T511 and T255 (bottom) for the first full forecast year (left) and the average of lead years 2–5 (right); hatching masks significant ACC ($p < 0.05$); ERA5-reanalysis as observational reference.

The ACC results for total precipitation show a much noisier picture. A benefit from resolution over the tropical and subtropical Pacific for the first forecast year is evident for precipitation, too. Here the HR-configuration leads to significant forecast skill over large parts of the Pacific basin (not shown). For lead years 2–5 there is some improvement in the subtropics over the Pacific basin. However, the general level of skill for total precipitation beyond lead year 1 is rather low and patchy, independent of resolution (not shown).

In addition to the general skill assessment of HR vs. SR for 2m air temperature and precipitation fields discussed above, we conducted a first assessment of the predictive skill for the temporal evolution of North Atlantic climate variability. The North Atlantic is a key region with large climate variability on different timescales and one of the source regions for decadal predictability in terms of unforced (and/or residual) internal variability. In addition to multi-annual to multi-decadal variations of the Atlantic Multidecadal Oscillation (AMO), large (multi-)annual temperature anomalies in the subpolar gyre (SPG) region gain increasing scientific interest regarding both, a potential source of predictability as well as the impacts of these anomalies on weather and climate in the North Atlantic region.

Figure 3.10.2 shows a comparison of predicted annual sea surface temperature (tos) anomalies for SPG and AMO for lead-years 1 and 5 in comparison to the reference dataset ORAS5 that was used for

initializing the hindcasts. Displayed are the ensemble means of SR (15 members) and HR (10 members) together with the ± 1 sigma ensemble spread. The SPG temperature anomalies represent the subpolar North Atlantic region (50–65°N; 60–20°W) as e.g. used in Robson et al. (2018). The AMO is defined following Trenberth and Shea (2006) as globally detrended basin-wide annual SST anomalies of the N-Atlantic (0–60°N; 80–0°E).

The SR predictions for lead-1y show a high agreement with ORAS5 including most interannual anomalies. For lead-5y, the decadal rise until ~2005 is predicted well but the prominent downward trend afterwards (Robson et al., 2016) is underestimated for the SPG and almost absent for the predicted AMO. The ensemble means of the HR-prediction conducted so far for the start-dates 1990–2004 show a very similar evolution as the SR-prediction for both, lead-1y and lead-5y (Fig. 3.10.2, red). For the SPG region, the strong warming trend from 1990 to 2006 is very well predicted at both model resolutions. The last simulated year for SPG and two last years for AMO of the HR-runs correctly capture the initial reversal of the positive trend even with a lead-time of 5 years. The continued steep cooling trend is well predicted by the SR-runs at lead-1y and lead-5y although the steepness of the decline is underestimated by the lead-5y predictions. Relative to ORAS5, the correlation of the HR ensemble mean for lead-1y (1991–2005) has improved by ~14% relative to SR ($r_{HR}=0.67$ vs. $r_{SR}=0.56$) but there is no skill or improvement for lead-5y (1995–2009). Despite the large difference in spatial resolution, it is noteworthy that there is remarkably high covariability of AMO predictions between SR and HR of ~48% for lead-1y and even ~62% for lead-5y that is also evident in Fig. 3.10.2.

To summarize these first analyses of the limited set of HR-hindcasts and the potential benefits from resolution, we can state that the HR-system seems to perform significantly better over the tropics and the whole Pacific basin in the first forecast year. For other regions and longer lead times, the benefit from resolution seems to be rather limited, or not existing, given the initialization approach taken for this system. This holds also for the extra-tropical North Atlantic, a region of special interest for EUCP and a hot-spot of decadal climate predictability where already the SR-system (EC-Earth3 contribution i2 to DCP) performs quite well in terms of temperature (SST and 2m air temperature) prediction. It has to be stated however, that the number of hindcasts performed with the HR-system is quite limited and all results are hence subject to significant uncertainty.

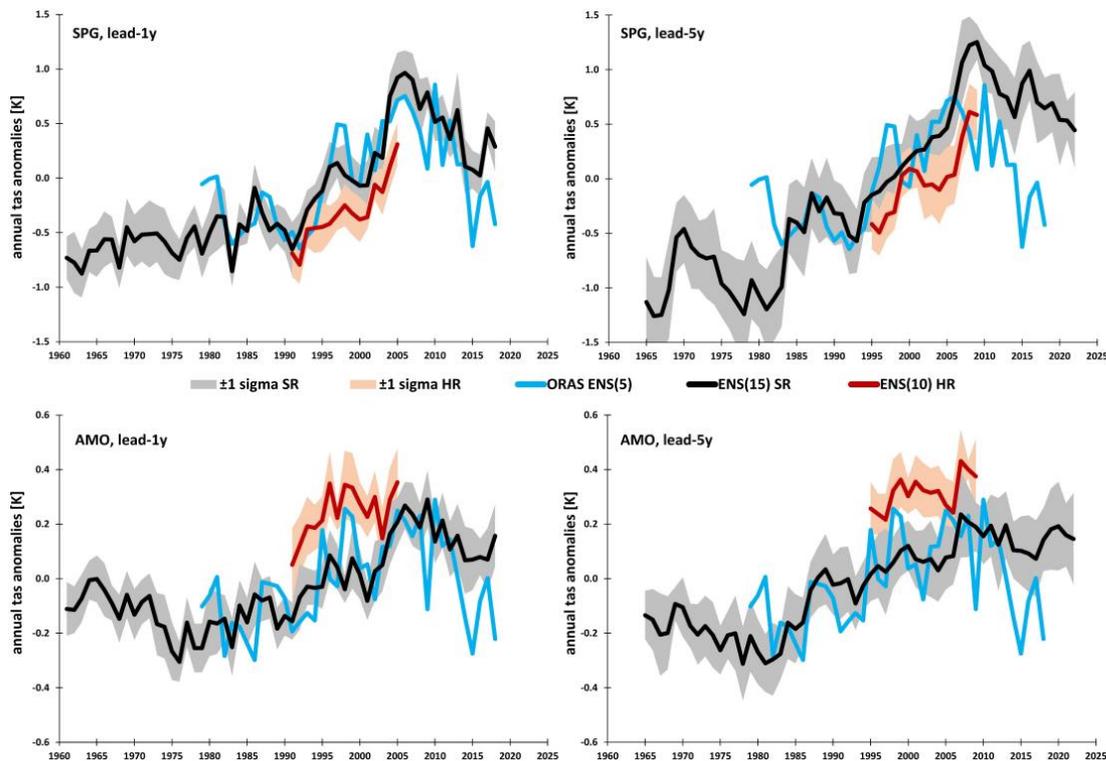


Figure 3.10.2: Predicted evolution of SPG and AMO temperature anomalies in comparison with ORAS5 reanalysis. Left: Ensemble prediction for lead-1y. Right: Ensemble prediction for lead-5y. The ± 1 sigma ensemble spread is shown based on 15 members for the standard resolution (SR, T255) in grey and 10 members for the high resolution (HR, T511) in light red. Anomalies are relative to the mean of 1981–2010. The AMO was globally detrended following Trenberth and Shea (2006).

Development of a new coupled assimilation approach for the next generation of decadal climate prediction at SMHI

SMHI works on developing a new improved climate prediction system based on EC-Earth3 (for now in standard resolution). The particularity of this system is the performance of coupled assimilation simulations to derive potentially improved initial conditions for decadal climate predictions. The setup currently found to yield promising results, assimilates monthly mean sea-surface temperatures (starting in 1900 based on HadISST1) for several decades before the start date of the first planned hindcasts in the ocean model component and 6-hourly instantaneous fields of low-level atmospheric vorticity and divergence which are spectral representations of the horizontal wind components. The assimilation of these atmospheric fields is started in 1950 and based on ECMWF's most recent reanalysis ERA5 (Hersbach et al., 2020) and its backward extension. All fields are assimilated as anomalies compared to a 30-year running climatology of the model and the respective reanalysis dataset. This is done to retain the assimilation results close to the preferred model attractor even in the light of a model climate sensitivity potentially differing from the real world. This is done to mitigate the risks of model drifts after starting free-running prediction simulations from initialization fields derived from the assimilation run.

A 5-member ensemble of these assimilation experiments has recently been completed, covering the period 1950–2020 and providing climate prediction initial fields for every 1 November. The ensemble was generated by starting the assimilation runs on 1 January 1900 from five different CMIP6-historical simulations, assimilating SST-anomalies only for 50 years and then continue with assimilating SST-anomalies plus anomalies of vorticity and divergence as described above.

The analysis of these assimilation experiments has just been started. An encouraging first result is described in the following. Figure 3.10.3 shows the Arctic sea-ice area in September (annual minimum) for the period 1992–2020 for a 10-member ensemble of historical simulations with EC-Earth3 (grey lines), the 5-member ensemble of assimilation runs (blue lines) and the observed equivalent derived from a high-resolution (12.5 km) satellite product (CERSAT) that was compiled and made available within the EU H2020 project INTAROS (grant no. 727890).

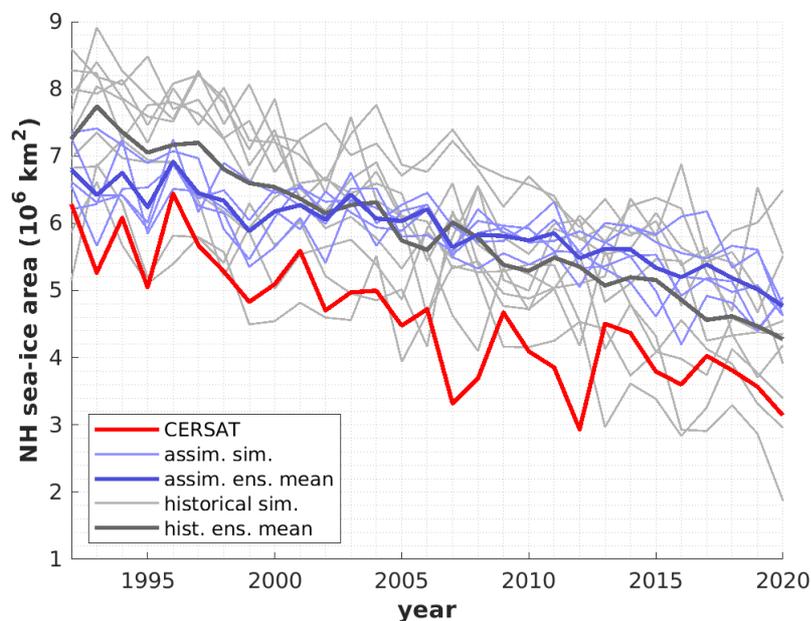


Figure 3.10.3: NH September sea-ice area according to the CERSAT observational product (red line) as well as an ensemble of free running EC-Earth simulations (grey lines; following CMIP6-historical protocol and SSP2-4.5 after 2014) and the new set of assimilation experiments (blue lines) that assimilate anomalies of SST and low-level winds in the atmosphere.

While the assimilation runs feature a somehow smaller trend ($-0.05 \cdot 10^6 \text{ km}^2/\text{a}$) over the analyzed period when compared to the CERSAT-observations ($-0.09 \cdot 10^6 \text{ km}^2/\text{a}$) and the historical ensemble ($-0.11 \cdot 10^6 \text{ km}^2/\text{a}$), the interannual variability of the assimilation ensemble mean matches the observations quite well, even though with substantially reduced magnitude. The correlation between the assimilation ensemble mean and CERSAT is 0.90 when considering the raw values (including the warming trends), compared with 0.82 for the historical ensemble. Focusing on the interannual variability by linearly detrending the data eliminates any skill of the historical ensemble (correlation of -0.16) while the assimilation ensemble still features a correlation of 0.58. This confirms our expectation that the assimilation of SST-anomalies (in ice-free regions only) and near-surface winds is sufficient to represent

the NH sea-ice area reasonably well in our model. The impact on other variables and domains that are not directly subject to data assimilation is currently analyzed. It further needs to be shown if this approach is sufficient to generate better initial conditions used for potentially improved decadal climate predictions. A respective scientific publication documenting the approach and its performance will be prepared.

3.11 — Representation of model uncertainty in multiannual predictions (UOXF)

Models are not perfect: they have biases and have difficulty representing some physical processes related to small-scale variability (e.g. clouds and turbulence). Model errors typically vary from model to model. To account for these model uncertainties, the multi-model ensemble (MME) has emerged as a pragmatic and much used approach. However, a major common source of uncertainty across model simulations is the inevitable need of approximations to solve the laws of physics in state-of-the-art climate models. Large uncertainty arises from sub-grid-scale uncertainties as these processes are often parameterized. Alternative approaches, which rely on stochastic parameterization schemes where the underlying deterministic subgrid parameterizations are replaced by an inherently stochastic formulation, to explicitly account for uncertainties due to unresolved physical processes have been developed over the last decades by the numerical weather prediction community but so far have not been used much in climate predictions.

In particular, stochastically perturbed physical tendency (SPPT) schemes for the atmosphere are now used routinely for global and regional numerical weather prediction (e.g., Berner et al., 2017; Leutbecher et al., 2017; Lock et al., 2019; Pegion et al., 2016). Such schemes are also increasingly used in seasonal forecast models, with several studies showing improvements of skill in the tropics. Furthermore, SPPT schemes can improve the reliability of tropical ocean sea surface temperatures (SST) by increasing the ensemble spread and reducing the forecast errors. Besides the proven positive impacts on seasonal time scales, decadal predictions are predominantly performed without stochastic parameterizations.

This study focuses on assessing the impact of SPPT on multi-annual timescales using targeted simulations conducted with ECMWF's coupled forecast model. Results from those simulations are further contrasted to an MME consisting of five state-of-the-art decadal prediction ensembles. This work has been published in Befort et al. (2021).

To test the impact of SPPT, two targeted simulations using ECMWF's coupled model CY46R1 have been carried out. The setup of both experiments is identical: (i) initialized each 1st November from 1981 to 2014, (ii) 10 ensemble members, (iii) 28-month forecasts, and (iv) atmospheric horizontal resolution Tco199 (approx. 50 km); 1° ocean resolution. The only difference between the runs is that the stochastic physics scheme SPPT has been switched off (*ECMWF-noSPPT*) in one experiment, whereas it is included in the other experiment (*ECMWF-SPPT*).

The stochastic physics approach is compared to a multi-model ensemble consisting of 5 models: NCAR-DPLE (Yeager et al., 2018) and four dcppA CMIP6 systems: EC-Earth (Bilbao et al., 2021; Doblas-Reyes et al., 2018; Haarsma et al., 2020), MPI-ESM1-2-HR (Mauritsen et al., 2019; Müller et al., 2018;

Pohlmann et al., 2019), MIROC6 (Kataoka et al., 2020), and HadGEM3-GC31-MM (Andrews et al., 2020; Williams et al., 2018).

Skill is assessed using the anomaly correlation coefficient (ACC) of the ensemble mean, whereas reliability is measured using the spread-over-error (SoE) relationship, which is defined as the ratio between the average ensemble spread and the root-mean-square-error (RMSE) of the ensemble mean. A SoE value of 1.0 indicates a perfectly reliable ensemble; SoE values below 1.0 indicate that the ensemble is under-dispersive and SoE values above 1.0 characterize an over-dispersive ensemble.

Previous studies have shown the positive impact of SPPT on tropical Pacific SSTs, e.g. the ENSO regions (e.g. Weisheimer et al., 2011). Thus, in this study SSTs over the NINO3 region (150–90°W; 5°S–5°N) of the eastern tropical Pacific for forecast times up to 28 months are analysed. Furthermore, the impact of SPPT on skill and reliability of the North Pacific index (180–120°W; 30–65°N), which is strongly influenced by tropical SSTs in the NINO3 region (O'Reilly, 2018), is assessed.

For all hindcasts lead-time-dependent biases are taken into account and anomalies are calculated using all initialization dates from 1981 until 2014. ERA5 SLP and SST reanalysis data are used as reference (Hersbach et al., 2020).

The main finding of this study is summarised in Fig. 3.11.1. The ACC skill of the different prediction ensembles for SSTs over the NINO3 region is shown in Fig. 3.11.1.a. Results suggest that SPPT increases skill compared to the ensemble without stochastic physics scheme (*ECMWF-noSPPT*) up to about 16–17 months lead time (second spring season). Especially, ACC skill from early first summer to early second winter is significantly improved when using SPPT. The skill of both ECMWF predictions is higher compared to the MME. After the second spring skill gets insignificant from zero in all ensembles. For reliability (Fig. 3.11.1.b) even bigger differences between the ensemble with and without SPPT are found. SSTs over the NINO3 region are statistically reliable in the *ECMWF-SPPT* for most times during the 28 months forecast, whereas the *ECMWF-noSPPT* hindcast is under-dispersive (not enough spread) for almost the entire time period. Similarly, SSTs over this region in the decadal prediction single model ensembles (SME) tend to be under-dispersive as well. The MME exhibits a much improved reliability compared to the underlying SMEs, thus confirming results from earlier studies showing the MME approach has positive impact on reliability (Palmer et al., 2004).

The ACC skill of the North Pacific SLP (NPI) index is shown in Fig. 3.11.1.c. All ensembles (*ECMWF-SPPT*, *ECMWF-noSPPT*, MME) exhibit similarly high levels of skill for the extended first winter season. Beyond the first summer season, none of the ensembles shows significant ACC skill for the NPI. However, significant skill returns during the extended second winter in the *ECMWF-SPPT* ensemble. As SLP anomalies in this region are strongly influenced by tropical Pacific SST anomalies through an atmospheric teleconnection, the increased skill of *ECMWF-SPPT* is broadly consistent with increased skill and reliability seen in the *ECMWF-SPPT* ensemble for the NINO3 SST index (Fig. 3.11.1.a).

Overall, results from this study indicate that stochastic physics schemes are an effective way to account for model uncertainty in a SME. Positive impact on multi-annual timescales is especially pronounced for reliability and also partly for skill. Given the low computational costs of these schemes, it motivates

their application for climate predictions on multi-annual to decadal time scales, in conjunction with the combination of SMEs into a MME.

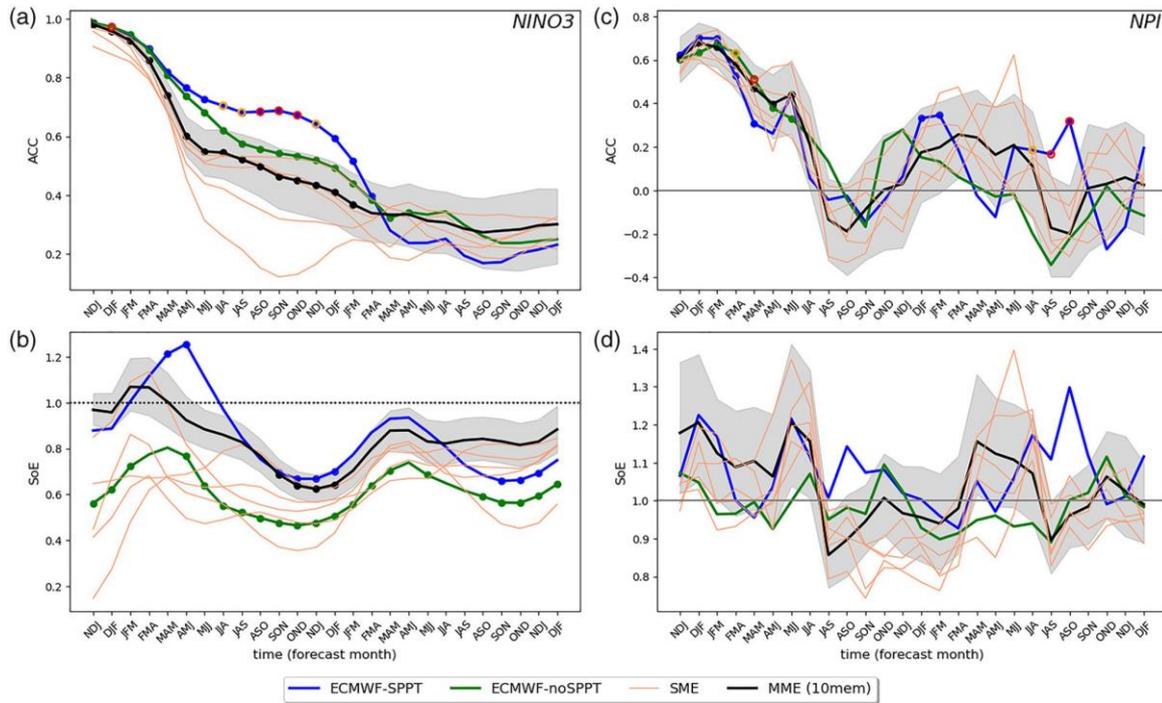


Figure 3.11.1: (a) Anomaly correlation coefficients for SSTs over NINO3 region using ERA5 as reference, (b) same as (a) but for SoE, (c) same as (a) but for the North Pacific index (NPI), (d) same as (c) but for SoE. Grey shading for the MME indicates 2.5 and 97.5 percentile derived from randomly sampling (10,000 samples) two members from each single-model ensemble. Dots in (a) and (c) indicate forecast times for which the respective ensemble is significantly larger than 0, whereas dots in (b) and (d) indicate forecast times for which the respective ensemble is significantly different from 1 (95% confidence, 10,000 samples). Samples have been generated by bootstrapping over years for ECMWF-SPPT and ECMWF-noSPPT ensembles and over both years and members for the MME. Orange and red circles in (a) and (c) indicate those forecast times for which the respective ECMWF ensemble shows significantly higher skill compared to the other ECMWF ensemble (orange: 10% and red: 5% significance level, following Siegert et al., 2017). SST, sea surface temperature; SoE, spread-over-error; MME, multimodel ensemble; SPPT, stochastically perturbed physical tendency. Figure taken from Befort et al. (2021); their Fig. 3.

3.12 — To what extent predicting the North Atlantic variability could improve our ability to predict climate over the globe? (BSC, CMCC, CNRS/IPSL, Met Office)

The North Atlantic sea surface temperature exhibits multi-decadal fluctuations that are referred to as the Atlantic Multi-decadal variability (AMV). Marked climate anomalies with substantial impacts upon human activities over many areas of the globe are phased with the AMV, suggesting a driving role of the latter in those fluctuations. These impacts include drought over Africa, North and South America substantial changes in Arctic sea-ice extent and tropical cyclone activity, and modulations in European climate conditions, with documented impacts in river streamflow and electricity production (e.g., Knight

et al., 2005; Schubert et al., 2009; Sutton and Dong, 2012). The decadal variability of the tropical Pacific, including the cooling trend over 1988–2012 that contributed decisively to the slowdown of global warming (the so-called hiatus), is also linked to the AMV through an atmospheric bridge (e.g., McGregor et al. 2014; Li et al., 2016). Given the numerous climate impacts of the AMV, correctly predicting the future AMV evolution is key to enable our predictive capacity over many regions of the world, and thus to provide valuable information for the energy/industry sectors and decision makers several years in advance.

Results from climate predictions highlight that the North Atlantic is the region where the initialization step provides higher predictive skill compared to non-initialized historical simulations, and for longer forecast horizons (up to a decade). The high predictive skill of the North Atlantic region is encouraging for the prospect of getting skillful decadal predictions all over the globe through the climate impacts of the AMV. In the current decadal predictions systems, the initialization from observations provides additional skill over Africa, Europe and the Middle East, three regions impacted by the AMV. However, over America and Asia, as over the tropical Pacific, the prediction skill is mostly only coming from the global warming trend due to anthropogenic activities. Given the observed climate impacts of the AMV over those regions, the absence of additional skill coming from ocean initialization reflects shortcomings in the decadal climate prediction systems that need to be solved. This lack of skill could be explained by any or a combination of the following reasons: (1) the AMV is not the primary driver of those observed variations, (2) the teleconnection mechanisms associated to the AMV are poorly simulated by climate models used for climate prediction, (3) the prediction systems underestimate the predictability of the AMV impacts because of their too chaotic behaviour, a problem often referred to as the signal-to-noise problem. Those different hypotheses need to be evaluated to understand the apparent lack of skill provided by initialization in the current decadal prediction systems outside of the North Atlantic region.

Here, we focus on the first two hypotheses. In particular, we assess the robustness and the strength of the impacts of the AMV on the tropical Pacific. We analyze 21 ensemble simulations from 13 climate models that follow the CMIP6/DCPP-C protocol. Following this protocol, the same observed AMV SST anomalies are imposed in the North Atlantic of each model to investigate the worldwide teleconnections associated with the observed AMV. Though models mostly agree on the sign of the tropical Pacific response, the magnitude of their response varies by an order of magnitude, from 0.01°C to -0.23°C , with a MMM of -0.12°C for a similar $\sim 0.2^{\circ}\text{C}$ tropical North Atlantic warming. This inter-model spread is a synonym of strong uncertainties for someone who wishes to predict the AMV impacts.

We explore the origins of this spread using an energy constraint approach. We find that the large inter-model spread is mainly driven by different amounts of moist static energy injection from the tropical Atlantic surface into the upper troposphere, which is mostly due to different models ITCZ positions and strengths. To a smaller extent, we find that inter-model spread is also due to different ascent anomalies simulated for the same imposed AMV forcing. Building a bi-linear regression model with those two variables as predictors, we capture as much as 73% of the inter-model variance in the wintertime NIÑO3.4 SST response (Figure 3.12.1.b).

Analytically correcting models for their mean precipitation biases, we reduce the inter-model uncertainty and we quantify that, following an observed 0.26°C AMV warming, the equatorial Pacific

cools by 0.11°C with an inter-model standard deviation of 0.03°C (Figure 3.12.1.c). Therefore, our analysis highlights the importance of reducing mean climate model biases in order to properly simulate and predict the global AMV impacts. More details about our work have been published in Ruprich-Robert et al. (2021).

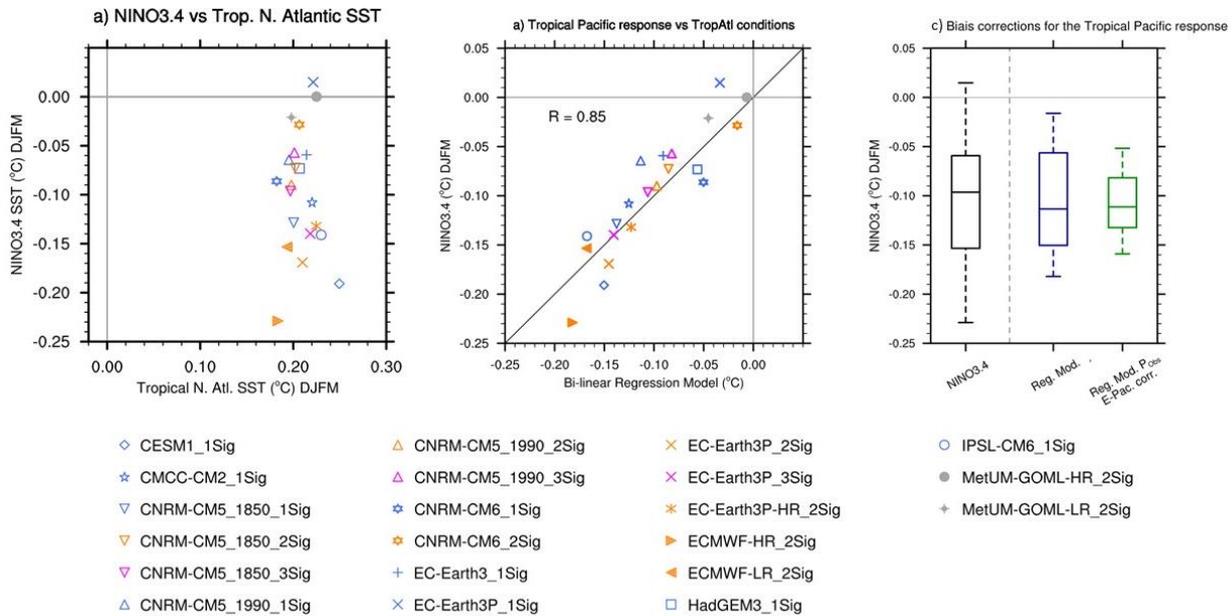


Figure 3.12.1: Inter-model relationship between several indices. Markers represent the 10-year averaged ensemble mean of the difference between AMV+ and AMV− simulations from individual experiments. (left) Winter NIÑO3.4 SST index versus winter tropical North Atlantic SST (averaged over 5°N – $20^{\circ}\text{N}/60^{\circ}\text{W}$ – 10°E). (middle) Inter-model relationship between the wintertime NIÑO3.4 SST index and the outputs of a bi-linear regression model built with the summertime Tropical Atlantic ascent and the injection of moist static energy from the Tropical Atlantic surface. (right) Whisker box plots indicating the minimum/maximum values, the 20th/80th percentiles, and the median from the inter-model distribution of several indices: the wintertime NIÑO3.4 SST index (black), the outputs of the regression model fed with summertime Tropical Atlantic ascent, and (blue) the injection of moist static energy from the Tropical Atlantic surface, (green) same as blue but with the injection of moist static energy corrected for the model precipitation climatology biases.

4. Lessons learnt

As mentioned in the introduction, this report documents an ensemble of varied analyses performed by the project partners. The material was organized per partner, and the lessons learnt listed below correspond to the presented studies. In addition, Section 3.9 provides some key priorities to be considered.

- High model resolution in both the ocean and the atmosphere, respectively $1/4^\circ$ and $1/2^\circ$, is required in order to reduce certain long-standing biases and successfully reproduce physical processes that are vital for the interaction between Gulf Stream SST front variability and North Atlantic storm track and jet variability.
- In decadal as well as in seasonal predictions, increasing the ensemble size of individual model systems and using multi-system ensembles monotonically increases the predictive skill in the extra tropics.
- By exploring Arctic sea-ice, we have shown that multi-model prediction ensembles can be a valuable resource to explore the predictive capacity on decadal timescales and the added value of model initialization on different variables and regions.
- We have shown that if multi-model ensembles are large enough, it is possible to exploit our knowledge related to teleconnection mechanisms, in order to enhance predictive capacity by selecting members that better represent physical reality. This technique can be extended to regions beyond the Arctic.
- The comparison of the new EC-Earth-HR seasonal forecasts with the ones using a previous, untuned HR version of EC-Earth shows the importance of a good calibration and tuning of the models to truly benefit from the added value of the high resolution on skill.
- Explosive volcanic eruptions cause climate impacts ranging from months-to-decades, and therefore need to be accounted for in climate predictions. In case of a future large volcanic eruption occurs, it is important to have the capability to quickly generate the associated volcanic forcing, so that new predictions can be timely produced including the volcanic effects, and benefiting from their predictive value.
- Analyses on longer lead times towards decadal prediction with an anomaly-initialized system reveal benefit from higher resolution mainly in the tropical regions and the Pacific basin. However, this benefit is most pronounced in the first forecast year and decreases afterwards. For the extra-tropical North Atlantic, no benefit from resolution could be found for predicting temperature variability there.
- Given the high computational cost to produce a decadal climate prediction system, especially at high resolution, it is useful to proceed with reduced cheaper setups that are still scientifically sound to determine the benefits enabled by the high-resolution on the representation of key ocean processes.
- Parallel work on (simple) coupled assimilation approaches yields promising results, as illustrated by the Northern hemisphere sea-ice area in September. It is still to be shown if this can be translated into improved decadal climate predictions.

- Stochastic physics schemes show positive impact on reliability and also partly for skill for tropical ocean SSTs on multi-annual time scales. The low computational costs of these schemes motivates their application for predictions on decadal time scales.
- Results from AMV experiments show that simulated responses to an observed North Atlantic response are functions of the mean model state, in particular of how models simulate mean precipitation. This demonstrates the importance of reducing mean model biases to simulate realistic teleconnections between the north Atlantic and remote regions, a prerequisite to improve climate prediction system skills.

5. Links built

- Collaboration within the EC-Earth consortium (in particular BSC and SMHI) in developing and tuning a new, high-resolution model to be used for both long-term climate simulations and predictions.
- Initiatives for future collaboration (in particular between Met Office and CMCC) in the direction of understanding model deficiencies, specifically referring to the signal-to-noise problem in seasonal-to-decadal predictions.
- Increased presence of European partners in international research activities like those promoted by the World Climate Research Programme.

6. Acronyms

BSC: BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION

CMCC: CENTRO EURO-MEDITERRANEO SUI CAMBIAMENTI CLIMATICI

CNRS: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS

IPSL: INSTITUT PIERRE SIMON LAPLACE

UOXF: UNIVERSITY OF OXFORD

SMHI: SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT

7. References

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., et al. (2020). Historical simulations with HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001995. <https://doi.org/10.1029/2019MS001995>

Acosta Navarro, J. C., et al. "Link between autumnal Arctic Sea ice and Northern Hemisphere winter forecast skill." *Geophysical Research Letters* 47.5 (2020): e2019GL086753.

Athanasiadis, P.J., Yeager, S., Kwon, Y.O. et al. Decadal predictability of North Atlantic blocking and the NAO. *npj Clim Atmos Sci* 3, 20 (2020). <https://doi.org/10.1038/s41612-020-0120-6>

Athanasiadis, P., Ogawa, F., Omrani, N-E., Keenlyside, N., Schiemann, R., Baker, A. J., Vidale, P. L., Bellucci, A., Ruggieri, P., Rein Haarsma, R., Roberts, M., Roberts, C., Lenka Novak, L. and S. Gualdi: Mitigating climate biases in the mid-latitude North Atlantic by increasing model resolution: SST gradients and their relation to blocking and the jet (in review for the *J. Clim.*). [Currently accessible via this link:](#)
<https://drive.google.com/drive/folders/1m5t9vtDETH7UamCl8b7qXGhWnqo78jzx?usp=sharing>

Befort, D. J., O'Reilly, C. H., & Weisheimer, A. (2021). Representing model uncertainty in multiannual predictions. *Geophysical Research Letters*, 48, e2020GL090059. <https://doi.org/10.1029/2020GL090059>

Berner, J., Achatz, U., Batté, L., Bengtsson, L., de la Cámara, A., Christensen, H. M., et al. (2017). Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3), 565–588. <https://doi.org/10.1175/BAMS-D-15-00268.1>

Bilbao, Roberto, et al. "Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth." *Earth System Dynamics* 12.1 (2021): 173-196.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, *Geoscientific Model Development*, 9, 3751–3777, 2016. <https://doi.org/10.5194/gmd-9-3751-2016>.

Bretherton, C. S., Widmann, M. & Dymnikov, V. P. et al. The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* 12, p.1990–2009, (1999).

Cherchi, A., and Coauthors, 2019: Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. *Journal of Advances in Modeling Earth Systems*, 11, 185–209, <https://doi.org/10.1029/2018MS001369>.

Czaja, A., Frankignoul, C., Minobe, S. & Vannièrè, B. Simulating the midlatitude atmospheric circulation: what might we gain from high-resolution modeling of air–sea interactions? *Curr. Clim. Change Rep.* 5, 390–406 (2019).

Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N. The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination, *Tellus A: Dynamic Meteorology and Oceanography* 57:3, 234-252 (2005) <https://doi.org/10.3402/tellusa.v57i3.14658>

Doblas-Reyes, F. J., Navarro, J. C. A., Batté, L., Volpi, D., Acosta, M., Bellprat, O., et al. (2018). Using EC-Earth for climate prediction research. *ECMWF Newsletter*, 154, 35–40. <https://doi.org/10.21957/fd9kz3>

Dobrynin, M., D. I.V. Domeisen, W. A. Müller, L. Bell, S. Brune, F. Bunzel, K. Fröhlich, H. Pohlmann, J. Baehr, 2018: Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, 45. <https://doi.org/10.1002/2018GL077209>

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégos, M., Miller, P. A., Moreno-Chamarro, E., Nieradzic, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6, *Geosci. Model Dev. Discuss.* [preprint], <https://doi.org/10.5194/gmd-2020-446>, accepted for publication in *Geosci. Model Dev.*, 2021.

Donegan, S., et al. "Conditioning ensemble streamflow prediction with the North Atlantic Oscillation improves skill at longer lead times." *Hydrology and Earth System Sciences* 25.7 (2021): 4159-4183.

Famoos Paolini, L., Athanasiadis, P. J., Ruggieri, P. and Bellucci, A. The atmospheric response to meridional shifts of the Gulf Stream SST front and its dependence on model resolution (in review for the *J. Clim.*) [Currently accessible via this link:](#)
<https://drive.google.com/drive/folders/1m5t9vtDETH7UamCl8b7qXGhWnqo78jzx?usp=sharing>

Goessling, H. F., et al. "Predictability of the Arctic sea ice edge." *Geophysical Research Letters* 43.4 (2016): 1642-1650.

Gutjahr, O., D. Putrasahan, K. Lohmann, J. H. Jungclaus, J. S. von Storch, N. BrÄijggemann, H. Haak, and A. Stossel, 2019: Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geosci. Model Dev.*, 12, 3241–3281, <https://doi.org/10.5194/gmd-12-3241-2019>

Haarsma, R., Acosta, M., Bakhshi, R., Bretonnière, P., Caron, L., Castrillo, M., & et al. (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – Description, model computational performance and basic validation. *Geoscientific Model Development*, 13, (8), 3507–3527. <http://dx.doi.org/10.5194/gmd-13-3507-2020>

Hermanson, L., Bilbao, R., Dunstone, N., Ménégos, M., Ortega, P., Pohlmann, H., Robson, J. I., Smith, D. M., Strand, G., Timmreck, C., Yeager, S., and Danabasoglu, G.: Robust Multiyear Climate Impacts of Volcanic Eruptions in Decadal Prediction Systems, *Journal of Geophysical Research: Atmospheres*, 125, 2020.

Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; Simmons, A.; Soci, C.; Abdalla, S.; Abellan, X.; Balsamo, G.; Bechtold, P.; Biavati, G.; Bidlot, J.; Bonavita, M.; De Chiara, G.; Dahlgren, P.; Dee, D.; Diamantakis, M.; Dragani, R.; Flemming, J.; Forbes, R.; Fuentes, M.; Geer, A.; Haimberger, L.; Healy, S.; Hogan, R. J.; Holm, E.; Janiskova, M.; Keeley, S.; Laloyaux, P.; Lopez, P.; Lupu, C.; Radnoti, G.; de Rosnay, P.; Rozum, I.; Vamborg, F.; Villaume, S. & Thepaut, J. N.: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999-2049, 2020.

Jianping, L. & Wang, J. X. L. A new North Atlantic oscillation index and its variability. *Adv. Atmos. Sci.* 20, 661–676 (2003).

Joyce, T. M., Y.-O. Kwon, and L. Yu, 2009: On the Relationship between Synoptic Wintertime Atmospheric Variability and Path Shifts in the Gulf Stream and the Kuroshio Extension. *J. Climate*, 22 (12), 3177–3192, doi:10.1175/2008JCLI2690.1.

Joyce, T. M., Y.-O. Kwon, H. Seo, and C. C. Ummerhofer, 2019: Meridional Gulf Stream Shifts Can Influence Wintertime Variability in the North Atlantic Storm Track and Greenland Blocking. *Geophys. Res. Lett.*, 46, 1702–1708, doi:10.1029/2018GL081087.

Kennedy, J., H. Titchner, N. Rayner, and M. Roberts, 2017: input4MIPs.MOHC.SSTsAndSeaIce.HighResMIP.MOHC-HadISST-2-2-0-0-0, version 20170505. Earth System Grid Federation, <https://doi.org/10.22033/ESGF/input4MIPs.1221>

Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., et al. (2020). Seasonal to decadal predictions with MIROC6: Description and basic evaluation. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002035. <https://doi.org/10.1029/2019MS002035>

Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann, 2005: A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.*, 32, L20708, doi:10.1029/2005GL024233.

Kohyama, T., Yamagami, Y., Miura, H., Kido, S., Tatebe, H. and Watanabe, M. (2021). The Gulf Stream and Kuroshio Current are synchronized. *Science*, v.374, p.341-346. <https://www.science.org/doi/abs/10.1126/science.abh3295>

Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., et al. (2017). Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2315–2339. <https://doi.org/10.1002/qj.3094>

Li, X., Xie, S. P., Gille, S. T. & Yoo, C. Atlantic-induced pan-tropical climate change over the past three decades. *Nat. Clim. Change* 6, 275–279 (2016).

Lock, S.-J., Lang, S. T. K., Leutbecher, M., Hogan, R. J., & Vitart, F. (2019). Treatment of model uncertainty from radiation by the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme and associated revisions in the ECMWF ensembles. *Quarterly Journal of the Royal Meteorological Society*, 145(S1), 75–89. <https://doi.org/10.1002/qj.3570>

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M earth system model version 1.2 (MPI-ESM1.2) and its response to increasing CO₂. *Journal of Advances in Modeling Earth Systems*, 11, 998–1038. <https://doi.org/10.1029/2018MS001400>

McGregor, S. et al. Recent Walker circulation strengthening and Pacific cooling amplified by Atlantic warming. *Nat. Clim. Change* 4, 888–892 (2014).

Merryfield, W. J., Baehr, J., Batté, L., et al. Current and Emerging Developments in Subseasonal to Decadal Prediction, *Bulletin of the American Meteorological Society* 101(6), E869-E896 (2020) <https://doi.org/10.1175/BAMS-D-19-0037.1>

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A higher-resolution version of the Max Planck institute earth system model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, 10, 1383–1413. <https://doi.org/10.1029/2017MS001217>

O'Reilly, C. H. (2018). Interdecadal variability of the ENSO teleconnection to the wintertime North Pacific. *Climate Dynamics*, 51, 3333–3350. <https://doi.org/10.1007/s00382-018-4081-y>

Ossó, A., Sutton, R., Shaffrey, L. and Dong, B. (2020) Development, amplification, and decay of Atlantic/European summer weather patterns linked to spring North Atlantic sea surface temperatures. *Journal of Climate*, 33 (14). pp. 5939-5951. ISSN 1520-0442 doi: <https://doi.org/10.1175/jcli-d-19-0613.1>

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Délecluse, P., et al. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85(6), 853–872. <https://doi.org/10.1175/BAMS-85-6-853>

Parfitt, R., A. Czaja, S. Minobe, and A. Kuwano-Yoshida, 2016: The atmospheric frontal response to SST perturbations in the Gulf Stream region. *Geophys. Res. Lett.*, 43, 2299–2306, doi:10.1002/2016GL067723.

Pegion, P., Whitaker, J., Hamill, T., Bates, G., Gehne, M., & Kolczynski, W., Jr. (2016). Stochastic parameterization development in the NOAA/NCEP global forecast system. Paper presented at Proceedings of the ECMWF/WWRP Workshop: Model Uncertainty. Reading, UK.

Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, J. (2019). Realistic quasi-biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophysical Research Letters*, 46, 14118–14125. <https://doi.org/10.1029/2019GL084878>

Roberts, C. D., R. Senan, F. Molteni, S. Boussetta, M. Mayer, and S. P. E. Keeley, 2018: Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43R1) for HighResMIP. *Geosci. Model Dev.*, 11, 3681–3712, <https://doi.org/10.5194/gmd-11-3681-2018>

Roberts, M. J., and Coauthors, 2019: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments. *Geosci. Model Dev.*, 12, 4999–5028, <https://doi.org/10.5194/gmd-12-4999-2019>

Robson, J., Ortega, P. & Sutton, R.: A reversal of climatic trends in the North Atlantic since 2005. *Nature Geosci* 9, 513–517 (2016). <https://doi.org/10.1038/ngeo2727>

Robson, J., Polo, I., Hodson, D.L.R. et al.: Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model. *Clim Dyn* 50, 921–937 (2018). <https://doi.org/10.1007/s00382-017-3649-2>

Ruprich-Robert Y., and Coauthors (2021) “Impacts of Atlantic Multidecadal Variability on the Tropical Pacific: a multi-model study”, *npj Climate and Atmospheric Science*, <https://doi.org/10.1038/s41612-021-00188-5>

Scaife, A. A., and Coauthors, 2014: Skillful long range prediction of European and North American winters. *Geophys. Res. Lett.*, 41, 2514–2519, <https://doi.org/10.1002/2014GL059637>

Scaife, A.A., J. Camp, R. Comer, P. Davis, N. Dunstone, M. Gordon, C. MacLachlan, N. Martin, Y. Nie, H-L. Ren, M. Roberts, W. Robinson, D. Smith and P-L. Vidale, 2019, Does increased atmospheric resolution improve seasonal climate predictions?, *Atm Sci Lett*, DOI: 10.1002/asl.922.

Scherrer, S. C., Croci-Maspoli, M. & Schwierz, C. et al. Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *Intern. J. Clim.* 26, 233–249 (2006).

Schiemann, R., and Coauthors, 2017: The resolution sensitivity of Northern Hemisphere blocking in four 25-km Atmospheric Global Circulation Models. *J. Climate*, 30, 337–358, <https://doi.org/10.1175/JCLI-D-16-0100.1>

Schiemann, R., and Coauthors, 2020: Northern Hemisphere blocking simulation in current climate models: evaluating progress from the Climate Model Intercomparison Project Phase 5 to 6 and sensitivity to resolution. *Wea. Clim. Dyn.*, 1, 277–292, <https://doi.org/10.5194/wcd-1-277-2020>

Schubert, S., and Coauthors, 2009: A U.S. CLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: Overview and results. *J. Climate*, 22, 5251–5272, doi:10.1175/2009jcli3060.1.

Sein, D. V., and Coauthors, 2017: Ocean modeling on a mesh with resolution following the local Rossby radius. *J. Adv. Model Earth Syst.*, 9, 2601–2614, <https://doi.org/10.1002/2017MS001099>

Siegert, S., Bellprat, O., Ménégoz, M., Stephenson, D. B., & Doblas-Reyes, F. J. (2017). Detecting improvements in forecast correlation skill: Statistical testing and power analysis. *Monthly Weather Review*, 145(2), 437–450. <https://doi.org/10.1175/MWR-D-16-0037.1>

Smirnov, D., M. Newman, M. A. Alexander, Y.-O. Kwon, and C. Frankignoul, 2015: Investigating the local atmospheric response to a realistic shift in the Oyashio sea surface temperature front. *J. Climate*, 28, 1126–1147, <https://doi.org/10.1175/JCLI-D-14-00285.1>

Smith, D. M., Eade, R., Scaife, A. A. et al. (2019). Robust skill of decadal climate predictions, *npj Clim. Atmos. Sci.* 2, 13. <https://doi.org/10.1038/s41612-019-0071-y>

Smith, D. M. et al, Robust but weak winter atmospheric circulation response to future Arctic sea ice loss, *Nature Communications* (accepted).

Smith D. et al. (2020). North Atlantic climate far more predictable than models imply. *Nature*, 583, 796-800. <https://doi.org/10.1038/s41586-020-2525-0>.

Sutton, R., and B. Dong, 2012: Atlantic Ocean influence on a shift in European climate in the 1990s. *Nat. Geosci.*, 5, 788–792, doi:10.1038/ngeo1595.

Swingedouw, D., Mignot, J., Ortega, P., Khodri, M., Menegoz, M., Cassou, C., & Hanquiez, V.: Impact of explosive volcanic eruptions on the main climate variability modes. *Global and Planetary Change*, 150, 24– 45, 2017. <https://doi.org/10.1016/j.gloplacha.2017.01.006>.

Tian, T., et al.: Benefits of sea ice initialization for the interannual-to-decadal climate prediction skill in the Arctic in EC-Earth3, *Geosci. Model Dev.*, 14, 4283–4305 (2021).

Tibaldi, S. & Molteni, F. On the operational predictability of blocking. *Tellus* 42, 343–365 (1990).

Tietsche, S., Balmaseda, M., Zuo, H. et al. The importance of North Atlantic Ocean transports for seasonal forecasts. *Clim Dyn* 55, 1995–2011 (2020). <https://doi.org/10.1007/s00382-020-05364-6>.

Trenberth, K.E. and Shea, D.J.: Atlantic hurricanes and natural variability in 2005. *Geophysical Research Letters* 33 (2006) <https://doi.org/10.1029/2006GL026894>.

Voltaire, A., and Coauthors, 2019: Evaluation of CMIP6 deck experiments with CNRM-CM6-1. *J. Adv. Model Earth Syst.*, 11, 2177–2213, <https://doi.org/10.1029/2019MS001683>.

Wallace, J. M. & Gutzler, D. S. Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.* 109, 784–812 (1981).

Weisheimer, A., Palmer, T. N., & Doblas-Reyes, F. J. (2011). Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophysical Research Letters*, 38, L16703. <https://doi.org/10.1029/2011GL048123>

Willison, J., W. A. Robinson, and G. M. Lackmann, 2013: The importance of resolving mesoscale latent heating in the North Atlantic storm track. *J. Atmos. Sci.*, 70, 2234–2250, <https://doi.org/10.1175/JAS-D-12-0226.1>

Wilks, D. S.: *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series, Academic Press 100, 301-394 (2011) <https://doi.org/10.1016/B978-0-12-385022-5.00008-7>

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, 10, 357–380. <https://doi.org/10.1002/2017MS001115>

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A. et al. A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model, *Bull. Am. Meteorol. Soc.* 99, 1867–1886 (2018), <https://doi.org/10.1175/BAMS-D-17-0098.1>.

Zhang, W., Kirtman, B., Siqueira, L., Clement, A. & Xia, J. Understanding the signal-to noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled 563 model. *Climate Dynamics* 56, 2895–2913 (2021).

List of figures

- **Figure 3.1.1:** Ensemble mean, pan-Arctic integrated ice edge error (IIEE, Goessling et al., 2016) as a function of forecast year for 6-model ensembles of CMIP6 in DCPP (55 members) and Historical experiments (55 members) in a) March and b) September. The IIEE is computed from the SIC absolute values. Difference in anomaly correlation of sea-ice concentration between DCPP and Historical 6-model ensemble means for c) March monthly-mean of 1st forecast year, d) September monthly-mean of 1st forecast year, e) March monthly-mean of 6th forecast year and d) September monthly-mean of 6th forecast year. Dots indicate statistical significance at the 95% confidence level. Observational reference: HadISSTv1.1.
- **Figure 3.2.1:** a) Strength of the Barents/Kara Sea Ice-Western Russia teleconnection in winter as defined in Acosta Navarro et al. (2020) and its link with the skill for Western Russia's atmospheric circulation (purple box in c) in the multi-model ensemble members selections (dots). Each one of the 1000 dots represents the average of 11 members randomly chosen. The strength of the link in observations is represented by the thick vertical black line (thin lines indicate its uncertainty), and for the individual forecast members with triangles. b) Increment in ACC skill for winter sea level pressure when using the subset of ensemble members with a more realistic representation of the teleconnection, as compared to the full multi-model ensemble. ACC scores are computed after linearly detrending the anomalies. Stippling represents significant values at the 95% confidence level. c) The same as for b), but for surface air temperature. Decadal predictions include 55 members from 6 different DCPP systems, and their skill is computed for forecast years 2-10. Observational references for evaluation are HadISSTv1.1 (a) and JRA55 (b-c).
- **Figure 3.3.1:** (left) Timeseries of the mixed layer depth in the Labrador Sea in March for the different high-resolution experiments with a fixed forcing of the year 1980 (coloured lines) compared to the standard-resolution experiments with a fixed forcing of the pre-industrial period (black dashed lines) and with the historical forcing (1850-2014) (grey lines). The mixed layer depth is a proxy of the deep ocean convection. (right) Timeseries of the AMOC strength at 26°N for the same experiments (only one standard resolution experiment with historical forcing - grey line). The decrease in the formation of deep water masses directly impacts the strength of the AMOC. The y-axis on the figures correspond to simulation years, not to calendar years.
- **Figure 3.3.2:** (left) JJA sea-ice concentration anomaly correlation coefficient between 10-member ensemble mean from EC-Earth3.2 and CERSAT for the period 1993-2015. (center) Same as (left) but for the new tuned EC-Earth3.3 configuration. (right) Added value of the new tuned version of the model shown as the difference in correlation coefficient between EC-Earth3.2-HR and EC-Earth3.3-HR, both against CERSAT.
- **Figure 3.4.1:** Global annual mean surface temperature (°C) response (dcpA-dcpC) to the eruption of a) Mount Agung (1963), b) El Chichón (1982), c) Mount Pinatubo (1991) and d) the mean of the three eruptions. Filled dots indicate statistically significant differences between the hindcasts with and without the volcanic forcing.
- **Figure 3.4.2:** Multi-model mean surface temperature (°C) response (dcpA-dcpC) the first year (June-May) following the eruptions of a) Mount Agung (1963), b) El Chichón (1982), c) Mount Pinatubo (1991) and d) the mean of the three volcanoes. Hatching indicates statistically significant anomalies and shading indicates model sign consistency.

- **Figure 3.5.1:** Predictive skill for high-latitude blocking and the NAO. 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 defined by the start lead-year (ordinate) and the end lead-year (abscissa). The cyan 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). Reproduced from Athanasiadis et al. (2020).
- **Figure 3.6.1:** Differences in wintertime absolute bias between “LR” and “HR” multi-model means (Table 3.5.1). Upper panel: SST biases (shading, in K) with the HadISST2 1950–2014 climatology in contours (°C). Lower panel: blocking frequency biases (shading, in % of blocked days) with the ERA-JOINT climatological blocking frequency in contours (% of blocked days). Details on the blocking detection method are provided in Athanasiadis et al. (in review).
- **Figure 3.7.1:** (Top) Winter-mean latitude of the GSF averaged in the range 50°–68°W. Red, yellow and blue stars represent years in which the GSF latitude falls, respectively, in the upper, middle and lower tercile categories (i.e. North, Middle and South position). The dashed line is the 10-year running mean applied to GSF latitude timeseries. (Bottom) SST (K; colour shaded) anomalies associated to “North” (left) and “South” (right) phases of the GSF in winter (DJF) in HadISST2 dataset. The “North” (“South”) phase has been defined as the upper (lower) tercile of GS SST front mean latitude. The respective climatological position of the SST front is indicated by the cyan dashed line. The black contours indicate winter SST climatology. Black dots denote anomalies that were found to be statistically significant at the 90% confidence level.
- **Figure 3.7.2:** 850 hPa zonal wind (m s^{-1} ; colour shaded) response to the GSF shift in winter (DJF). (a) ERA5. (b, c) EC-Earth. (d, e) MOHC. (f, g) ECMWF. Beside institution name, the model nominal resolution in km is reported. Black dots denote anomalies that were found to be statistically significant at the 90% confidence level. Green contours indicate the winter climatology of zonal wind at 850 hPa every 2 m s^{-1} from 4 m s^{-1} . The winter climatological position of the GSF is indicated by the black dashed line.
- **Figure 3.8.1:** Evolution of the SST in the Labrador Sea (40–60°W, 45–65°N) in the HadISST observation in red, in the nudged simulation including (in blue) or not (in black) the Greenland ice sheet melting for a) the hindcasts with the melting and b) the hindcasts including the melting. The thick lines represent the ensemble mean of the different starting dates, while the overlap represents one standard deviation of the spread. A 3-year running mean has been applied to all time series.
- **Figure 3.8.2:** Root Mean Square Error (RMSE) for the SST averaged for the 7 starting dates of the ensemble mean of the hindcasts over 10 years. a) for the hindcasts that do not include the melting and b) for the hindcasts including it. c) Difference of RMSE between the Melting and Reference hindcasts.
- **Figure 3.10.1:** ACC for the annual mean 2m-temperature calculated over initializations s1990-s2004 for Ec-Earth3 standard resolution (T255) (top), high resolution (T511) (middle) and the differences between T511 and T255 (bottom) for the first full forecast year (left) and the average of lead years 2–5 (right); hatching masks significant ACC ($p < 0.05$); ERA5-reanalysis as observational reference.

- **Figure 3.10.2:** Predicted evolution of SPG and AMO temperature anomalies in comparison with ORAS5 reanalysis. Left: Ensemble prediction for lead-1y. Right: Ensemble prediction for lead-5y. The ± 1 sigma ensemble spread is shown based on 15 members for the standard resolution (SR, T255) in gray and 10 members for the high resolution (HR, T511) in light red. Anomalies are relative to the mean of 1981–2010. The AMO was globally detrended following Trenberth and Shea (2006).
- **Figure 3.10.3:** NH September sea-ice area according to the CERSAT observational product (red line) as well as an ensemble of free running EC-Earth simulations (grey lines; following CMIP6-historical protocol and SSP2-4.5 after 2014) and the new set of assimilation experiments (blue lines) that assimilate anomalies of SST and low-level winds in the atmosphere.
- **Figure 3.11.1:** (a) Anomaly correlation coefficients for SSTs over NINO3 region using ERA5 as reference, (b) same as (a) but for SoE, (c) same as (a) but for the North Pacific index (NPI), (d) same as (c) but for SoE. Gray shading for the MME indicates 2.5 and 97.5 percentile derived from randomly sampling (10,000 samples) two members from each single-model ensemble. Dots in (a) and (c) indicate forecast times for which the respective ensemble is significantly larger than 0, whereas dots in (b) and (d) indicate forecast times for which the respective ensemble is significantly different from 1 (95% confidence, 10,000 samples). Samples have been generated by bootstrapping over years for ECMWF-SPPT and ECMWF-noSPPT ensembles and over years and members for the MME. Orange and red circles in (a) and (c) indicate those forecast times for which the respective ECMWF ensemble shows significantly higher skill compared to the other ECMWF ensemble (orange: 10% and red: 5% significance level, following Siegert et al., 2017). SST, sea surface temperature; SoE, spread-over-error; MME, multimodel ensemble; SPPT, stochastically perturbed physical tendency. Figure taken from Befort et al. (2021); their Fig. 3.
- **Figure 3.12.1:** Inter-model relationship between several indices. Markers represent the 10-year averaged ensemble mean of the difference between AMV+ and AMV– simulations from individual experiments. (left) Winter NIÑO3.4 SST index versus winter tropical North Atlantic SST (averaged over 5°N–20°N/60°W–10°E). (middle) Inter-model relationship between the wintertime NIÑO3.4 SST index and the outputs of a bi-linear regression model built with the summertime Tropical Atlantic ascent and the injection of moist static energy from the Tropical Atlantic surface. (right) Whisker box plots indicating the minimum/maximum values, the 20th/80th percentiles, and the median from the inter-model distribution of several indices: the wintertime NIÑO3.4 SST index (black), the outputs of the regression model fed with summertime Tropical Atlantic ascent, and (blue) the injection of moist static energy from the Tropical Atlantic surface, (green) same as blue but with the injection of moist static energy corrected for the model precipitation climatology biases.