H2020 - Research and Innovation Action

APPLICATE

Advanced Prediction in Polar regions and beyond: Modelling, observing system design and LInkages associated with a Changing Arctic climaTE
Grant Agreement No: 727862

Deliverable No. 5.3

Report on individual impacts of improved process-representation, treatment of snow, ensemble generation and increased resolution on the weather and climate prediction performance

Submission of Deliverable

Work Package	WP5
Deliverable No	5.3
Deliverable title	Report on individual impacts of improved process-representation, treatment of snow, ensemble generation and increased resolution on the weather and climate prediction performance
Version	1
Status	Final
Dissemination level	Public
Lead Beneficiary	Met Norway
Contributors	■ 1 – AWI ■ 2 – BSC ■ 3 - ECMWF □ 4 – UiB □ 5 – UNI Research ■ 6 – MET Norway □ 7 – Met Office ■ 8 – UCL □ 9 - UREAD □ 10 – SU ■ 11 – CNRS-GAME □ 12 - CERFACS □ 13 – AP □ 14 – UiT □ 15 - IORAS □ 16 - MGO
Due Date	31 October 2019
Delivery Date	31 October 2019
Coordinating author	Morten Køltzow (famo@met.no)
Contributing authors	Niramson Azouz (CNRS), Lauriane Batté (Meteo-France), Eric Bazile (Meteo-France), Adrien Napoly (Meteo-France) Ilona Välisuo (CNRS), Juan Camilo Acosta Navarro (BSC), Pablo Ortega (BSC), Eduardo Moreno-Chamarro (BSC), Rafael Grote (MET Norway), Jonathan Day (ECMWF), Tido Semmler (AWI), François Massonnet (UCLouvain), Leandro Ponsoni (UCLouvain)

This project has received funding from the European Union's Horizon 2020 Research & Innovation programme under grant agreement No. 727862.

Table of Contents

EXECUTIVE SUMMARY

This document provides an overview of individual impacts of improved description of Arctic processes, model resolution and ensemble generation in weather and climate predictions based on work in APPLICATE work package 5.3. The report provides recommendations on which of the developments to be included in the enhanced (stream 2) predictions performed for Deliverable 5.4, whose purpose is to assess the added-value of APPLICATE on weather and climate predictions.

This document shows that for short- and medium-range predictions enhancement in prediction skill can be achieved by introducing a sea ice model in the Meteo France weather prediction systems (improve near-surface temperature) and a multi-layer snow scheme in the European Center for Medium Weather Forecasting (ECMWF) forecast systems (improve near-surface temperature and snow depth). These changes are recommended to be included in their respective systems. The superiority of surface assimilation compared to dynamical downscaling of global forecasts in regional forecast systems further underlines the importance of the surface processes.

Regional high-resolution systems will benefit from both a further increase in resolution and introduction of Ensemble Prediction Systems (EPSs) to account for the uncertainty in the predictions. Both these enhancements are recommended, but require a substantial increase in operational computer power and should therefore be considered depending on the use of the prediction systems.

Improving the oceanic and atmospheric resolution in seasonal predictions give rather inconclusive results. Both positive and negative impacts are found in different systems. More studies are needed before recommendations on operational use of increased computer power can be given.

In climate prediction systems finer oceanic resolution can improve the representation of the Arctic ocean with realistic atmospheric forcing. However, large-scale atmospheric circulation biases in the fully coupled system deny such an improvement.

No substantial improvements and small sensitivity in seasonal prediction skill is found by introducing a more realistic description of sea ice melt ponds. This may be related to the existing tuning of the sea ice albedo which is reduced to account for the missing melt ponds. It is therefore not recommended to include a more realistic melt pond description without any further tuning of the sea ice models.

The description of the sea ice thickness distribution in seasonal prediction systems has also been studied. A sensitivity is found on the number of sea ice categories, but

no evident benefit from including additional categories beyond the default configuration (5 levels) which it is recommended to keep.

Stochastic perturbations to represent the uncertainty in a seasonal prediction system show a neutral impact in the ocean model, but a deterioration in the atmosphere. It is therefore not recommended to use these perturbations scheme as more work, e.g. on tuning, is needed.

1. INTRODUCTION

1.1. Background and objectives

This deliverable provides an overview of work done in APPLICATE task 5.3. Model enhancements by improved physical description, resolution, ensemble and limited area model configurations are presented and assessed in a prediction framework. These enhancements are tested individually and also in coordinated exercises combining several APPLICATE prediction systems. A main objective of APPLICATE is to contribute strongly to the next generation of weather and climate prediction systems. The results presented here, together with model development work reported in other APPLICATE deliverables (like D2.3) will certainly serve as recommendations for future updates of APPLICATE (near-)operational prediction systems. Some of the recommendations will be applied in APPLICATE stream 2 experiments to quantify the added value of the project (described in more detail in Milestone 3).

Task 5.3.1 evaluates several changes in model physical descriptions that are important for the Arctic. The GELATO sea ice model has been introduced in the Meteo France weather forecasting models (AROME and ARPEGE) and evaluated. In the ECMWF Integrated Forecast System (IFS) a new multi-layer snow scheme are evaluated. This is done by traditional metrics, but also by introducing two new types of diagnostics related to the sensitivity of the surface skin temperature. In the sea ice components of the CNRM-CM6 (CNRM) and EC-Earth 3.2 (BSC) seasonal prediction systems, the impact of new melt pond descriptions and number of sea ice thickness categories on the representation of past sea ice variability have been evaluated in forecast mode and also in historical atmospheric-forced NEMO/LIM experiments. Furthermore, the sensitivity to the discretization of the sea ice thickness distribution in the NEMO3.6-LIM3 global ocean-sea ice model (UCLouvain) has also been explored and discussed.

The use of regional models can, compared with global models, add value by the use of optimized physics for the targeted area and finer resolution. In task 5.3.2, the focus is on the configuration of such regional high-resolution short-range weather forecast systems in the Arctic. The impacts of improved resolution (2,5 km to 1,25 km

horizontal resolution combined with more vertical levels in the boundary layer) is evaluated in AROME-Arctic (MET-Norway) and MF-AROME (CNRS, Meteo France). Furthermore, initialization strategies, deterministic or ensemble approach, domain location and size, lateral boundary quality and added value as a function of lead time are discussed based on AROME-Arctic experiments.

Potential improvements in seasonal prediction skill with higher resolution in the ocean/ice and atmosphere/land surface components are discussed in task 5.3.3. A set of ensemble seasonal reforecasts similar to the APPLICATE stream 1 for CNRM-CM6 (CNRM) and EC-Earth3.2 (BSC) have been performed with increased horizontal resolution in the ocean and atmosphere. Both models have increased the ocean horizontal resolution from the ORCA1° to the ORCA1/4° grid. The CNRM-CM6 atmospheric resolution has increased from ~1.4° to 0.5° (30 members) while the EC-Earth3.2 atmospheric resolution has increased from ~0.7° to ~0.35° (10 members). Analysis with a focus on the impact on the sea ice and Arctic atmospheric variables are presented. Climate experiments with resolution refinements in the ocean model FESOM (4,5 km compared to 24 km) in the coupled FESOM-ECHAM6 system (AWI) have also been assessed.

To date, no regional high-resolution EPS with focus on the Arctic are applied for operational short-range weather forecasting. In task 5.3.4, the added value of a regional high-resolution AROME-Arctic EPS (MET-Norway) is discussed together with the impact of sea surface temperature perturbations to improve the ensemble spread. Furthermore, an ensemble approach is essential for representing different types of uncertainty also in seasonal predictions. Stochastic perturbations of temperature and salinity in the ocean model NEMO 3.6 in CNRM-CM6-1 (CNRM) are therefore introduced to better represent the uncertainties in the ocean equation of state. This experiment and the associated results are presented and discussed.

1.2. Organisation of this report

Results from task 5.3 are presented sub task by sub task. The impact of process descriptions relevant for the Arctic in prediction systems are presented in section 3.1 (task 5.3.1). The impact of higher resolution and configurations of short range weather forecasts are presented in section 3.2 (task 5.3.2), while the impact of resolution for seasonal and climate time-scales are shown in section 3.3 (task 5.3.3). In section 3.4, aspects of Arctic EPS are discussed (task 5.3.4). Conclusions from all parts are summarized in section 4.

Within each section (task) the work is organized from shorter- to longer prediction time scales. All headlines are followed by the name of the APPLICATE partners contributing to the particular work presented.

2. METHODOLOGY

This report is based on a set of numerical experiments with different model systems. Most of these model systems are already described in section 2 in APPLICATE D5.2 (Batte et al., 2018) together with most of the observational data and verification metrics and diagnostics considered in this deliverable. Key information on the prediction systems, the sensitivity experiments and results is presented in the corresponding tasks. In a few cases new metrics have been introduced when appropriate.

The predictions systems;

- AROME-MF-Arctic: experiment configuration based on the previous Météo-France AROME configurations (Seity et al. 2011) with a dynamical adaptation from the Meteo-France operational global model ARPEGE. Used for comparison with AROME-Arctic and to evaluate the coupling with the sea-ice model GELATO-1D.
- AROME-Arctic; for short-range Arctic weather predictions a configuration of the Applications of Research to Operations at Mesoscale forecast system (AROME) which is in operational use at MET-Norway (Müller et al. 2017, Batte et al., 2018).
- ARPEGE; Météo-France global model used for weather predictions with a 4Dvar assimilation cycle (Pailleux et al. 2014). Used also to provide initial and lateral boundary conditions to AROME-MF-Arctic.
- ECMWF-IFS: ECMWF forecasts are produced with the ECMWF Integrated Forecasting System (IFS). The dynamical core of the model is hydrostatic and spectral, and employs a two−time−level semi−implicit semi−Lagrangian scheme combined with a spectral transform technique for the horizontal discretization and a finite−element method for the vertical discretization. Land-surface processes, including snow physics, are included in the Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL), which is coupled within the IFS. The simulations contained here use IFS Cycle 45r1 the main aspects of which are summarised here: https://www.ecmwf.int/en/forecasts/documentation-and-support/evolution-ifs/cycles/summary-cycle-45r1.
- EC-Earth 3: is an European Community Earth system model. Its atmospheric component is based on the Integrated Forecast Model (IFS) cy36r4 that has been used operationally by ECMWF for making seasonal predictions; IFS solves the primitive equations of motion for a hydrostatic atmosphere, horizontally on a spectral grid with spatial resolutions of T255/T551, and a finite-element scheme discretized in 91 levels. The ocean component of the EC-Earth model is the version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO; Madec et al., 2015). NEMO3.6 includes the ocean model

OPA (Ocean PArallelise), and the Louvain la Neuve sea ice model LIM3 (Rousset et al. 2015). EC-Earth uses ORCA1 with 75 vertical levels, with an upper level of about 3m and 10 levels distributed over the upper 100m. A high-resolution configuration with 0.25° resolution is also available, using the same vertical distribution. LIM3.6 is based on the Arctic Ice Dynamics Joint EXperiment (AIDJEX) framework, combining the ice thickness distribution framework, the conservation of horizontal momentum, an elastic-viscous plastic rheology, and energy-conserving halo-thermodynamics.

- CNRM-CM6-1: global coupled model developed jointly by CNRM and CERFACS for CMIP6 (Voldoire et al. 2019), used here in seasonal re-forecast mode initialized in May and/or November. These benchmark re-forecasts are compared to CNRM-CM6 including melt ponds parameterization in GELATO, stochastic physics in the NEMO ocean component, and higher resolution ocean (ORCA 0.25° L75) and atmosphere (tl359l91r).
- AWI-CM 1.1: global coupled model consisting of the Finite Element Sea ice Ocean Model (FESOM 1.4) developed at AWI and the atmospheric model ECHAM 6.3 developed at Max-Planck-Institute for Meteorology in Hamburg. A model description and analysis of benchmark simulations is provided in the following publications: Sidorenko et al., 2015; Rackow et al., 2016; Rackow et al., 2019. The model has been used within CMIP6 and the first data from DECK, ScenarioMIP, PAMIP, HighResMIP are published at the Earth System Grid Federation (ESGF).

3. RESULTS AND DISCUSSION

3.1 Enhanced sea ice models and air-sea interactions in weather and climate prediction (Task 5.3.1)

Environmental prediction systems for the polar regions are less developed than elsewhere (Jung et al., 2016). One of several reasons are the many local physical processes that are not well represented by the prediction systems. Historically, lower priority and lack of in situ observations have contributed to this. However, to improve our Arctic prediction capabilities a better description of Arctic processes, e.g. connected to atmosphere, sea ice and ocean interactions needs to be improved. In the following a set of experiments with more realistic description of snow and sea ice properties in weather and climate prediction are presented and discussed.

3.1.1 Impact of the GELATO1D sea ice model in AROME/ARPEGE (CNRS and Meteo France)

In the Météo-France operational Numerical Weather Prediction (NWP) models (ARPEGE and AROME) the sea-ice temperature is usually fixed during the model forecast and initialized with a climatological value. With the recent introduction of the SURFEXv8 platform in the Météo-France system, the use of the GELATO1D sea ice model (Salas y Mélia, 2002) in AROME and in ARPEGE has become possible. First, the coupling with GELATO1D has been implemented in the AROME-MF-Arctic model and evaluated with the SOP1 data. For the initialization of the GELATO variables (sea-ice thickness, temperature, enthalpy etc) to start an AROME forecast at 00 UTC every day it was decided to use the 24h GELATO1D forecast fields from the previous forecast (more details in D2.3). However this configuration was not optimum as seen with a spin-up problem at 00 UTC due to the dynamical adaptation from the ARPEGE model which does not use the GELATO1D sea-ice model. In fact, at 00 UTC, near the surface, the boundary layer comes from the ARPEGE model with a prescribed sea-ice temperature that could be significantly different from the one computed in AROME-GELATO. Figure 3.1.1.1 shows the positive impact of the coupling with the ARPEGE-GELATO with a significant reduction of the spin-up in the AROME-GELATO forecast (blue curve vs green curve).

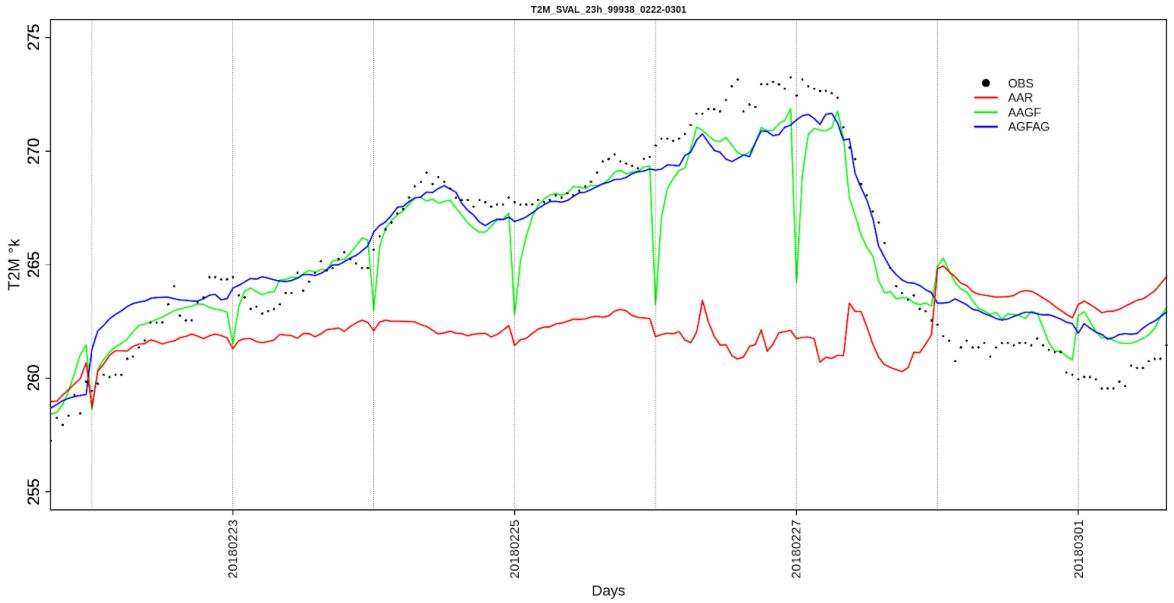


Figure 3.1.1.1 Time evolution of the 2m temperature at the KVITØYA station in Svalbard for the observation black dot,red line: AROME-MF-Arctic (AAR), green line: AROME-MF-Arctic with GELATO coupled with ARPEGE (AAGF) and blue line : AROME-MF-Arctic with GELATO coupled with ARPEGE-GELATO (AGFAG)

Finally, the implementation of the GELATO1D sea-ice model in the ARPEGE/AROME forecast system improves significantly the forecasted 2m temperature for the SOP1 period (Figure 3.1.1.2). The impact over mountains and inland stations is relatively weak, as expected: sea-ice variations poorly influence these geographical regions. Two regions are strongly affected by the sea-ice during the SOP1 in the AROME domain: fjord and Svalbard and not surprisingly, the impact of GELATO1D is very important with a reduction of the average cold bias by 1.4 and 0.8 K respectively.

Coupling ARPEGE with GELATO1D is really necessary to provide accurate initial conditions, for AROME-GELATO in a dynamical adaptation mode. Then, we can expect a combination of an increase of the horizontal resolution (1.25 km) and a coupling of GELATO1D with AROME, to improve to a larger extent the Svalbard and fjord areas.

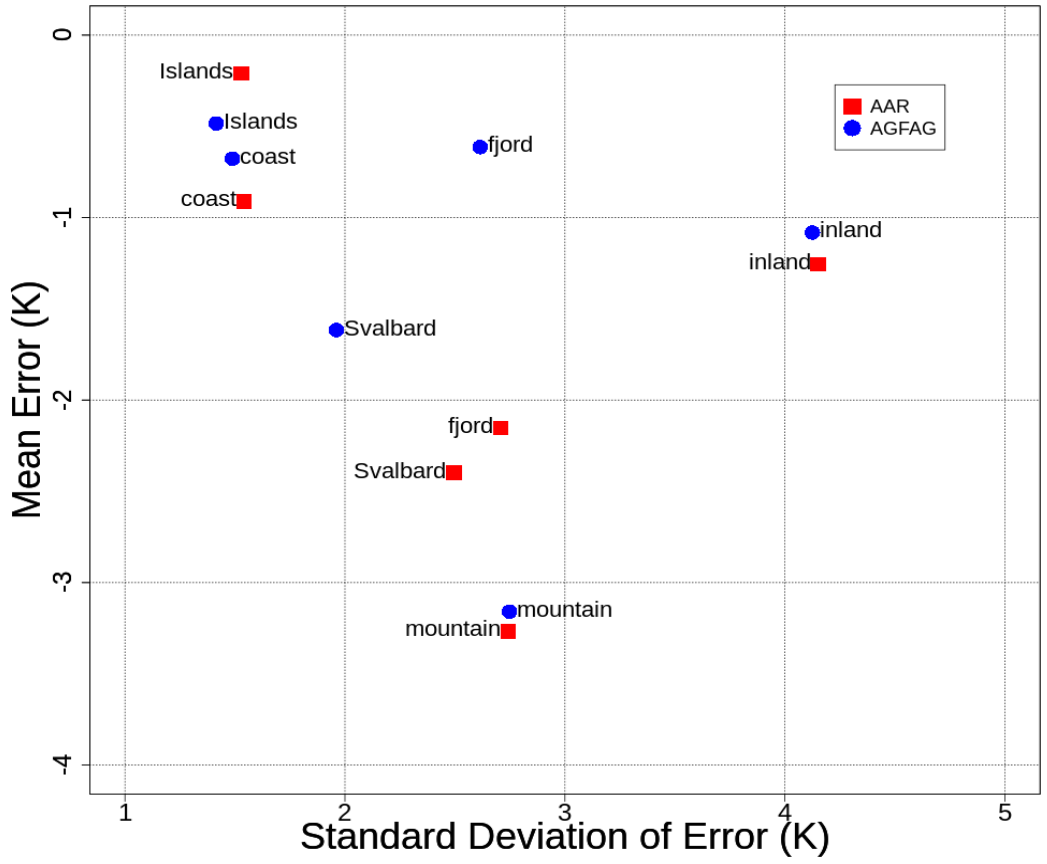


Figure 3.1.1.2: average bias and standard deviation error for the 2m temperature in several areas defined in Koltzow et al (2019). Red square for AROME-MF-Arctic (AAR), blue dot : AROME-MF-Arctic with GELATO coupled with ARPEGE-GELATO (AGFAG)

3.1.2 Understanding the impact of changes in surface physics from an energy balance perspective: Snow in Arctic winter. (ECMWF)

A new multi-layer snow model was recently introduced within the ECMWF Integrated Forecast System (IFS) as part of APPLICATE (see Arduini et al. (in review) and APPLICATE deliverables 2.3 and 2.5 for full details). It replaces a single layer scheme, that is thought to be one of the causes of 2m temperature forecast errors in the Arctic (Haiden et al., 2018). At night during clear-sky conditions the top of the snowpack cools rapidly as a result of longwave cooling at the surface and the reduced heat input from the ground underneath because of the snow insulation properties. The correct representation of these processes is challenging for NWP systems using a single-layer snow scheme, because of the large thermal inertia associated with a deep snowpack. Therefore, the implementation of multi-layer snow

schemes is currently going through a process of testing before inclusion in operational forecasts.

In Arduini et al. (in review) and D2.3 a detailed evaluation of the impact on forecasts of snow parameters was presented. Here we present an evaluation of the impact of including this model component on the skill of temperature forecasts in Northern Europe and the wider Arctic region. This study also demonstrates the use of new diagnostic techniques that could be used at any supersite (such as Sodankylä, Finland) to understand the responses of a weather or climate model to any given model physics change both in the boundary, surface and subsurface layers and determine whether the change has led to an improvement at the process level.

These diagnostics provide insights into how a given physics package is influencing the overall balance of processes within the near-surface layers of the atmosphere and subsurface and can help to anticipate and isolate any compensating errors which might become dominant when the bias they were opposing is removed. Otherwise, such compensating errors could lead to an overall decrease in skill at some locations despite improved physical realism in the model component being changed.

Evaluation against conventional weather stations (SYNOP)

An anticipated outcome of incorporating the multi-layer snow scheme is that the mean error in 2m temperature forecasts over snow-covered surfaces improves. An evaluation of the change in 2m-temperature forecast skill between the two model formulations is performed in Northern Europe and in the Arctic.

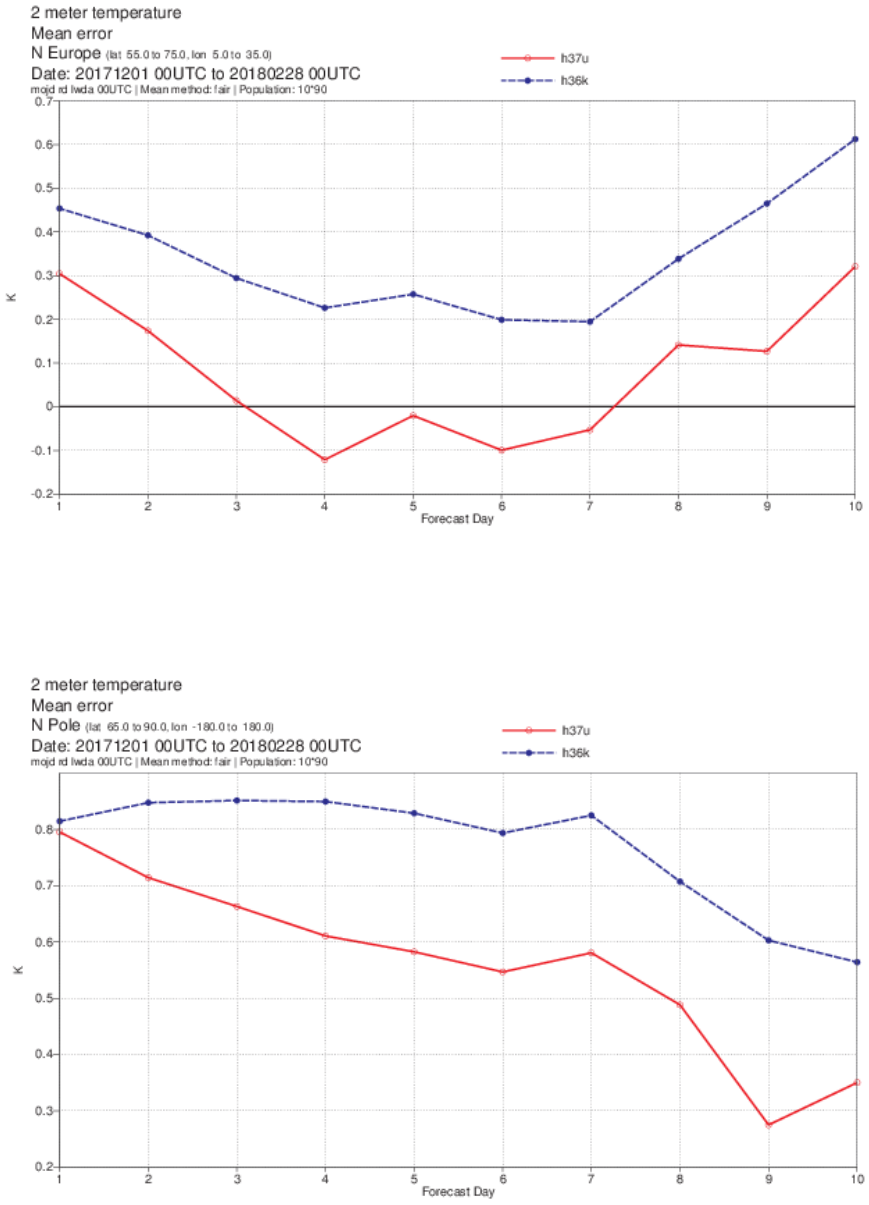


Figure 3.1.2.1. 00 UTC 2m temperature mean error for Northern Europe (top) and The North Pole (bottom), compared to SYNOP, for 10-day forecasts with single layer snow (blue: h36k) and multi-layer snow (red: h37u) for DJF 2017/18.

There is a clear reduction in the winter warm bias of 0.2-0.3 degrees when moving from the single layer control to multi-layer snow (Figure 3.1.2.1). For spatial maps of the change see Arduini et al. (in review).

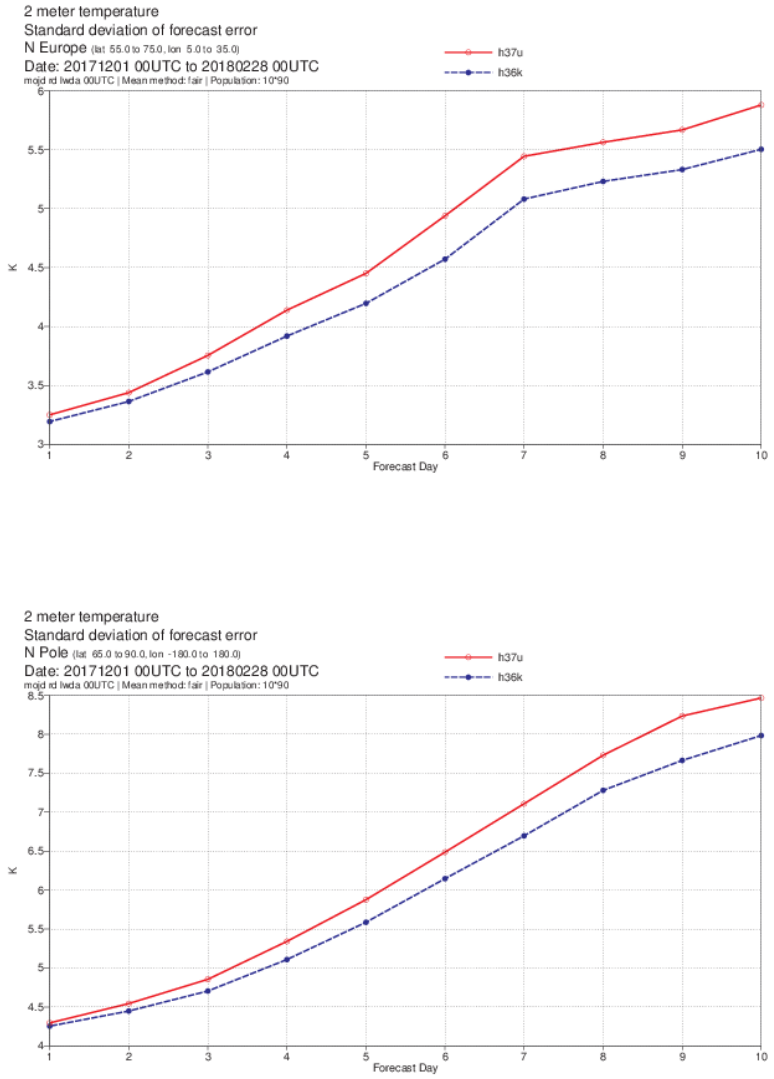


Fig 3.1.2.2. 00UTC 2m temperature standard deviation of forecast error for Northern Europe (top) and The North Pole (bottom), compared to SYNOP, for 10-day forecasts with single layer snow (blue: h36k) and multi-layer snow (red: h37u) for DJF 2017/18.

Interestingly, in the deterministic sense the standard deviation of the 2m temperature error becomes larger with the introduction of the multi-layer snow (Figure 3.1.2.2). This is because the surface becomes more sensitive to errors in synoptic conditions, e.g. cloud forcing errors. In the context of the ECMWF ensemble, this is likely to be a positive step. Bauer et al. (2016) showed that 2m temperature forecasts from the ensemble were under-dispersive in the Arctic during winter, so this increase in the standard deviation should contribute to improved forecast reliability. This improvement in the skill of ensemble forecasts can be seen in the lower values of the Continuous Ranked Probability Score for ensemble forecasts with multi-layer snow, compared to those with the single layer scheme (Figure 3.1.2.3)

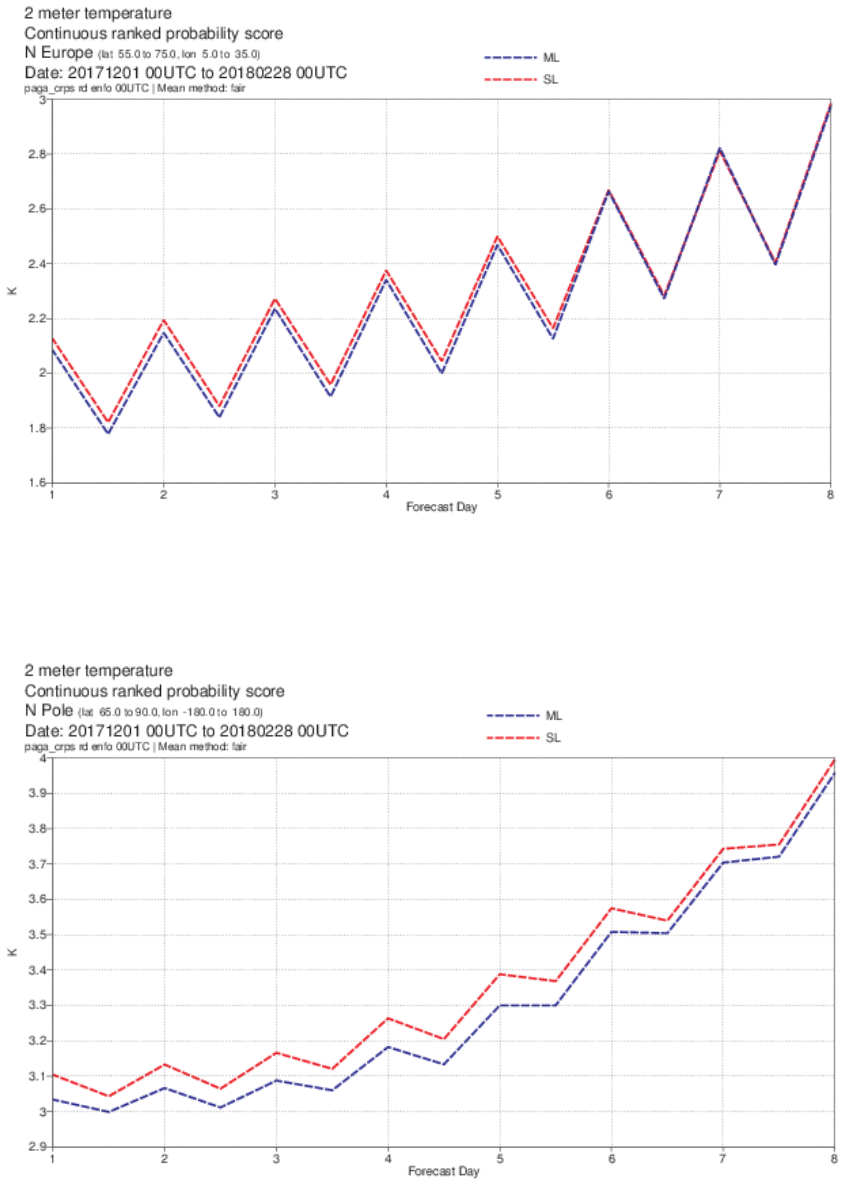


Figure 3.1.2.3. 00 and 12UTC 2m temperature Continuous Ranked Probability Score (CRPS) for Northern Europe (top) and The North Pole (bottom), compared to SYNOP, for 10-day forecasts with single layer snow (red) and multi-layer snow (blue) for DJF 2017/18.

Process based evaluation at the Sodankylä supersite

Evaluation against conventional SYNOP stations can indicate areas where a given model change has improved or degraded the forecast overall. However, due to a limited set of observed parameters, use of this observing network cannot tell us much about how the model change is affecting things at the process level. Supersites, such as Sodankylä, in Finland, collect a much wider set of observations, so that one can

explore the impact on a range of processes (Essery et al., 2016; Kangas et al., 2016; Leppänen et al., 2016). This allows detailed investigation of how a given model change affects the balance of terms in the energy, momentum and moisture budgets. Sodankylä is particularly useful for evaluating the new snow module since detailed surveys of snow thermodynamic and mechanical properties are conducted regularly at the same location as atmospheric measurements (Fig 3.1.2.4).

A simple inspection of the meteogram (Figure 3.1.2.4) shows some basic properties of the meteorology on the site. Because of the low insolation, surface and near-surface temperature largely follow changes in downwelling long-wave radiation, reflecting changes in the synoptic situation. Snow temperature also follows, but with an amplitude that decreases and a lag that increases with depth. This difference in the timescale and amplitude of the response in snow temperature with depth clearly demonstrates the motivation for including multi-layer snow.

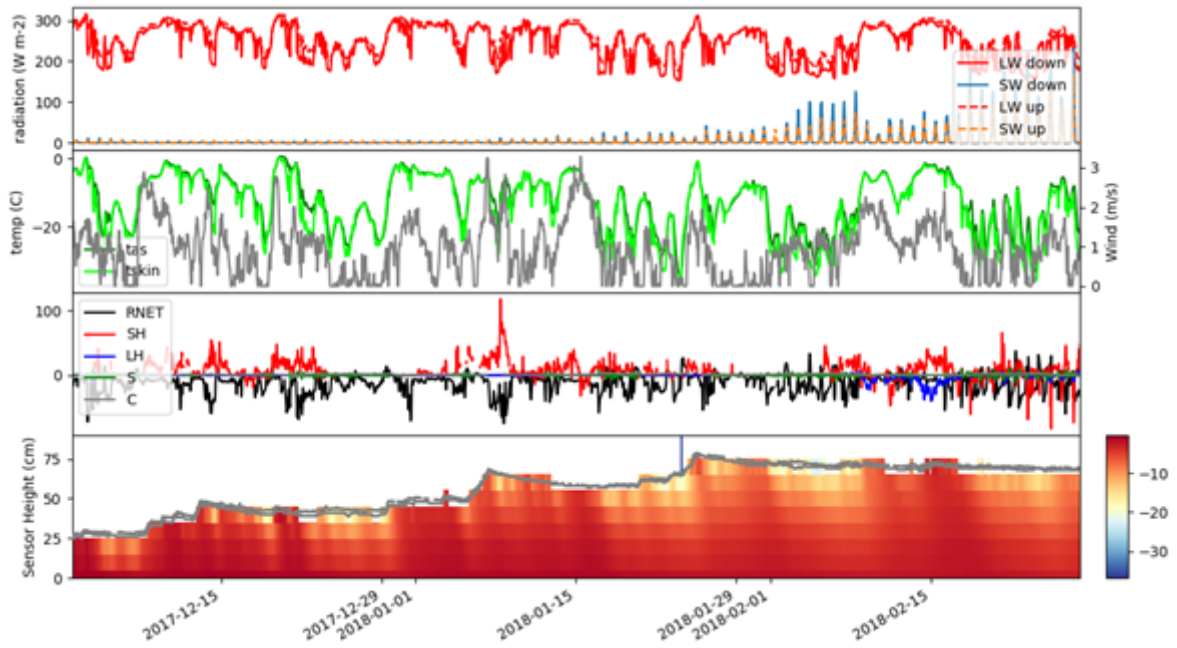


Fig 3.1.2.4. Observed meteogram for winter (DJF201718) at Sodankylä showing (top-to-bottom) radiation terms, basic meteorology, energy balance terms and snow temperature.

Representativeness of Sodankylä for Northern Europe

An important first step is to determine whether a given supersite provides a perspective on the modelled change which is consistent with the impact in the wider region. If, for example, the impact of multi-layer snow on 2m temperature in Sodankylä had a different sign to the Northern Europe region, as a whole (seen in the evaluation against SYNOP; Figure 3.1.2.1), then one would be sceptical that one can understand the regional change through understanding the impact at the Sodankylä

site. Conversely, if the sign and approximate magnitude is the same, one can be more confident that any understanding gained from diagnostic analysis at an individual site is translatable to the wider region.

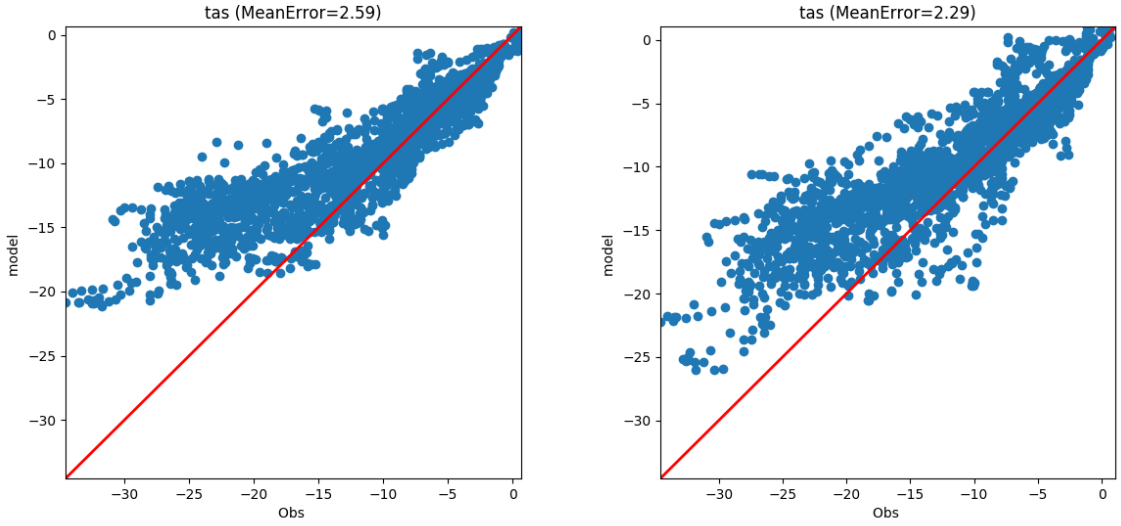


Figure 3.1.2.5. Hourly observed vs forecast (during day-2) 2m temperature for Sodankyla with single layer snow (left) and multi-layer snow (right) for DJF 2017/18.

The 2.6 degree mean warm bias in 2m temperature (at day 2) in Sodankylä is reduced by 0.3 degrees and the standard deviation of the error goes up with the introduction of multi-layer snow (Figure 3.1.2.5). This is consistent with the evaluation against SYNOP in Northern Europe, suggesting that Sodankylä is a representative site which can be used to better understand the impact of the new snow scheme from a process perspective.

Impact on vertical profiles and energy budget in cloudy and clear boundary layer regimes

It is known that in Arctic winter the boundary layer adopts distinct bimodal states associated with the transformation from marine air into polar continental air masses (Pithan et al., 2014; Stramler et al., 2011). Similarly, the boundary layer in Sodankylä exhibits such bimodality (Figure 3.1.2.6; left). In the radiatively clear state, the atmospheric profile is free of liquid-containing cloud and characterised by a boundary layer temperature inversion (Figure 3.1.2.6). In the radiatively-opaque cloudy state, which is the more common state in Sodankylä, the boundary layer is warmer, more humid, less stable and characterised by higher downwelling longwave radiation (Figure 3.1.2.6 & 3.1.2.4).

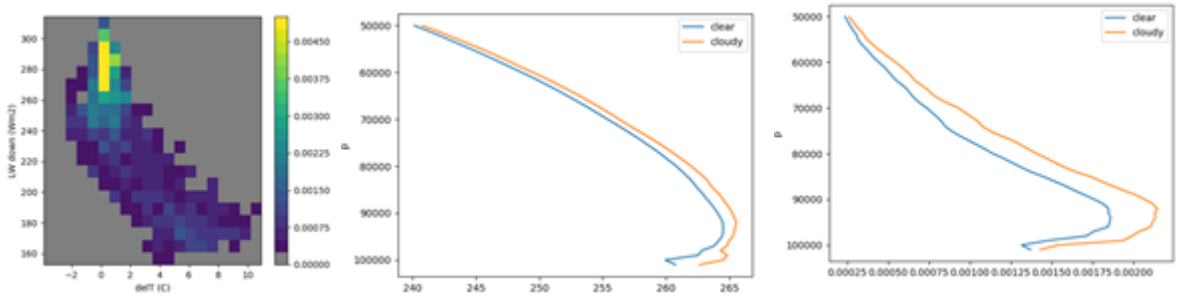


Figure 3.1.2.6. 2d histogram of downwelling longwave radiation and static stability (10m temp – skin temp) (left). Mean radiosonde temperature profiles for the clear (LWdown<210) and cloudy (LWdown>210) states (middle). (right) as (middle) but for specific humidity. All data are from Sodankylä, Finland (DJF2017/18).

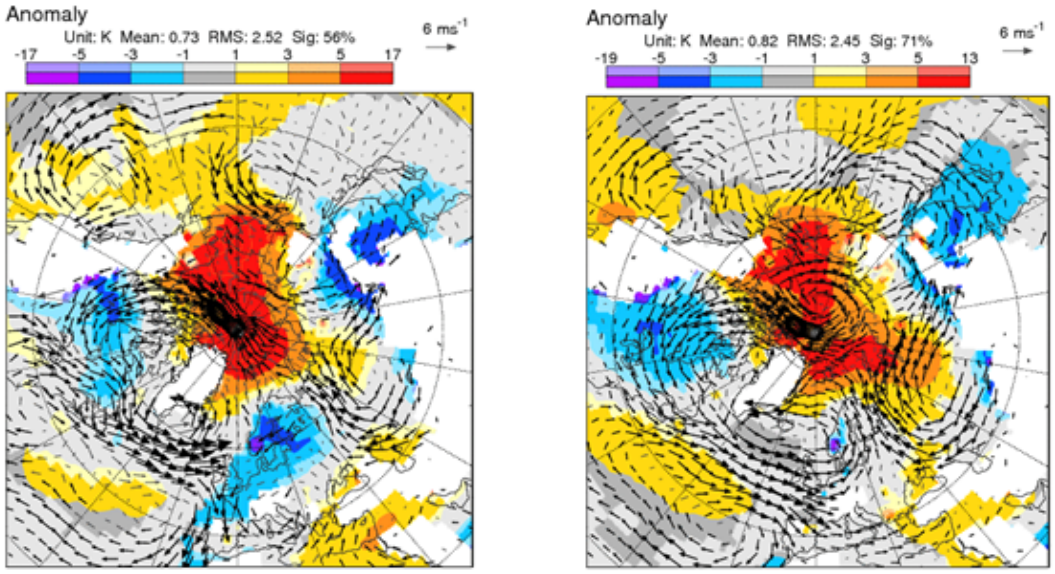


Figure 3.1.2.7. Composite of 925hPa mean wind and temperature anomalies during for periods where conditions at Sodankylä are defined as clear the clear (LWdown<210; left) or cloudy (LWdown>210; right) during DJF2017/18.

The mean spatial anomalies during these regimes put these states at Sodankyla into the context of the regional circulation (Figure 3.1.2.7). The clear state is characterised by cold anomalies across Northern Europe and a slight north-easterly flow. The cloudy state is characterised by southerly flow across Finland and Eastern Europe.

Due to the bimodal character of the winter boundary layer at Sodankylä, it seems natural to evaluate the impact of the ML snow in these states, rather than to simply evaluate the mean.

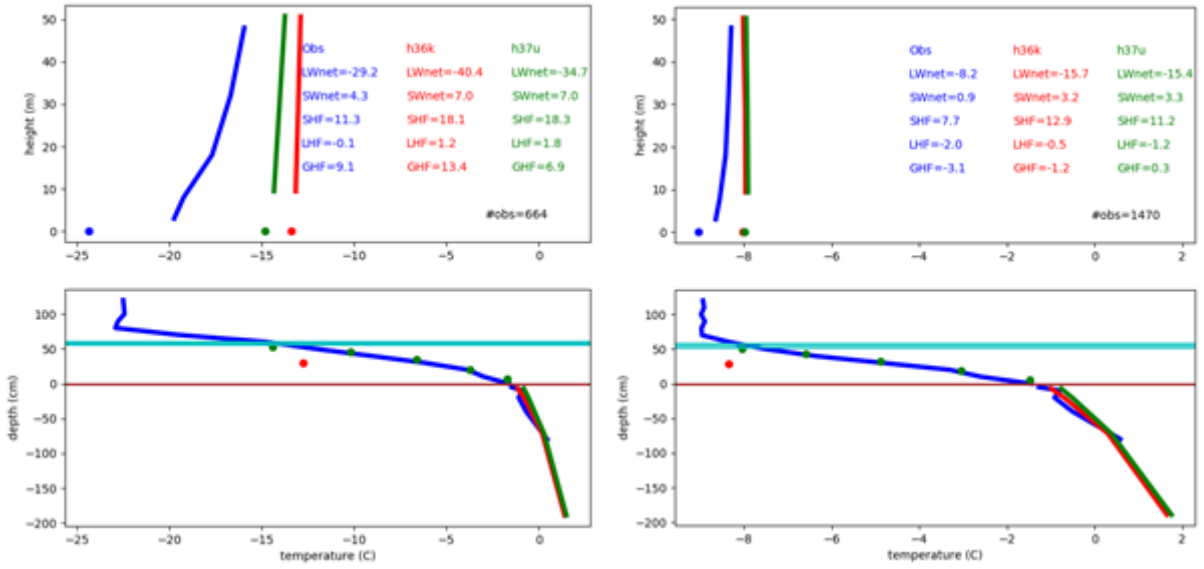


Figure 3.1.2.8. Mean vertical profiles of atmospheric temperature (top) and snow/soil temperature (bottom) for the clear (left) and cloudy (right) state at Sodankyla (DJF201718). The observed profile is shown in blue and the single-layer snow control is in red and the multi-layer experiment is in green. The mean values for the radiative and turbulent fluxes are stated in the figure. All fluxes are positive towards the surface (i.e. downwards for atmospheric terms, upwards for the GHF).

In the “clear” state atmospheric temperatures from the Sodankyla tower show that the inversion seen in the radiosondes (Figure 3.1.2.6) get “steeper” as one approaches the surface (Figure 3.1.2.8). It also goes hand in hand with a sharp gradient through the snow, with a temperature gradient of 15°C through the snowpack. In this situation longwave cooling of the surface is balanced in approximately equal measure by the sensible heat flux (SHF) from the atmosphere and the ground heat flux (GHF) between the snowpack and the atmosphere, corresponding to strong temperature gradients and in the atmosphere and the snow respectively.

The “cloudy” state is characterised by a much weaker inversion strength in the atmosphere, weaker gradient through the snow pack and less-negative net-longwave resulting in a lower SHF and GHF terms respectively.

It is clear from Figure 3.1.2.8 that in both regimes the multi-layer snow follows the observed snow-temperature profile well. The largest difference between the simulations occurs in the clear state, where the multi-layer experiment has cooler air-temperatures both at the surface and through the lowest model levels. This cooling is fairly uniform over the lower layers and does not result in a much sharper inversion, as is seen in observations. This increased cooling at the surface is due to the inclusion of a thin (5cm) top snow layer, which reduces the thermal inertia of the surface compared to the single layer where the whole depth of snow is assumed to cool uniformly. Therefore, transitions in between the cloudy and clear states in

forecasts with the new formulation would lead to a larger change in 2m temperature than with the single layer scheme. Interestingly the SHF is too large in both experiments and in both regimes, despite the difference between the lowest model level and skin temperature being too small.

A temperature sensitivity diagnostic

In order to formalise this idea of thermal-inertia of the surface it is natural to consider the surface energy balance. One can separate the energy budget into driving and response terms. In this framework the incoming radiation (LW_{down}+SW_{net}), which is affected by the synoptic situation (air mass temperature, cloud properties, etc.), can be considered a driving term for the boundary layer-land-surface system as a whole (Miller et al., 2017). Looking at the relationship (and calculating the gradient) between this forcing term and the skin temperature is one way to define the thermal inertia of the surface. Although this approach could be easily applied to other surface or near-surface parameters, like 2m temperature or the turbulent fluxes (Miller et al., 2017) to provide a perspective on the sensitivity of surface or near-surface conditions to changes in synoptic regime, which is relevant to the overall skill in forecasting these parameters.

Comparing the skin-temperature sensitivity in forecasts with observations provides a diagnostic for the whole boundary layer-land-surface system, which is independent of errors in the synoptic circulation, cloud position, or the representation of radiative properties in the atmosphere (Miller et al., 2018). It provides a framework for assessing the impact of the multi-layer snow or other physical changes to the land surface or boundary layer.

Comparing the observations to the control run at Sodankylä one can see that when cold skin temperatures are observed, which coincide with low values of LW_{down}+SW_{net}, the forecast is too warm (Fig 3.1.2.9). The gradient of a line of best fit provides a measure of the temperature sensitivity of the surface temperature to changes in radiative forcing. One can see that the sensitivity is too low (0.10°C/Wm^{-2} compared to 0.20°C/Wm^{-2}). This shows that there would be an error in the skin temperature forecast even if the radiative input to the surface was perfect. The multi-layer snow results in a reduction of the temperatures during the coldest periods, reducing the warm bias, and increasing the skin temperature sensitivity to 0.13°C/Wm^{-2}.

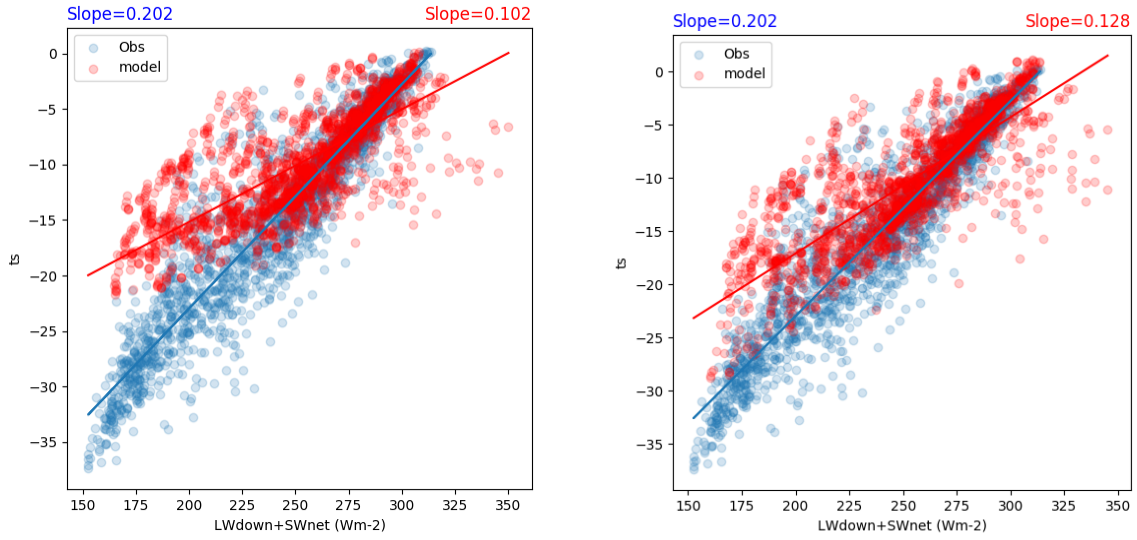


Fig 3.1.2.9. Scatter plot of skin temperature sensitivity vs LW_{down}+SW_{net} in observations and hourly forecast output from day2 from the single layer control (left) and the multi-layer experiment (right). Lines of best fit are shown for each as is the gradient of this line which gives the sensitivity (in °C/Wm^{-2})

This metric shows that overall the multi-layer formulation of the snow leads to a better representation of winter temperature in this region. However, differences in the sensitivity are affected not only by the snow thermodynamics, but also the coupling and the balance of processes within the boundary layer as a whole.

To take a more land-surface focussed approach, one can perform the same type of sensitivity analysis, but using the the sum of the net radiation at the surface and the turbulent heat fluxes (F_{net}=LW_{net}+SW_{net}+SHF+LHF≈GHF), also known as the ground heat flux (when the budget is closed, which may not be the case in observations) as the driving term. The sensitivity of the surface to GHF provides a different perspective on the thermal inertia of the surface, but one which is only dependent on land-surface processes, thereby removing boundary layer processes as a source of sensitivity.

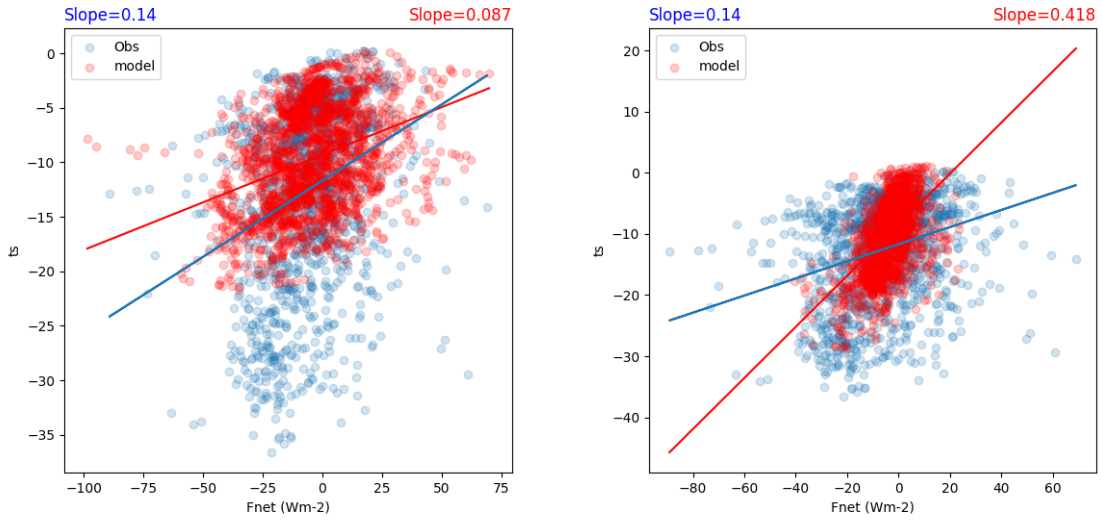


Figure 3.1.2.10. Scatter plot of skin temperature sensitivity vs F_{net} in observations and hourly forecast output from day2 from the single layer control (left) and the multi-layer experiment (right). Lines of best fit are shown for each as is the gradient of this line which gives the sensitivity (in °C/Wm^{-2})

In the control simulation the skin temperature sensitivity to F_{net} is too low (0.087°C/Wm2 compared to 0.14°C/Wm^{-2} in observations, Figure 3.1.2.10). This demonstrates that insensitivity of the land-surface to F_{net} is contributing to the overall insensitivity to changes in incident radiation seen in the previous metric.

The inclusion of the multi-layer snow increases the skin temperature sensitivity to F_{net} to 0.418°C/Wm^{-2}, which is nearly three times the observed sensitivity. Nevertheless, the skin temperature is still too insensitive to the radiative forcing, demonstrating that in the multi-layer formulation the land surface model is compensating for a lack of sensitivity coming from the boundary layer scheme.

Concluding remarks

1. Inclusion of multi-layer snow in the IFS leads to an improvement in forecasts of 2m temperature and snow depth over the Arctic by cooling the near-surface and thereby reducing the mean bias.
2. Two new diagnostics of surface sensitivity are presented which quantify the thermal inertia of the surface. The first quantifies the skin-temperature sensitivity to changes in incoming radiation, the second quantifies the skin-temperature sensitivity to changes in the ground heat flux.
3. They are used to show that the multi-layer snow scheme has increased the overall sensitivity of the surface temperature to changes in incoming radiation, but that this has been done by making the sensitivity to the

ground heat flux too high. This indicates that other processes in the boundary layer are contributing to this lack of sensitivity to radiative forcing.

The temperature sensitivity diagnostics in this study are presented without any hypothesis testing. Creating such a test for these requires some thought, since the uncertainty in the observations should be considered. The lack of closure in observations of the surface energy budget (see Foken 2008 for an overview) is an additional complication. Therefore this will be future work.

The overall improvement in forecasts of snow depth and 2m temperature motivates the inclusion of multi-layer snow in the APPLICATE stream-2 experiments.

3.1.3 On the discretization of the ice thickness distribution in the NEMO3.6-LIM3 global ocean-sea ice model (UCL)

The ice thickness distribution (ITD) is one of the core constituents of modern sea ice models, since it accounts for the unresolved spatial variability of sea ice thickness within each grid cell. While there is a general consensus on the added physical realism brought by the ITD, how to discretize it remains an open question. Within the context of **Task 5.3.1**, we have conducted **three sets of experiments (Figure 3.1.3.1) aiming at understanding the processes driving the model response to changes in the ice thickness distribution (ITD)**. Based on that, the overall objective of the experiments was to **examine the sensitivity of Arctic and Antarctic sea ice, as simulated by the global ocean–sea ice general circulation model (NEMO3.6-LIM3), to the discretization of the ITD.** Previous studies using coupled (Bitz et al., 2001; Holland et al., 2006) and uncoupled (Massonnet et al., 2011; Uotila et al., 2017; Ungermann et al., 2017; Hunke, 2014) models suggested the potential role of the ITD discretization on the simulate mean sea ice state. **With our experiments, we aim at understanding the physical processes behind the model responses.**

Our results, recently published by Massonnet et al. (2019), have shown that increasing the number of categories leads to an increase of winter sea ice volumes, which persists in summer in the Arctic (Figure 3.1.3.2). In both hemispheres, the summer extents are sensitive to the number of categories only for fewer than five categories (Figure 3.1.3.2). Higher winter ice volumes are caused by higher thicknesses due to enhanced bottom growth, which is related to the ice thickness distribution discretization via the conductive heat flux through the ice. Our results also indicate that the inclusion of a very large number of ice thickness categories does not systematically improve the realism of the simulations against available observational references and reanalyses (Figure 3.1.3.2). However, these sensitivity experiments have not been tuned (unlike the reference experiment). In addition, verification data are uncertain: for sea ice extent, variations among products can reach values as high

as 1 million km^2 (Meier and Stewart, 2018). The sea ice volume values provided for reference in our figures are even more uncertain, being estimated from reanalyses. No strict convergence of ice volumes is achieved with less than 10 categories and the following observations can be made. First, it is required to have categories with lower bounds around 4 and 2 m in the Arctic and the Antarctic, respectively. When this is not the case, the thick ice produced by ridging is blended with thinner ice, increasing its thickness, reducing the bottom growth and eventually decreasing the total ice volume. This confirms and explains the importance of thick ice categories already noted for the Northern Hemisphere by Hunke (2014). The existence of these thick categories is critical to host deformed ice and to let thin ice, which is subject to high basal growth rates in winter, occupy a sufficient fraction of the grid cells. Second, refining the ice thickness distribution discretization in the thin range (below 4 and 2 m for the Arctic and Antarctic, respectively) causes hemispheric ice volumes to keep growing, though a very large number of categories (at least 33) is necessary to detect a significant increase. We stress that, by design, our experimental protocol ignores possible feedbacks between the atmosphere and the ice–ocean system, which could enhance or dampen the responses seen in our results.

One important criterion when choosing the ice thickness distribution discretization is the associated computing cost. Compared to a reference case with one category, computing time increases by 2 %–6% when five categories are used, by 42 % when 17 categories are used and by 210 % when 50 categories are used (Figure 3.1.3.3). However, as discussed above, the gains in terms of convergence of modeled sea ice volumes are weak for such a number of categories. **Hence, using five categories, with sufficiently thick categories, appears to be an appropriate compromise for global experiments**: the ice extent converges in both hemispheres, while a reasonable level of convergence is reached for ice volume. Simulations of the Southern Ocean sea ice may require fewer categories, while applications needing a very detailed representation of the thick Arctic sea ice should use a much finer ice thickness distribution discretization. **Thus, for large-scale climate applications with NEMO3.6-LIM3, we recommend using the default ITD discretization** (experiment S1.05; Figure 3.1.3.1).

It is finally important to place the results of the sensitivity tests conducted in this study in a broader context. Specifically, one should investigate how the sea ice volume and extent responses seen in this study compare to other influences. For instance, the net increase of $\sim 3\times10^3$ km^3 in annual mean Arctic sea ice volume, seen in Fig. 2 when changing from S1.05 to S1.50, lies in the $2–5\times10^3$ km^3 range of interannual variability noted by Olonscheck and Notz (2017), who analyzed the output from climate models participating in the fifth phase Coupled Model Intercomparison Project (CMIP5). The response is in addition much smaller than the range obtained in the sensitivity tests conducted by Urrego-Blanco et al. (2016) to various parameters in the CICE model. The response is also small compared to the range of sea ice

volumes estimated by state-of-the-art sea ice reanalyses (Chevallier et al., 2016), which are supposed to be among the best constrained estimates on this quantity. In conclusion, choices related to the ITD discretization should always be put in the perspective of other competing influences, such as parameter tuning and background internal variability (Notz, 2015), the choice of atmospheric forcing (Barthélemy et al., 2017; Hunke, 2010) and the choice of observational references or reanalyses (Massonnet et al., 2018) used to evaluate the outcome of such sensitivity tests.

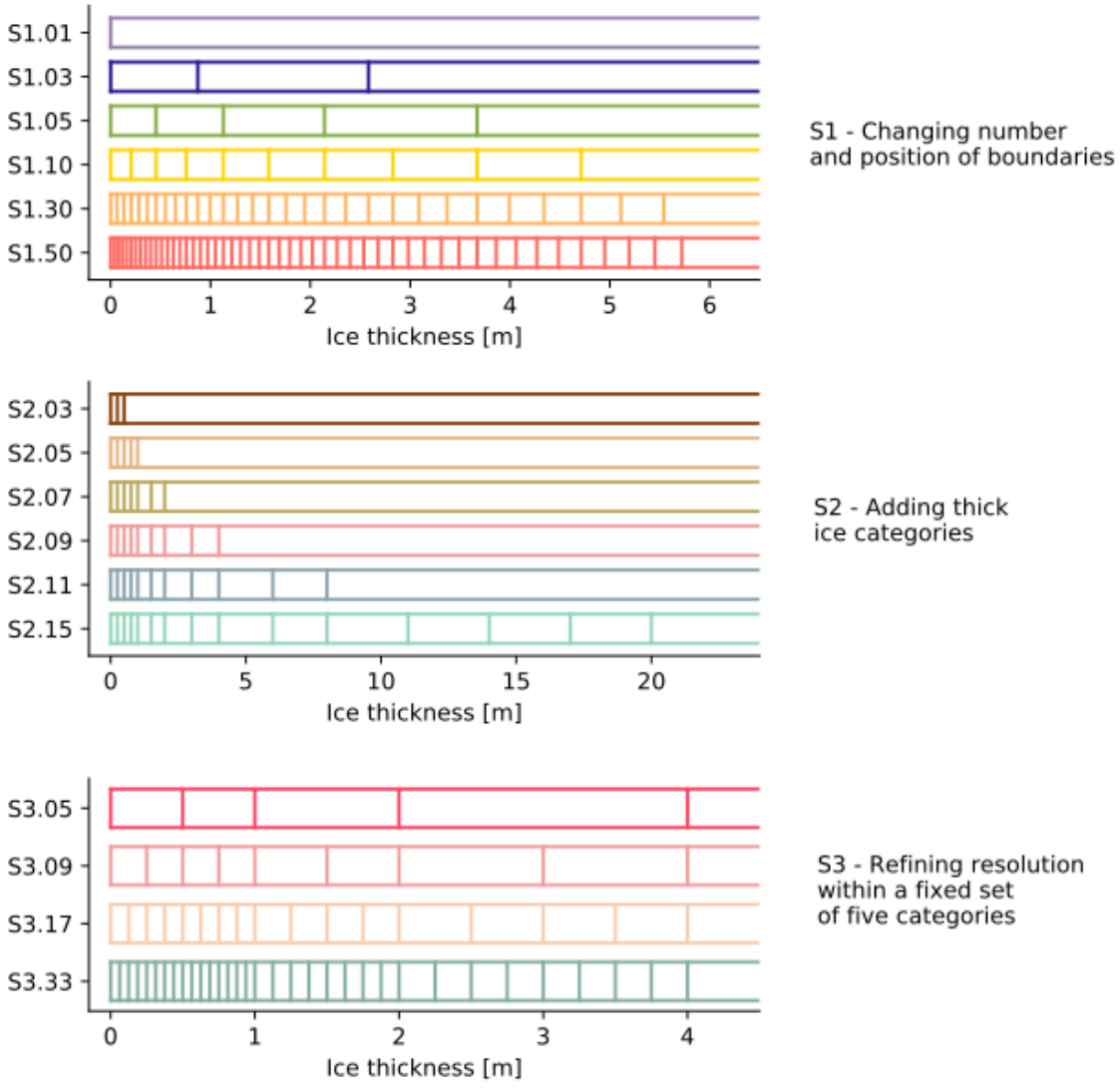


Figure 3.1.3.1 Ice thickness category boundaries in the three sets of sensitivity experiments. The upper boundary of the last category is always set to 99.0 m. Note that the ice thickness scale is different in the three panels. Because the ITD discretization in the third set of experiments (S3) branches from experiment S2.09 of the second set, that experiment is repeated in the list but labeled as S3.09.Average monthly distribution sea-ice thickness from May to October on Greenland and

Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with May start dates.

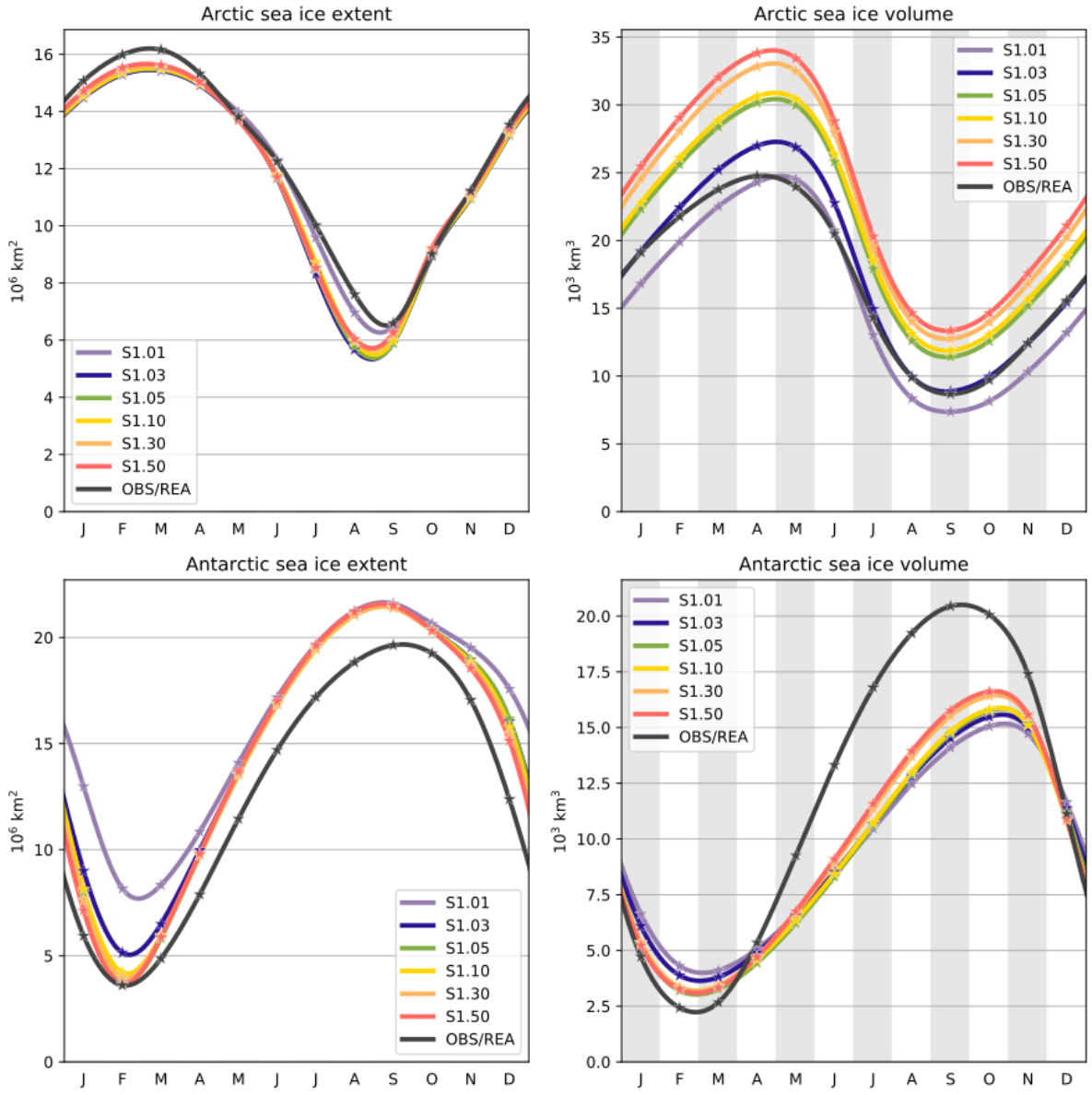


Figure 3.1.3.2 Mean seasonal cycles of Arctic (a, b) and Antarctic (c, d) sea ice extents (a, c) and volumes (b, d), over 1995–2014, in the first set of sensitivity experiments. Ice extents derived from the Ocean and Sea Ice Satellite Application Facility (OSISAF) sea ice concentration observational product (OSI-409a; EUMETSAT, 2015) are also shown, as well as Arctic and Antarctic ice volumes derived from the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) and Global Ice Ocean Modeling and Assimilation System (GIOMAS)reanalyses, respectively (Schweiger et al., 2011; Zhang and Rothrock, 2003). The stars show the monthly data and the curves are cubic interpolations between the data points.

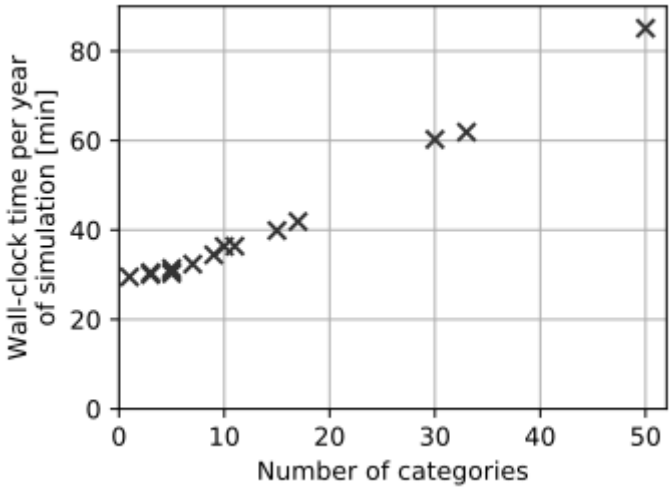


Figure 3.1.3.3 Wall-clock time required for 1 year of simulation as a function of the number of ice thickness categories. The coupled ocean–sea ice model is run on 260 cores. The computing times indicated in this figure correspond to the average over the first 5 years of each simulation.

3.1.4 Impact of the number of sea ice thickness categories on the representation of recent sea ice variability with NEMO3.6-LIM3 (BSC)

We explore the impact of different configurations of the sea ice thickness distribution (ITD) on the Arctic sea ice variability, in a complementary analysis to Massonnet et al. [2019, also included in section 3.1.3], where the mean state was studied. We use the same experiments with NEMO3.6-LIM3 detailed in the previous Section 3.1.3., run on the ORCA1 grid and forced by DFS5.2 over the period 1979–2014, in which both the number and boundaries of the ice thickness categories are changed (Figure 3.1.3.1).

The model simulations are compared with three satellite observational products of sea ice concentration (SIC): HadISSTv2.2 (Titchner and Rayner, 2014), NSIDC (Cavalieri et al., 1996), and OSISAF (EUMETSAT SAF, 2016). In both model and satellite data, we analyze interannual variability applying k-means clustering to the Arctic SIC. K-means clustering is an alternative method of dimension reduction to other, more widely used, such as principal component analysis. The analysis focuses on two seasons, January to March (JFM), and August to October (ASO), the seasons in which cluster coherence is strongest between individual months (evaluated through spatial correlations between the respective cluster centroids; not shown). Variability is characterized by three clusters extracted via the k-means approach. Three clusters is the optimal number, as derived from a suite of 10 validity indices

(Duda–Hart, Ratkowsky–Lance, Ball–Hall, SD, cubic clustering criterion, traceCovW, Rubin, Beale, Scott, and Marriot; Charrad et al. [2014]) that allow determining the most robust choice of the number of clusters.

Clusters in OSISAF are shown in Figure 3.1.4.1 in both seasons. Each cluster is characterized by a spatial pattern of SIC anomalies, a percentage of occurence over the period, and a time series of cluster occurrence, which indicates which cluster is the closest to the anomaly pattern in a year together with their Euclidean distance (defined as the root-mean-square error difference). In winter (JFM), the leading cluster resembles a quadrupole described by previous literature (e.g., Close et al., 2017), whereas the third one reflects the NAO imprint of the SIC (e.g., Bader et al., 2011). The three summer (ASO) clusters reflect a long-term trend of melting sea ice, albeit with different spatial weightings.

To evaluate the impact of the ITD configuration on the SIC, spatial correlation coefficients are computed between the observed and simulated clusters (Figure 3.1.4.2). In winter, the impact is small and most simulations capture well the observed variability; however, there is a slight decrease in the correlation coefficients in all the clusters as the number of categories increases, something observed for all the configurations. In summer, spread in model–data agreement is larger than in winter. The lowest coefficients are found for the second cluster across all the configurations. This is likely because, overall, it is a noisy pattern with predominant non-significant anomalies, which is difficult to be accurately captured by the model. By contrast, anomalies in the first and second clusters take larger values over a larger area and therefore are more easily simulated in a realistic way. Model–data correlation coefficients are little impacted by the ITD configuration for the first and third cluster but decrease with a larger number of thin ice categories for the second cluster in the S1 and S3 configurations. Although increasing the number of thick categories in the S2 configuration has no major impact on model–data correlation coefficients, the S2.07 shows a drop in correlation values in all the clusters. This suggests that variability is slightly differently distributed across the clusters in this configuration. The configuration with one single category, S1.01, shows systematically the lowest correlation coefficients.

Our results do not allow to draw any particular conclusions on which is the best configuration or number of categories beyond that one category tends to perform the worst, and that increasing the number of sea ice categories does not have a clear beneficial impact in the representation of past SIC variability.

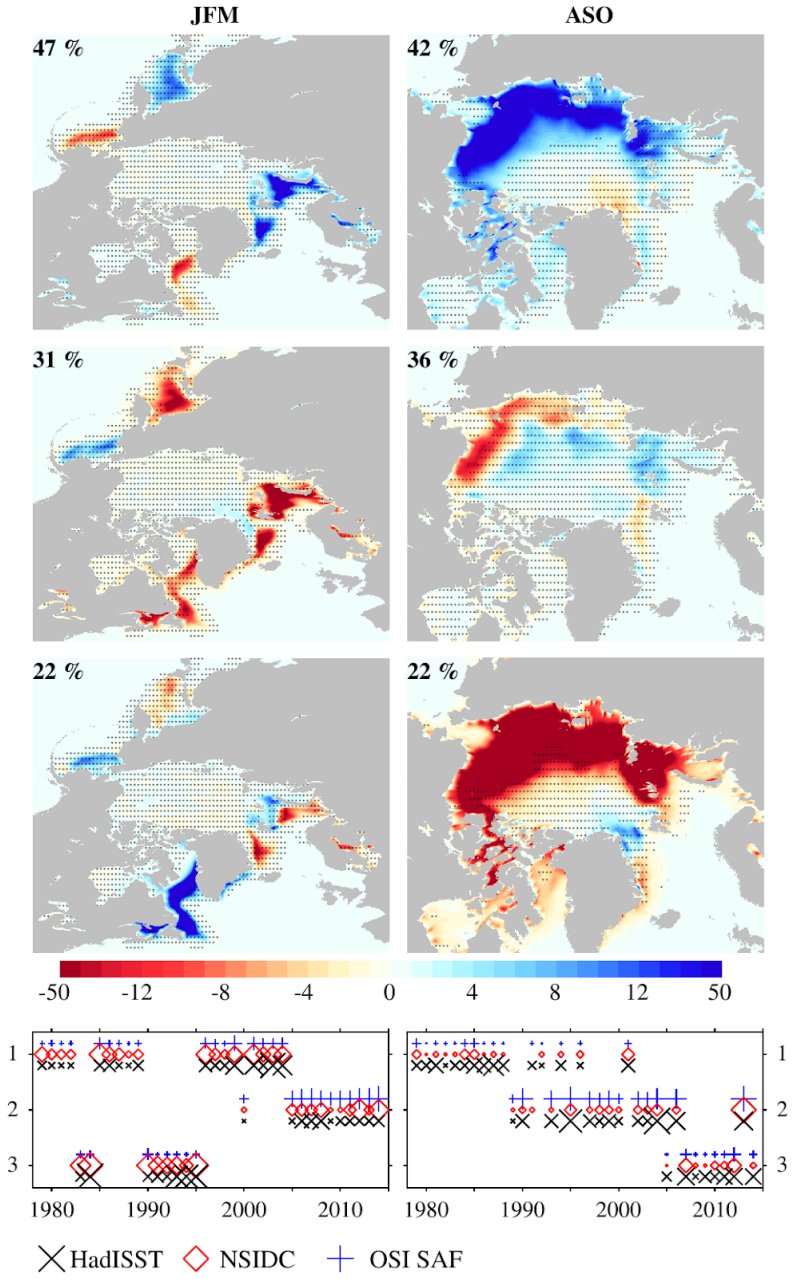


Figure 3.1.4.1: *First three rows: cluster patterns of Arctic SIC anomalies (shading; in % of area) in OSISAF in JFM (left) and ASO (right). Stippling masks statistically non-significant anomalies at the 5% level; p values at each grid point are computed through a t-test that accounts for serial autocorrelation (Manubens et al., 2018). Each cluster's percentage occurrence over the period 1979–2014 is indicated in each panel. The shading color scale is adapted for a better view of the anomalies in the range ±15%. The area is zoomed in ASO (right) for a better view of the central Arctic. Fourth row: time series of cluster occurrence in HadISST (black crosses), NSIDC (red diamonds), and OSISAF (blue pluses). The larger the symbol size, the larger the Euclidean distance (root-mean-square difference) between a pattern of anomalies and the associated cluster in a particular year (the maximum symbol size is shown in the legend).*

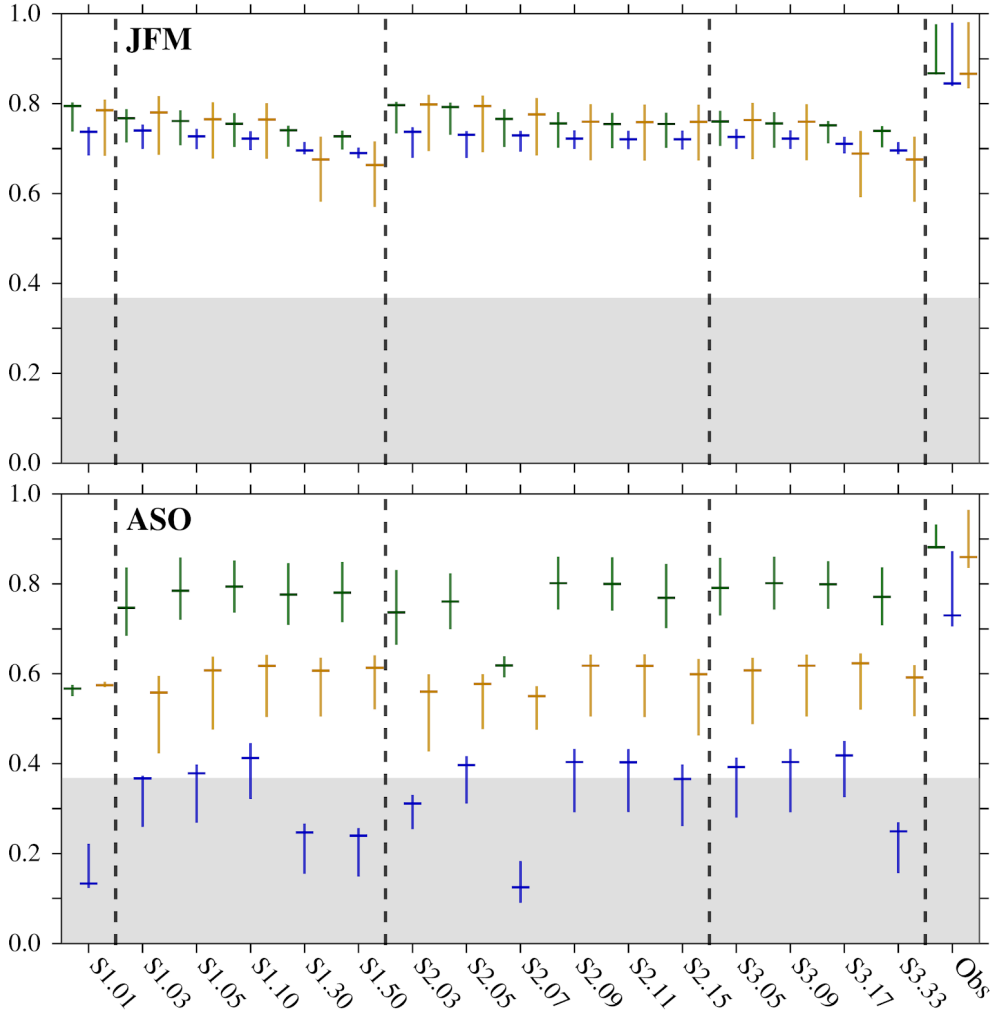


Figure 3.1.4.2: *Spatial correlation coefficients between the simulated and observed clusters and across the three satellite observational products (marked as Obs) of Arctic SIC in JFM (top) and ASO (bottom). For each case, the vertical line spans the maximum and minimum correlation coefficients, and the horizontal line marks the middle one; green, blue, and orange indicate that the first, second, and third clusters are correlated. Gray shading masks statistically non-significant coefficients below 1/e value. Dashed vertical lines separate results for the simulation with one single category (S1.01), the different ITD configurations (S1, S2, and S3), and the observations. Note that the configurations S2.09 and S3.09 are the same.*

3.1.5 Impact of activation of melt ponds on the representation of recent sea ice variability with NEMO3.6-LIM3 (BSC)

Three different melt pond parametrization are tested in LIM3 (implemented by Olivier Lecomte at UCLouvain for EC-Earth), and their results are compared with

observations of sea ice concentration (SIC) and reanalysis data of sea ice volume. The melt pond configurations are the following (David Docquier's contribution):

- MP1: a prescribed parameterization, in which two constant parameters (melt-pond fraction and depth) are prescribed;
- MP2: an empirical parameterization, in which a fraction of the melt water accumulates in pond reservoirs; the volume in the ponds is cleared when the ice thickness is below 0.1m, or released exponentially when the surface freezes; this parameterization is based on Holland et al. (2012);
- MP3: a topographic parameterization, in which melt-pond evolution is computed from the ice topography as inferred from the ice-thickness distribution, based on Flocco and Feltham (2007) and Flocco et al. (2010).

For each configuration, a NEMO-standalone simulation is performed, run on the global ORCA1 grid, which features a nominal horizontal resolution of 1deg and 75 vertical levels. NEMO3.6-LIM3 is forced by the DRAKKAR atmospheric Forcing Set version 5.2 (DFS5.2; Dussin et al., 2016) over the period 1958–2014. For comparison, a simulation with no melt pond parametrization is implemented, and one with the MP3 configuration and 15 ice thickness categories.

Results of the simulations (Figure 3.1.5.1) suggest that all the melt pond configurations have a small impact on the Arctic and Antarctic sea ice extent and Antarctic ice volume in both March and September (months of the maximum and minimum climatological values; Fig. BSC-1). The largest impact is on the Antarctic ice volume for the MP2 and MP3 configurations, for which both the March and September values are reduced by about 2 to 5 10^3 km^3. Although in March this reduction brings the model closer to reanalysis estimates in PIOMAS and GIOMAS, in September it moves simulated values quite far from reanalysis ones, reaching nearly zero ice volume by the early 21st century.

These results suggest that including melt pond parametrizations in NEMO3.6-LIM3 decreases model realism. We therefore suggest not to include them directly without further tuning of the sea ice model.

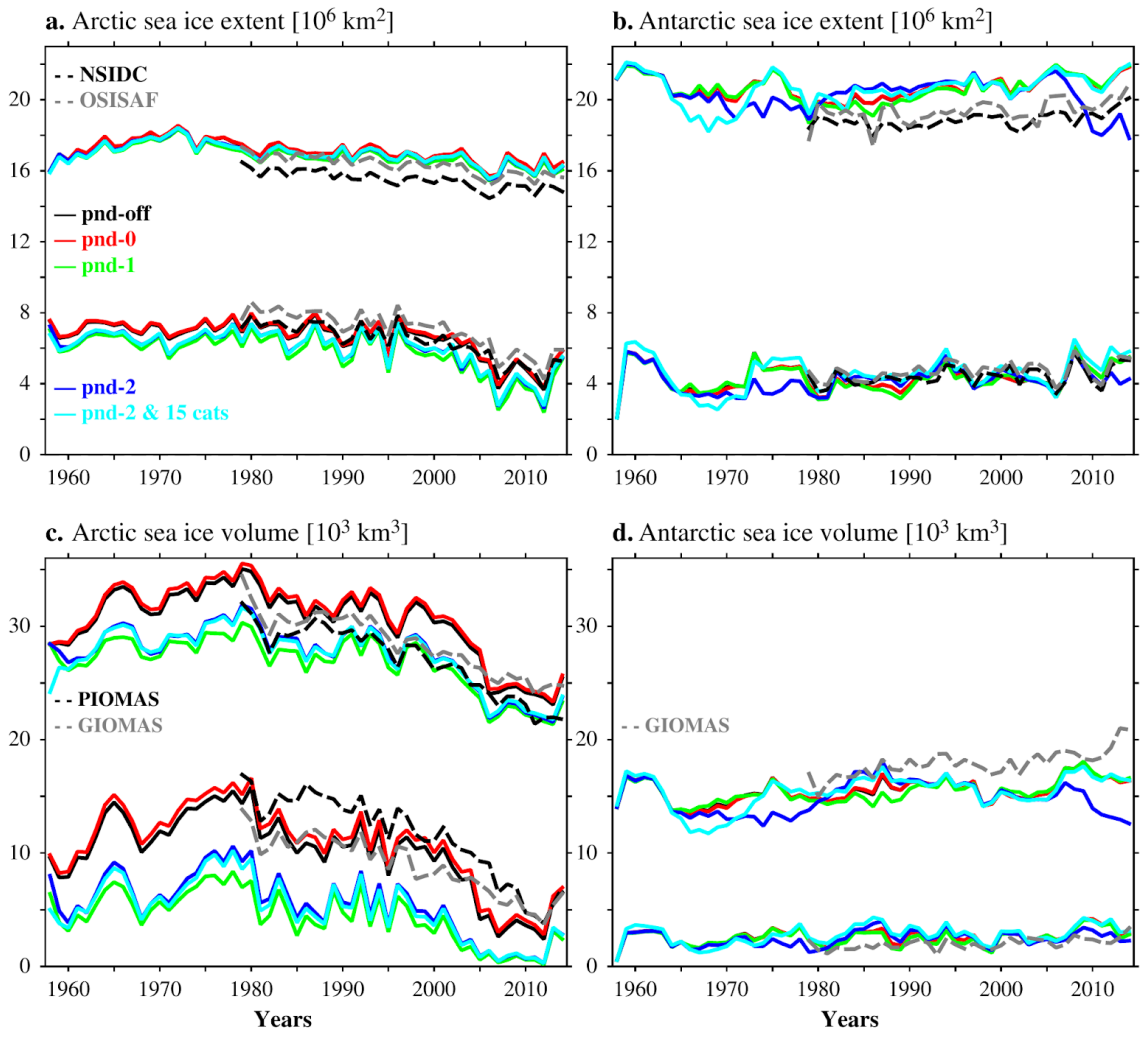


Figure 3.1.5.1: *Arctic (a,c) and Antarctic (b,d) sea ice extent (a,b; in 10^6 km²) and volume (in 10^3 km³) in the NEMO3.6-LIM3-standalone simulations without melt ponds (pnd-off; solid black line) and with melt pond parametrizations MP1, MP2, and MP3 (pnd-0, pnd-1, pnd-2, solid red, green, and dark blue lines respectively). A simulation using MP3 and a larger number of sea ice thickness categories (i.e., 15) is also shown (solid, light blue line). Observed sea ice extent from NSIDC and OSISAF is shown in a,b (dashed black and gray lines respectively). Ice volume in PIOMAS and GIOMAS is shown in c,d (dashed black and gray lines respectively).*

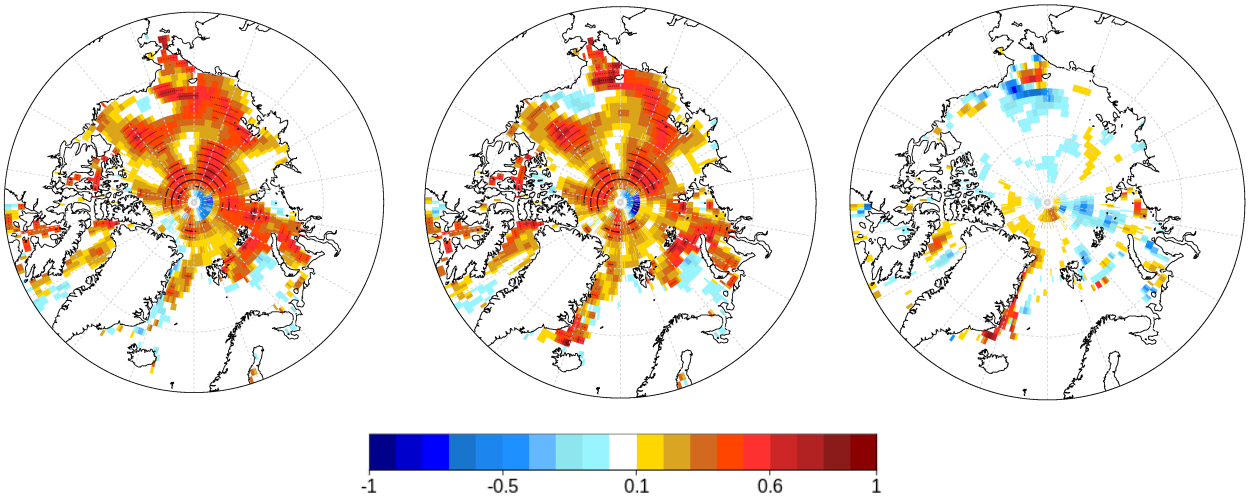
3.1.6 Impact of activation of melt ponds on climate prediction skill with CNRM-CM6 (CNRM)

The sea ice model Gelato v6 of the CNRM-CM6 contains an optional melt pond parametrization. This empirical parametrization relates the evolution of melt pond area and depth to the surface melt-water flux, using a similar approach as Holland et al. (2012). In this approach a certain fraction of surface melt water and rainfall is collected to a virtual reservoir: the melt pond volume. The melt pond depth and pond

fraction are solved using an empirical function. The melt ponds are allowed to change the surface albedo in the bare ice portion, but are not allowed to show through the snow-covered portion of the ice category. Once the sea ice becomes too thin (0.10m), the melt ponds are emptied to mimic the draining (Holland et al. 2012).

The melt pond parameterization was not used in the CNRM-CM6-1 experiments. To test the effect of the melt pond parameterization another seasonal re-forecast experiment (CNRM-CM6 MP) was carried out. CNRM-CM6 MP is almost identical to the CNRM-CM6-1, but with the melt pond scheme activated. We run this test for May start dates (1993–2014) with 10 ensemble members. Activating the melt pond parametrization affects the albedo of the sea-ice. When the melt pond scheme is activated, the albedo of bare melting ice is set to 0.65. When melt ponds form, they reduce the total albedo by replacing a fraction of the bare ice with melt ponds that have a low albedo. When the melt pond scheme is not active, the albedo of bare melting ice is set 0.56 to mimic the effect of melt ponds without parametrizing them.

In terms of mean sea ice concentration biases over the Arctic region, little to no differences were found over the 1993-2014 period, thereby confirming that the role of the melt ponds scheme on the ensemble mean monthly sea ice concentration is limited at a pan-Arctic scale.



(a) CNRM-CM6-1 Sep. SIC *(b) CNRM-CM6 MP Sep. SIC* *(c) Difference*

Figure 3.1.6.1 September sea ice concentration grid point correlation with NSIDC data for re-forecasts initialized in May with (a) CNRM-CM6-1 and (b) CNRM-CM6-1 with melt ponds scheme. (c) Difference in correlation with and without the melt ponds scheme.

Figure 3.1.6.1 shows the grid point correlation of September mean sea ice concentration in the CNRM-CM6-1 re-forecasts initialized in May with and without

activation of the melt ponds scheme, and the difference between both. Only small differences are found in most areas, although correlation is improved over parts of the Greenland Sea. Over the Beaufort and Chukchi seas, an area with no skill (or even slightly negative skill) gets larger with the melt ponds activated. However, results should be interpreted with caution in the marginal seas, since some grid points are ice free over most of the re-forecast period, and a limited ensemble size and re-forecast period are used.

The regional analyses of Greenland Sea and Chukchi Sea are in line with the Pan-Arctic results. The average annual cycle of sea-ice concentration is very similar in both experiments (Figure 3.1.6.2), the main difference being the slightly larger spread in the melt pond –experiment. Both experiments underestimate the sea ice concentrations in these regions, which was also evident on the bias maps (not shown). The annual cycles of sea ice thickness and albedo (Figure 3.1.6.3 and 3.1.6.4) indicate only minor differences between the experiments: for both variables, the melt pond experiment produces a larger inter annual variability. This is especially true for the late summer albedo.

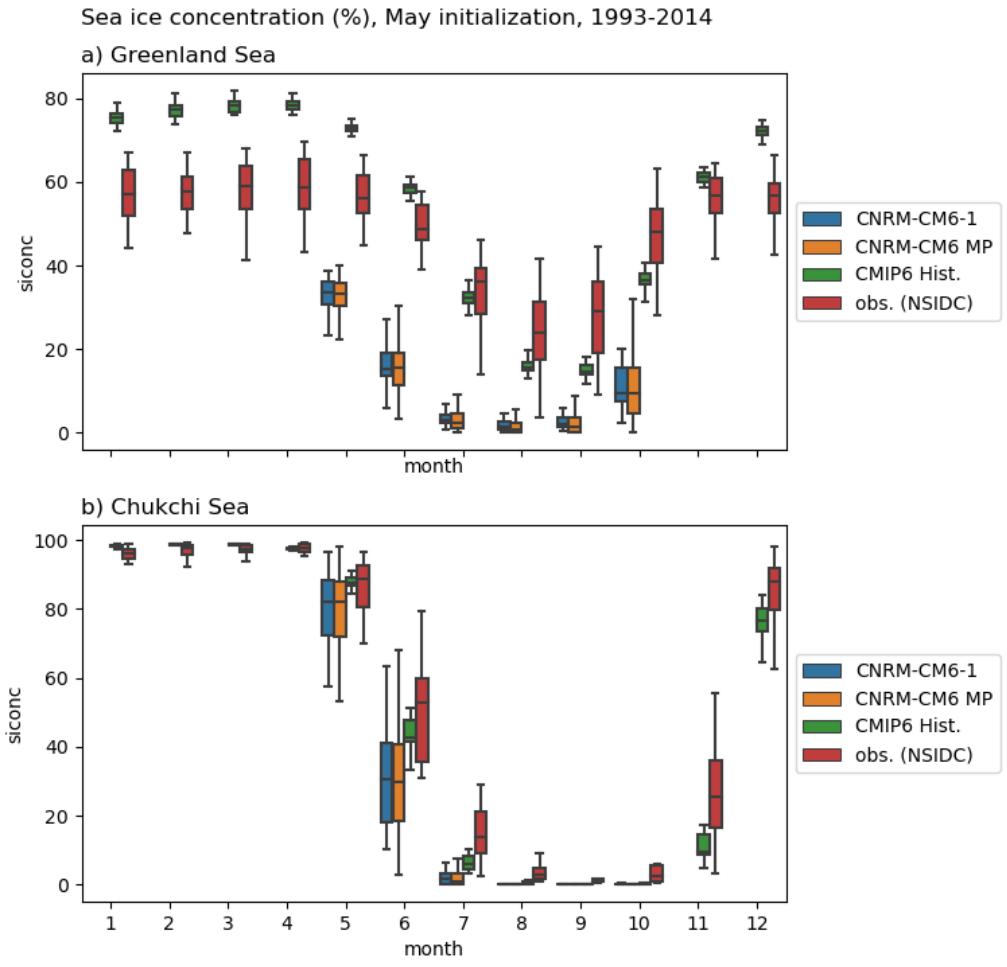


Figure 3.1.6.2: Average annual cycle of sea ice concentration (%) in (a) Greenland Sea and (b) Chukchi Sea. "CNRM-CM6-1" and "CNRM-CM6 MP" are the seasonal re-forecast with the melt pond scheme not active and active, respectively. "CMIP6 Hist." represent the fully forced historical climate simulation with the CNRM-CM6 model. "obs. (NSIDC)" is the observed sea ice concentration.

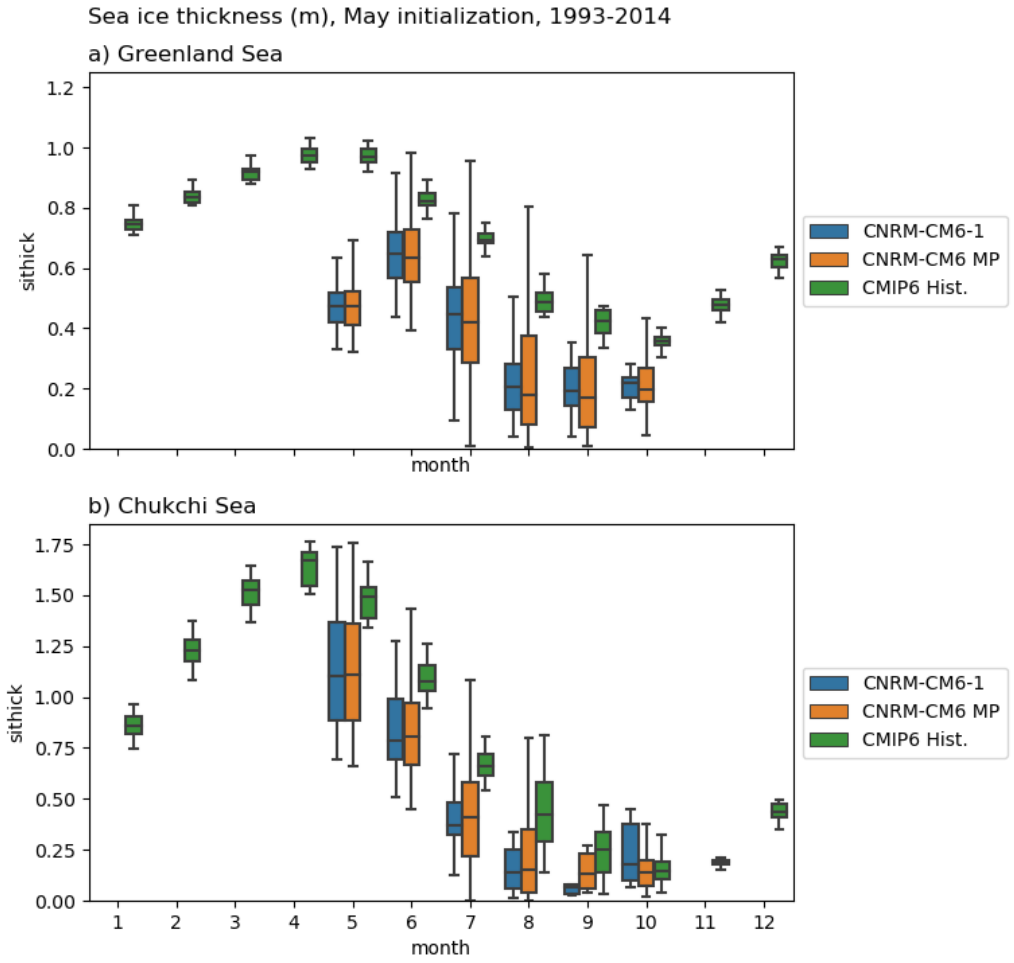


Figure 3.1.6.3 Average annual cycle of sea ice thickness (m) in (a) Greenland Sea and (b) Chukchi Sea. "CNRM-CM6-1" and "CNRM-CM6 MP" are the seasonal re-forecast with the melt pond scheme not active and active, respectively. "CMIP6 Hist." represent the fully forced historical climate simulation with the CNRM-CM6 model.

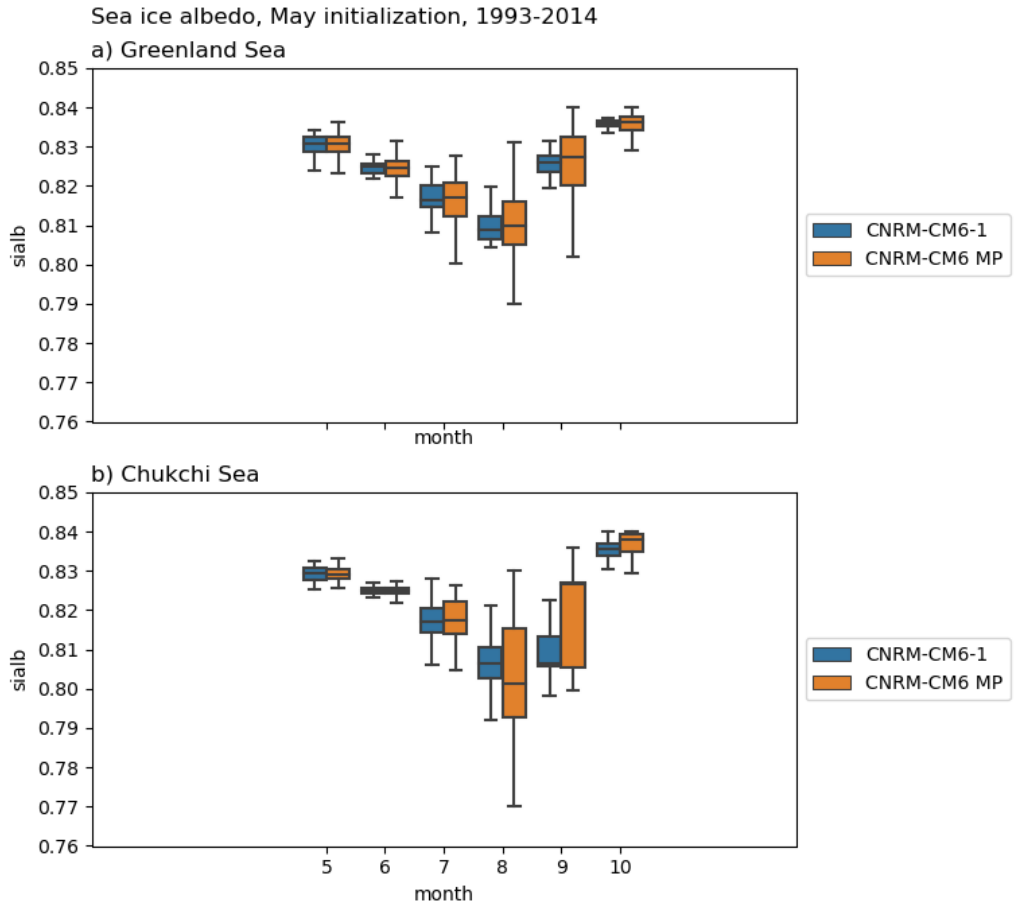


Figure 3.1.6.4 Average annual cycle of sea ice albedo in (a) Greenland Sea and (b) Chukchi Sea. "CNRM-CM6-1" and "CNRM-CM6 MP" are the seasonal re-forecast with the melt pond scheme not active and active, respectively.

The average thickness distribution inside Greenland or Chukchi Sea is not much affected but the melt pond scheme during May, June and July (Figure 3.1.6.5). In August, September and October, we see that the sea ice concentrations reach higher values in when the melt pond scheme is turned on. Similarly, the thickness distribution in the late summer is smoother, because higher thicknesses are present, when the melt pond scheme is active (Figure 3.1.6.6).

The small differences between CNRM-CM6-1 and CNRM-CM6 MP experiments are potentially related to the tuning of the albedo in Gelato v6: when the melt pond scheme is not active, the albedo of bare melting ice is reduced to account for the missing melt ponds.

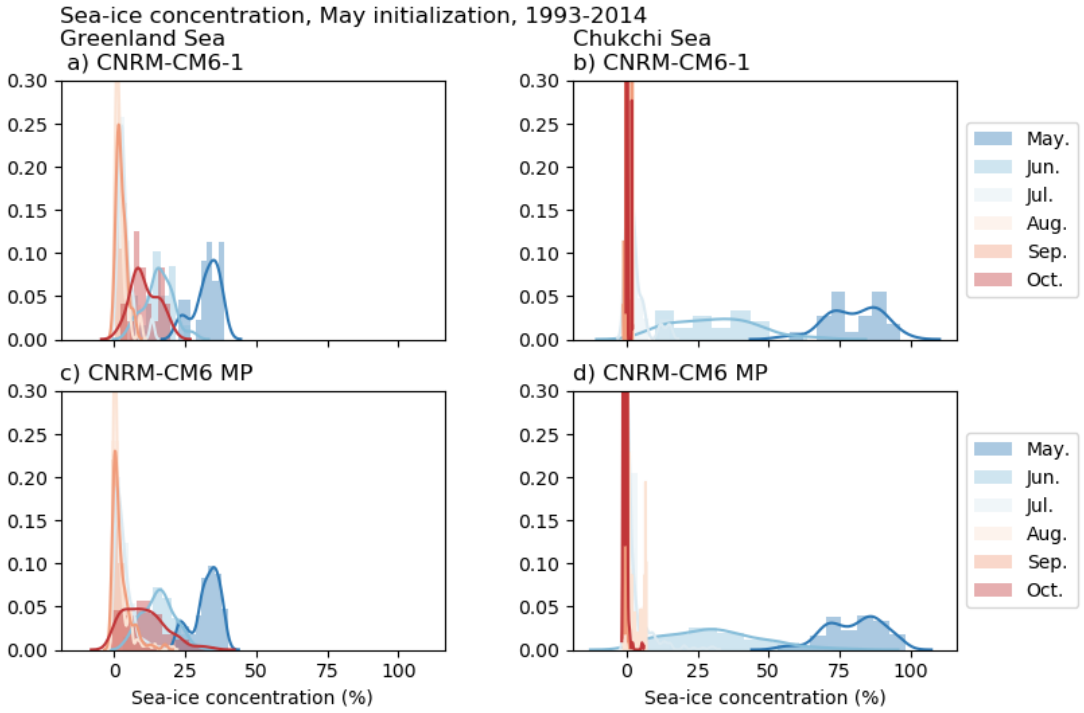


Figure 3.1.6.5 Average monthly distribution sea-ice concentration from May to October on Greenland and Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with May start dates.

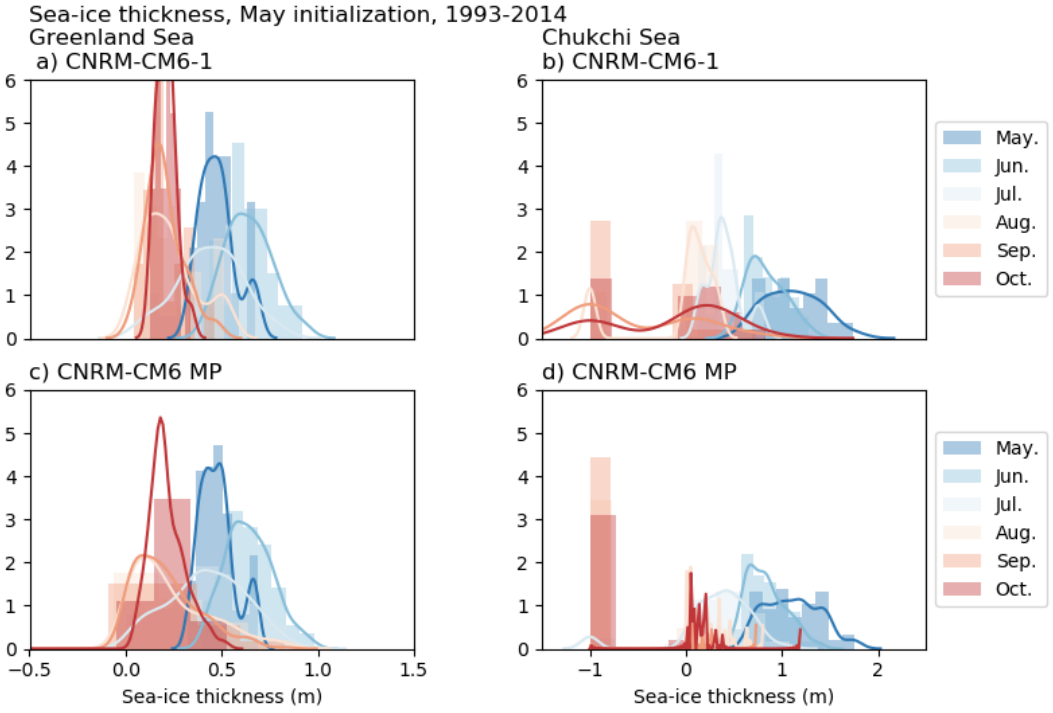


Figure 3.1.6.6 Average monthly distribution sea-ice thickness from May to October on Greenland and Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with May start dates.

3.2 Increased atmospheric resolution in weather prediction (Task 5.3.2) (MET-Norway, CNRS and Meteo France)

Introduction

The use of regional models can, compared with global models, add value by the use of optimized physics for the targeted area and finer resolution (Jung et al., 2016). Such added value (AV) have been demonstrated in many studies and also for Arctic short-range weather predictions (e.g. Muller et al., 2018, Yang et al., 2018, Køltzow et al., 2019, Batte et al., 2018). In general, the increase in model resolution has played an important part in the improved forecast accuracy seen over the last decades (Bauer et al., 2015). However, the added value of resolution varies, e.g. by the jump in resolution, the presence of (surface) small scale forcing, the weather type, lead time and user needs to mention some. For regional models the AV may further be modified through the choices of regional domain size and location (e.g. Kristiansen et al., 2010, Wang et al., 2016) and the treatment and quality of the Lateral Boundary Conditions (LBCs) may be an issue (Warner et al. 1997, Davies, 2014, Chikhar and Gauthier, 2017). In addition, by the advent of operational regional Ensemble Prediction Systems (EPSs) with convection permitting models (e.g. Frogner et al., 2019a, Hagelin et al., 2017) an important question is whether to use operational computer power on resolution, domain size, lead time or ensemble members.

In the Regional Climate Modelling community there is a high awareness on issues related to the configuration of regional systems and their AV (e.g. Laprise et al., 2008, Feser et al., 2011, Rummukainen, 2016). The operational Numerical Weather Prediction (NWP) community are also aware of these issues, but less systematic investigations have been performed. Historically, weather services has chosen a pragmatic approach; a deterministic run for the area and lead time covering forecast obligations. Then the highest possible resolution affordable can be found within the limitations of computer power and operational deadlines.

The aim of the presented work is to discuss the impact of configuration choices (e.g. initialisation methods, resolution, domain size and location, lead time, EPS or deterministic) in *operational regional Arctic weather prediction*. This work is mainly done in Task 5.3.2, but the EPS experiments are performed in Task 5.3.4 and discussed in more detail in section 3.4.

Experiments

A set of experiments for the last part of Year of the Polar Prediction Special Observing Period Northern Hemisphere (YOPP-SOP-NH1), 8. March 2018 - 31.

March 2018, based on the configuration of the Applications of Research to Operations at Mesoscale (AROME-Arctic) forecast system in operational use at MET-Norway (Müller et al. 2017) have been performed. All experiments use the same model description (dynamics and physics) as AROME-Arctic, but have differences in their configurations for initialisation, resolution, domain location and size, quality of the lateral boundary conditions and ensemble members (summarized in Table 3.2.1). The regional integration domains are shown in Figure 3.2.1. The observations are taken from the quality controlled observation data base at MET Norway for Mean Sea Level Pressure (MSLP), 2m air temperature (T2m), 10m wind speed (S10m), 2m Relative Humidity (RH2m), one hour accumulated precipitation (precip1) and Total Cloud Cover (TCC).

For horizontal and vertical resolution an additional set of experiments are performed with the Meteo-France AROME model (Seity et al. 2011) for the entire YOPP-SOP-NH1. The change in resolution for this experiment is similar to the HIGHRES experiment in Table 3.2.1 with horizontal grid spacing increasing from 2,5 to 1,25km and an increase from 60 to 90 vertical levels. The runs are dynamical adaptations of the global ARPEGE model and employ the large domain in Figure 3.2.1.

In addition, analysis of already existing simulations with 1) the global high-resolution version of the ECMWF Integrated Forecasting System (IFS-HRES), 2) the AROME-Arctic operational at MET Norway, 3) the AROME version of Meteo-France (MF-AROME) employed on same integration domain as AROME-Arctic and 4) the Canadian Arctic Prediction System (CAPS) are performed. The AV of the three latter models over the coarser resolution IFS-HRES are already described in Køltzow et al. (2019) and APPLICATE Deliverable 5.2 (Batte et al., 2018) and will only briefly be referred to in the following.

Comparison methods

To ease the comparison between all experiments we present a general overview of the outcome using Mean Absolute Error Skill Score (MAESS). Mean Absolute Error (MAE) is calculated for a set of parameters for each experiment and then compared with a reference experiment (usually CNTRL in Table 3.2.1 if nothing else is noted) by (1- MAE_{exp}/MAE_{ref}). Positive (negative) values shows that the experiment performs better (worse) than the reference. MAESS is not necessarily the best metric for all parameters, but allow a useful inter-comparison between different configurations and parameters and reflect results obtained by other metrics (not shown). However, for the added value of EPS also probabilistic scores are calculated. The MAESS and the probabilistic scores are a part of the observational AV, i.e.AV by predictions being more similar to observations . Other types of AV following Di Luca et al. (2015) are also discussed, but more in brief.

Assimilation

The AV of surface assimilation compared to dynamical downscaling (CNTRL vs DD, Figure 3.2.2) are substantial and show clear improvements in MSLP, T2m (largest) and RH2m and S10m. With the exception of S10m, the advantages are most pronounced for the shorter lead times. However, we notice that surface assimilation only have negligible impact on precip1 and TCC. The impact of surface assimilation is sensitive to differences in the surface schemes, i.e. the dynamical downscaling is much more challenging when different surface schemes, as in our case, are used in the global and regional model and spin-up is necessary for the regional model.

Upper-air assimilation (UAASS vs CNTRL, Figure 3.2.3) has a positive impact on short lead times for MSLP (< ~ 6 h), TCC (< ~ 6 h) and T2m (< ~12 h). A degradation of RH2m and precip1 in the first few hours are also noticed indicating problematic issues in the assimilation of moisture which should be addressed. Only a small or no significant impact of upper-air assimilation is found for S10m and RH2m. It should be noted that even if the CNTRL don't do a upper-air assimilation it is initialized from a 6h old IFS-HRES run where an upper-air assimilation has been done and thereby inherit assimilated information.

The importance of initialisation has also been tested in MF AROME runs by taking surface and upper-air analysis from the operational AROME-Arctic and from the global ARPEGE model. The results presented in Køltzow et al. (2019) show; (1) high resolution surface analysis improve the quality of near surface variables (T2m errors reduced with 15-20%) and (2) the forecast of a meso-scale polar disturbance is improved significantly with high-resolution analysis.

Resolution

The effect of resolution is investigated by comparing HIGHRES and CNTRL (Figure 3.2.4). A clear positive impact is seen for T2m, S10m and RH2m and precip1h. The added value for T2m and RH2m show a diurnal cycle with a maximum AV during day time, while AV related to precip1 increase with lead time. The only parameter that are not sensitive to the resolution jump is TCC (MSLP is not available for inter-comparison from HIGHRES due to technical problems).

In addition, the effect of the increased resolution has also been tested with the Meteo-France AROME model (Seity et al. 2011) coupled with ARPEGE for the lateral boundary conditions and initial conditions. The domain of this high resolution AROME (MF-1.25km) is the same as the large domain in Figure 3.2.1 but with a horizontal resolution set to 1.25km and 90 vertical levels with the height of the first level at 5m. A reference run MF-2.5km with 2,5km and 60 levels has also been performed. The comparison have been done for the entire SOP1 period (01/02/2018-31/03/2018).

The added value of the MF-1.25 configuration is mainly on the 2m temperature thanks to a more accurate orography and an improvement of vertical resolution near the surface (Figure 3.2.5). The bias reduction is significant for several regions: fjords, mountains and Svalbard. The horizontal structure of the precipitation is in better agreement with the observations (Figure 3.2.5 right), however for some categorical scores the improvement is not clear (not shown).

The AV of increased resolution in short-range weather forecasting is also discussed in APPLICATE Deliverable 5.2 (Batte et al., 2018) and in Køltzow et al. (2019). Køltzow et al. (2019) compare three high-resolution models (AROME-Arctic, MF-AROME and CAPS) with the coarser resolution IFS HRES. The AV varies with parameters (less for MSLP and TCC, larger for T2m and S10m), regions (larger in the presence of complex topography/coast lines), seasons (larger when the errors are largest, i.e. during winter) and lead time (larger for short lead times).

Domain size and location

Two experiments with small domains (Little Brothers), one covering Svalbard (LB-Svalbard) and one covering Northern Norway (LB-NN) are compared with the CNTRL (Figure 3.2.6 and 3.2.7, respectively). Only observation sites well inside the small domains (> 100 km from the lateral boundaries) are used in the calculation of MAESS. The impact is relatively small, but for some parameters and lead times significant. Precip1 and TCC show little sensitivity to domain size, while T2m and RH2m show a small, but not always significant degradation with the smaller domains. Interestingly, for the Svalbard domain MSLP and S10m are partly worse, while the opposite is true for the North Norway domain.

Lateral boundary condition quality

The impact of LBC quality are assessed by three sets of experiments, i.e. we discuss how the impact of LBC quality may vary with domain size and location. For the large domain (PLBC vs CNTRL, Figure 3.2.8), the MSLP forecast is improved together with the S10m forecasts for lead times larger than ~ 12 h. However, T2m, RH2m, precip1 and TCC show only minor sensitivity to LBC quality. More sensitivity is seen for the smaller domains (PLBC-Svalbard vs LB-Svalbard, Figure 3.2.9 and PLBC-NN vs LB-NN, Figure 3.2.10). A large positive impact for MSLP, but also T2m, RH2m and S10m show a pronounced positive impact for some lead times. However, again there are no significant impact on precip1 and TCC. There are also some differences on the impact between the two smaller domains, e.g. improved LBCs are more beneficial at Svalbard than in Northern Norway for these domains and period.

Ensemble Prediction Systems

The AV of high-resolution EPS is investigated following two pathways; 1) by construction of a deterministic forecast by the ensemble mean and calculate MAESS against the EPS control run. This makes it possible to compare the added value with the previous results. In addition, 2) by a probabilistic verification of high-impact weather including calculating probabilistic forecasts from deterministic runs.

MAESS comparing ensemble mean and ensemble control is shown in Figure 3.2.11. AV is seen for T2m (lead times > 18 h), S10m, RH2m and precip1. For MSLP and TCC a neutral impact is seen with the exception of a deterioration during the first 12 h for TCC. The AV are in the same order as seen for HIGHRES, with the exception of precipitation where HIGHRES provide larger AV.

With EPS a huge amount of output are available and a common way to use this data is to make probabilistic forecasts. We therefore look at AV by probabilistic forecasting of high-impact / relatively rare events. We follow the approach of Frogner et al. (2019b) and probabilities for an event are calculated by the number of members with exceedance of a given threshold in a grid box (i.e. probabilities from the deterministic runs are binary, either 0 or 100%). Then the Brier Skill Score (BSS) is calculated for the EPS experiment with the control run as a reference (Table 3.2.2). In a similar way, BSS for the HIGHRES (CNTRL as reference) is calculated for comparison. The AV from the EPS in forecasting high and low temperatures, high S10m and precipitation/ no-precipitation and a relatively high hourly precipitation rate are large and significant in terms of a positive BSS. In comparison, The AV by EPS is also considerable higher than what is provided by HIGHRES. The inter-comparison period is relatively short and the chosen thresholds (based on observation percentiles for the period) are not very extreme. However, the results clearly indicate that there are substantial AV provided by EPS.

Lead time

High-resolution models add predictability on small scale features. However, the predictability of such small scale features are also rapidly lost. In Køltzow et al. (2019), it is shown that Arctic model errors in high-resolution models grow faster than in the coarser resolution IFS-HRES model. The added value is therefore lead time dependent. This is also noticed in the verification of high-resolution EPS compared to the coarser resolution IFS-Ensemble in section 3.4.1.

Additional Added value

The results so far have been based on observational AV (better correspondence with observations), but following Di Luca et al. (2015) other types of AV can be identified; conjectural AV (supported by theoretical considerations or by relating to other studies

in similar circumstances supported by observations), potential AV (differences are needed to add value and all differences include a potential added value), and user specific AV (different use of predictions imply differences in AV). These types of AV should also be considered in the process of configuration of weather prediction systems, and particular in data-sparse regions as in the Arctic where observational AV come with uncertainties (Køltzow et al., 2019). However, in this report we limit this part to just a few examples.

From the discussion above, TCC is the least sensitive parameter to the different configuration choices, e.g. the effect of increased horizontal and vertical resolution is small in the verification statistics. However, TCC is observed at few stations and less frequently than the other parameters. A possibility is therefore to use the potential AV concept for further analysis. In Figure 3.2.12, the mean TCC for the inter-comparison period for CNTRL, HIGHRES and their differences are shown. Although, not large there is a systematic decrease in TCC in HIGHRES. This will further increase differences between AROME-Arctic TCC over the ocean and sea ice and TCC in MF-AROME and IFS-HRES, which on average have more clouds as reported in Køltzow et al. (2019). However, the potential AV reveal differences that should be studied further, e.g. compared with satellite based cloud products.

The fjords inside Sørøya, situated in the northern part of Norway (~70.5N, 22.4E, marked C in Figure 3.2.13), are known for situations with strong wind channeling in the fjords. In Figure 3.2.13, S10m and direction from CNTRL and HIGHRES in an offshore wind situation are shown. A more pronounced channeling and higher wind speeds are seen in HIGHRES at Sørøysundet (marked A) and in Stjernsundet towards Lopphavet (marked B). The correctness of this can not easily be confirmed with observations, but wind channeling is a well known phenomena and the presence of such features in the region are confirmed by ferries and fisher boats. These differences are therefore an illustration of conjectural AV. Another example of conjectural AV is the additional information about uncertainty provided by EPS. When only employing deterministic forecasts it is only possible to guess on the uncertainty in a particular situation.

A weather forecast acquire value first when it is used and early delivery time is an example of user specific AV. At MET-Norway, the operational regional AROME-Arctic is available 3-4 h before the same forecast cycle of IFS-HRES (longer observational cut-off, assimilation process, forecast production and distribution). Some users are also not able or willing to use probabilistic approaches. AV in form of (better) EPS systems are therefore not necessarily straightforward transferred to the end user. Some configuration choices therefore also require more investment in the communication and distribution part of the weather forecasting chain (e.g. Fundel et al., 2019).

Summary

In the experiments, the prediction skill is sensitive to configuration choices (initialisation, resolution, domain size and location, lead time, LBC quality and EPS) in regional weather prediction systems. Qualitatively the results are as expected, upper-air and surface assimilation is important, increased resolution increase in general the quality, small integration domains may deteriorate the quality (but also improve), and the importance of LBC quality increase with smaller integration domains. It is also shown that a substantial AV can be found by employing EPS systems. In addition, the AV differs between parameters. However, the analysis also add quantitative information to be used as guidance for future operational configurations. For example, HIGHRES and EPS require approximately the same amount of increase in computer power and both add substantial AV. However, comparing them, and in particular for high-impact forecasting it would be beneficial to choose EPS in an operational setting. The results found here are based mainly on experiments with one model system. Most likely many of the findings are also valid to other regional weather forecasting systems. However, this should be tested as results may be sensitive to model system.

Experiment	UPPER-AIR initialisation	SURFACE Assimilation	Resolution	Domain	Lateral Boundary conditions	EPS members	Cost
CNTRL	blending	yes	2,5 km, 65L	large	Oper IFS	1	1.0
UA-ASS	**3D-var**	yes	2,5 km, 65L	large	Oper IFS	1	~ 1.1
DD	**no**	**no**	2,5 km, 65L	large	Oper IFS	1	~ 0.925
HIGHRES	blending	yes	**1.25km, 90L**	large	Oper IFS	1	~ 10
LB-Svalbard	blending	yes	2,5 km, 65L	**Svalbard**	Oper IFS	1	~ 0.1
LB-NN	blending	yes	2,5 km, 65L	**N. Norway**	Oper IFS	1	~ 0.15
PLBC	blending	yes	2,5 km, 65L	large	**analysis / fc**	1	1.0
PLBC-Svalbard	blending	yes	2,5 km, 65L	**Svalbard**	**analysis / fc**	1	~ 0.1
PLBC-NN	blending	yes	2,5 km, 65L	**N. Norway**	**analysis / fc**	1	~ 0.15

EPS	**3D-Var**	yes	2,5 km, 65L	large	Oper IFS	**1+10**	~ 12

*Table 3.2.1. Summary of experiments for the period 8. March - 31. March 2018; CNTRL, UA-ASS (upper-air assimilation), DD (Dynamical Downscaling), HIGHRES (higher horizontal and vertical resolution), LB-Svalbard ("Little Brother" on a small domain around svalbard), LB-NN ("Little Brother" on a small domain around Northern Norway), PLBC (Perfect Lateral Boundary Conditions), PLBC-Svalbard (Perfect LBCs on the small Svalbard domain), PLBC-NN (Perfect LBCs on the small North Norway domain) and EPS (1+10 member Ensemble Prediction System, see details under Task 5.3.4). Notice that the UA-ASS experiment is the stream 1 runs from the operational AROME-Arctic which is not a total clean experiment when compared to the others since differences in the initial spin-up may be present. Upper-air initialisation varies by **no** assimilation (use a short IFS-HRES forecast), **blending** (hydrostatic parameters from IFS-HRES and non-hydrostatic parameters and hydrometeors from previous model cycle) and **3D-var** (full assimilation of conventional and remote sensing data). Surface assimilation is based on optimal interpolation (**yes**) and adaptations of surface fields from IFS-HRES (**no**). The latter being a non-trivial task due to differences in surface models, parameters and resolution. All experiments with assimilation are running a 3 hourly cycling, i.e. assimilation is done every third hour, but only 00 UTC runs are run for +48 h and compared. Notice that the upper-air assimilation is omitted in most experiments as it require generation of new structure functions for the different grids/resolution. More details about the assimilation procedures can be found in Muller et al. (2017). The regional **domains** can be seen in Figure 3.2.1, the "large" domain is the operational AROME-Arctic domain. The lateral boundary conditions are taken from the previous operational IFS-HRES run (**Oper IFS**) in a similar way as in operational use and by only using analysis or short forecasts from IFS HRES to have LBCs every hour (**analysis / fc**). The **cost** is estimated based on operational experience, and consideration about grid points, levels and time step.*

High-impact cases	Observation Quantile	threshold value	Brier Skill Score EPS vs cntrl	Brier Skill Score Highres vs cntrl
2m air temperature	0.05	<= -18.6C	0.24 (0.23-0.25)	0.02 (0.01-0.04)
	0.95	>= +0.8C	0.22 (0.21-0.23)	0.07 (0.06-0.09)
10m wind speed	0.95	>= 12.0 m/s	0.28 (0.27-0.29)	0.10 (0.09-0.12)

	0.99	>= 16.2 m/s	0.23(0.21-0.25)	0.13 (0.10-0.16)
hourly precipitation	-	>= 0.05 mm/h	0.36(0.35-0.36)	0.04(0.02-0.05)
	0.95	>= 0.5 mm/h	0.31(0.30-0.32)	0.03(0.01-0.05)

Table 3.2.2. Brier Skill Score (including 95th percentile confidence interval by bootstrapping) for exceedance of thresholds defined by observation percentiles and precipitation /no-precipitation. For EPS experiment the control run is used as a reference while the HIGHRES experiment use the CNTRL run as reference.

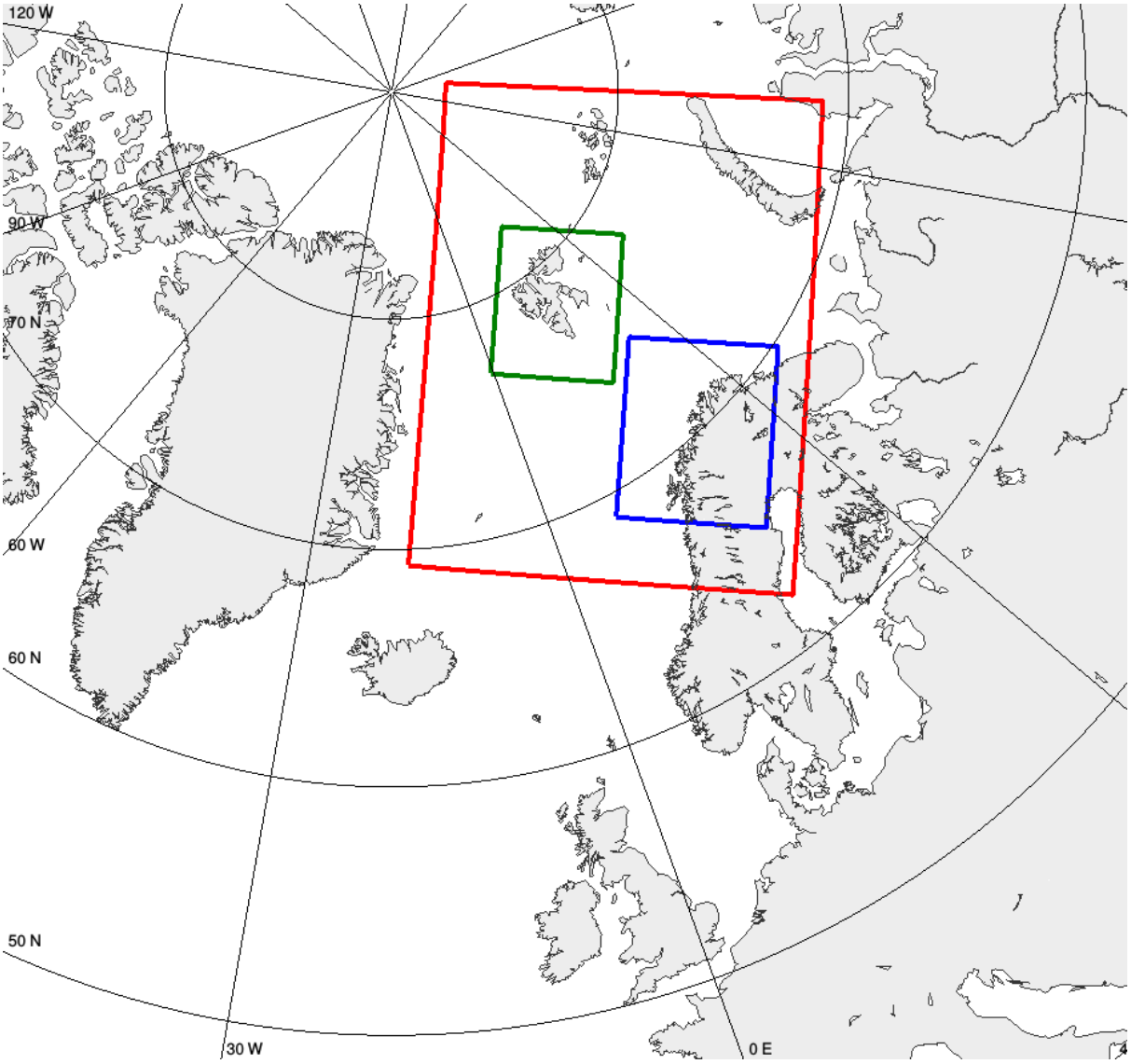


Figure 3.2.1. Regional integration domains used in experiments. AROME-Arctic / CNTRL in the large red domain, Svalbard domain (LB-Svalbard) in green and North Norway domain (LB-NN) in blue.

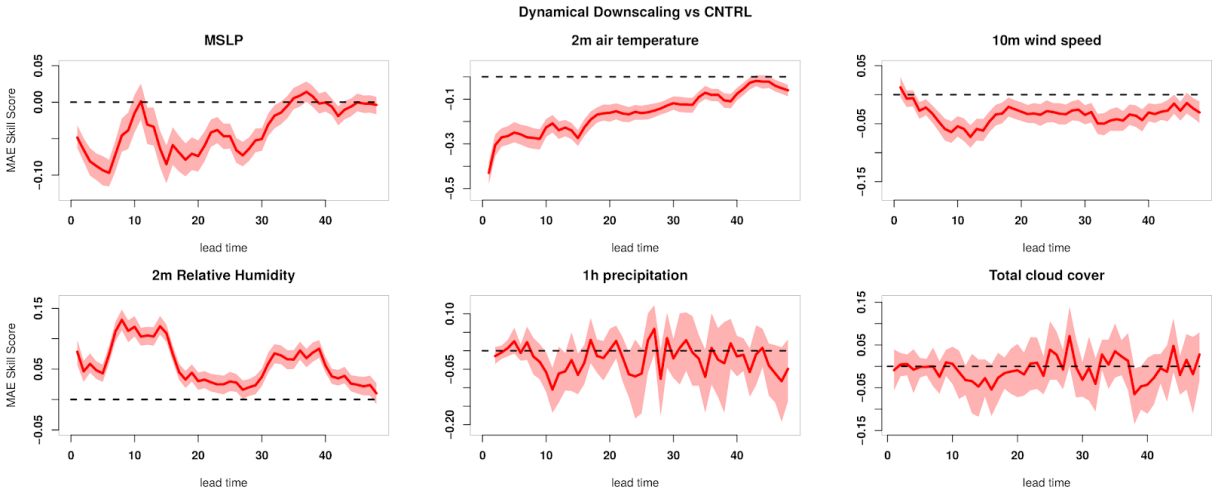


Figure 3.2.2. Impact of surface assimilation. MAESS comparing dynamical downscaling (DD) and surface assimilation (CNTRL). Positive values indicate added value of DD compared to CNTRL, while negative values show deterioration by DD compared to CNTRL. The shaded area indicate 95% confidence interval calculated by bootstrapping.

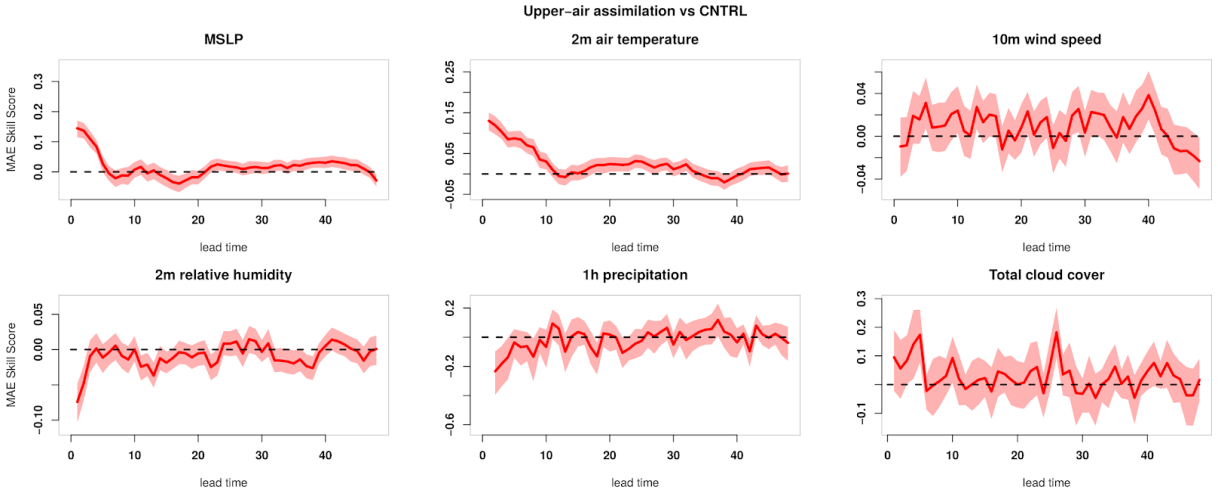


Figure 3.2.3. Impact of upper-air assimilation in regional modelling. Same as Figure 3.2.2, but comparing UAASS and CNTRL.

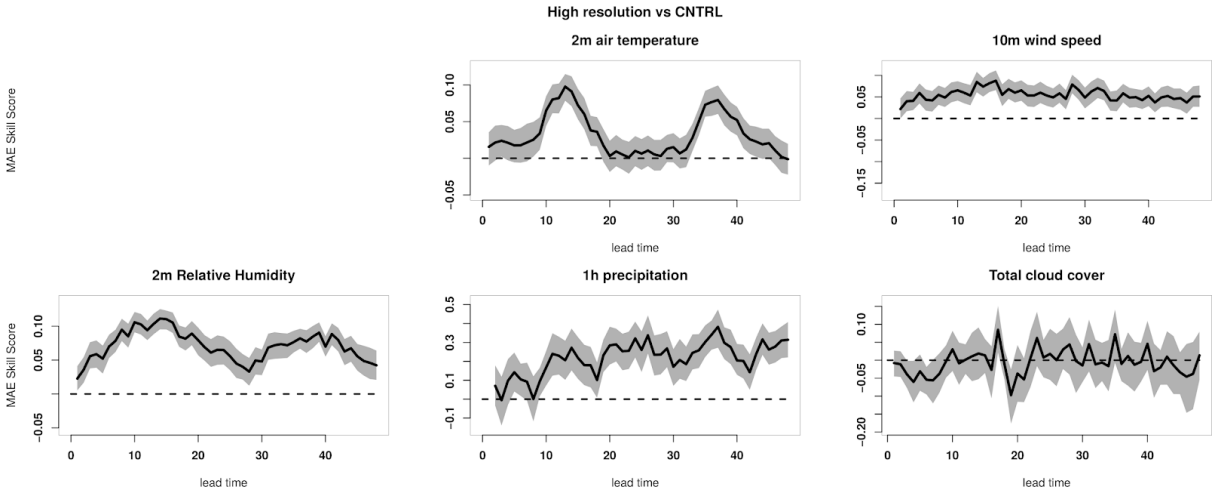


Figure 3.2.4. Impact of increased resolution. Same as Figure 3.2.2, but comparing HIGHRES and CNTRL.

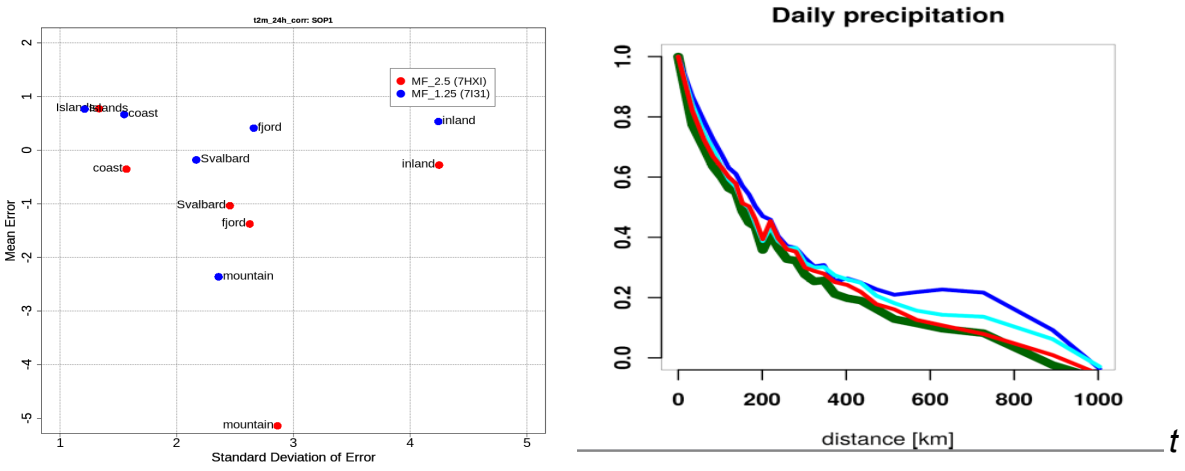


Figure 3.2.5 : Left: 2m temperature errors for several areas (defined in Koltzow et al (2019). Right: precipitation variogram for day 2 (Green: observation, cyan: MF-2.5km, red=MF-1.25km, blue=UAASS (AROME-Arctic)).

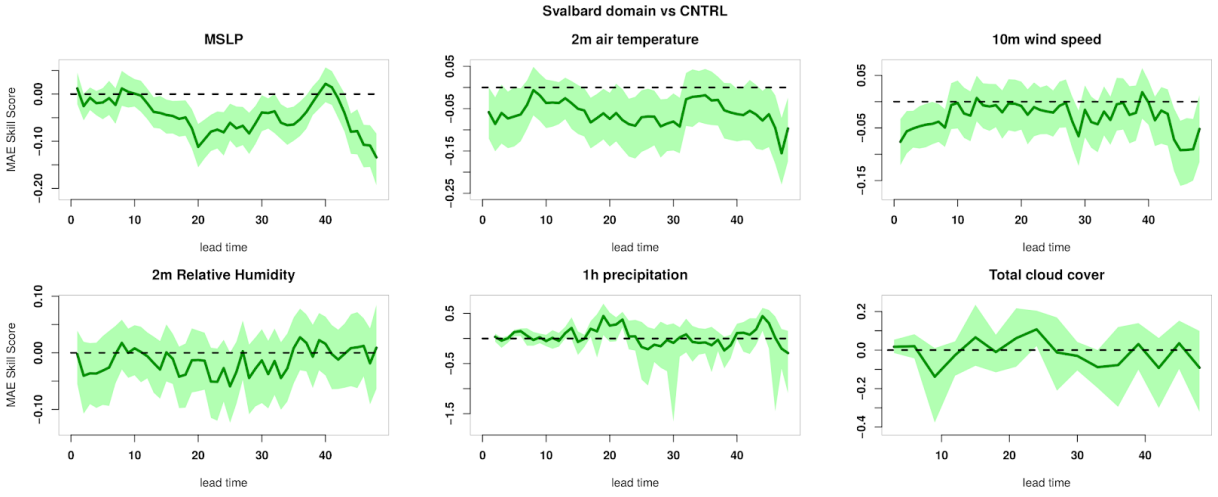


Figure 3.2.6. Impact of domain size and location. Same as Figure 3.2.2, but comparing LB-Svalbard and CNTRL.

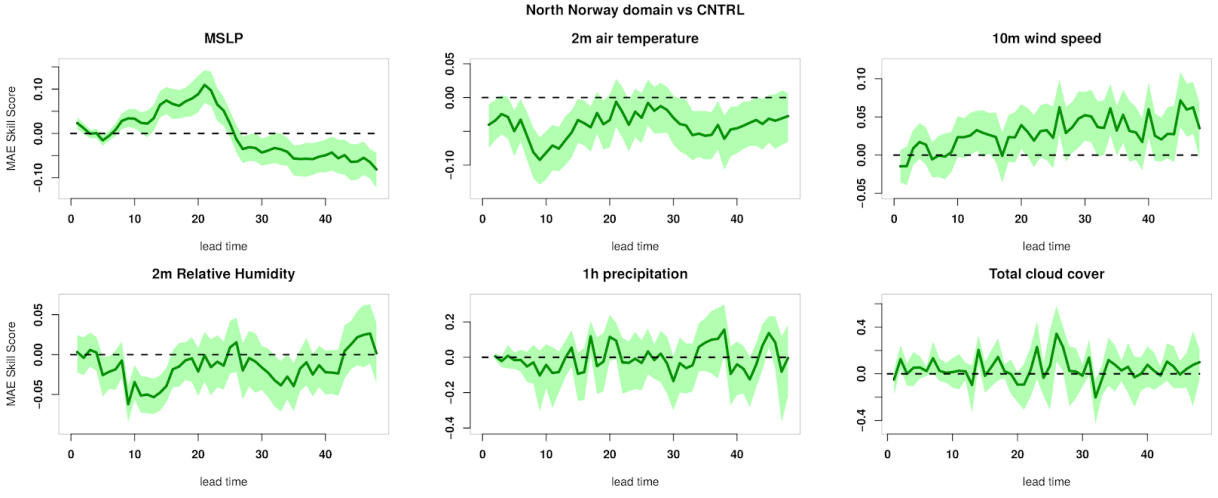


Figure 3.2.7. Impact of domain size and location. Same as Figure 3.2.2, but comparing LB-NN and CNTRL.

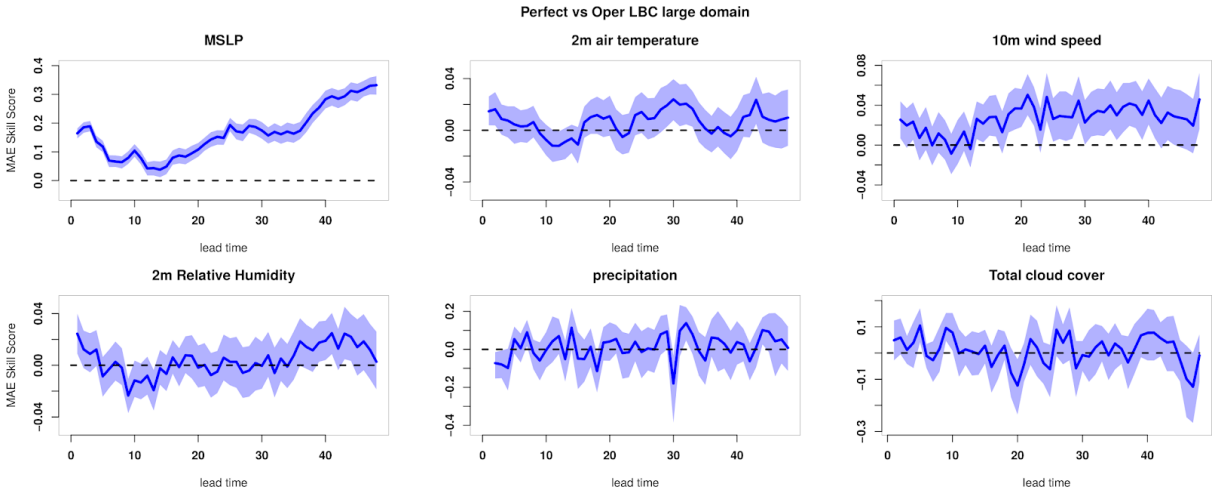


Figure 3.2.8. Impact of LBC quality. Same as Figure 3.2.2, but comparing PLBC and CNTRL.

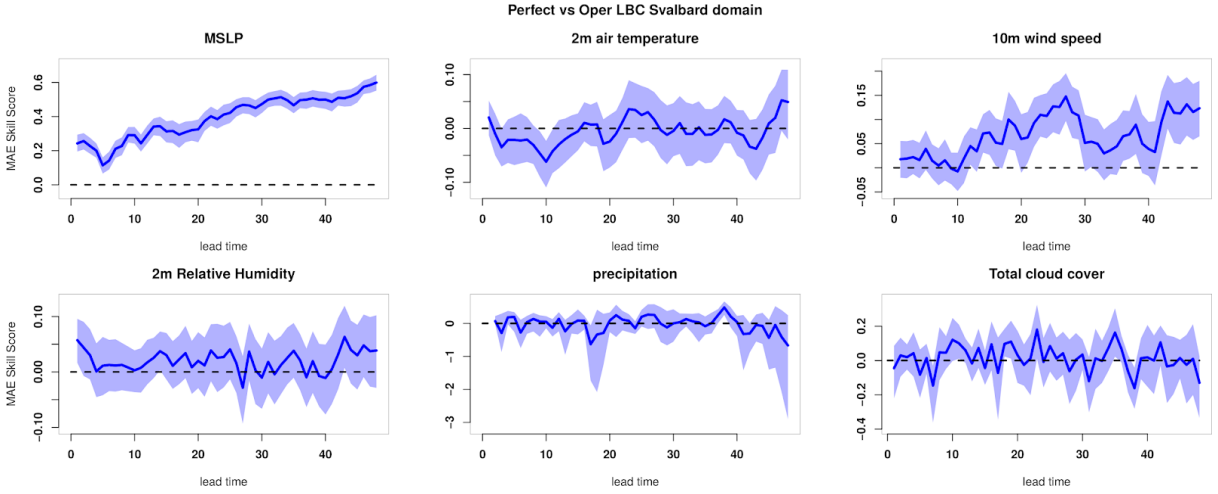


Figure 3.2.9. Impact of LBC quality. Same as Figure 3.2.2, but comparing PLBC-Svalbard and LB_Svalbard.

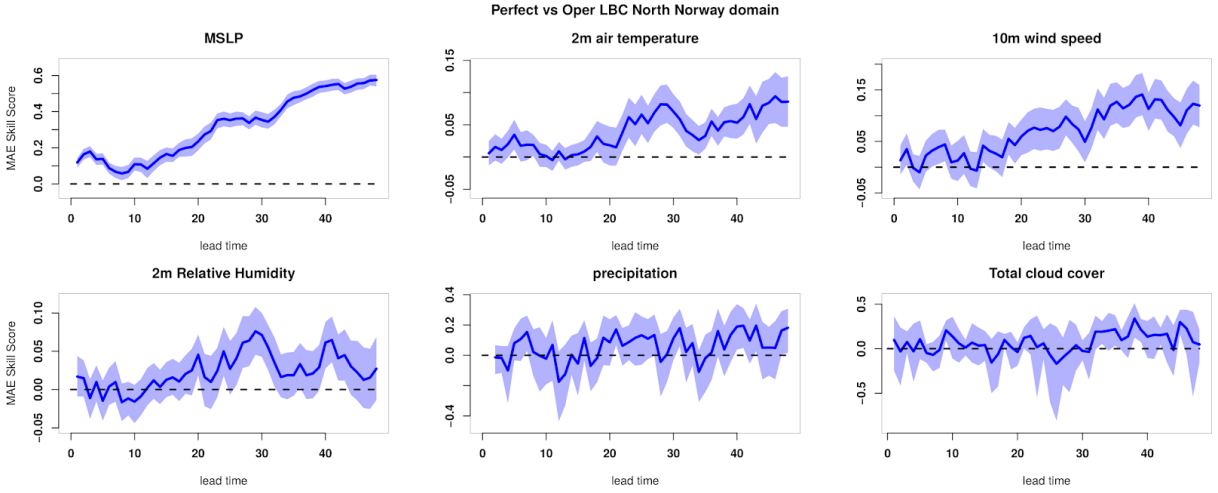


Figure 3.2.10. Impact of LBC quality. Same as Figure 3.2.2, but comparing PLBC-NN and LB-NN.

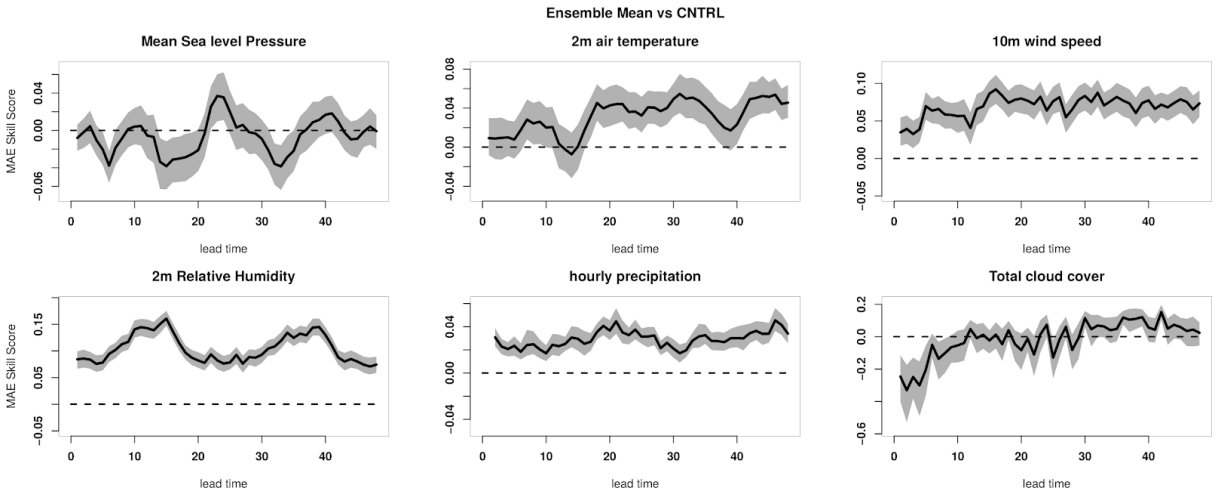


Figure 3.2.11. Impact of Ensemble Prediction Systems. Same as Figure 3.2.2, but comparing EPS and CNTRL.

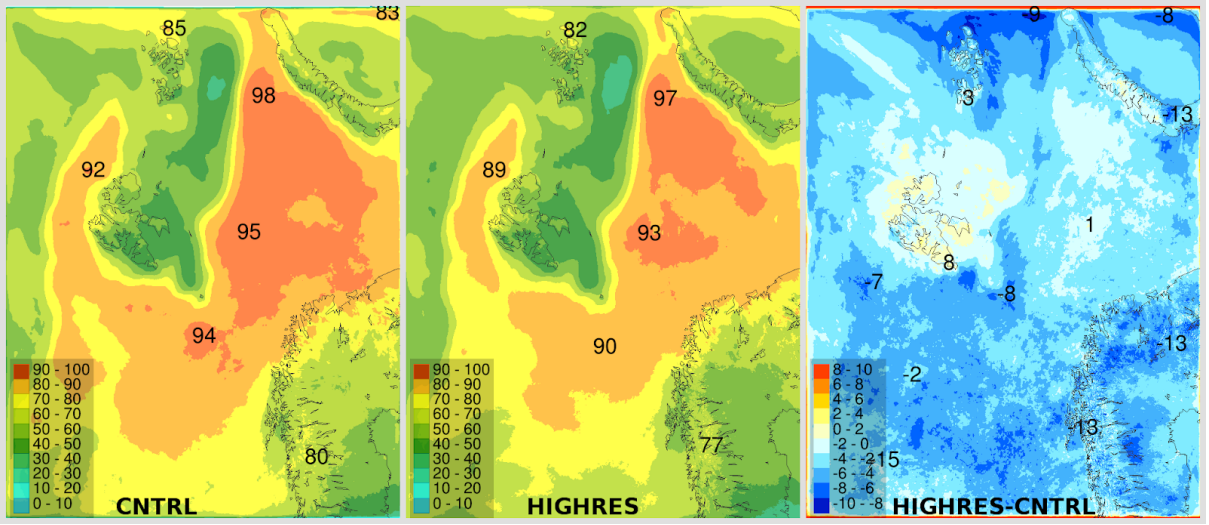


Figure 3.2.12. Average TCC for all lead times in CNTRL (left) and HIGHRES (mid) experiments initialized in the period 8. March - 31.March 2019. The Potential AV in bias, the difference between HIGHRES and CNTRL to the right.

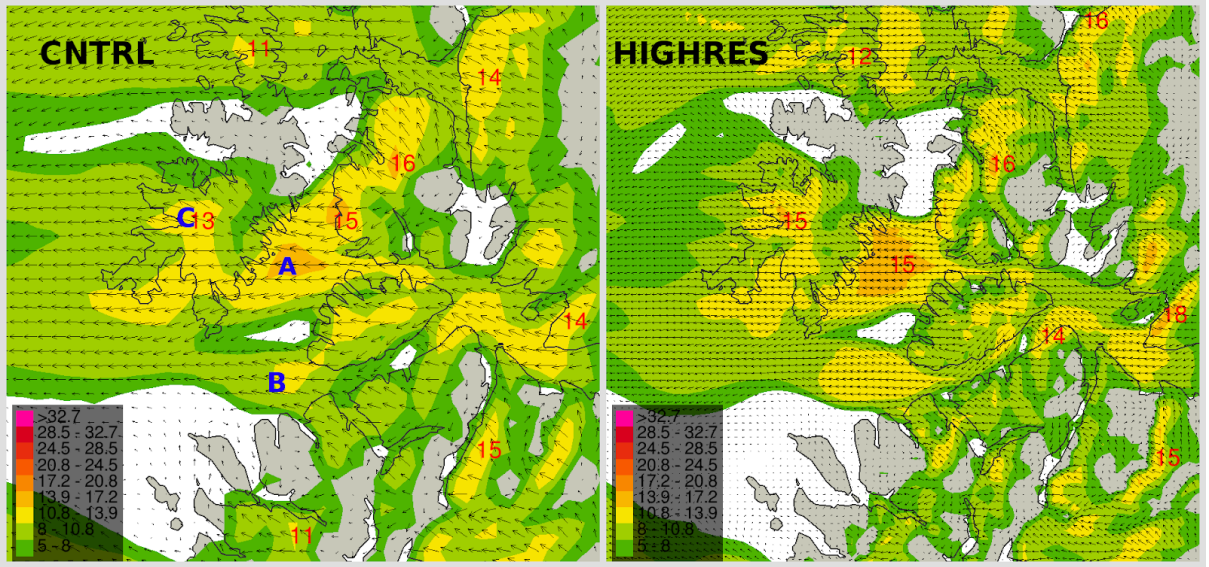


Figure 3.2.13 Wind speed and direction on 9.March 2018 03 UTC in fjords close to Sørøya Northern Norway. Sørøysundet (A), Stjernsundet (B) and Sluskfjellet (C) are known areas of high wind speeds.

3.3 Increased atmospheric and oceanic resolution in seasonal prediction (Task 5.3.3)
(BSC, CNRM, AWI)

Introduction
The impact of atmospheric and oceanic resolution in seasonal prediction is yet to be fully apprehended and understood; past studies have investigated this impact with contrasting conclusions, suggesting some model-dependency in results (e.g. Jia et al. 2015, Prodhomme et al. 2016).

In the framework of APPLICATE we investigate the possible improvements in seasonal forecast quality with higher resolution ocean/ice and atmosphere/land surface components using both CNRM-CM6 and EC-Earth3. In addition, high-resolution (4,5 km) oceanic experiments in the AWI-CM climate prediction system is discussed.

Experiment and model description
At CNRM, ensemble re-forecasts initialized in November and May 1993-2014 were run following the same protocol as the stream 1 experiments with CNRM-CM6-1, but using a higher resolution version of the coupled model. For EC-Earth the same experiments had been performed during the summer but an outage in the BSC archive system caused a major corruption in the HR simulation outputs and the reforecasts had to be repeated. Only the November initialized ones were produced on time for this deliverable. Results from the May initialized reforecasts will be reported in subsequent deliverables.

CNRM-CM6 HRES comprises a tl359l91r resolution atmosphere, corresponding to approximately 50 kilometers in horizontal resolution, over twice the resolution as in stream 1 with CNRM-CM6-1. With respect to the CMIP6 version, beyond resolution, some changes have been introduced to the ARPEGE atmospheric model regarding the formulation of the surface moment flux due to orography; instead of using orographic roughness length, the formulation based on Beljaars et al. (2004) takes into account a standard deviation of sub-grid scale orography and introduces an analytical dependency to sinusoidal orography.
In the ocean component, the same model version of NEMO3.6 and GELATO v6 sea ice are used, but on a ORCA 0.25° grid. Small changes in the tuning of the turbulent vertical diffusion were introduced in this high resolution version with respect to CMIP6, increasing the factors of thermodynamic and turbulent kinetic energy terms with respect to the dynamic.

Note that the change of resolution in the model components implied updating the initial condition fields used for these re-forecasts with respect to stream 1 experiments. The atmosphere was initialized with ERA5 instead of ERA-Interim, with minor impacts on the model biases and skill in the first month (not shown here); the ocean is initialized from an upscaled NEMO-GELATO standalone run forced by ECMWF fluxes and constrained towards Mercator Ocean International GLORYS 1/12° reanalysis.

Model outputs are interpolated onto a regular 1° horizontal grid for fair comparison between the high resolution and standard resolution runs. In some diagnostics, we also include seasonal re-forecasts with the operational Météo-France System 6 for the same nominal start date. This system uses a model version close to the CMIP6 and high resolution versions, with minor namelist and ARPEGE version changes, but a high resolution atmosphere and standard resolution ocean (NEMO 3.6 eORCA1° model). It can therefore be considered as an “intermediate” experiment to investigate the impact related to higher resolution atmosphere only.

EC-Earth3.3 is the global coupled model developed by the EC-Earth consortium for CMIP6 (Doblas-Reyes et al., 2018). It is based on the ECMWF's atmospheric circulation model IFS, cycle 36r4 and the land surface model H-Tessel. The ocean component is a recent version of the ocean model NEMO3.6 and the sea-ice model is a recent version of the Louvain-la-Neuve Sea Ice Model (LIM3). The different components communicate via the coupler OASIS-3. EC-Earth3.3 is used for seasonal re-forecast experiments initialized in November (the May initialized are still in production but will be used in future deliverables). Re-forecasts were performed for the (~ 1°x1°) and high (~ 0.25°x0.25°) resolution version. More detail on the specific components used in each case are included in Table 3.3.1.

Model/System	CNRM-CM6-1-LR	CNRM-CM6-1-HR	EC-Earth 3.3-LR	EC-Earth 3.3-HR
Atmosphere	ARPEGE 6.3	ARPEGE 6.4	IFS Cy36r4	IFS Cy36r4
Ocean	NEMO 3.6	NEMO 3.6	NEMO 3.6	NEMO 3.6
Sea ice	GELATO v6	GELATO v6	LIM3	LIM3
Atmospheric resolution	tl127l91r (~ 1.4°)	tl359l91r (~ 0.5°)	tl255l91r (~ 0.7°)	tl511l91r (~ 0.35°)
Ocean resolution	eORCA1 L75	ORCA025L75	ORCA1L75	ORCA025L75
Ocean and sea ice Initial	Upscaled GLORYS 1/4°	Upscaled GLORYS	Forced NEMO	Forced NEMO simulation (3D

conditions	(Mercator)	1/12° (Mercator)	simulation (3D Nudging + surface restoring to ECMWF's ORAS4)	Nudging + surface restoring to ECMWF's ORAS5)
Ensemble size	30	30	10	10

Table 3.3.1: Seasonal re-forecasts experiments included in the analysis presented in section 3.3. All systems are initialized in the atmosphere with ERA-Interim. All re-forecasts are initialized on the 1st of the month.

Impact of resolution on summer Arctic sea ice forecast quality

The impact of resolution on Arctic sea ice was investigated first in terms of root mean square error and correlation for May starts (figure 3.3.1). With respect to stream 1, both MF System 6 and CNRM-CM6 HR exhibit improved Root Mean Square Error (RMSE) and correlation especially from month 3 (July) onwards. Note however that this improvement could also be linked to the initialization technique, since stream 1 and MF System 6 use the same ocean model version but different nudging strength in the Arctic upper ocean to derive the initial conditions from Mercator GLORYS2V3.

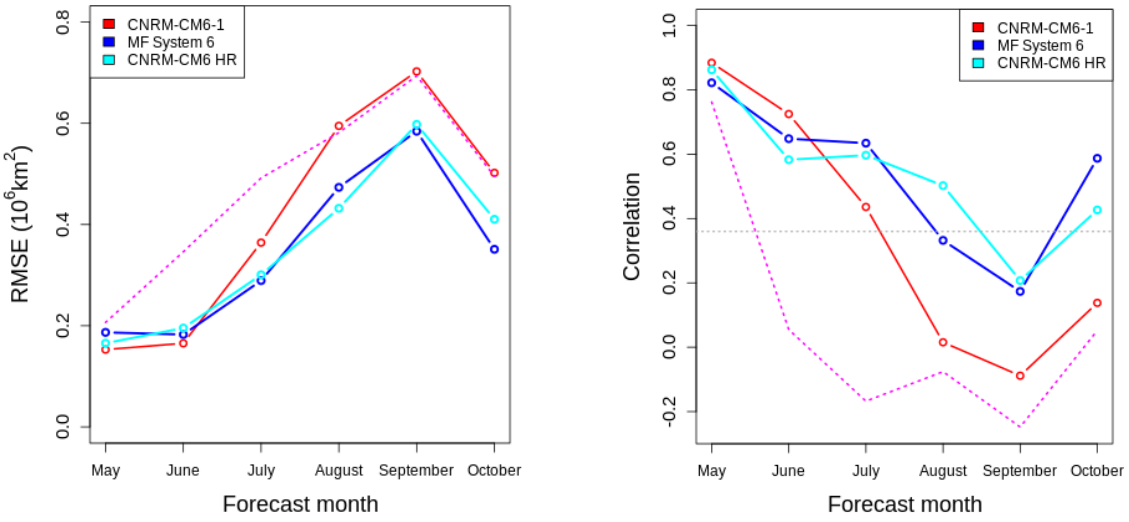
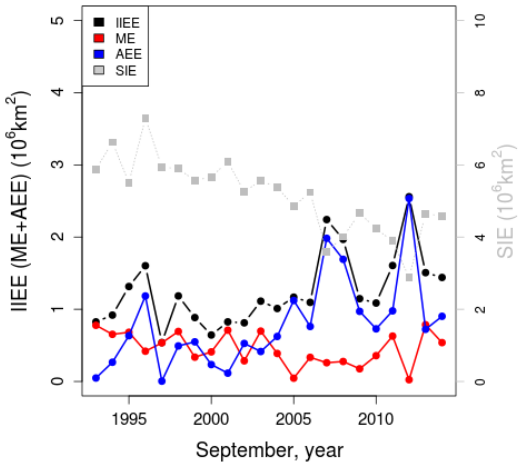


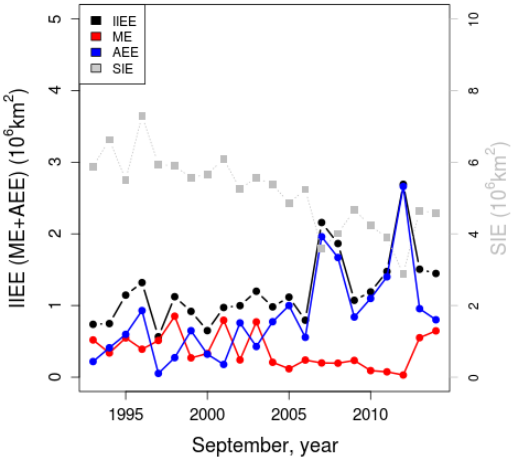
Figure 3.3.1: Evolution with forecast time of RMSE and correlation of pan-Arctic sea ice extent (15% SIC threshold) using NSIDC as a reference, for CNRM-CM6-1 (APPLICATE stream 1), Météo-France operational System 6, and high resolution ocean-atmosphere re-forecasts with CNRM-CM6 (HR).

Further improvement related to higher resolution ocean is unclear at a pan-Arctic scale, the only clear improvement being for August sea ice extent, for which the HR re-forecast has significant correlation (higher than 0.5).

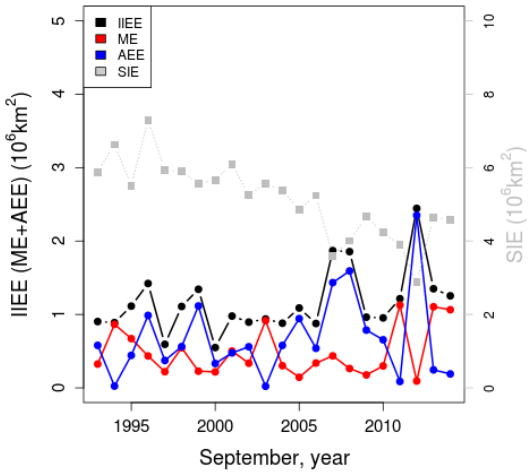
As for Deliverable 5.2, we use the Integrated Ice Edge Error metric (IIEE, Goessling et al. 2016) to determine if and how higher resolution changes the forecast quality for the sea ice edge. This metric is described in detail in Deliverable 5.2. Figure 3.3.2 (a-c) shows the IIEE and its Misplacement Error (ME) and Absolute Extent Error (AEE) components according to the re-forecast year, for September over the Arctic (regions from 45°N to 85°N), using NSIDC sea ice concentration data as a reference, for all three model versions considered. Overall, higher resolution atmosphere and ocean do improve the total IIEE, as shown in Figure 3.3.2 (d) for the 1993-2014 average IIEE according to forecast time. The largest improvements, consistent with RMSE and correlation, are found for August.



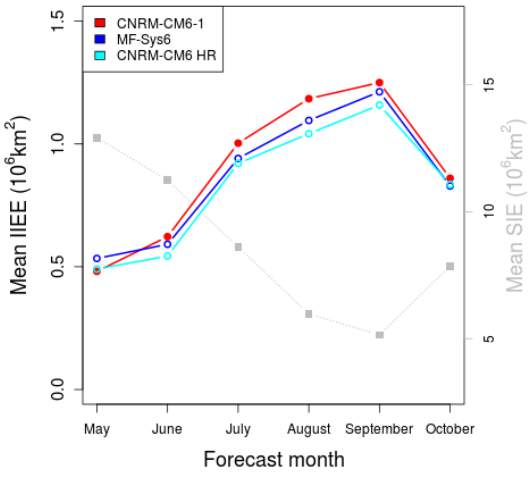
(a) CNRM-CM6-1



(b) MF System 6



(d) CNRM-CM6 HR



(e) IIEE vs forecast time

Figure 3.3.2: (a-c) IIEE and components (AEE in blue and ME in red) according to the year for 1993-2014 September sea ice ensemble re-forecasts initialized in May with CNRM-CM6-1, Météo-France System 6 and CNRM-CM6 HR, respectively. (d) Evolution of the mean 1993-2014 IIEE according to forecast time in the three CNRM re-forecasts initialized in May. In all four plots, NSIDC reference SIE (in millions of km²) is shown in grey with the right y-axis.

Analysis of sea ice concentration and thickness seasonal cycle over Arctic marginal seas

The seasonal cycle of sea-ice concentration and thickness have important spatial differences. To study these differences, we look at the sea-ice concentration and thickness in marginal sea regions in Greenland Sea and Chukchi Sea. Greenland Sea represents a marginal sea in the Atlantic sector of the Arctic Ocean. It is strongly affected by the Atlantic inflow that brings warm water masses to the region, and by the transpolar drift, that exports ice from the Central Arctic to the Greenland Sea. Chukchi Sea is a marginal sea at the Pacific sector of the Arctic. It experiences some inflow of Pacific water masses through the Bering strait, and ice transport by the Beaufort Gyre.

We have selected the data over Greenland Sea and Chukchi Sea using a mask that corresponds to the definitions by the National Snow and Ice data Centre (NSIDC). We compare the CNRM-CM-1 and the CNRM-CM6-HR seasonal re-forecast experiments addressing the ensemble means. Figures 3.3.3 and 3.3.4 show the evolution of sea-ice concentration and thickness from May to October. Similarly, Figures 3.3.5 and 3.3.6 show the evolution from November to April. The higher resolution (CNRM-CM6-HR) favors higher sea-ice concentration and thickness on Greenland and Chukchi Seas throughout the year.

Figure 3.3.3 (a) show the evolution of sea-ice concentration from the May starts of the CNRM-CM-1 re-forecasts. The histograms and the KDE-lines (for Kernel Density Estimates) present the multiyear monthly mean distribution of sea-ice concentration. We see a rapid decline in sea ice concentrations from May to April, with a median decreasing from about 35% to about 15 %. In June, July and August, the mode of the sea-ice concentration is close zero, however a higher sea ice concentration remains in part of the Greenland Sea. The concentration starts to increase again in October. In the CNRM-CM6 HR experiment (Fig 3.3.3 (c)) there is a larger spread in the sea-ice distribution in each month, and the concentrations are overall larger. The sea is decreasing more slowly during the melt season, from a median of c. 45% in May, to c. 35% in June, further to c. 18% in July. Still in August and September, the sea-ice concentration remains in the order of 10%. As ice growth picks up in October, we observe an increase in sea ice concentrations at a large range between 0 and 50 % concentrations, unlike in the CNRM-CM6-1, where the concentration increased more

on the lower end of the concentration range. The sea-ice thickness distribution (Fig. 3.3.4 (a) and (c)) is affected by the resolution (and other differences such as initial conditions) between the CNRM-CM6-1 and CNRM-CM6 HR. The HR experiment produces clearly thicker ice throughout the year.

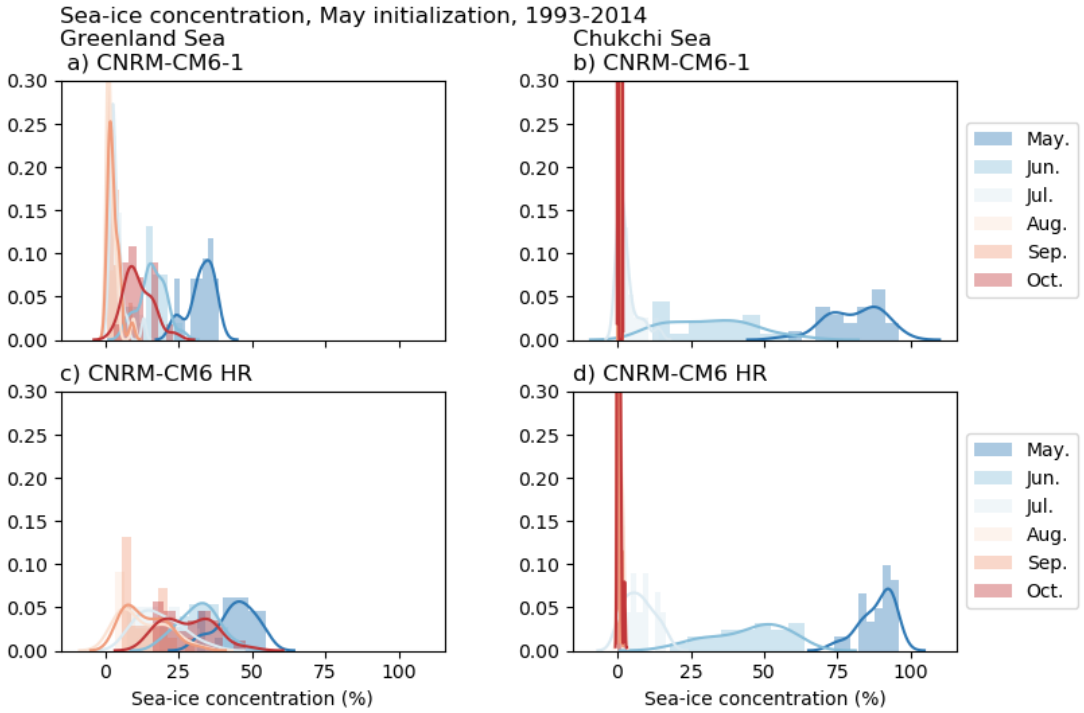


Figure 3.3.3 Average monthly distribution sea-ice concentration from May to October on Greenland and Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with May start dates.

In Chukchi Sea the sea-ice concentration has a strong annual cycle from mostly ice covered to completely ice free (Fig. 3.3.3 (c) and (d)). The thawing reduces the sea ice concentration smoothly from May to July. From August to October the sea is essentially ice-free. The general characteristics are similar between the CNRM-CM-1 and CNRM-CM HR, except that the HR experiment favors higher sea-ice concentrations in May-July and a more visible increase in sea-ice concentration in October. Late in the melt season, the HR experiment conserves more thick ice in Chukchi Sea than CNRM-CM-1 experiment (Fig. 3.3.4 (b) and (d)).

From November to April we see a striking difference in the sea-ice thickness between the CNRM-CM-1 and CNRM-CM HR experiments (Fig. 3.3.6). In the HR experiment the thicknesses range from 0.5 to 1.5m (between November and April), whereas in the CNRM-CM6-1 experiment, the range was only from c. 0.25 to 0.75 m.

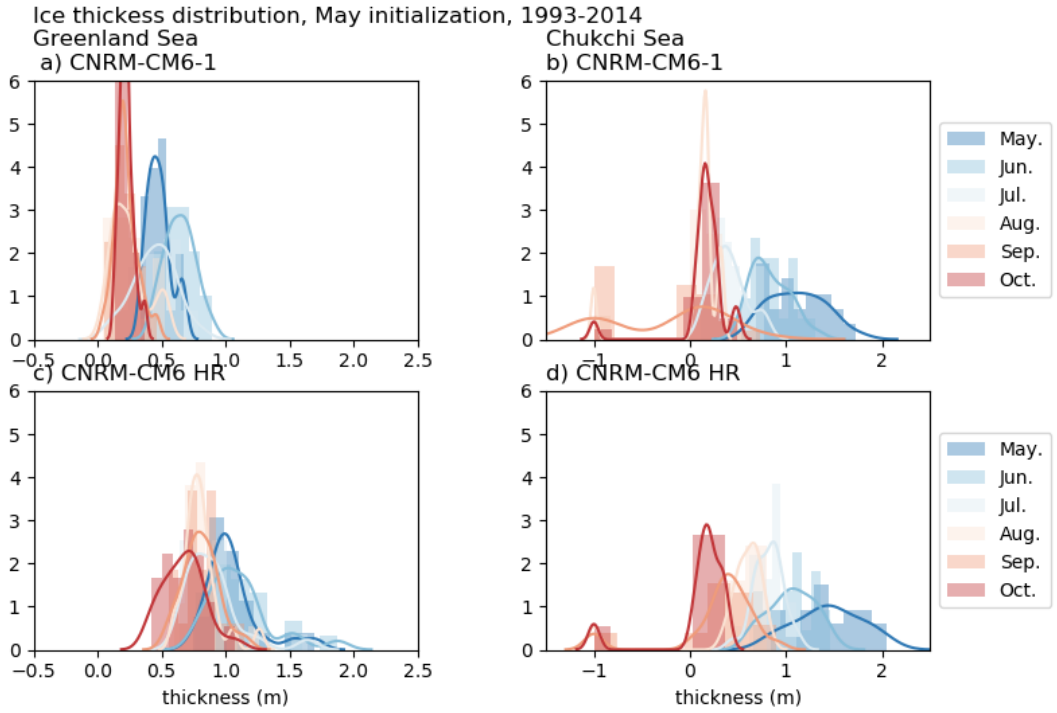


Figure 3.3.4 Average monthly distribution sea-ice thickness from May to October on Greenland and Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with May start dates. On panels (b) and (d) -1 indicates missing values.

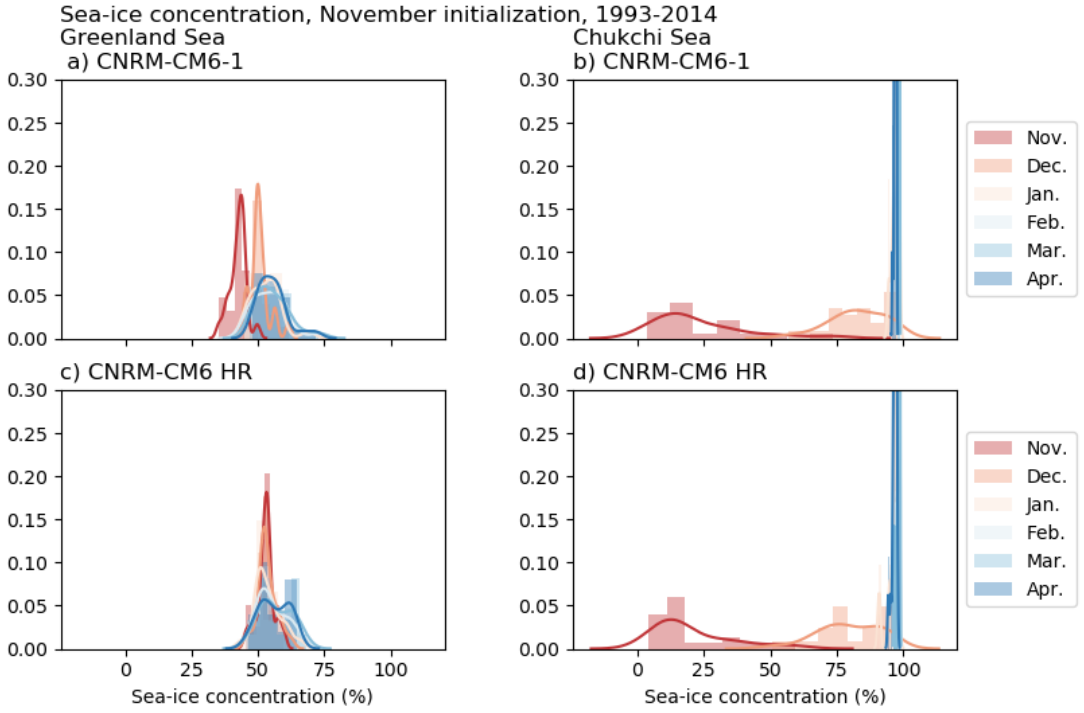


Figure 3.3.5 Average monthly distribution sea-ice concentration from November to April on Greenland and Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with November start dates.

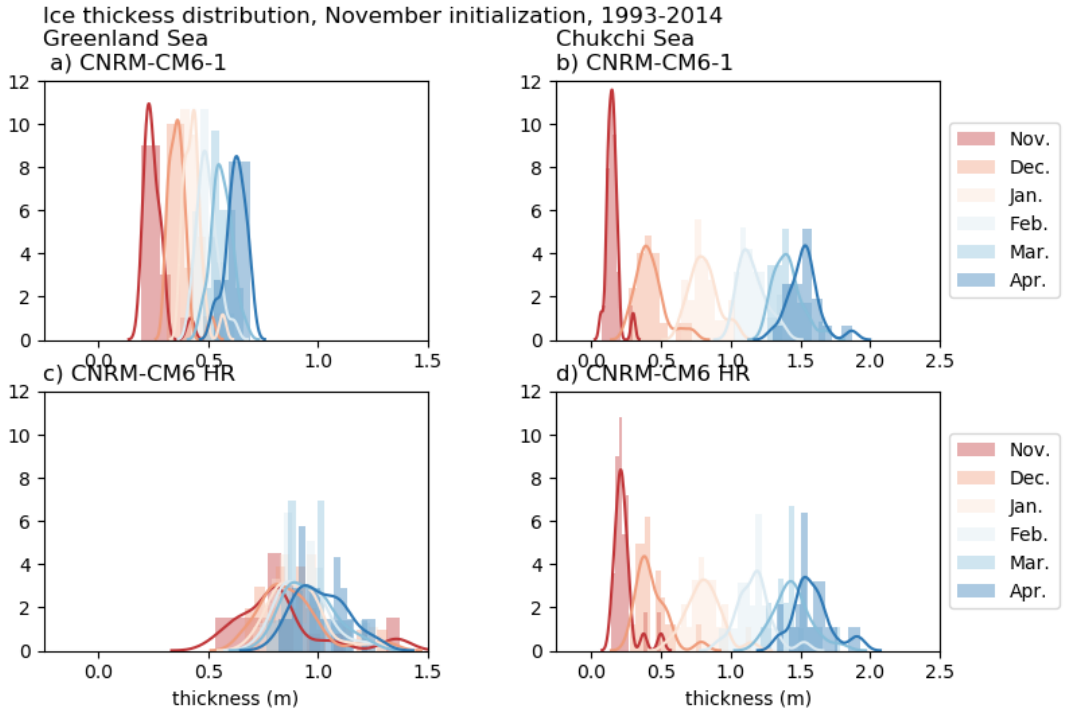


Figure 3.3.6 Average monthly distribution sea-ice thickness from November to April on Greenland and Chukchi Sea in the CNRM-CM6-1 and CNRM-CM6 HR re-forecast experiments with November start dates.

The sea-ice concentrations in the HR experiment are closer to the observations of the NSIDC (Fig. 3.3.7), but in the lack of extensive observations on sea ice thickness, it is hard to assess whether the thicknesses are also improved. Eicken et al. (2001) show based on the Surface Heat Budget of the Arctic Ocean (SHEBA) drifting ice camp in the Chukchi sea in 1998 that the ice thickness in June was 2.13 m and in August c. 1.24m, with standard deviations of 0.53 and 0.54, respectively. Oikkonen and Haapala (2011) assessed the ice draft distribution in Chukchi sea, and obtained mean and modal values of 2.4 m and 1.5 m in spring, and 1.4 m and 0.3 m in autumn over 1988-2000. As all these values are larger than either of the CNRM-CM6 experiment produced, it seems that the thicker ice in the HR experiment would be more realistic than the thinner ice in the CNRM-CM6-1 experiment.

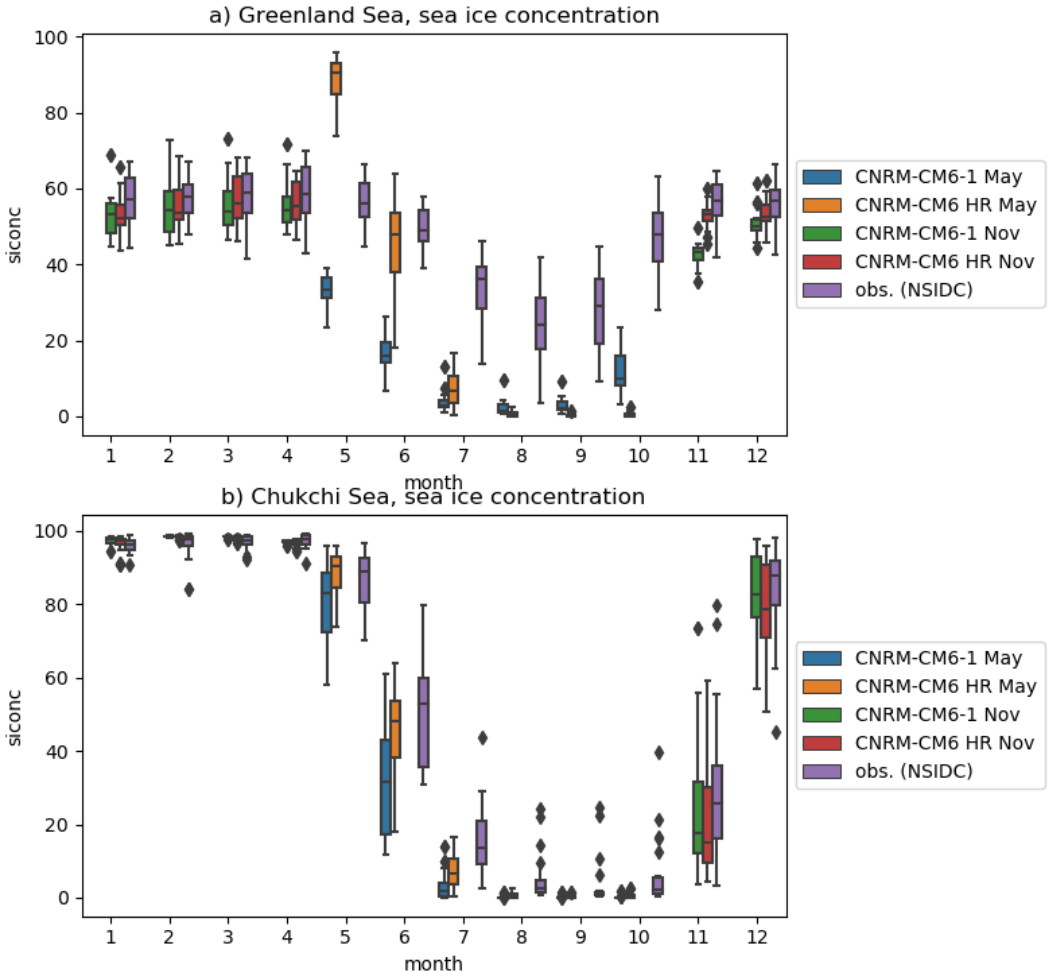
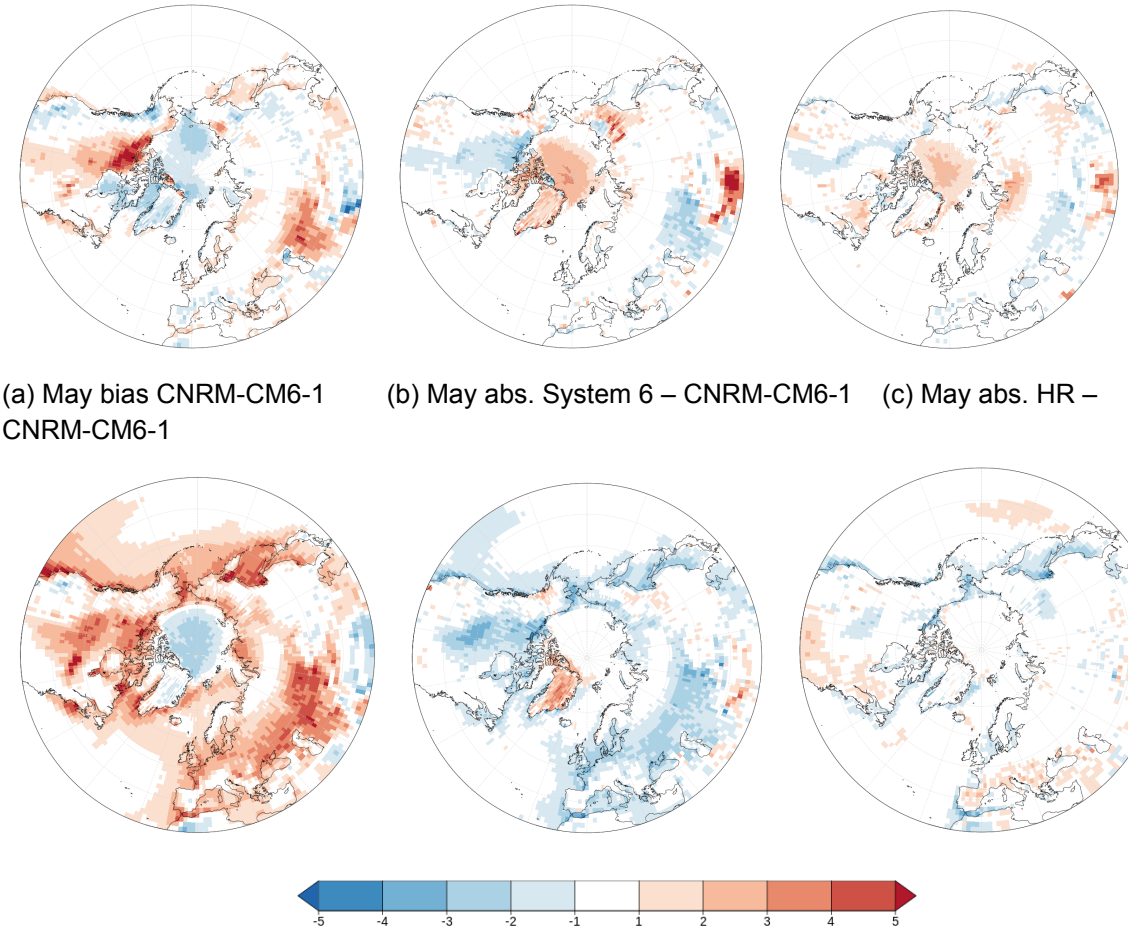


Figure 3.3.7 Average annual cycle of sea-ice concentration on Greenland Sea and Chukchi sea in the CNRM-CM6-1 and CNRM-CM6 HR experiments initialized in May and November, and according to the observations by NSIDC.

Impact of resolution on Northern hemisphere summer forecast quality

In this section we investigate the impact of higher atmosphere and ocean resolution on summer forecast quality with CNRM-CM runs initialized in May, comparing ensembles using CNRM-CM6-1 (standard resolution), System 6 (intermediate resolution: high resolution atmosphere, 1° ocean) and CNRM-CM6 HR (high resolution atmosphere and ocean).

First we assess the differences in mean biases during the first month (May) and three-month season encompassing forecast months 2 to 4 (JJA) of the 1993-2014 re-forecasts. Results for near-surface temperature (T2m) are shown in figure 3.3.8.



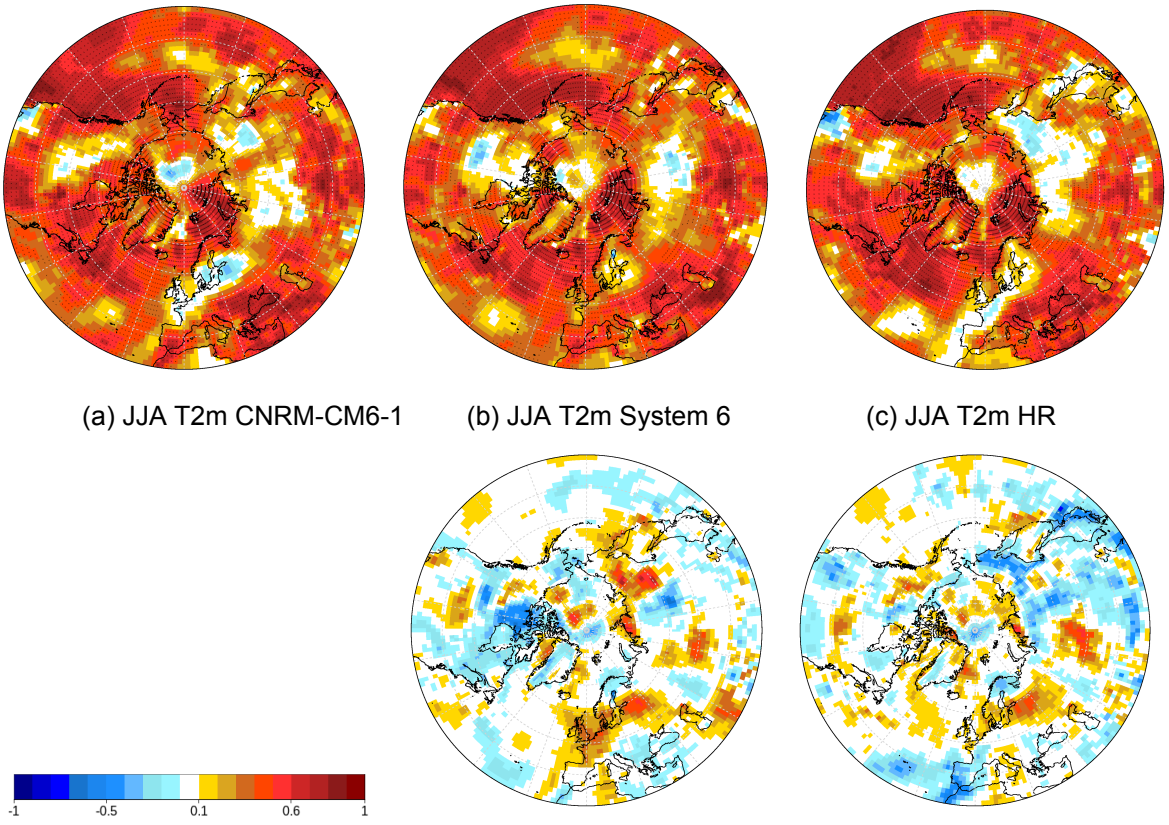
(a) May bias CNRM-CM6-1 (b) May abs. System 6 – CNRM-CM6-1 (c) May abs. HR – CNRM-CM6-1

(d) JJA bias CNRM-CM6-1 (e) JJA abs. System 6 – CNRM-CM6-1 (f) JJA abs. HR – CNRM-CM6-1

Figure 3.3.8: Mean 2m temperature bias with respect to ERA-Interim over the 1993-2014 re-forecast period for May (a) and JJA (d) with CNRM-CM6-1 initialized in May; absolute bias difference with CNRM-CM6-1 re-forecasts in System 6 (b,e) and HR runs (c,f). Values are in °C.

Results in terms of temperature biases can be summarized as follows: Northern Hemisphere summer re-forecasts with CNRM-CM6-1 develop a pronounced warm bias over continents with forecast time, already present in some regions from month 1. Over the Arctic region, a cold bias is found. When evaluating the differences in absolute biases (blue regions show where the bias is improved, disregarding the sign of the bias), it appears that this bias increases with a higher resolution atmosphere. However the strong biases over the high and mid-latitudes continental regions are generally improved in System 6. When increasing the ocean resolution, some of these improvements are lost. A possible explanation is the change in initial conditions in the ocean, leading to an overall warmer model climate in terms of surface temperature.

So as to assess the impact of higher resolution on seasonal forecast quality in summer, we focus here on near-surface temperature re-forecasts with three model versions of the CNRM coupled model (including operational seasonal forecasting system 6). Results for correlation over 1993-2014 for JJA near surface temperature are shown in Figure 3.3.9. A first conclusion is that correlation is not improved everywhere – some regions exhibit a clear degradation of skill in the higher resolution versions. Although part of these differences may be due to chance and higher ensemble sizes may be needed to draw robust conclusions, we find some similarities in the correlation differences with stream 1 when comparing System 6 and the CNRM-CM6 HR re-forecast. For instance, apart from the Central Arctic, T2m over peripheral seas in the Arctic basin is generally improved. This could be related to the improvements in the representation of sea ice during the melt season.



(a) JJA T2m CNRM-CM6-1 (b) JJA T2m System 6 (c) JJA T2m HR

(d) JJA T2m System 6 - Ref (e) JJA T2m HR - Ref

Figure 3.3.9: Anomaly correlation of summer (JJA) 2m temperature with ERA-Interim in re-forecasts initialized in May with (a) CNRM-CM6-1 (APPLICATE stream 1) (b) Météo-France operational System 6, and (c) CNRM-CM6 (HR). (d) and (e) show the difference in correlation with stream1: (b-a) and (c-a), respectively.

Over the North Atlantic, T2m is improved only with higher resolution in the ocean. We note however that System 6 had higher skill for summer T2m over Western Europe, and more generally over the Eurasian continent, than both other model versions.

Impact of resolution on Northern hemisphere winter forecast quality

Table 3.3.1 presents information on the seasonal re-forecasts evaluated in this deliverable. They cover the period 1993-2014. Both CNRM-CM6-1 and EC-Earth3.3 re-forecasts were used in their respective low (~ 1°x1°) and high (~ 0.25°x0.25°) resolution versions.

The high resolution version of CNRM-CM6 shows enhanced skill with respect to its low resolution version in sea ice concentration when comparing the RMSE: Okhotsk, Bering, Labrador and GIN Seas show a decrease in RMSE, but there is also a region with lower skill at the North of the Barents and Kara Seas . The Anomaly Correlation Coefficient (ACC) shows a clear improvement over the Baffin and Greenland Seas (Figure 3.3.10).

EC-Earth3.3 shows some improvement over the Labrador and Okhotsk Seas in terms of RMSE reduction (Figure 3.3.10), but a large degradation near Iceland and over the Barents Sea which is likely caused by a large negative sea surface temperature bias over the North Atlantic sector in the high resolution version of EC-Earth3.3 (not shown). A deficient model tuning in its high resolution version is the most likely explanation for the strong biases.

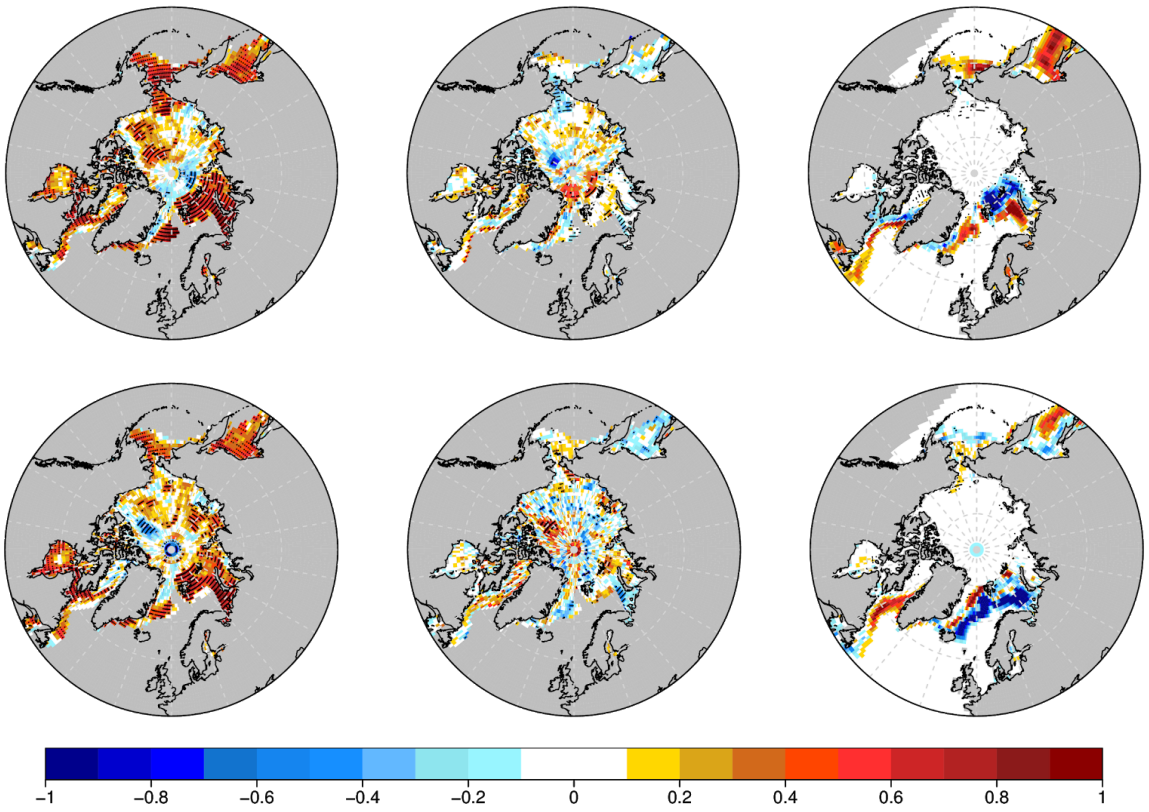


Figure 3.3.10: Top left: DJF sea ice concentration anomaly correlation coefficient between 30-member ensemble mean from CNRM-CM6-LR and CERSAT for the period 1993-2014. Top center: added value of an increase in resolution shown as the difference in correlation coefficient between CNRM-CM6-HR , and CNRM-CM6-LR , both against CERSAT. Top right: added value of an increase in resolution shown as the difference in root mean square error (times 10) between CNRM-CM6-LR and CNRM-CM6-HR , both against CERSAT. Bottom: The same as the top panel for EC-Earth3.2 using ten members.

Arctic sea ice area shows similar levels of skill (ACC and RMSE) for each model in its high and low resolution configurations (Figure 3.3.11), with a slight improvement in the high resolution versions. The general decreasing trend in sea ice extent found in observations and models explains a large fraction of the skill. This can be seen after linearly detrending the sea ice area series (Figure 3.3.12), that skill is substantially reduced for both models, with an especially strong decrease in the EC-Earth3.3-HR version (correlation changing from 0.75 to 0.18). This detrending leads to a lower skill (both in terms of RMSE and ACC) in EC-Earth3.3-HR with respect to EC-Earth3.3-LR, probably reflecting the impact of the large biases in the former previously identified over the Barents Seas. CNRM-CM6-HR still performs slightly better than CNRM-CM6-LR after detrending the series, which confirms the findings for individual marginal seas discussed above.

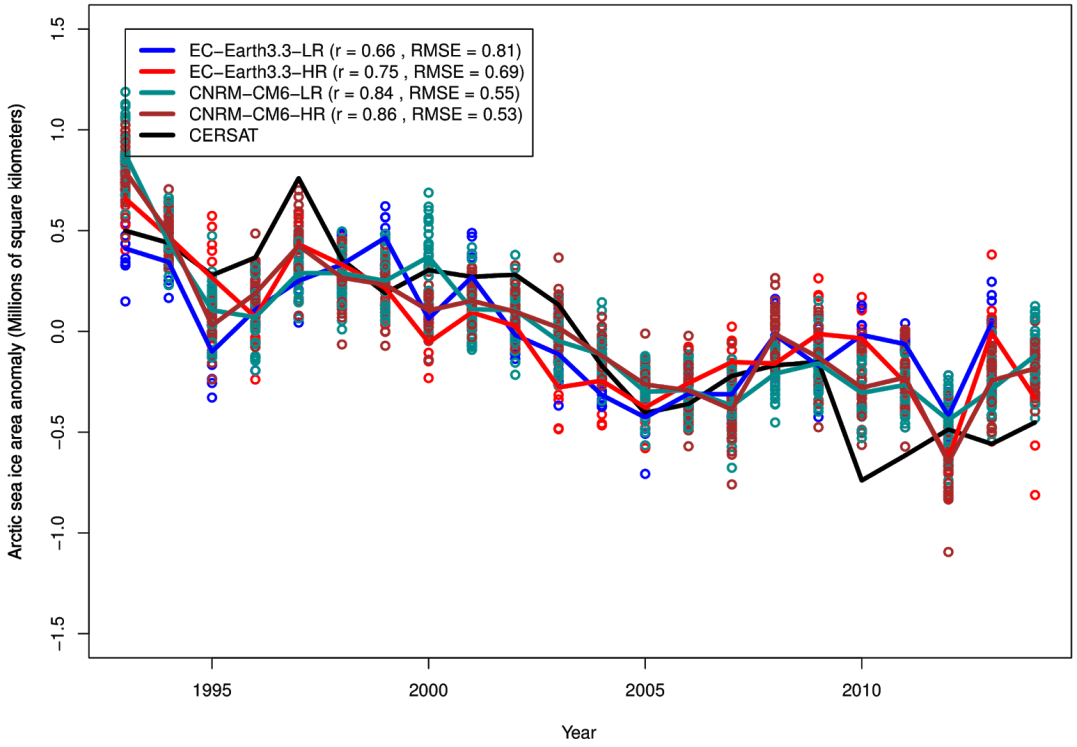


Figure 3.3.11: DJF (1993-2014) Arctic sea ice area anomaly from EC-Earth3.3 (LR and HR), CNRM-CM6 (LR and HR) and CERSAT. Thick colored lines represent ensemble means and circles show individual member values.

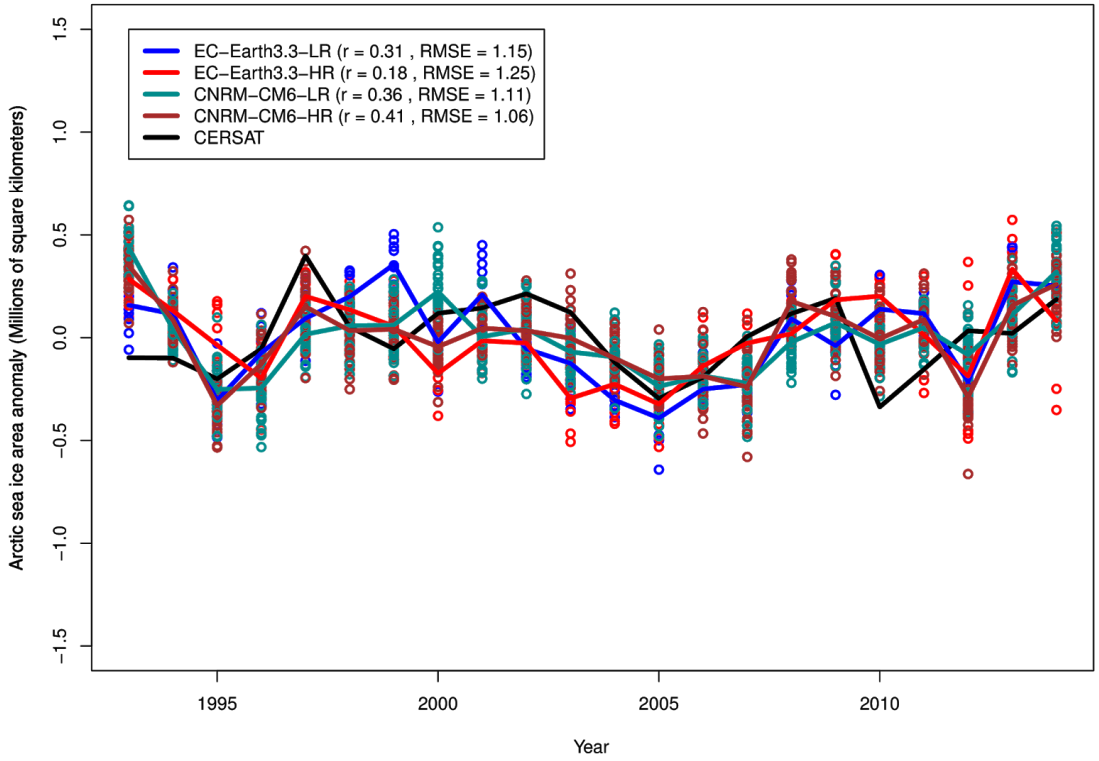


Figure 3.3.12: Linearly detrended DJF (1993-2014) Arctic sea ice area anomaly from EC-Earth3.3 (LR and HR), CNRM-CM6 (LR and HR) and CERSAT. Thick colored lines represent ensemble means and circles show individual member values.

Predictability of winter climate in the Northern Hemisphere

Mean winter sea level pressure shows both regional increases and decreases of skill due to changes in model resolution in CNRM-CM6 and EC-Earth3.3 (Fig 3.3.13). CNRM-CM6-HR shows statistically significant higher skill than CNRM-CM6-LR over the Beaufort Sea and lower skill over the western North Atlantic, both in terms of ACC and RMSE. This loss of skill over the North Atlantic has an impact on the predictive skill of in the North Atlantic Oscillation, with values of correlation coefficient changing from 0.44 to 0.23 in the low and high resolution versions, respectively (Fig 3.3.14). By contrast, the Pacific North American (PNA) index shows a slight improvement in CNRM-CM6-HR. In EC-Earth3.3-HR large improvements appear over the Mediterranean/North Africa, Northwest Pacific regions, contrasting with a degradation of skill over the Aleutian region. In terms of NAO and PNA skill (Figs 3.3.14 and 3.3.15), EC-Earth3.3-HR is always outperformed by EC-Earth3.3-LR. However, due to the small ensemble size, EC-Earth3.3 results should in general not be overinterpreted. To illustrate this we took the first ten members of the CNRM-CM6 forecasts and repeated the analysis for sea level pressure (Fig 3.3.13). The noise is clearly larger when using ten members as compared with 30 members. Hence ten-member ensembles might not be enough to detect real skill improvement or degradation.

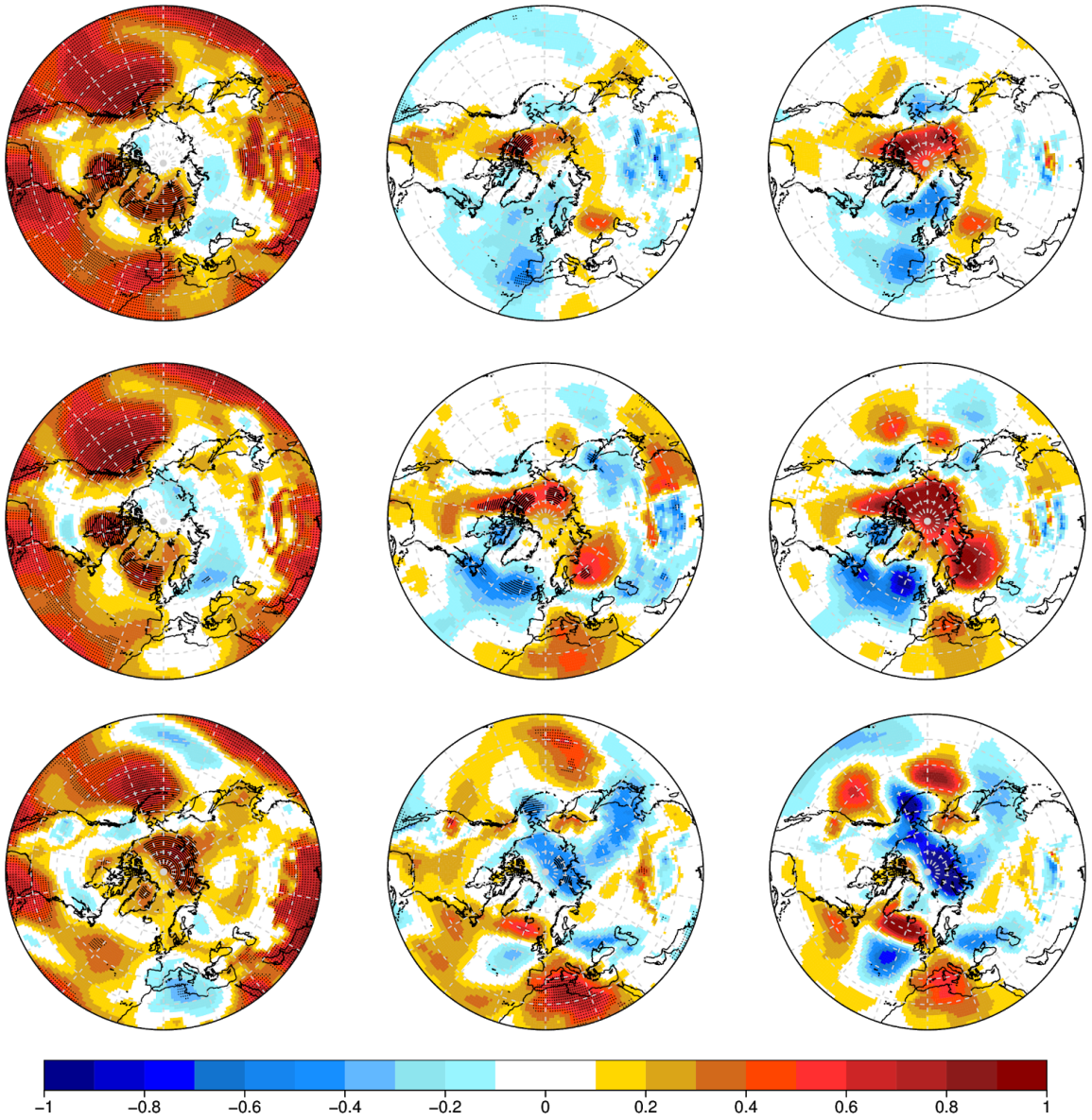


Figure 3.3.13: Top left: DJF sea level pressure anomaly correlation coefficient between 30-member ensemble mean from CNRM-CM6-LR and ERA-Interim for the period 1993-2014. Top center: added value of an increase in resolution shown as the difference in correlation coefficient between CNRM-CM6-HR , and CNRM-CM6-LR , both against ERA-Interim. Top right: added value of an increase in resolution shown as the difference in root mean square error (hPa) between CNRM-CM6-LR and CNRM-CM6-HR , both against ERA-Interim. Middle: the same as the top panel, but using only ten members. Bottom: The same as the panel in the middle for EC-Earth3.2 using ten members.

Surface temperature forecast skill is improved in CNRM-CM6-HR when compared to CNRM-CM6-LR over northeastern North America, Aleutian region and the GIN Seas,

while there is a very large degradation over eastern Europe and the Mediterranean (Fig 3.3.16). These results are consistent, regardless of the skill metric considered (ACC or RMSE). The improvement in surface temperature over the GIN Seas is likely a result of the higher sea ice concentration skill over the region. EC-Earth3.3-HR displays a strong surface temperature skill gain in North Africa, Northwest Atlantic, the Eurasian Arctic, with large losses over the Northeast Atlantic and large parts of Eurasia and the Pacific Ocean with respect to EC-Earth3.3-LR (Fig 3.3.16). The skill gain over the Labrador Sea/ loss over the GIN Seas in EC-Earth3.3-HR is probably stemming from enhanced/reduced skill capacity of sea ice concentration in the respective regions.

The response of precipitation to increased resolution both in CNRM-CM6 and EC-Earth3.3 (Fig 3.3.17) is quite noisy which hinders its interpretation. The clearest feature identified both for the ACC and RMSE metrics is an increase in precipitation skill over the Mediterranean Sea and along the Kuroshio current in EC-Earth3.3-HR as a result of better skill in atmospheric circulation over the region (Fig 3.3.13).

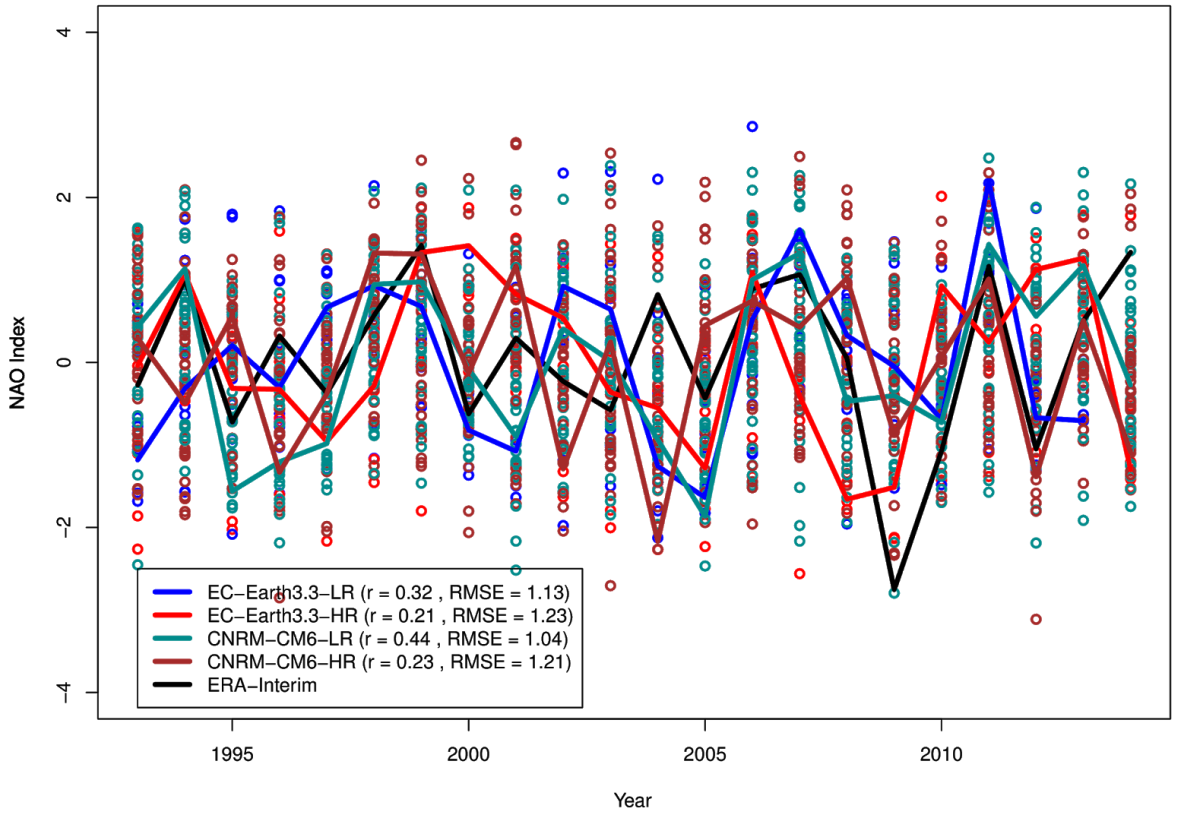


Figure 3.3.14: DJF (1993-2014) North Atlantic oscillation (as in Stephenson et. al, 2003) from EC-Earth3.3 (LR and HR), CNRM-CM6 (LR and HR) and ERA-Interim. Thick colored lines represent ensemble means and circles show individual member values. All values are standardized.

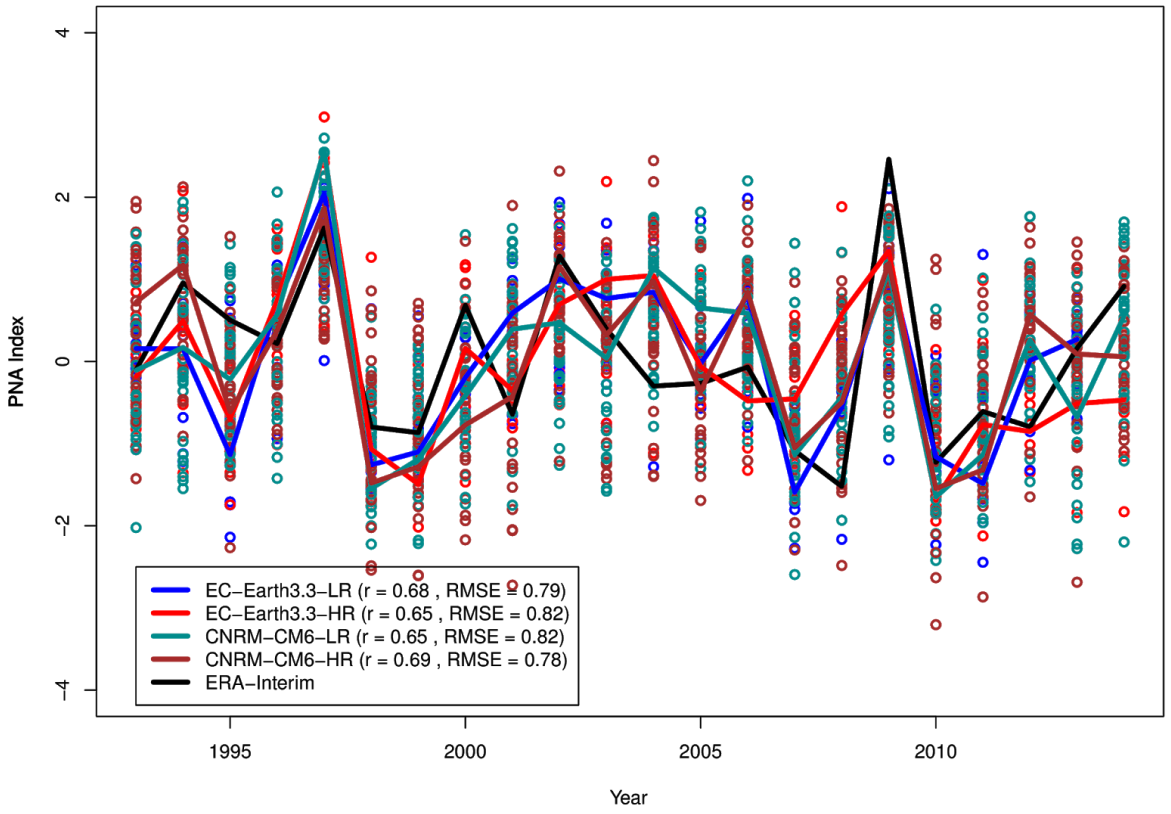


Figure 3.3.15: DJF (1993-2014) Pacific North American pattern (as in Wallace and Gutzler, 1981) from EC-Earth3.3 (LR and HR), CNRM-CM6 (LR and HR) and ERA-Interim. Thick colored lines represent ensemble means and circles show individual member values. All values are standardized.

Increased oceanic resolution in a climate prediction system (AWI)

From the ocean and sea-ice point of view, improvements in the simulation of the high-latitude ocean hydrography and the realization of small scale leads in Arctic sea ice can be seen from simulations with FESOM (Finite Element Sea-ice Ocean Model) when increasing the horizontal resolution from 24 km to 4.5 km in the Arctic.

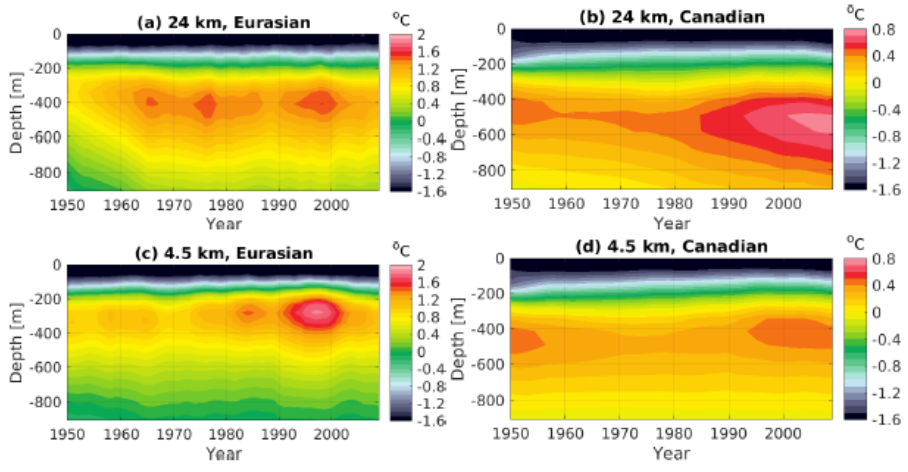


Fig. 3.3.16 Hovmöller diagram of mean potential temperature for the (a) Eurasian Basin and (b) Canadian Basin obtained in a simulation with 24 km horizontal resolution. Panels (c, d) are the same as (a, b) but for a simulation with 4.5 km resolution.

The issue of too deep and thick Atlantic Water (AW) layer in coarse ocean models reported in previous studies can be significantly alleviated by increasing horizontal resolution. The basin mean temperature shows a very different temporal evolution in two simulations with different horizontal resolution (Fig. 3.3.16). In the Eurasian Basin, the warm AW layer thickens with time during the first 15 model years in the low-resolution simulation (LOW), while the layer thickness remains quasi-steady (up to interannual variability) in the high-resolution simulation (HIGH). After initial spin-up, the depth of temperature maxima is located at about 400m in LOW, while in HIGH it remains at about 300 m, the observed depth suggested by the hydrographic climatology. In the Canadian Basin, the thickening and deepening of the AW layer is also very obvious in LOW. In this simulation, the core of the AW layer deepens by about 100 m, changing from about 450 to 550m during the 60 model years. The model drift in the AW layer occurring during model spin-up is irreversible afterwards. In HIGH, no thickening and deepening trend is found in the Canadian Basin.

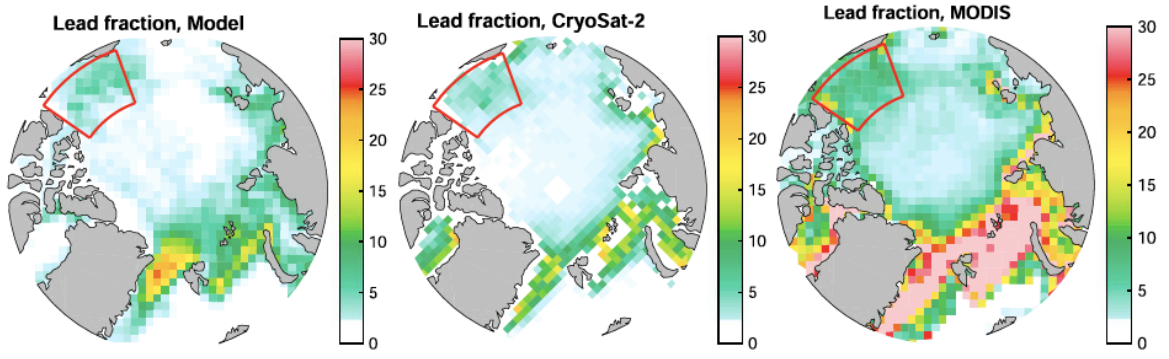


Fig. 3.3.17 Comparison of mean sea ice lead area fraction (%) between model simulation and satellite observations: model (left), CryoSat-2 (middle), and MODIS (right). The results are averaged from January to March of the period 2011–2014. In the model simulation the horizontal resolution is 4.5 km.

Sea ice leads in the Arctic are important features that give rise to strong localized atmospheric heating; they provide the opportunity for vigorous biological primary production, and predicting leads may be of relevance for Arctic shipping. By increasing horizontal resolution, one can reasonably simulate the spatial and temporal variation of lead area fraction in comparison to satellite remote sensing data. For example, Fig. 3.3.17 shows that the spatial distribution of leads obtained in

a high-resolution model simulation is consistent with different satellite observations. Both the simulations and observations reveal that leads are mainly located in marginal seas in wintertime, including the Beaufort, Laptev, Kara and Barents Seas and near the Fram Strait.

Unfortunately, when coupling FESOM to the atmosphere model ECHAM6.3 no improvements in the Arctic ocean hydrography can be shown with higher ocean model resolution. The same is valid for higher atmosphere model resolution. The reason for this probably is a strong atmospheric large scale circulation bias which occurs especially in the LR (T63), but also in the HR (T127) version of the atmospheric model ECHAM6.3 (Fig. 3.3.18). Tests with another atmosphere model, the OpenIFS, in higher horizontal resolution, are underway. It is clear that the next step is to tackle the atmospheric bias so that we can improve the ocean hydrography and realization of Arctic sea ice also in the coupled system.

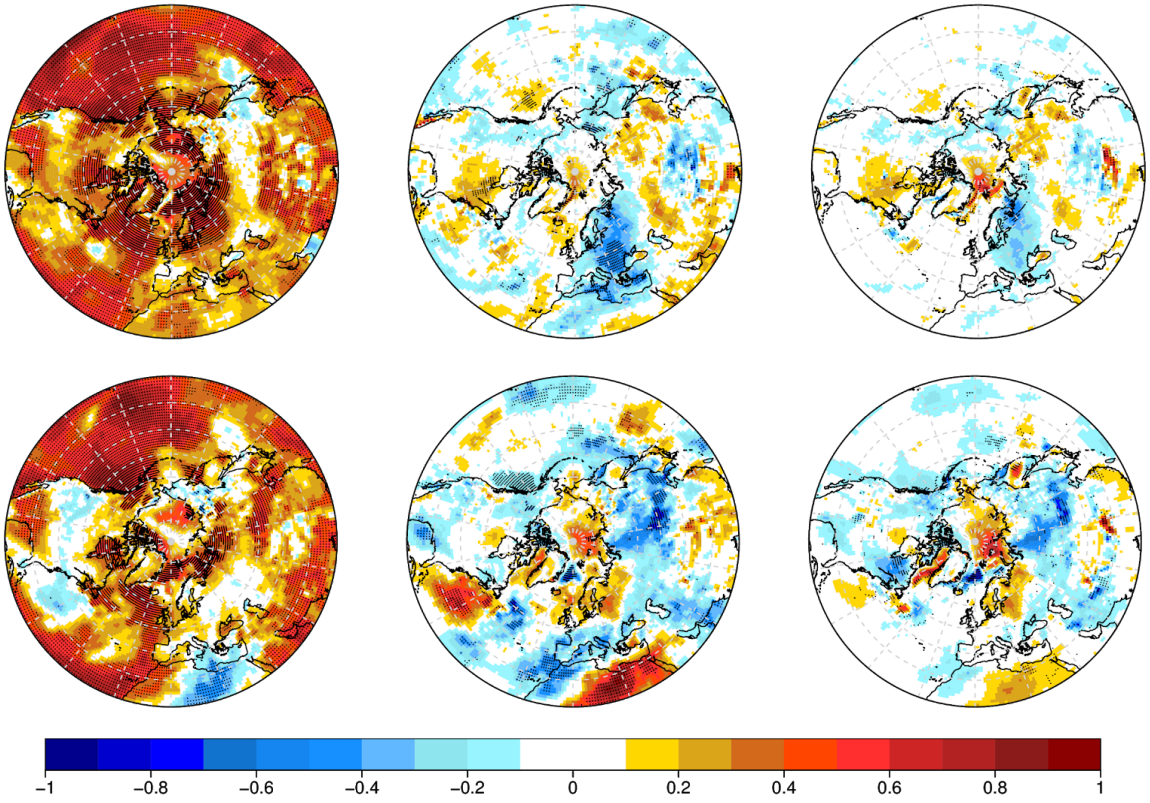


Figure 3.3.16: Top left: DJF 2-meter temperature anomaly correlation coefficient between 30-member ensemble mean from CNRM-CM6-LR and ERA-Interim for the period 1993-2014. Center: added value of an increase in resolution shown as the difference in correlation coefficient between CNRM-CM6-HR, and CNRM-CM6-LR , both against ERA-Interim. Top right: added value of an increase in resolution shown as the difference in root mean square error (K) between CNRM-CM6-LR and CNRM-CM6-HR , both against ERA-Interim. Bottom: The same as the top panel for EC-Earth3.2 using ten members.

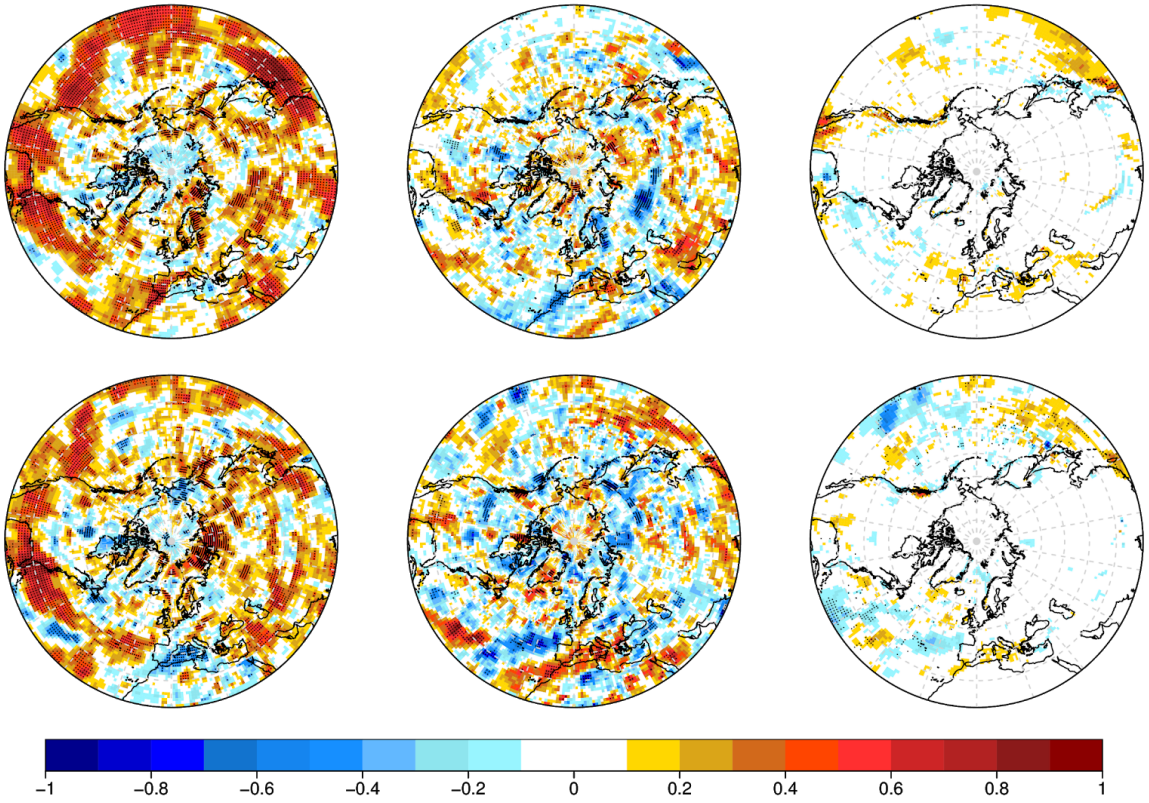


Figure 3.3.17: Left: DJF precipitation anomaly correlation coefficient between 30-member ensemble mean from CNRM-CM6-LR and GPCP v2.2 for the period 1993-2014. Center: added value of an increase in resolution shown as the difference in correlation coefficient between CNRM-CM6-HR ,and CNRM-CM6-LR , both against GPCP v2.2. Right: added value of an increase in resolution shown as the difference in root mean square error (mm/day) between CNRM-CM6-LR and CNRM-CM6-HR , both against GPCP v2.2. Bottom: The same as the top panel for EC-Earth3.2 using ten members.

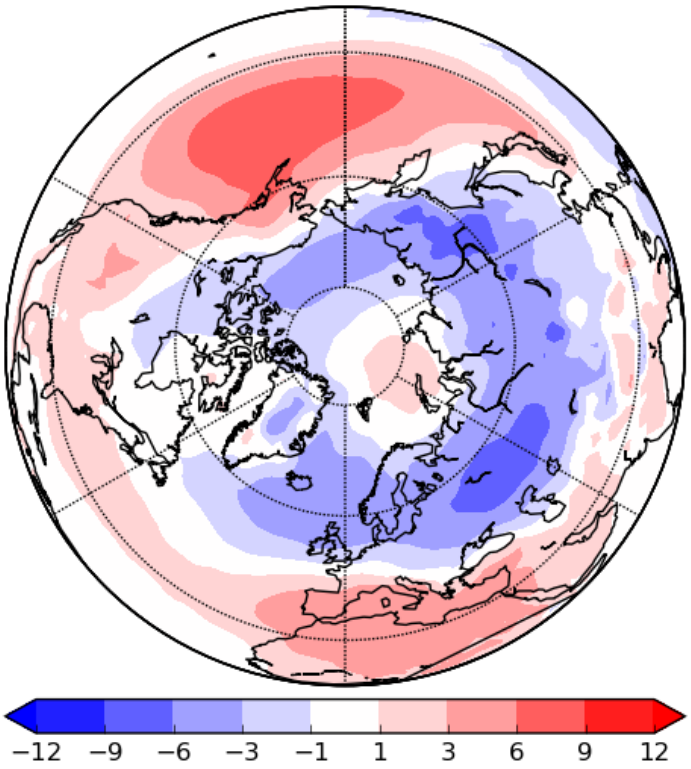


Fig. 3.3.18: Mean sea level pressure bias as an average over 100 DJF seasons from ECHAM6.3 in the LR configuration (T63 corresponding to around 200 km horizontal resolution).

Conclusions

There is a rather limited improvement on northern hemisphere winter (DJF) extratropical forecast skill in the climate models CNRM-CM6 and EC-Earth3.3 due to increased oceanic and atmospheric resolution. In particular, there seems to be an improvement in sea ice representation in CNRM-CM6-HR in the Atlantic and Pacific sectors leading to a small improvement of the integrated sea ice area. However, the sea ice improvement does not necessarily translate into better predictions in the mid-latitude regions. In EC-Earth3.3-HR the sea ice skill is possibly degraded due to an unsatisfactory tuning phase (hampered by the high computational demands required to cover efficiently the space of hyperparameters). Other differences between the high and low resolution versions in EC-Earth can be attributed to the different ocean reanalyses employed to produce the initial conditions. Additionally any interpretation of the EC-Earth3.3-HR results is difficult due to the use of only ten members. Ideally future forecasts should include more members. For AWI-CM, it has been found that the representation of the Arctic ocean hydrography and the Arctic sea ice can be improved by increasing the resolution of the ocean-sea-ice component to 4.5 km in ocean-sea-ice only simulations, but without tackling the

Arctic atmosphere large-scale circulation bias this does not translate into any improvement of the Arctic ocean hydrography in the coupled system.

3.4 Improved ensemble generation techniques for weather and climate predictions (Task 5.3.4) (MET-Norway, CNRM)

Uncertainties are present in all parts of a prediction system and should be taken into account to provide reliable predictions. A prediction should therefore also include estimates on the uncertainty of the prediction itself. Ensemble Prediction Systems (EPSs) are the most commonly used method to account for such uncertainties. This yields for different time-scales even if the predictability limit is reached faster for small−scale features than for large−scale features. In recent years, convection−permitting ensemble prediction systems (CPEPSs) have become an integrated part of operational weather forecasting (e.g. Hagelin et al., 2017, Frogner et al., 2019a). However, to our knowledge no high-resolution EPS have yet been used for operational weather forecasting in the Arctic. In 3.4.1 we present such a system, the regional high-resolution AROME-Arctic Ensemble Prediction System (AAEPS). Since a major part of the regional domain of AAEPS is over ocean also results from an experiment with additional perturbations of the sea surface temperature is discussed.

Also at the seasonal time scale, the use of ensembles is a key in representing different sources of uncertainty. In seasonal prediction experiments these origin from initial condition uncertainty, model uncertainty and uncertainties arising from the change in climate over the course of the re-forecast period. For Arctic seasonal predictions the ocean regions are crucial and in section 3.4.2, results from experiments with stochastic perturbations of temperature and salinity in the ocean part (NEMO 3.6) of CNRM-CM6-1 are presented and discussed. A possible better description of forecast uncertainties, as will be discussed in the following, will contribute to improve present day weather and seasonal prediction capabilities.

Regional high-resolution EPS for short-range weather forecasting

It is well known that higher-resolution deterministic systems add value to coarser resolution systems, but in order to justify the cost of running a regional CPEPS also a comprehensive evaluation of the EPS-part is needed. We therefore first describe the configuration of AAEPS, before a comparison with operational forecasts from the coarser resolution global IFS-ENS from ECMWF is presented. In the end, also results from an experiment with perturbations of the sea surface temperature are discussed. Note that the added value of AAEPS over it's control run is already discussed in Task 5.3.2/section 3.2 and show a substantial added value.

AROME-Arctic Ensemble Prediction System (AAEPS)

The system are based on the operational configuration of the deterministic AROME-arctic with respect to domain (see Figure 3.2.1), resolution (2,5 km horizontal, 65 vertikal layers), assimilation (3D-var upper-air and optimal interpolation for surface assimilation), model physics and dynamics (for more details see Müller et al. (2018), Batte et al., 2018). This configuration is combined with the operational settings of the MetCoOp EPS (MEPS), a shared operational high-resolution EPS for Scandinavia in cooperation between Finnish, Swedish and Norwegian MET services (Frogner et al. 2019a, 2019b).

The system include 1 unperturbed control and 10 perturbed ensemble members. Initial and lateral boundary condition perturbations are constructed by the "scaled lagged average forecasting" method (SLAF: Ebisuzaki and Kalnay, 1991; Hou et al., 2001) using the difference between two IFS-HRES forecasts 6 h a part but valid at the same time. Different members are perturbed based on different lead times (e.g. difference between IFS-HRES +6h and +12 h valid at the same time, and between +12 and +18 h etc). These perturbations are then scaled so all members have on average the same magnitude. Furthermore, random perturbations of sea−surface temperature (SST), soil water and temperature, roughness and albedo with a given correlation length follows the approach of Bouttier et al. (2016). An example of initial perturbations of air, sea surface and soil temperatures is shown in FIgure 3.4.1. Notice that the SST perturbations are smaller in amplitude than soil surface and upper-air perturbations and that upper-air perturbations contain more small scale structures than the surface perturbations. The soil surface perturbations are generated from a smoothing of a random field with a correlation length scale of 150km, whereas in the upper air the spatial scale of the perturbations are that of IFS-ENS. Smaller spatial scales for the surface perturbations has been tested, but they have less positive impact on the forecast performance. More detailed description and discussion of the different EPS configurations can be found in Frogner et al. (2019a, 2019b).

Added Value of AAEPS compared to IFS-ENS

AAEPS is verified against observations and compared with IFS-ENS (1+50 members). In Figure 3.4.2, CRPS and the spread-skill relations for both systems are shown for Mean Sea Level Pressure (MSLP), 10m wind speed (S10m) and 2m air temperature (T2m). For S10m and T2m, AAEPS have smaller RMSE and larger spread than IFS-ENS. This results in better (lower) CRPS and a substantial added value of AAEPS to the global EPS. However, both systems have to little spread and the added value of AAEPS is reduced for longer lead times. For the large scales, i.e. MSLP, AAEPS has better spread-skill relationship than IFS-ENS for shorter lead times, but a larger RMSE and no added value is measured. These

results are similar to what is found for MEPS over Scandinavia (Frogner et al., 2019b).

Running a 1+10 member high-resolution EPS is expensive in terms of computer power. An important question is therefore how many ensemble members are needed to provide AV, e.g. compared to IFSENS. In Figure 3.4.3 and 3.4.4 this is measured in terms of CRPS for T2m and S10m, respectively. Even reducing the AAEPS ensemble size to 1+2 members perform better for T2m than using all IFSENS members during day time and for the shortest lead times. This result indicates that a substantial part of the AV arise from the higher resolution in AAEPS. Furthermore, the CRPS score are very similar when applying 1+6, 1+8 and 1+10 ensemble members in AAEPS indicating that a small AAEPS ensemble also provide AV compared to IFSENS for S10m. For lead times between 12 and 30 h, reducing the AAEPS ensemble size to only 1+2 members keep approximately half of the AV compared to IFSENS. However, for such a small ensemble size there are no AV after ~ 30 h. These analyses show that even a very small ensemble may give AV compared to coarser resolution EPS and the numbers of members should be a part of the consideration when operational configurations are considered.

The 18. March 2018 00 UTC an intense low pressure system was situated just south of Svalbard. The resulting +48 h MSLP and S10m forecasts shown in Figure 3.4.5 illustrates some similarities and differences between AAEPS and IFS-ENS. The MSLP forecast in both systems control runs are very similar and so are also the patterns of high S10m over the ocean. However, over both Svalbard and Norway, higher S10m are seen in AAEPS. This is further reflected in the calculated probabilities (proportion of members to exceed a given threshold in a given grid point) for near gale and strong gale. While the probabilities for near gale in both systems are quite similar over the ocean, they are very different over land. IFS-ENS have almost zero probability for all land points, while AAEPS have clear patterns of high probabilities on the lee-side of the Svalbard topography. For the higher threshold, strong gale, the differences between the system are further enhanced. In general, the behavior of AAEPS looks more like a realistic result of the large scale situation. This is confirmed by looking at the wind speed measurements, for this particular time, at the relatively sheltered observation sites at the lee-side of Svalbard; 14,9 m/s (Ny-Ålesund), 17,7 m/s (Akseløya), 13,6 m/s (Hornsund), 15,4 m/s (Svalbard airport), 14,1 m/s (Isfjorden radio). On the east side of Svalbard the highest measured wind speed was 23,4 m/s at KARL XII island. As in deterministic runs, the higher-resolution AAEPS better capture the high wind speeds and spatial variability than the coarser resolution global IFS-ENS.

Impact of Sea Surface Temperature perturbations

A large part of the AAEPS integration domain is covered by sea. The sea surface temperature (SST), which is kept constant during the integration, is regridded from IFS-HRES, which at this time was based on the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Donlon et al. 2012) from the Met Office. The uncertainties in SST include subgrid scale variability and weaknesses in the SST product and perturbations are done following the random field approach of Bouttier et al. (2016). A comparison of SST from IFS-HRES and Multi-scale Ultra-high Resolution Sea Surface Temperature Analysis (MUR) for the year 2018 show that differences are largest from May to August. The standard deviation of the SST difference between IFS-HRES and MUR averaged over the domain was in the range of half a degree for August to April, but up to threefold during May to August. Based on this we employ smaller SST perturbations than originally used by Bouttier et al. (2016) for our winter period.

On the 24. March a polar meso-scale system developed in the Barents Sea. The system is seen in the analysis from the operational AROME-Arctic in Figure 3.4.6a and is reasonable well forecasted in the AAEPS control run, +36 h ahead. However, Køltzow et al. (2019) found this particular system to be challenging to forecast in some NWP systems and uncertainty are present. The mean MSLP and S10m and their variability over the ensemble members in the Barents Sea illustrate this (Figure 3.4.6). We chose this case to look in more detail on the effect of perturbations in general and SST perturbations in particular. In Figure 3.4.7, the result of SST perturbations are compared with the result of the sum of perturbations of initial conditions, lateral boundary conditions and in the soil. For the synoptic scales (i.e. in MSLP) the SST have only a minor impact, while the sum of the other perturbations have a clear impact on the location of the meso-scale system. The impact of SST perturbations in wind speed are more clearly seen, but only locally connected to the low pressure system. In comparison with the sum of the other perturbations, the impact of SST perturbations are less in both spatial extent and amplitude. This result is as expected, but show that some potential impact are present.

The largest impact of SST perturbations are expected to be found in coastal regions, close to the areas of SST perturbations. Only negligible changes are found in objective verification on other parameters than T2m so we limit ourselves to present results for T2m. In Figure 3.4.8, verification for Svalbard, Norwegian islands and coastal stations are presented separately. A small positive impact (reduced RMSE and small increase in spread) is seen at Svalbard, but the opposite is true for for the Norwegian islands and coastal stations. The decrease in quality for the Norwegian islands and coast is not yet understood and needs further

analysis. It should be noted that SST perturbations are done in areas with little observations available for verification and therefore difficult to verify. Even if some positive results are seen from SST perturbations, the results also show that perturbations of the sea surface is not straightforward. Ideally, also the sea ice temperatures should be perturbed. However, after the introduction of a new sea ice scheme in AROME-Arctic which simulates the sea ice temperatures in a more realistic way (SICE; Batrak et al., 2018, done outside APPLICATE) this is more challenging. It has therefore not been possible to include perturbations of sea ice within the time constraints of APPLICATE, but this will be a part of future research. However, we argue that sea ice temperature perturbations performed in a similar way as SST perturbations will give similar results.

Summary short-range high resolution EPS

This work has been part of the first steps toward an operational high-resolution (AROME-)Arctic EPS, the AAEPS. The results from a winter period are mainly positive with substantial added value compared 1) to the deterministic control run and 2) to the coarser resolution global IFS-ENS. However, AAEPSs are under dispersive for most parameters and further work on the representation of the forecast uncertainty is needed.

In this work we have tested a method by Bouttier et al. (2016) to perturb the sea surface temperature. The results show that this is a non-trivial task and in the present experiments it is not possible to identify an improvement over all regions by standard probabilistic verification. Furthermore, comparison of single members with and without SST perturbations show that little impact can be expected on the synoptic scale (i.e. MSLP), while local changes on smaller scales can be expected (e.g. 10m wind speed). Further work on SST and sea ice temperature perturbations, which build on the presented results here are planned.

The computer demands for AAEPS is approximately 11 times higher than today's operational deterministic AROME-Arctic. Given that such an amount of operational computer power become available in future HPC upgrades an operational implementation of AAEPS will substantially improve the Arctic forecast capabilities. A compromise, if necessary computer power is not available, is to reduce the number of ensemble members. Substantial AV is present compared to IFS-ENS even with smaller ensemble sizes. However, the proper ensemble size should be further investigated. This study recommend the operational implementation of AAEPS, but it should be combined with investment in the communication and distribution part of the weather forecasting chain (e.g. Fundel et al., 2019).

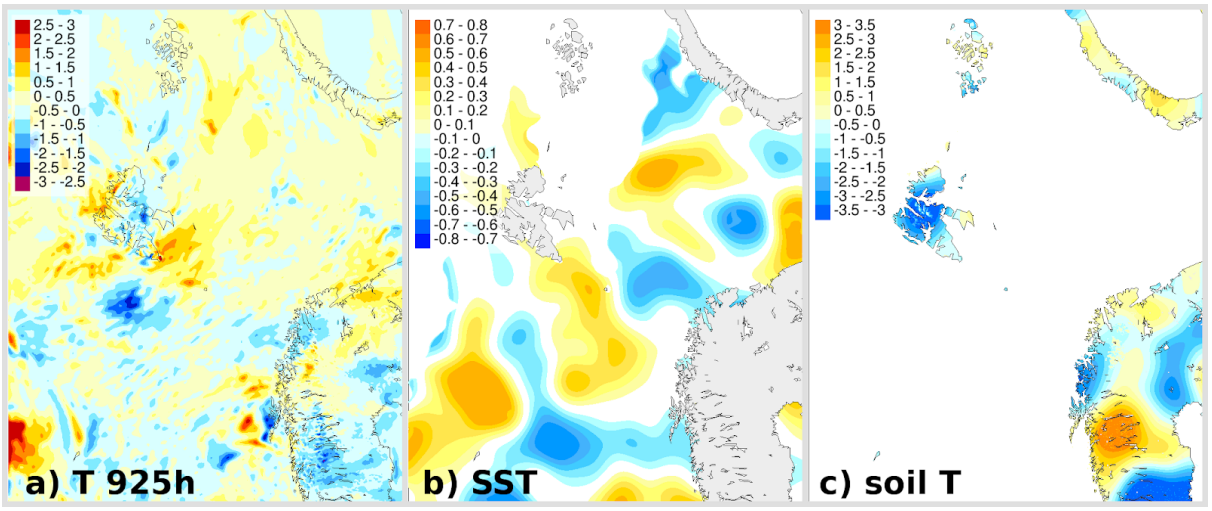


Figure 3.4.1. Example of perturbations. Temperature perturbations [degC] for member 1 in AAEPS for a) atmospheric temperature in 925 hPa (interpolated from model levels), b) sea surface temperature and c) soil top temperature on 8.March 2018 00 UTC.

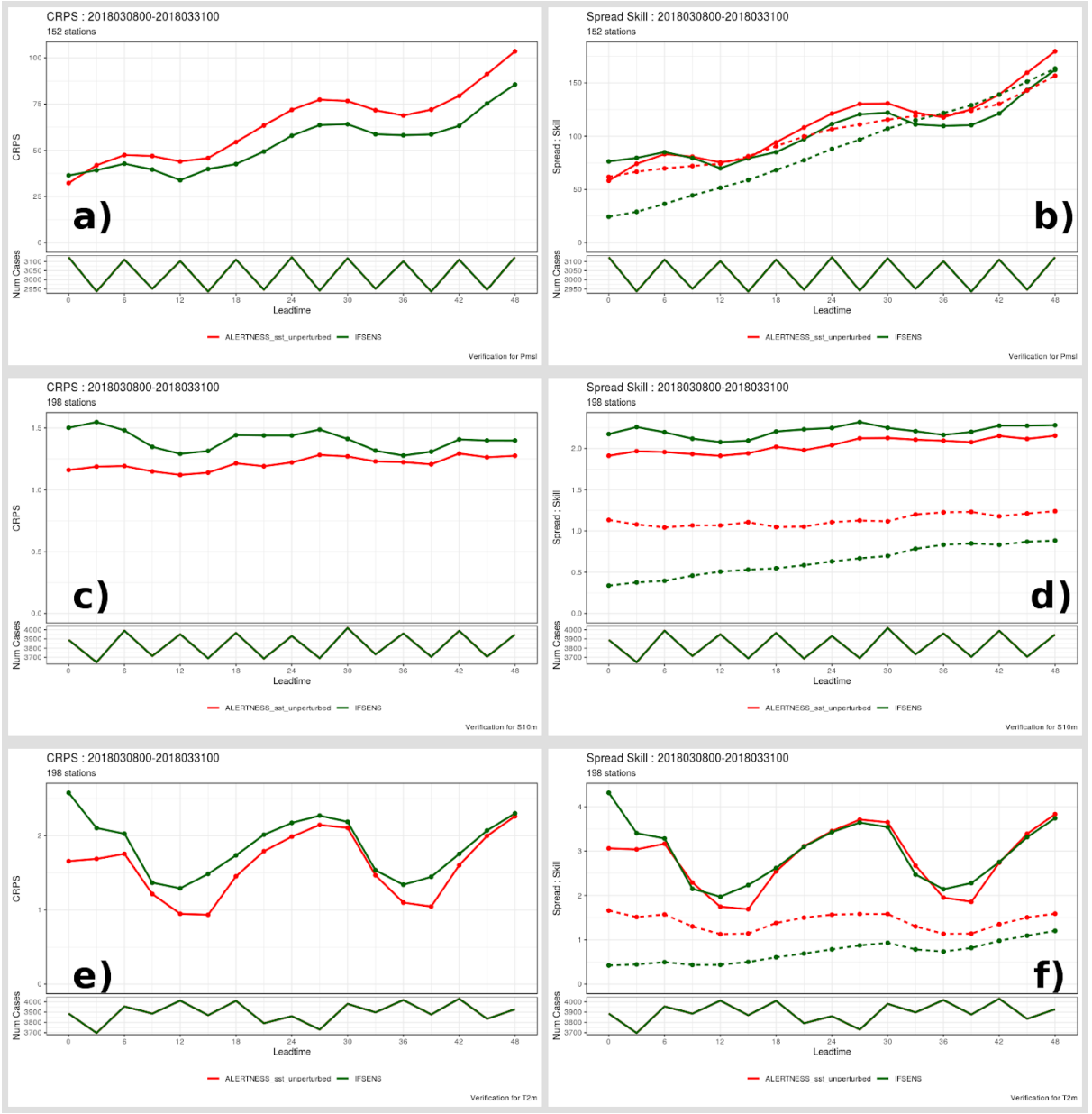


Figure 3.4.2. Probabilistic verification of AAEPS (red) and IFS-ENS (dark green) by a) CRPS for MSLP, b) Spread-skill relation MSLP, c) CRPS for 10m wind speed, d) spread-skill relation for 10m wind speed, e) CRPS for 2m air temperature and f) spread-skill relation for 2m air temperature. Period is 8.March - 31.March 2018 and statistics include all available observation stations.

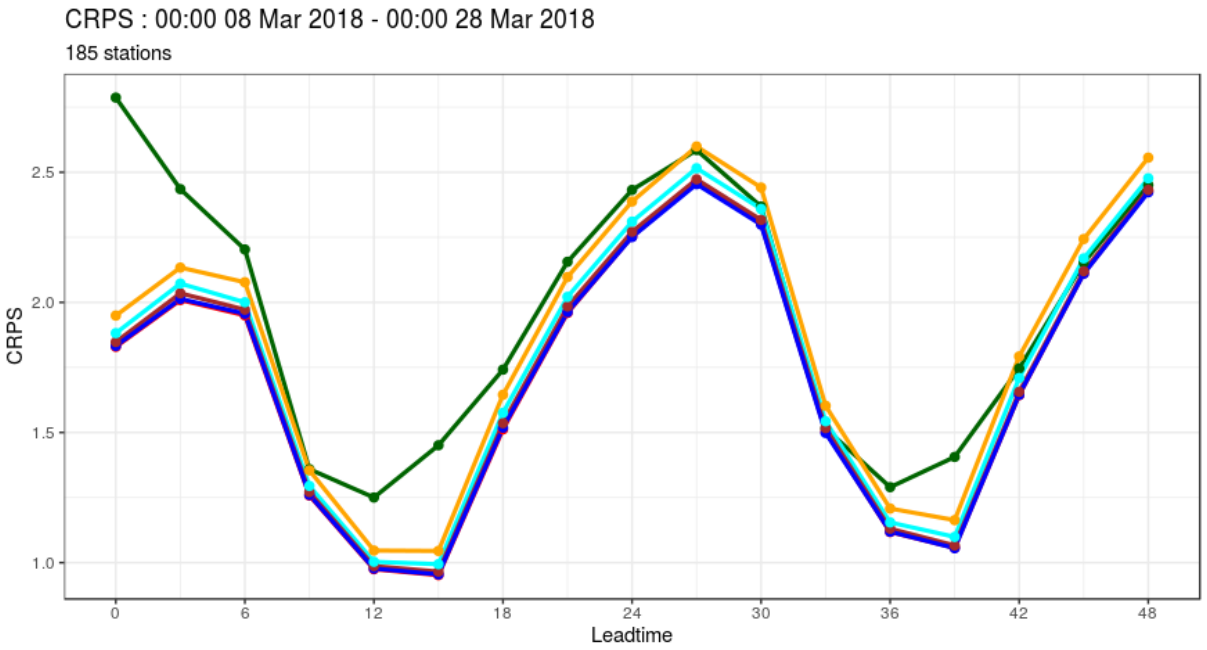


Figure 3.4.3. CRPS for T2m forecasts by IFSENS, 1+50 members (dark green), AAEPS 1+10 members (red), AAEPS 1+8 members (blue), AAEPS 1+6 members (brown), AAEPS 1+4 members (cyan) and AAEPS 1+2 members (orange) as a function of lead time.

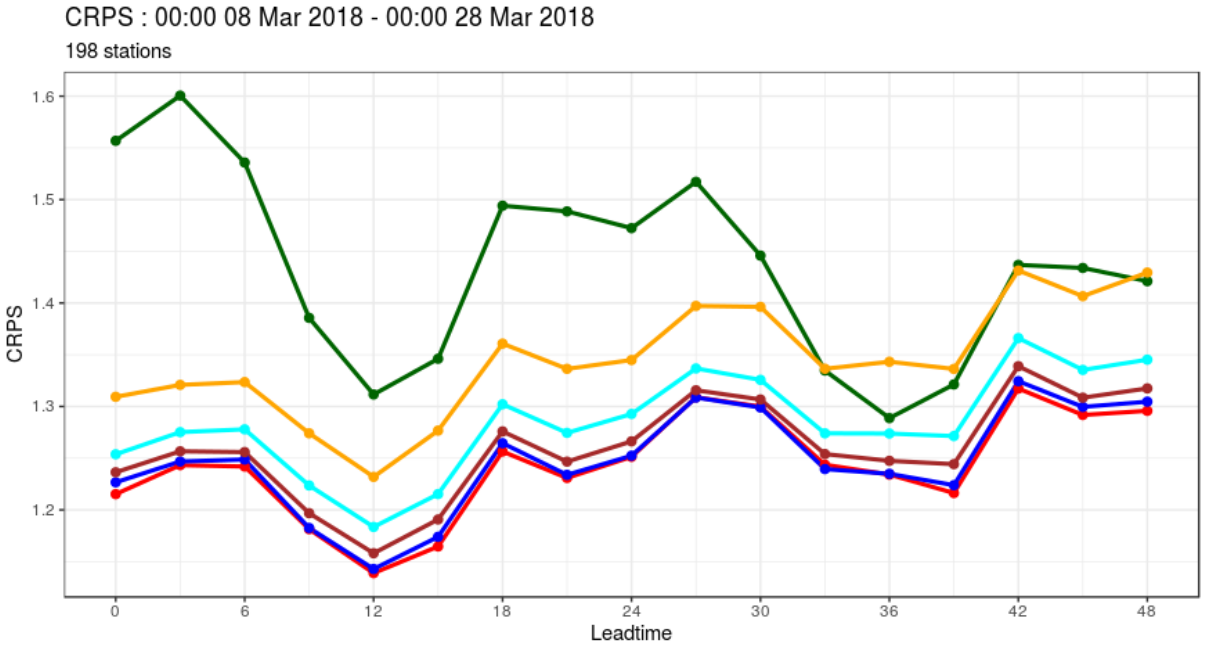


Figure 3.4.4. Same as Figure 3.4.3, but for S10m

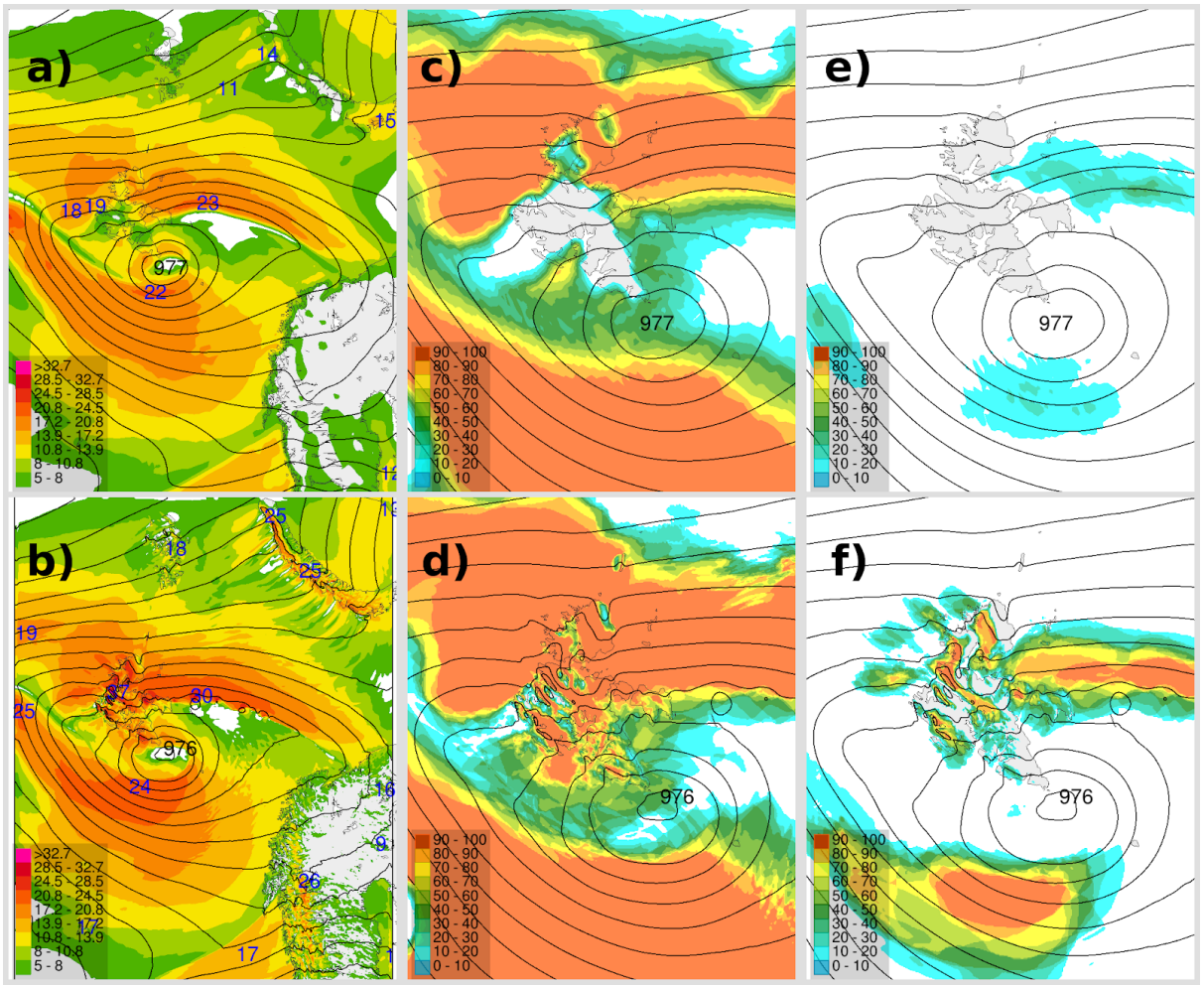


Figure 3.4.5. Mean Sea Level Pressure and 10m wind speed 18. March 2018 00 UTC (+48 h forecasts) from a) IFS-ENS control run and b) AAEPS control run during a low pressure centered just south of Svalbard. Probability of near gale wind speeds from c) IFS-ENS and d) AAEPS and for strong gale e) IFS-ENS and f) AAEPS.

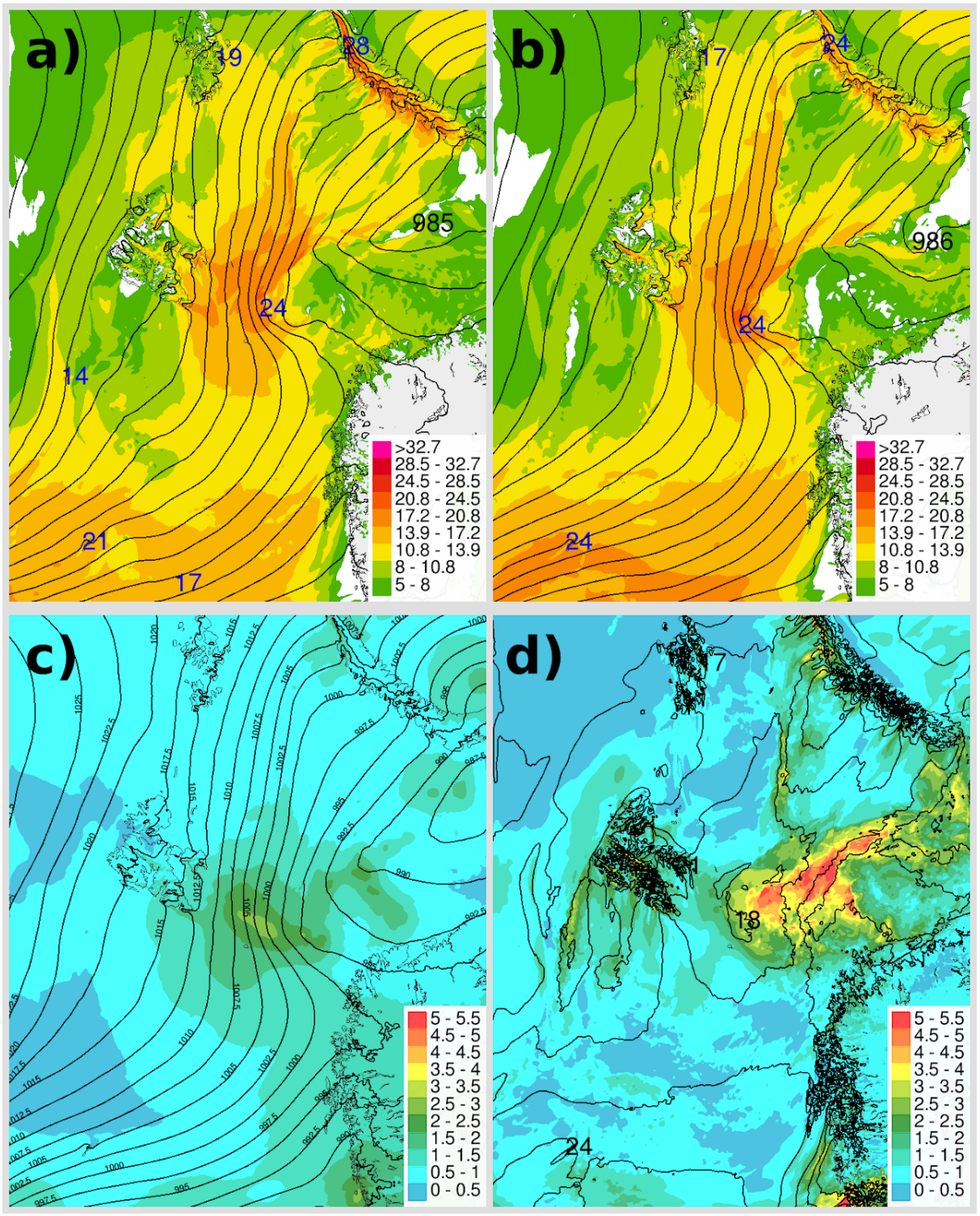


Figure 3.4.6. Meso-scale system develops in the Barents Sea on 24. March 2018 12 UTC; a) MSLP and 10 m wind speed in analysis from operational AROME-Arctic, b) +36h forecasts for MSLP and 10 m wind speed from AAEPS control, c) +36 h MSLP mean and standard deviation in AAEPS and d) +36 h 10 m wind speed mean and standard deviation in AAEPS.

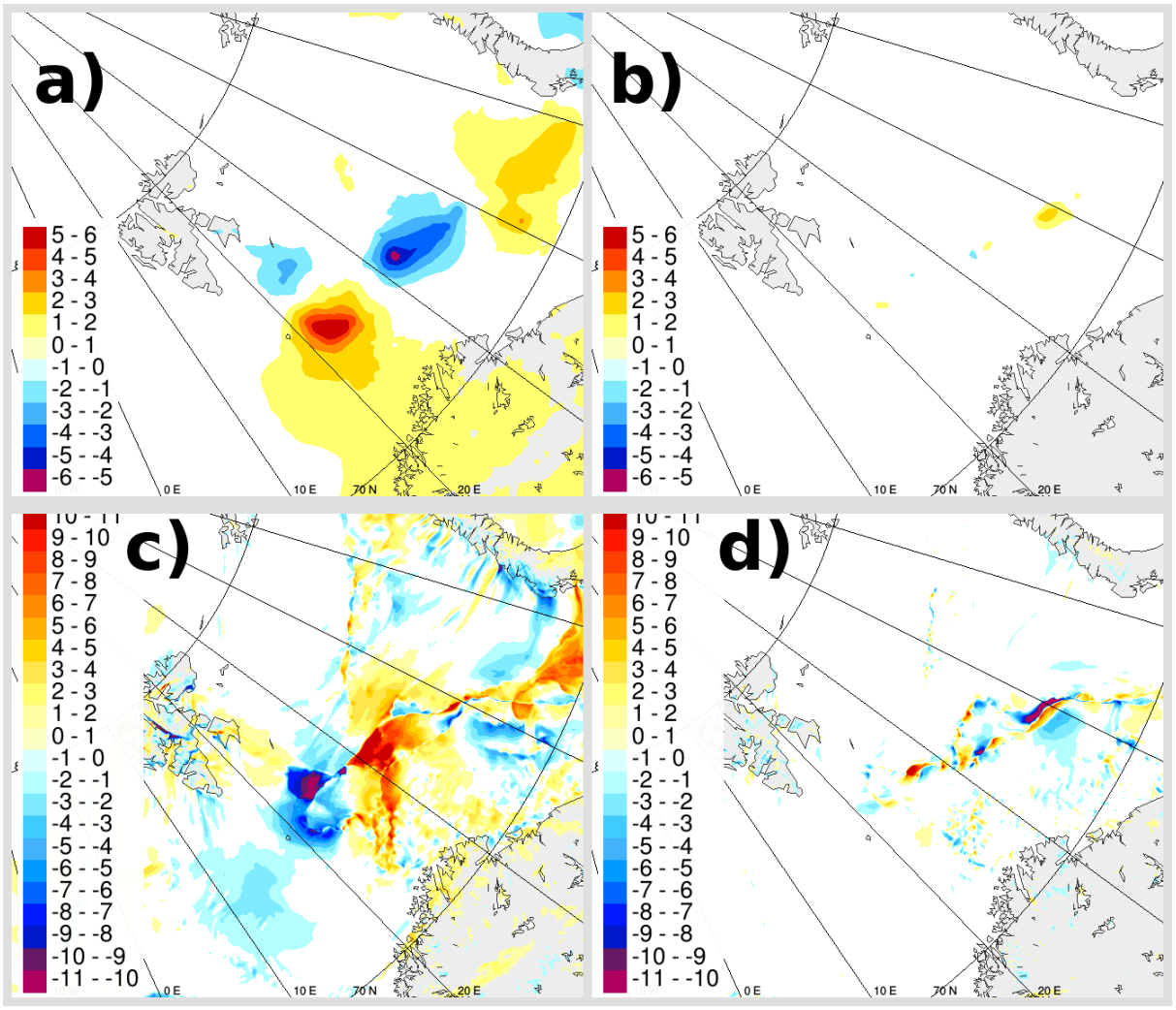


Figure 3.4.7. Impact of perturbations in +36 h forecasts for member 1 in AAEPS; a) changes in MSLP due to initial condition and lateral boundary perturbations, b) changes in MSLP due to sea surface temperature perturbations, c) changes in 10m wind speed due to initial condition, lateral boundary conditions and soil perturbations.

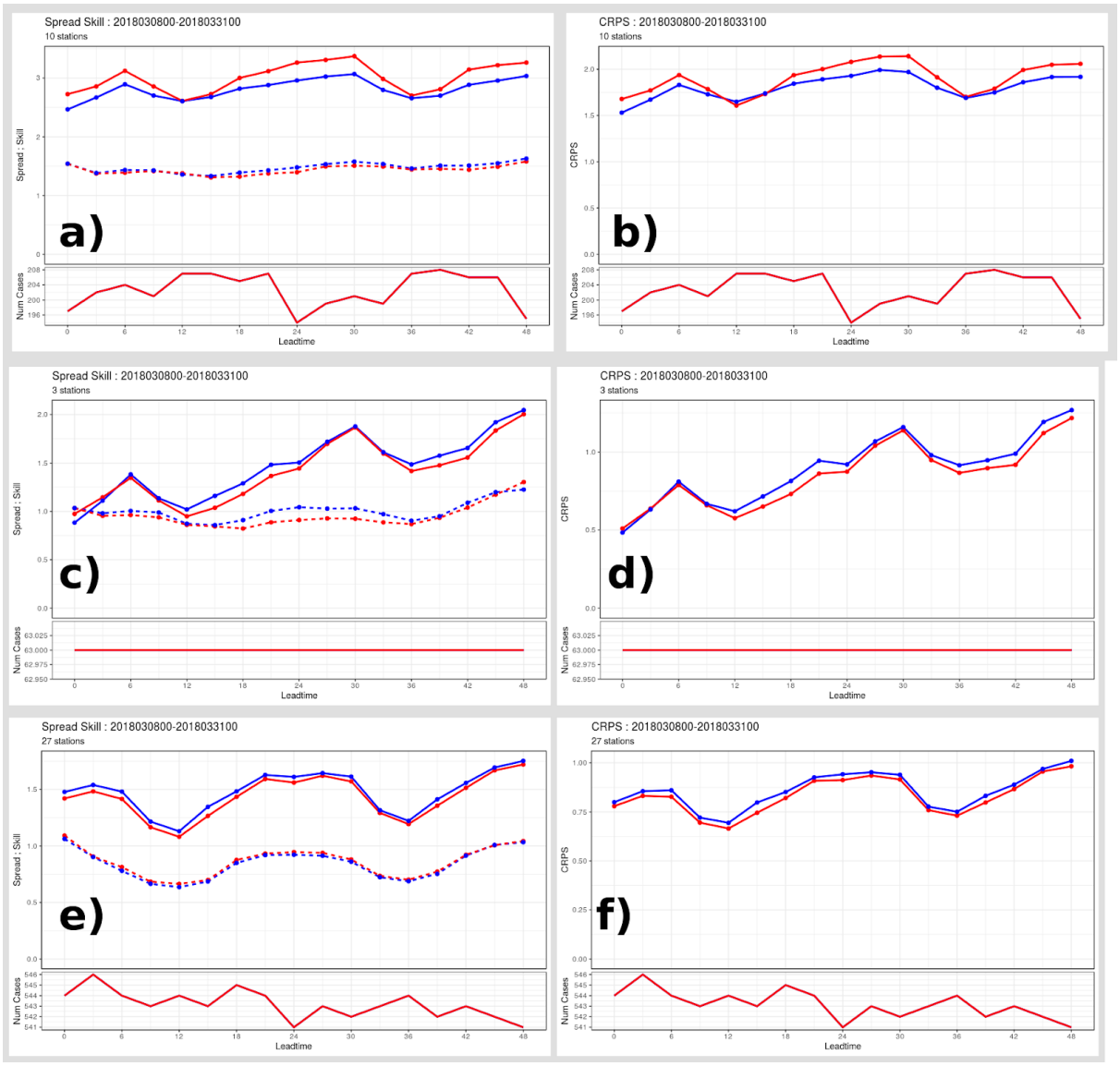


Figure 3.4.8 Impact of SST perturbations. No SST perturbations in red and SST perturbations in blue. T2m verification Spread-skill relation (a, c and e) and CRPS (b, d and f) for Svalbard (a and b), Islands (c and d) and coast stations in northern Norway (e and f).

Impact of the ensemble generation technique on climate predictions with CNRM-CM6

Several groups now resort to stochastic perturbations as a means of generating their ensemble forecasts. Indeed, even some *ad hoc* approaches have proven quite efficient in generating ensemble spread and increasing both the reliability and skill of seasonal predictions (see e.g. Berner et al. 2017 for a review on stochastic methods).

In the framework of the APPLICATE project, stochastic perturbations in the ocean model NEMO 3.6 were activated in the 1° ocean component of CNRM-CM6-1 so as to better represent the uncertainties in the ocean equation of state, following Brankart et al. (2015). The perturbations are applied to temperature and salinity fields as the scalar product between the local T, S gradients and a random walk using an autoregressive AR-1 process.

As many *ad hoc* stochastic perturbation methods, this method implies setting a number of parameters controlling the characteristics of the perturbations, and some tuning is involved. The set of parameters tested in APPLICATE was based on previous work with NEMO 3.6 at 1° resolution, excluding some parameter options that led to a too unstable model. Note however that with respect to the reference CNRM-CM6-1 model used for stream 1, an increase in the instability in the coupled model was found when running the re-forecast experiments, although remaining reasonably low enough so that these ensembles could be run.

Among the parameters to be set, there are the number *N* of random walks, the horizontal *stdxy* and vertical *stdz* standard deviation of the random walks (in grid points), the time correlation *tcor* of the random walks, and a limitation factor to avoid too high perturbations. Seasonal re-forecasts with CNRM-CM6-1 were run with the following values:

N= 4 independent random walks, *stdxy* = 0.7 grid points, *stdz* = 0.2 grid points, and *tcor* = 480 timesteps (corresponding to 10 days). The limitation factor was set to 2 instead of 3 by default.

The choice of these parameters was guided by both stability issues and previous experience with NEMO 3.6 on the eORCA1° grid at CERFACS and Mercator Ocean International with including these perturbations (Sanchez-Gomez et al., personal communication). Ensemble generation was therefore ensured by using slightly different atmosphere and ocean initial conditions (as for stream 1 experiments with CNRM-CM6-1), on top of which these random perturbations of the ocean equation of state came into play.

We ran a 10-member ensemble initialized in November over the 1993-2014 re-forecast period, for up to 6 months forecast time. This ensemble, called

CNRM-CM6 STO, is compared to the reference stream 1 re-forecast with the same model called CNRM-CM6-1. For the latter we select only the first 10 members for a direct comparison (exact same initial conditions for both ensembles).

Figure 3.4.9 shows the evolution of ensemble spread and root mean square error for surface temperature over the North Atlantic – Europe region according to forecast time. The reference data used is ERA-Interim reanalysis data; we find a very limited (non-significant) impact on the ensemble spread. This is also true for other surface parameters (such as sea level pressure) and other regions of the globe, suggesting that on average, the settings chosen for the stochastic ocean perturbations do not impact the atmospheric fields in terms of ensemble spread.

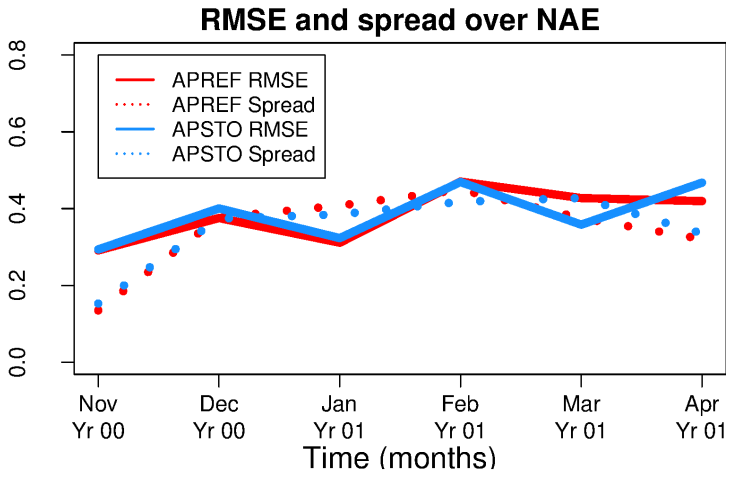
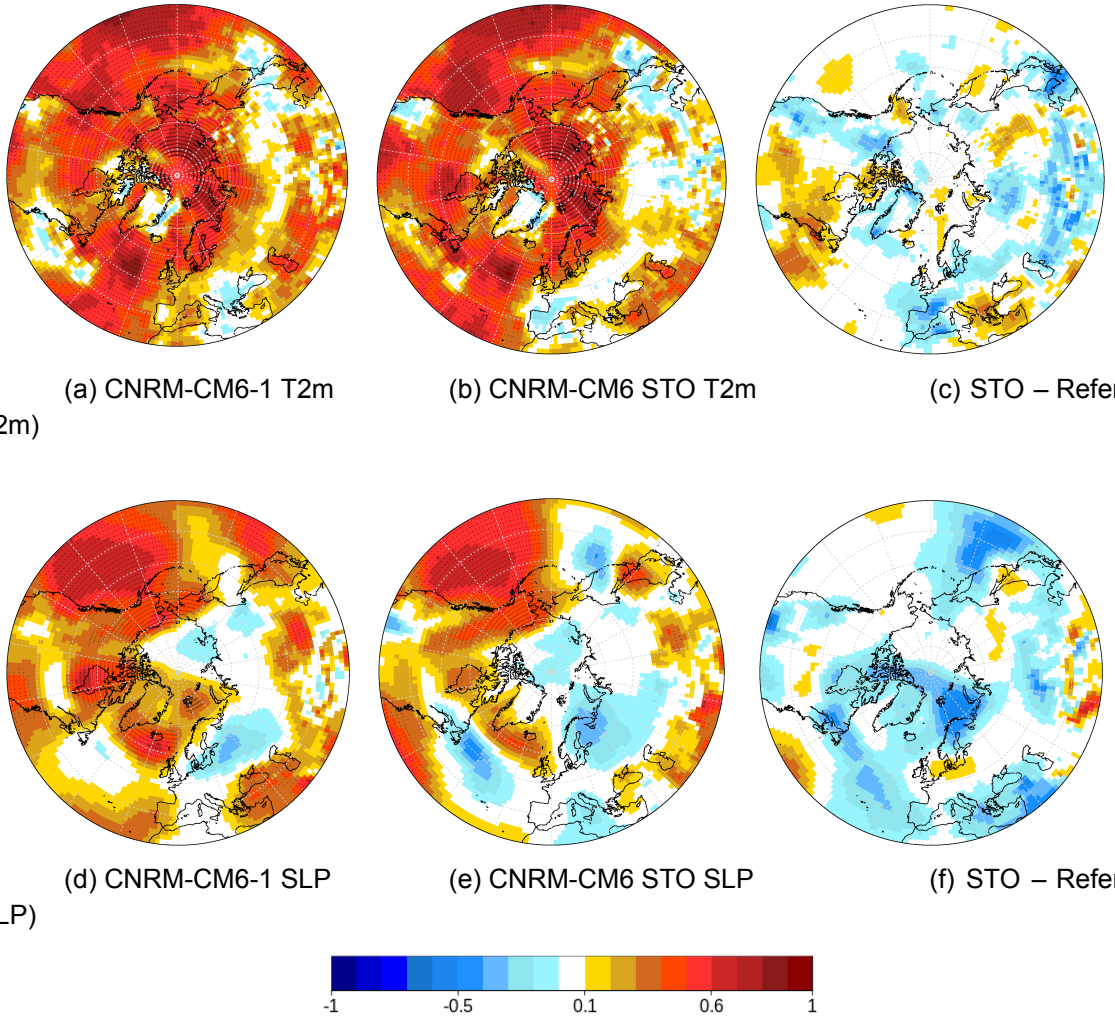


Figure 3.4.9: Evolution of surface temperature RMSE and spread (measured as the standard deviation around the ensemble mean) according to forecast time in CNRM-CM6-1 (APREF) and CNRM-CM6 STO (APSTO) re-forecasts initialized in November 1993-2014. Reference data is ERA-Interim.

Stochastic perturbations have been shown in the past to incur shifts in the model mean state. We therefore examined the differences between both runs in terms of mean biases for near-surface atmosphere variables, and for sea ice concentration. For atmospheric variables (temperature, sea level pressure, precipitation), virtually no differences in mean bias were found; this implies that the stochastic perturbations did not shift the mean model climate at the ocean-atmosphere interface. For sea ice, some small shifts in the sea ice concentration biases start to appear in the later months of the forecast, although these are very small compared to interannual variability (not shown).

We then assessed the changes in skill (gridpoint correlation) for atmospheric variables and sea ice concentration, and in IIEE over regional seas (based on sea ice concentration exceeding a 0.15 threshold). Figure 3.4.10 shows gridpoint correlation with ERA-Interim for DJF re-forecasts initialized in November: if for near-surface

temperature some regions with improvements and some shifts in patterns of skill over the mid-latitudes can be found, the introduction of stochastic perturbations clearly degrades the skill for sea level pressure with many regions exhibiting lower skill than in the reference run (see figure 3.4.10 f).



(a) CNRM-CM6-1 T2m (b) CNRM-CM6 STO T2m (c) STO – Reference (T2m)

(d) CNRM-CM6-1 SLP (e) CNRM-CM6 STO SLP (f) STO – Reference (SLP)

Fig 3.4.10: Gridpoint correlation with ERA-Interim for DJF 2-meter temperature (top) and sea-level pressure (bottom) in 10-member re-forecasts initialized in November using CNRM-CM6-1 (left), CNRM-CM6 STO (center); (right) difference in correlation between CNRM-CM6 STO and CNRM-CM6-1.

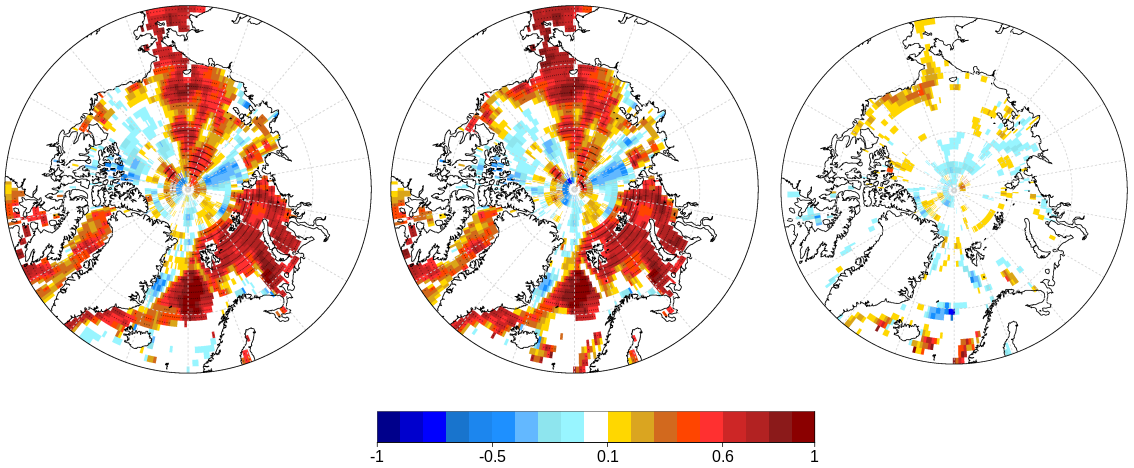
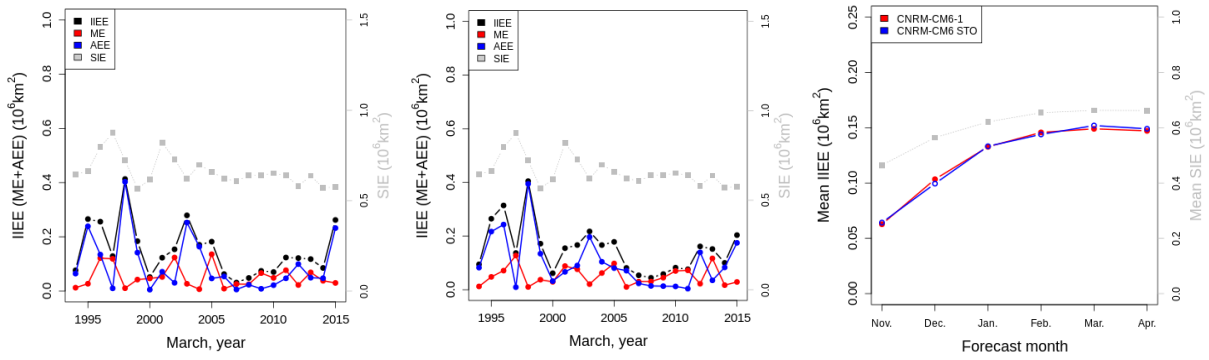
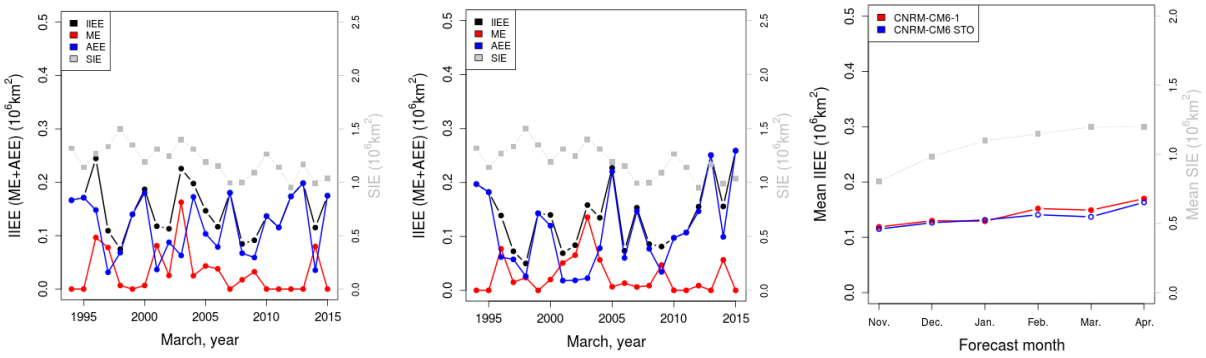


Fig 3.4.11: Gridpoint correlation with NSIDC for DJF sea ice concentration (SIC) in 10-member re-forecasts initialized in November using CNRM-CM6-1 (a), CNRM-CM6 STO (b); (c) difference in correlation between CNRM-CM6 STO and CNRM-CM6-1.

For boreal winter sea ice concentration (Fig. 3.4.11) we find some improvements in the Beaufort sea and in parts of the GIN seas, although these are quite limited; however this leads to an extension of the area where significant skill is found. Some other areas (with limited skill) show some degradation with the stochastic perturbations.



(a) CNRM-CM6-1 (Greenland Sea) (b) CNRM-CM6 STO (Greenland Sea) (c) IIEE vs time (Greenland Sea)



(d) CNRM-CM6-1 (BK Seas) (e) CNRM-CM6 STO (BK Seas) (f) IIEE vs time (BK Seas)

Fig 3.4.12: IIEE and decomposition for March in 10-member re-forecasts initialized in November using CNRM-CM6-1 (a,d), CNRM-CM6 STO (b,e); (c,f) evolution of IIEE with forecast time. Top row (a-c) is for Greenland sea, bottom row (d-f) is for Barents and Kara seas.

In terms of reproducing anomalies in the sea ice edge position, the stochastic perturbations seem to have, as could be expected from previous results, a very limited effect. Figure 3.4.12shows the evaluation of Integrated Ice Edge Error (IIEE) with respect to NSIDC for winter over an extended Greenland sea region (top) and the Barents and Kara seas (bottom). For Greenland sea, there are very few changes between both model versions. The evolution of average IIEE with forecast time is very similar in both cases (figure 3.4.12 c). For the Barents and Kara sector, the IIEE in the CNRM-CM6 STO experiment is slightly lower on average in March (figure 3.4.12 f), but year-to-year values show a very high variability, and larger errors in the later years of the re-forecast period. We also evaluated over the Arctic basin the probabilistic version of the IIEE, the Spatial Probability Score (see D5.2). The CNRM-CM6 STO re-forecasts have a slightly higher (ie, worse) SPS value for all forecast times from November to April (not shown).

The original intention in this task was to combine this approach with stochastic perturbations (SPPT) in the atmospheric component ARPEGE. ARPEGE was found to be highly unstable with SPPT (at least SPPT in combination with the new physics scheme in ARPEGE v6.3) and therefore no seasonal hindcasts were run with both stochastic methods activated.

Given the degradation in skill noted for atmospheric variables and lack of improvement in model mean state, we conclude that the settings tested for the stochastic equation of state in NEMO3.6 are improper for use in a seasonal forecasting framework. It could well be that these perturbations need more integration time to be efficient. Other attempts to tune the method with different sets of parameters for seasonal timescales (with possibly stronger perturbations and higher decorrelation spatial scales) led to high instabilities in the – otherwise very stable –

coupled model. Efficient tuning would require a number of experiments and results so far are not very convincing. We therefore recommend not to activate these parameterizations as such in the stream 2 runs with CNRM-CM.

4. CONCLUSIONS AND OUTLOOK

In this report, possible enhancements in Arctic prediction skill due to improved process description, model resolution and system configurations are investigated.

For the short-range (~days) predictions it is shown that the introduction of the GELATO sea ice model in Meteo France systems AROME and ARPEGE give a substantial improvement in forecast skill. The more realistic sea ice description show a pronounced improvement in near-surface temperature close to sea ice covered areas (e.g. at Svalbard). The results highly recommend the GELATO sea ice model to be used instead of climatology to describe sea ice properties in operational/ APPLICATE stream 2 configurations of AROME and ARPEGE.

In the ECMWF IFS, the inclusion of a multi-layer snow scheme leads to an improvement in short- and medium range weather forecasts of near-surface temperature and snow depth by cooling the near-surface and reduce the mean bias. In addition, two new diagnostics of skin-temperature sensitivity are presented showing that the multi-layer snow scheme has increased the overall sensitivity of the surface temperature to changes in incoming radiation, but that this has been done by making the sensitivity to the ground heat flux too high. This indicates that other processes in the boundary layer are contributing to this lack of sensitivity to radiative forcing. However, the overall improvement in forecasts of snow depth and 2m temperature motivates the inclusion of multi-layer snow in the APPLICATE stream-2 experiments.

The description of the sea ice thickness distribution (ITD) in seasonal prediction systems is evaluated with respect to the sea ice mean state (Massonet et al., 2019, UCLouvain) and sea ice variability (Moreno-Chamarro et al., In preparation, BSC). It is shown that the description of ITD has a significant influence on the model mean state, e.g. winter ice volumes increase with the number of ice thickness categories due to an increase in basal ice growth rates. However, it is also found that the current default discretization of the NEMO3.6-LIM3 model is sufficient for large-scale present-day climate applications for the mean state. Further analysis on sea ice concentration variability are inconclusive regarding the optimal configuration or number of sea ice thickness categories to be used beyond that one category tends to perform the worst. Altogether, both analysis suggest that there is no evident benefit from including additional categories to the default configuration (with 5 levels).

Three different melt pond schemes have been tested in NEMO3.6-LIM3 and compared with observations of sea ice concentration and reanalysis data sets for sea ice volume (BSC). The results suggest that including melt pond parametrizations decreases model realism. A similar experiment was also performed with the sea ice model Gelato v6 in the CNRM-CM6 system (CNRM). However, also in this experiment a relatively small impact of turning on melt pond schemes is noticed. The small sensitivity and lack of improved predictions using a melt pond scheme may potentially be related to the tuning of the sea ice albedo which in case of no explicit melt pond description, is reduced to account for the missing melt ponds. Based on the results presented it is therefore not recommended to include more realistic melt pond description directly without any further tuning of the sea ice models.

Variations in configuration (e.g. initialization, resolution, deterministic or ensemble approach, domain size and location and lead time) of Arctic regional high-resolution short-range weather forecasting systems have been evaluated with AROME-Arctic (MET-Norway) and MF-AROME (CNRS and Meteo France). Qualitatively the results confirm what is reported elsewhere showing 1) the importance of upper-air and surface assimilation, 2) the added value of further increase in resolution, 3) the added value of an EPS approach (compared with deterministic), 4) that small integration domains may deteriorate the forecast quality (but also improve), and 5) the importance of LBC quality increase with smaller integration domains. It should also be noted that the different configuration choices impact different parameters differently. However, the presented analysis also add quantitative information on the relative importance of the different choices. In the specific settings tested here, the initialization of the surface is more important than upper-air assimilation for forecast quality of near-surface parameters. The added value by a further increase in resolution (2,5 km 60/65 vertical layers to 1,25 km and 90 vertical layers including additional layers in the boundary layer) is present, but not as large as by introducing a regional high-resolution EPS with similar computational costs. The regional high-resolution Arctic EPS add also substantial value compared to the global coarser resolution IFSENS. Both further increase in resolution and in particular the introduction of a high-resolution EPS is therefore recommended.

The impact of increased oceanic (1° to 1/4°) and atmospheric resolution in CNRM-CM6 (1.4° increased to 0.5°) and EC-Earth3.2 (0.7° to 0.35°) for seasonal predictions has been evaluated. In CNRM-CM6 an improvement is found in sea ice concentration over the Atlantic and Pacific sectors leading to an overall improved integrated sea ice extent. In EC-Earth3.2 a possible degradation in skill in the representation of sea ice is identified. Possibly, this is due to an unsatisfactory tuning of the higher resolution version. However, differences in high and low resolution runs in EC-Earth can also be attributed to the differences in the difference in origin of initial conditions (ERA5 vs ERA-Interim) and interpretation is further hampered by the

limited number of members performed (only 10) in the high resolution runs. Further investigation into the origin of improvements and degradations in skill found, with fully consistent setups (e.g. initialized with the same products) and with sufficiently large ensemble sizes, is therefore needed before concluding on the interest in increasing resolution for the seasonal prediction models considered in this study. Other factors such as identifying the best initial conditions and performing an adequate model tuning are key in improving forecast quality.

It has been found that the representation of the Arctic ocean hydrography and the Arctic sea ice in climate predictions by AWI-CM, can be improved by increasing the resolution of the ocean-sea-ice component to 4.5 km in ocean-sea-ice only simulations. However, in the full coupled system biases in the atmospheric part destroy the achieved improvement in the ocean part.

For seasonal forecasting stochastic perturbations of temperature and salinity has also been tested in an ensemble with the ocean model NEMO 3.6 in CNRM-CM6-1. The original intention was to combine the ocean perturbations with stochastic perturbations (SPPT) in the atmospheric component (ARPEGE). However, ARPEGE was found to be highly unstable with SPPT, and SPPT could not be applied in these experiments. The impact of perturbations in the ocean model is somewhat neutral on the surface ocean and sea ice at the seasonal time scale, but a deterioration in the atmosphere is seen, in particular for MSLP. A recommendation for APPLICATE stream 2 is therefore not to activate these perturbations scheme as more work, e.g. on tuning is needed.

The results presented in this report points toward some ways to enhance future prediction systems, while other results indicate that more work is needed on specific topics. However, the work contributes to the next generation of weather and climate prediction systems and thereby to one of the objectives of the work package. Some of the recommendations will further be a part of the APPLICATE stream 2. APPLICATE stream 2 are described in more detail in Milestone 3 and will, when compared to stream 1, take part in assessing the added value of the APPLICATE project.

5. REFERENCES

Arduini, G. et al.: Impact of a multi-layer snow scheme on near-surface weather forecasts, JAMES, in review.

Bader, J., Mesquita, M. D., Hodges, K. I., Keenlyside, N., Østerhus, S., & Miles, M. (2011). A review on Northern Hemisphere sea-ice, storminess and the North Atlantic Oscillation: Observations and projected changes. Atmospheric Research, 101(4), 809–834.

Barthélemy, A., Goosse, H., Fichefet, T., and Lecomte, O.: On the sensitivity of Antarctic sea ice model biases to atmospheric forcing uncertainties, Clim. Dynam., 51, 1585–1603, https://doi.org/10.1007/s00382-017-3972-7, 2017.

Batte, L. & co-authors, The APPLICATE project Deliverable No. 5.2, "Strengths and limitations of state-of-the-art weather and climate prediction systems in the Arctic".

Bauer, P., Magnusson, L., Thépaut, J.-N. and Hamill, T. M.: Aspects of ECMWF model performance in polar areas, Q. J. R. Meteorol. Soc., 142(695), 583–596, doi:10.1002/qj.2449, 2016.
Bauer P, Thorpe A, Brunet G. 2015 The quiet revolution of numerical weather prediction. Nature 525, 47–55.

Batrak, Y., Kourzeneva, E., and Homleid, M., 2018: Implementation of a simple thermodynamic sea ice scheme, SICE version 1.0-38h1, within the ALADIN–HIRLAM numerical weather prediction system version 38h1. Geosci. Model Dev. 11, 3347–3368, https://doi.org/10.5194/gmd-11-3347-2018

Berner et al. (2017) Towards a new view of weather and climate models. B. Am. Meteorol. Soc., 565-587, doi:10.1175/BAMS-D-15-00268.1

Bouttier, F., Raynaud, L., Nuissier, O. and Ménétrier, B. (2016) Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. Quarterly Journal of the Royal Meteorological Society, 142, 390–403. https://doi.org/10.1002/qj.2622.

Bitz, C. M., Holland, M. M., Weaver, A. J., and Eby, M.: Simulating the ice-thickness distribution in a coupled climate model, J. Geophys. Res., 106, 2441–2463, https://doi.org/10.1029/1999JC000113, 2001.

Brankart et al. (2015) A generic approach to explicit simulation of uncertainty in the NEMO ocean model. Geosci. Model Dev., 8, 1285–1297, doi: 10.5194/gmd-8-1285-2015

Cavalieri, D. J., Parkinson, C. L., Gloersen, P., & Zwally, H. J. (1996). Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. National Snow and Ice Data Center, Boulder, CO, USA.

Charrad, M., Ghazzali, ., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, 61(6), 1–36.

Chevallier, M., Smith, G. C., Dupont, F., Lemieux, J.-F., Forget, G., Fujii, Y., Hernandez, F., Msadek, R., Peterson, K. A., Storto, A., Toyoda, T., Valdivieso, M., Vernieres, G., Zuo, H., Balmaseda, M., Chang, Y.-S., Ferry, N., Garric, G., Haines, K., Keeley, S., Kovach, R. M., Kuragano, T., Masina, S., Tang, Y., Tsujino, H., and Wang, X.: Intercomparison of the Arctic sea ice cover in global ocean–sea ice reanalyses from the ORA-IP project, Clim. Dynam., SI, 1–30, https://doi.org/10.1007/s00382-016-2985-y, 2016.

Chikhar, K. and P. Gauthier, 2017: Impact of Lateral Boundary Conditions on Regional Analyses. Mon. Wea. Rev., 145, 1361–1379, https://doi.org/10.1175/MWR-D-16-0245.1

Close, S., Houssais, M. N., & Herbaut, C. (2017). The Arctic winter sea ice quadrupole revisited. Journal of Climate, 30(9), 3157–3167.

Davies, T. (2014), Lateral boundary conditions for limited area models. Q.J.R. Meteorol. Soc., 140: 185-196. doi:10.1002/qj.2127

Doblas-Reyes, F., Acosta Navarro, J., Acosta, M., Bellprat, O., Bilbao, R., Castrillo, M., ... & Prodhomme, C. (2018). Using EC-Earth for climate prediction research, In ECMWF Newsletter (ECMWF, 2018).

Dussin, R., Barnier, B., Brodeau, L., & Molines, J. M. (2016). DRAKKAR FORCING SET DFS5.

Ebisuzaki W and Kalnay E. (1991) Ensemble experiments with a new lagged average forecasting scheme. WMO Research Activities in Atmospheric and Oceanic Modelling. Report 15, 6.31–6.32. Geneva, Switzerland: WMO.

Eicken, H., Tucker, W., & Perovich, D. (2001). Indirect measurements of the mass balance of summer Arctic sea ice with an electromagnetic induction technique. Annals of Glaciology, 33, 194-200, doi:10.3189/172756401781818356

Essery, R., Kontu, A., Lemmetyinen, J., Dumont, M. and Ménard, C. B.: A 7-year dataset for driving and evaluating snow models at an Arctic site (Sodankylä, Finland), Geosci. Instrum. Methods Data Syst., 5(1), 219–227, doi:10.5194/gi-5-219-2016, 2016.

Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn, 2011: Regional Climate Models Add Value to Global Model Data: A Review and Selected Examples. Bull. Amer. Meteor. Soc., 92, 1181–1192, https://doi.org/10.1175/2011BAMS3061.1

EUMETSAT Ocean and Sea Ice Satellite Application Facility. Global sea ice concentration reprocessing dataset 1978-2015 (v1.2, 2015), [Online]. Norwegian and Danish Meteorological Institutes.

Flocco, D., & Feltham, D. L. (2007). A continuum model of melt pond evolution on Arctic sea ice. Journal of Geophysical Research: Oceans, 112(C8).

Flocco, D., Feltham, D. L., & Turner, A. K. (2010). Incorporation of a physically based melt pond scheme into the sea ice component of a climate model. Journal of Geophysical Research: Oceans, 115(C8).

Frogner, I., U. Andrae, J. Bojarova, A. Callado, P. Escribà, H. Feddersen, A. Hally, J. Kauhanen, R. Randriamampianina, A. Singleton, G. Smet, S. van der Veen, and O. Vignes, 2019a: HarmonEPS - the HARMONIE ensemble prediction system. Wea. Forecasting, 0, https://doi.org/10.1175/WAF-D-19-0030.1

Frogner, I–L, Singleton, AT, Køltzow, MØ, Andrae, U. Convection–permitting ensembles: Challenges related to their design and use. Q J R Meteorol Soc. 2019b; 145 (Suppl. 1): 90– 106. https://doi.org/10.1002/qj.3525

Foken, T.: The Energy Balance Closure Problem: An Overview, Ecological Applications, 18(6), 2008.

Fundel, VJ, Fleischhut, N, Herzog, SM, Göber, M, Hagedorn, R. Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end–users. Q J R Meteorol Soc. 2019; 145 (Suppl. 1): 210– 231. https://doi.org/10.1002/qj.3482

Hagelin, S., Son, J., Swinbank, R., McCabe, A., Roberts, N.M. and Tennant, W., 2017, The Met Office convective–scale ensemble, MOGREPS–UK. Quarterly Journal of the Royal Meteorological Society, 143, 2846–2861. https://doi.org/10.1002/qj.3135.
Haiden et al.: Addressing biases in near-surface forecasts, ECMWF Newsletter 157 (2018)

Holland, M. M., Bitz, C. M., and Tremblay, B.: Future abrupt reductions in the summer Arctic sea ice, Geophys. Res. Lett., 33, L23503, https://doi.org/10.1029/2006GL028024, 2006.

Holland, M.M., D.A. Bailey, B.P. Briegleb, B. Light, and E. Hunke, 2012: Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice. J. Climate, 25, 1413–1430, https://doi.org/10.1175/JCLI-D-11-00078.1

Hou, D., Kalnay, E. and Droegemeier, K.K. (2001) Objective verification of the SAMEX'98 ensemble forecasts. Monthly Weather Review, 129, 73–91. https://doi.org/10.1175/1520–0493(2001)129<0073:OVOTSE>2.0.CO;2.

Hunke, E. C.: Thickness sensitivities in the CICE sea ice model, Ocean Modell., 34, 137–149, https://doi.org/10.1016/j.ocemod.2010.05.004, 2010.

Hunke, E. C.: Sea ice volume and age: Sensitivity to physical parameterizations and thickness resolution in the CICE sea ice model, Ocean Modell., 82, 45–59, https://doi.org/10.1016/j.ocemod.2014.08.001, 2014.

Jia et al. (2015) Improved seasonal prediction of temperature and precipitation over land in a high resolution GFDL climate model. J. Climate, 28, 2044–2062, doi:10.1175/JCLI-D-14-00112.1.

Jung, T., and Coauthors, 2016: Advancing polar prediction capabilities on daily to seasonal time scales. Bull. Amer. Meteor. Soc., 97, 1631–1647, https://doi.org/10.1175/BAMS-D-14-00246.1.

Kangas, M., Rontu, L., Fortelius, C., Aurela, M. and Poikonen, A.: Weather model verification using Sodankylä mast measurements, Geosci. Instrum. Methods Data Syst., 5(1), 75–84, doi:https://doi.org/10.5194/gi-5-75-2016, 2016.

Kristiansen, J. , Sørland, S. L., Iversen, T. , Bjørge, D. and Køltzow, M. Ø. (2011), High−resolution ensemble prediction of a polar low development. Tellus A, 63: 585-604. doi:10.1111/j.1600-0870.2010.00498.x

Køltzow, M., B. Casati, E. Bazile, T. Haiden, and T. Valkonen, 2019: An NWP Model Intercomparison of Surface Weather Parameters in the European Arctic during the Year of Polar Prediction Special Observing Period Northern Hemisphere 1. Wea. Forecasting, 34, 959–983, https://doi.org/10.1175/WAF-D-19-0003.1

Laprise, R., de Elía, R., Caya, D. et al. Challenging some tenets of Regional Climate Modelling, Meteorol Atmos Phys (2008) 100: 3. https://doi.org/10.1007/s00703-008-0292-9

Leppänen, L., Kontu, A., Hannula, H.-R., Sjöblom, H. and Pulliainen, J.: Sodankylä manual snow survey program, Geosci. Instrum. Methods Data Syst., 5(1), 163–179, doi:https://doi.org/10.5194/gi-5-163-2016, 2016.

Manubens, N., Caron, L. P., Hunter, A., Bellprat, O., Exarchou, E., Fučkar, N. S., ... & Batté, L. (2018). An R package for climate forecast verification. Environmental Modelling & Software, 103, 29–42.

Madec, G. and NEMO System Team: NEMO Ocean Engine, Issue 27, Scientific Notes of Climate Modelling Center, Institut Pierre-Simon Laplace (IPSL), ISSN 1288-1619, 2015.

Massonnet, F., Fichefet, T., Goosse, H., Vancoppenolle, M., Mathiot, P., and König Beatty, C.: On the influence of model physics on simulations of Arctic and Antarctic sea ice, The Cryosphere, 5, 687–699, https://doi.org/10.5194/tc-5-687-2011, 2011.

Massonnet, F., Vancoppenolle, M., Goosse, H., Docquier, D., Fichefet, T., and Blanchard-Wrigglesworth, E.: Arctic sea-ice change tied to its mean state through thermodynamic processes, Nat. Clim. Change, 8, 599–603, https://doi.org/10.1038/s41558-018-0204-z, 2018.

Massonnet, F., Barthélemy, A., Worou, K., Fichefet, T, Vancoppenolle, M., Rousset, C. and Moreno-Chamarro, E.: On the discretization of the ice thickness distribution in the NEMO3.6-LIM3 global ocean–sea ice model, Geosci. Model Dev., 12, 3745–3758, https://doi.org/10.5194/gmd-12-3745-2019, 2019.

Meier, W. N. and Stewart, J. S.: Assessing uncertainties in sea ice extent climate indicators, Environ. Res. Lett., 14, 35005, https://doi.org/10.1088/1748-9326/aaf52c, 2018.

Miller, N. B., Shupe, M. D., Cox, C. J., Noone, D., Persson, P. O. G. and Steffen, K.: Surface energy budget responses to radiative forcing at Summit, Greenland, The Cryosphere, 11(1), 497–516, doi:10.5194/tc-11-497-2017, 2017.

Miller, N. B., Shupe, M. D., Lenaerts, J. T. M., Kay, J. E., de Boer, G. and Bennartz, R.: Process-Based Model Evaluation Using Surface Energy Budget Observations in Central Greenland, J. Geophys. Res. Atmospheres, 123(10), 4777–4796, doi:10.1029/2017JD027377, 2018.

Moreno-Chamarro, E., Ortega, P., & Massonnet, F. Impact of the ice thickness discretization on the sea ice concentration variability in the NEMO3.6-LIM3 global ocean–sea ice model (in preparation)

Müller, M., Y. Batrak, J. Kristiansen, M. A. Køltzow, G. Noer, and A. Korosov, 2017: Characteristics of a convective-scale weather forecasting system for the European Arctic. Mon. Wea. Rev., 145, 4771–4787, https://doi.org/10.1175/MWR-D-17-0194.1.

Notz, D.: How well must climate models agree with observations?, Philosophical Transactions of the Royal Society A: Mathematical, Phys. Eng. Scie., 373, 20140164, https://doi.org/10.1098/rsta.2014.0164, 2015.

Oikkonen, A. and J. Haapala (2011): Variability and changes of Arctic sea ice draft distribution –submarine sonar measurements revisited, The Cryosphere, 5, 917–929, doi:10.5194/tc-5-917-2011

Olonscheck, D. and Notz, D.: Consistently Estimating Internal Climate Variability from Climate Model Simulations, J. Climate, 30, 9555–9573, https://doi.org/10.1175/jcli-d-16-0428.1, 2017.

Pailleux J., Geleyn J-F, Hamrud M., Courtier P., Thepaut J.-N. Rabier F., Anderson E., Burridge D., Simmons A., Salmond D., El Khatib R., Fischer C., 2014. Thenty-five years of IFS/ARPEGE. ECMWF Newsletter, 141,22-30

Pithan, F., Medeiros, B. and Mauritsen, T.: Mixed-phase clouds cause climate model biases in Arctic wintertime temperature inversions, Clim. Dyn., 43(1–2), 289–303, doi:10.1007/s00382-013-1964-9, 2014.

Prodhomme et al. (2016) Benefits of Increasing the Model Resolution for the Seasonal Forecast Quality in EC-Earth. Journal of Climate, 29, 9141-9162, doi :10.1175/JCLI-D-16-0117.1

Rackow, T., Goessling, H.F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D. and Handorf, D., 2016. Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability. Climate Dynamics, pp.1-26.

Rackow, T., Sein, D. V., Semmler, T., Danilov, S., Koldunov, N. V., Sidorenko, D., Wang, Q., Jung, T., 2019: Sensitivity of deep ocean biases to horizontal resolution in prototype CMIP6 simulations with AWI-CM 1.0. Geophysical Model Development. doi: 10.5194/gmd-12-2635-2019.

Rousset, M. Vancoppenolle, G. Madec, T. Fichefet, S. Flavoni, A. Barthélemy, R. Benshila, J. Chanut, C. Levy, S. Masson, and F. Vivier: The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities, Geosci. Model Dev., 8, 2991–3005, 2015.

Rummukainen, M. (2016), Added value in regional climate modeling. WIREs Clim Change, 7: 145-159. doi:10.1002/wcc.378

Salias y Mélia D., 2002 : A global coupled sea ice-ocean model. Ocean Modelling 4, 137-172.

Seity, Y. , P. Brousseau, S. Malardel, G. Hello, P. Benard, F. Bouttier, C. Lac and V. Masson, 2011; The AROME-France Convective Scale Operational Model. Mon. Weat. Rev. 139-976-991, DOI: 10.1175/2010MWR3425.1

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi ,D., Danilov, S., Dethloff ,K., Dorn, W., Fieg, K., Goessling, H. F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D. V., Wang, Q. 2015. Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate. Climate Dynamics, 44(3-4), pp.757-780.

Stephenson, D., & Pavan, V. (2003). The North Atlantic Oscillation in coupled climate models: a CMIP1 evaluation. Climate Dynamics, 20(4), 381-399.

Stramler, K., Del Genio, A. D. and Rossow, W. B.: Synoptically Driven Arctic Winter States, J. Clim., 24(6), 1747–1762, doi:10.1175/2010JCLI3817.1, 2011.

Yang, X., and Coauthors, 2018: IGB, the upgrade to the joint operational HARMONIE by DMI and IMO in 2018. ALADIN-HIRLAM Newsletter, No 11, ALADIN Consortium, Brussels, Belgium, 93–96, http://www.umr-cnrm.fr/aladin/IMG/pdf/nl11.pdf.

Ungermann, M., Tremblay, L. B., Martin, T., and Losch, M.: Impact of the ice strength formulation on the performance of a sea ice thickness distribution model in the Arctic,

J. Geophys. Res.-Oceans, 122, 2090–2107, https://doi.org/10.1002/2016JC012128, 2017.

Uotila, P., Iovino, D., Vancoppenolle, M., Lensu, M., and Rousset, C.: Comparing sea ice, hydrography and circulation between NEMO3.6 LIM3 and LIM2, Geosci. Model Dev., 10, 1009–1031, https://doi.org/10.5194/gmd-10-1009-2017, 2017.

Urrego-Blanco, J. R., Urban, N. M., Hunke, E. C., Turner, A. K., and Jeffery, N.: Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model, J. Geophys. Res. Oceans, 121, 2709–2732, https://doi.org/10.1002/2015jc011558, 2016.

Titchner, H. A., & Rayner, N. A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. Journal of Geophysical Research: Atmospheres, 119(6), 2864–2889.

Voldoire, A., Saint‐Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., ... & Nabat, P. (2019). Evaluation of CMIP6 DECK Experiments With CNRM‐CM6‐1. Journal of Advances in Modeling Earth Systems, 11(7), 2177-2213.

Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the Northern Hemisphere winter. Monthly Weather Review, 109(4), 784-812.

Wang X., P. Steinle, A. Seed, and Y. Xiao, “The Sensitivity of Heavy Precipitation to Horizontal Resolution, Domain Size, and Rain Rate Assimilation: Case Studies with a Convection-Permitting Model,” Advances in Meteorology, vol. 2016, Article ID 7943845, 20 pages, 2016. https://doi.org/10.1155/2016/7943845.

Warner, T.T., R.A. Peterson, and R.E. Treadon, 1997: A Tutorial on Lateral Boundary Conditions as a Basic and Potentially Serious Limitation to Regional Numerical Weather Prediction. Bull. Amer. Meteor. Soc., 78, 2599–2618, https://doi.org/10.1175/1520-0477(1997)078<2599:ATOLBC>2.0.CO;2

6. ACRONYMS

AAEPS	AROME-Arctic Ensemble Prediction System
ACC	Anomaly correlation coefficient
ASO	August, September, October
AV	Added Value
BSS	Brier Skill Score
CMIP6	6th Coupled Model Intercomparison Project
CRPS	Continuous Rank Probability Score
CPEPS	Convection−Permitting Ensemble Prediction System
EPS	Ensemble Prediction System
GIN	Greenland, Icelandic and Norwegian
ITD	Ice Thickness Distribution
JFM	January, February, March
LBC	Lateral Boundary Conditions
MAE	Mean Absolute Error
MAESS	Mean Absolute Error Skill Score
MEPS	MetCoOp Ensemble Prediction System
MSLP	Mean Sea Level Pressure
MUR	Multi-scale Ultra-high Resolution Sea Surface Temperature Analysis
NWP	Numerical Weather Prediction
OSTIA	Operational Sea Surface Temperature and Sea Ice Analysis
Precip1	1h accumulated precipitation
RH2m	Relative Humidity at 2m height
RMSE	Root Mean Square Error
S10m	10m wind speed
SIC	Sea ice concentration
SLAF	Scaled Lagged Average Forecasting
SST	Sea Surface Temperature
T2m	2m air temperature
TCC	Total Cloud Cover
YOPP-SOP-NH1	Year of the Polar Prediction Special Observing Period Northern Hemisphere