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. 4.1

Initial assessment of the added value of observations in existing long time-series datasets – guidance for dedicated observing system experiments

Submission of Deliverable

Work Package	WP4 Support for Arctic observing system design		
Deliverable No	4.1		
Deliverable title	Initial assessment of the added value of observations in existing long time-series datasets – guidance for dedicated observing system experiments		
Version	1		
Status	Final		
Dissemination level	PU - Public		
Lead Beneficiary	11 - CNRS-GAME		
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	31/10/2018		
Delivery Date	20 November 2018		
Coordinating author	Matthieu Chevallier (matthieu.chevallier@meteo.fr)		
Contributing authors	François Massonnet (francois.massonnet@uclouvain.be) Leandro Ponsoni (leandro.ponsoni@uclouvain.be) Irina Sandu (irina.sandu@ecmwf.int)		

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

The assessment of state-of-the-art long-term datasets, such as analyses and reanalyses, and short-range weather forecasts was carried out for Arctic regions. It was demonstrated that these datasets can be used for understanding the Arctic climate system variability. It was shown that the quality of (re)analyses in Arctic regions has increased over the last years, despite the challenges related with the modelling, observation usage and data assimilation encountered in Arctic regions. This further indicates that Numerical Weather Prediction systems can be used for numerical experimentation aimed at informing the design of future observing systems in the Arctic.

The use of atmospheric observations in the Arctic region was further analysed for the ECMWF NWP system. The analysis highlights that for global NWP, the polar regions are a particularly data-rich area in terms of satellite sounding observations from polar-orbiting satellites, whereas conventional observations are sparse north of 70N. However, the use of the satellite radiances is hampered during the winter period due to problems over snow and sea-ice in the forward modelling of the satellite radiances and cloud detection, combined with potentially larger errors in the forecast model. There is evidence of considerable bias, originating from the forecast model (especially around 200 hPa), the observation operator for microwave radiances (especially for surface-sensitive radiances over snow and sea-ice), and the cloud screening applied to infrared radiances. As these biases are not accounted for during the assimilation, biases in the resulting analyses are very likely in polar regions. The crucial role of conventional observations in identifying these biases has been highlighted. In addition, our analysis shows that background errors used in the analysis over the Arctic are likely to be under-estimated for the lower troposphere and the upper-troposphere lower-stratosphere. To improve the use of satellite data in the troposphere, improvements in the forward model and/or cloud detection would be needed, whereas to improve the use of near-surface data, it would be necessary to increase the background errors to avoid limiting the adjustments observations can make to the short-range forecasts in this area during the assimilation.

An initial assessment of the impact of observations has been performed using adjoint-based techniques that estimate the contribution of the observations to the reduction on the overall short-range forecast error. This suggests that conventional Arctic observations appear to contribute the most to reducing the global short-range forecast error during winter, while microwave observations from polar orbiting satellites contribute the most in summer. To further assess the impact of different observation types, particularly on improving the medium-range forecasts and on interactions with the mid-latitudes, observing system experiments (OSE's) are needed where different observation types are removed from the Arctic and the effect analysed. These will be conducted in task 4.2.2, with a view to provide further guide the design of future atmospheric observing systems.

In terms of Arctic sea ice, fourteen state-of-the-art reanalyses were analysed in order to understand the added value of sea ice concentration (SIC) assimilation for the sea ice thickness (SIT) field. First, it was aimed to identify whether or not reanalyses built with SIC assimilation are closer to observed values of SIT, in contrast to the products built with no sea ice data assimilation. All reanalyses were compared against several sources of observations such as moored upward-looking sonars, submarines, airbornes, satellites and ice boreholes. The best performing reanalyses (the ones which SIT values are closer to the observations) indeed assimilates SIC data (eg. TOPAZ1, C-CLORS05 and PIOMAS). However, the SIC data assimilation does not necessarily improve the reproduction of the SIT fields in all reanalyses. In those cases, the reanalyses do not take the best advantage of the covariances between SIC and SIT. Second, the impact of SIC assimilation in two aspects of the SIT variability, the time and length scales, was analysed. The results clearly show that SIC assimilation plays a role in both scales, which are considerable shorter for reanalyses making use of SIC data assimilation. This is likely due to the fact that when a reanalysis assimilates SIC information, the system is forced towards the assimilated conditions, differently from what occurs with free-running models, likely bringing the scales towards a more realistic values.

A first-time-ever systematic assessment of the representation of Arctic Ocean hydrographic properties in ocean reanalyses has been conducted as well. Few hydrographic observations are assimilated in the Arctic Ocean, thus the drivers of the main biases are poorly represented physical processes, biases in the Arctic sea ice (in spite of SIC data assimilation) and in the North Atlantic Ocean hydrography which propagate through Fram Strait and Barents Sea Opening.

1. INTRODUCTION

1.1. Background and objectives

Task 4.1 'Evaluation of existing data sets with a polar focus' is a stepping stone towards achieving the overarching goal of WP4 ('Support for Arctic observing system design'), which is to provide guidance for a better exploitation of observational datasets and for the optimal design of the future observing system in the Arctic region. The main objective of Task 4.1 is to perform **a comprehensive assessment of the added value, quality and relevance of existing datasets for understanding the Arctic climate system's variability and for improving predictions**. The focus is on existing state-of-the-art operational analyses (which provide the initial conditions for weather predictions, see Sect 3 for more details) and long-term reanalyses. This first key part of WP4 consisted of understanding:

1. What is the quality of existing analyses and reanalyses in the Arctic?
2. What are the key variables to monitor?
3. What is the added value of the current observing systems in these existing long time-series datasets?

This deliverable summarizes the outcomes of this assessment performed in Task 4.1, and provides guidance for the numerical experiments to be performed in Task 4.2 and in the framework of the Year of Polar Prediction (YOPP), as well as for future reanalyses.

1.2. Organisation of this report

This report is organised as follows. Section 2 illustrates that the quality of operational weather forecasts (with a focus on short to medium-range), analyses and reanalyses for various components of the Earth System (atmosphere, ocean, sea ice) has improved over time. This is important as it increases the confidence in the use of these long time-series datasets for documenting the variability of the Arctic climate. These improvements also justify the use of numerical experimentation based on numerical weather and climate prediction systems (Tasks 4.2 and 4.3) to derive conclusions about the optimal design of the future Arctic observing system.

Section 3 describes the assessment of **operational atmospheric analyses** used as initial conditions for the weather forecasts performed by the European Centre for Medium-Range Weather Forecasts (ECMWF). Advanced data assimilation diagnostics are employed to evaluate key components of the data assimilation system and to assess the role of observations in improving the initial conditions and thereby forecast skill in Arctic regions (with a focus on short to medium range, i.e. day-1 to day-10). Insights gained here are used to guide the design of the numerical experiments to be performed in Task 4.2 where the importance of existing observing systems for short to medium-range predictive skill is assessed.

Section 4 summarizes the assessments of sea ice and ocean properties in the Arctic ocean as simulated in ocean and sea ice reanalyses, especially those used to initialize climate predictions (here seasonal-to-decadal) run with fully coupled models. This task, performed by partners from UCL, Météo France (third party of CNRS-GAME) and UREAD, has two objectives: (1) document the main observational sources that are currently assimilated in ocean-sea ice reanalysis systems and (2) assess how the assimilation of ocean observations and other constraints act on the mean states, when compared to unconstrained ocean-sea ice models. As planned, this task has been done in collaboration with the Polar Ocean-ReAnalysis Intercomparison Project (PORA-IP), a project in which the contribution of the APPLICATE team has been instrumental and substantial. Section 4 summarizes the main qualities and defects of current ocean-sea ice reanalyses, focusing on sea ice properties (sea ice concentration, sea ice thickness), and how those properties impact ocean properties reconstructed in the reanalyses. All results summarized in this section are explored and discussed in details in published or submitted papers : a paper entitled "**An assessment of ten ocean reanalyses in the polar regions**" published in

Climate Dynamics, and a manuscript entitled "**On the time and length scales of the Arctic sea ice thickness anomalies: a study based on fourteen reanalyses**", under review in *The Cryosphere*.

Section 5 summarizes the most important findings of the work undertaken in Task 4.1 and draws recommendations for future reanalysis and for the set-up of numerical experimentation in Task 4.2.2.

2. BETTER FORECASTS, ANALYSES AND REANALYSES IN ARCTIC REGIONS

The Numerical Weather Prediction (NWP) systems used to produce weather forecasts, analyses and reanalyses have massively improved over the last decades. Advances in NWP represent a quiet revolution because they have resulted from a steady accumulation of advances in science (numerical techniques, physical parameterizations, data assimilation methodologies), utilisation of observations and supercomputing capacities and technologies (Bauer et al., 2015). The accuracy of medium-range weather forecasts in the ECMWF Integrated Forecast System (IFS) has increased in the mid-latitudes by approximately one day per decade, making today's forecasts of the weather at six days ahead as good as forecasts at four days ahead twenty years ago (Bauer et al., 2015).

Recent evaluations performed at ECMWF in Task 4.1.2 have shown that the accuracy of medium-range weather forecasts in the Arctic has increased at a similar pace as in the mid-latitudes, albeit remaining generally lower (Figure 1

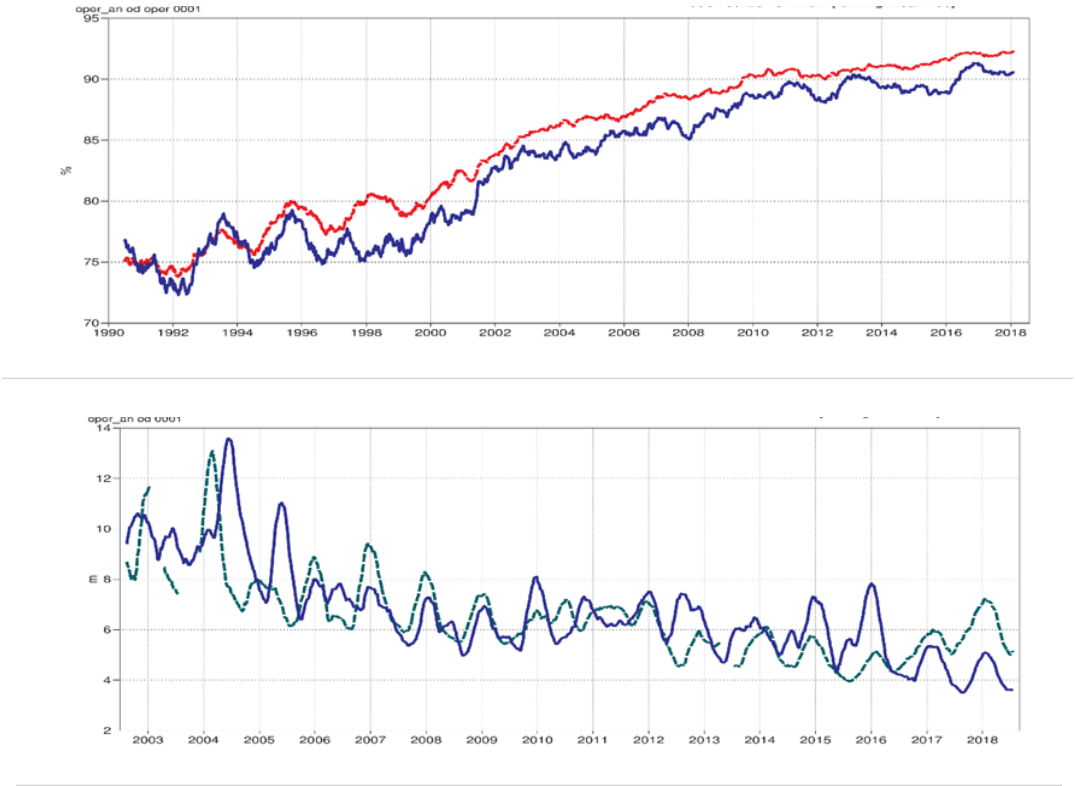

Figure 1a) when applying WMO standard skill metrics (see also Bauer et al. 2014). This is partly the result of the quality of the initial conditions of weather forecasts having also increased over time in the Arctic. This is suggested by the fact that the gap between the mid-tropospheric analyses from different operational NWP centres are in a much better agreement with each other over the Arctic than 15 years ago (Figure 1

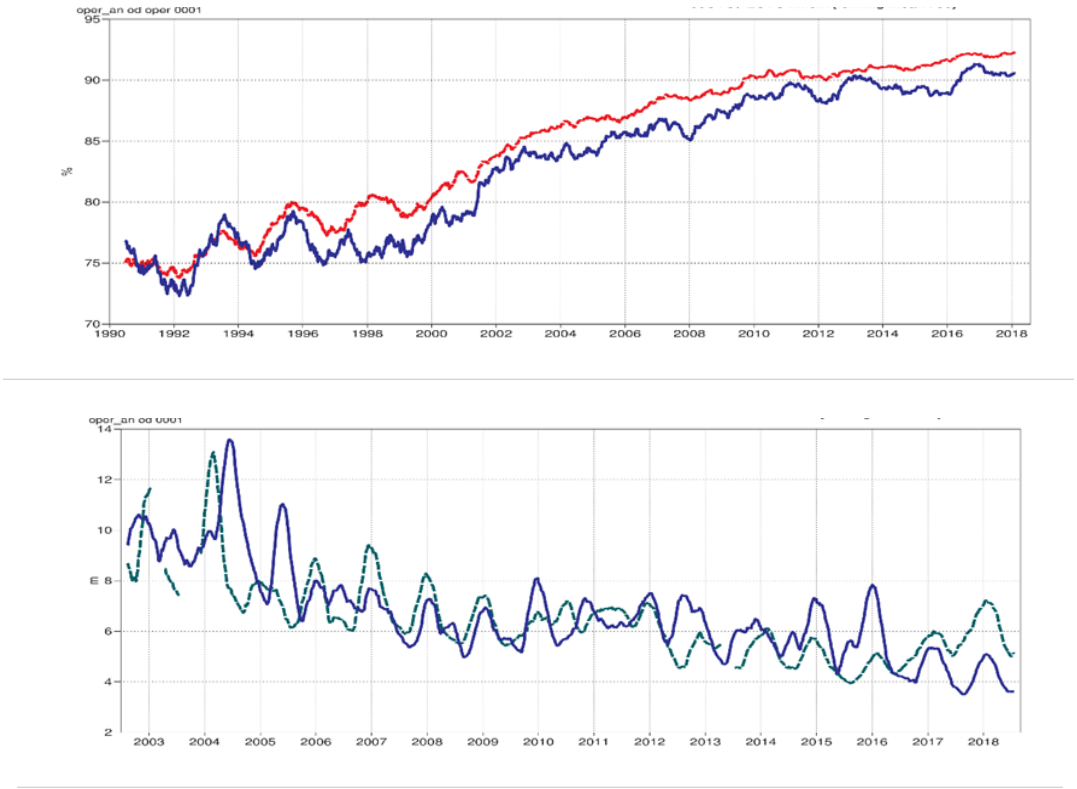

Figure 1b).

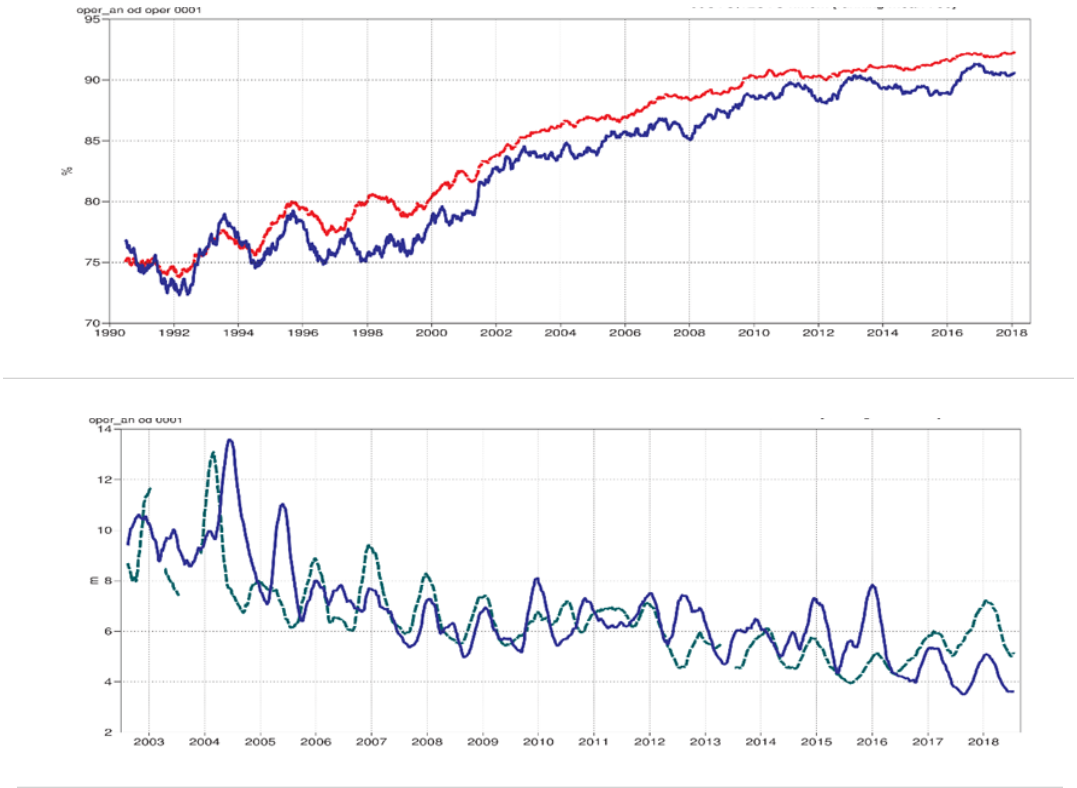

Figure 1: a. Evolution of the anomaly correlation of 500 hPa geopotential height at forecast day 5 (against the analyses from which the forecasts are initialized, at the forecast verification time, i.e. 0UTC), for the Northern Hemisphere (red) and for the Arctic (blue) (60-90N) in the ECMWF operational system. b. Evolution of the root mean square differences in 500hPa geopotential height in the Arctic between the operational analyses of various NWP centres: NCEP-ECMWF (dashed green) and UK MetOffice - ECMWF (blue).

The lower forecast skill in the Arctic (Figure 1

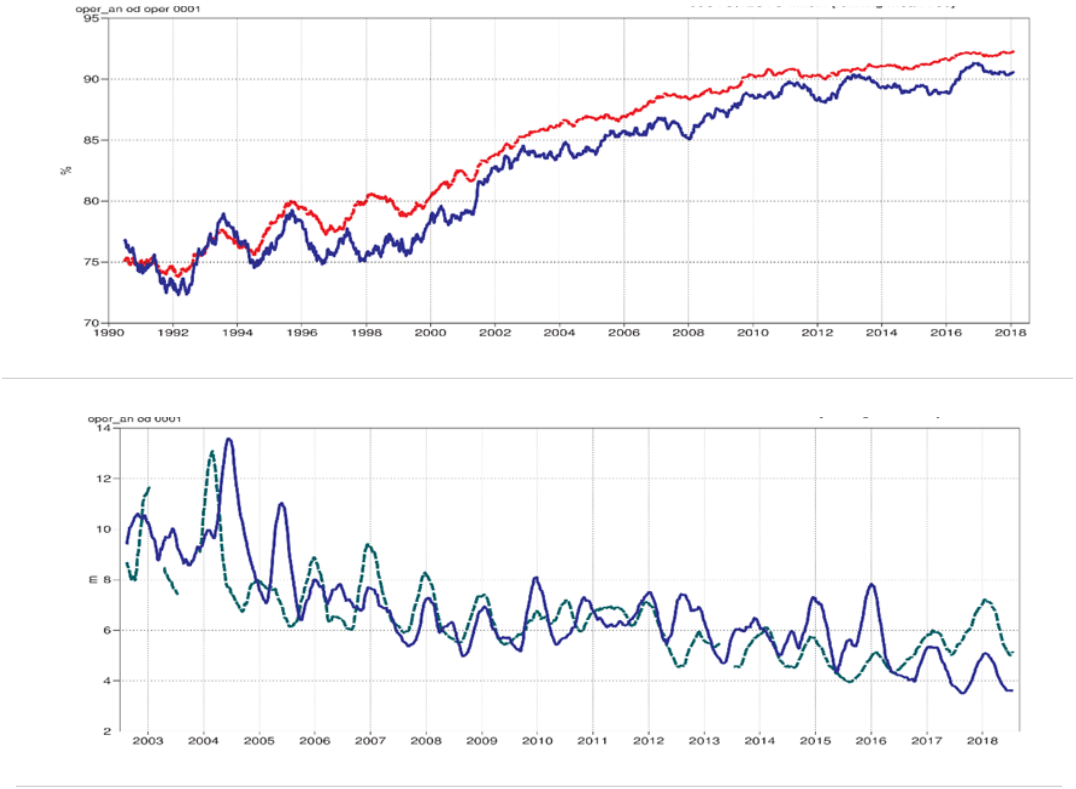

Figure 1a) is likely to be related to known challenges at high latitudes in all key aspects of NWP: coupled modelling, data assimilation and the effective use of observations. Indeed, model errors are large, typically in aspects related to the representation of stable boundary layers, mixed-phase clouds, snow and sea ice, as well as the coupling of these different elements. Moreover, in-situ observations are sparse, and while satellite radiance observations from polar orbiting satellites are plentiful, some are more difficult to use in data assimilation due to specific problems related to cloud detection and radiative transfer modelling over snow and ice surfaces, and limited vertical resolution provided by satellite soundings in the shallow polar atmosphere. In Section 3 we discuss some of these observation-related challenges and evaluate the performance of some of the key components of the ECMWF data assimilation system (used to produce initial conditions, i.e. analyses, for the operational medium-range weather forecasts), with a particular focus on the Arctic.

Atmospheric reanalyses over the Arctic have similarly improved over the last decades, benefiting from the enhancements seen in operational NWP systems, and reanalyses such as the latest ERA5 produced by ECMWF for the Copernicus Climate Change Service (C3S; run at 32 km resolution) can now be used to monitor the Arctic climate and how it is changing. Figure 2 illustrates, as an example, that the representation of one of the key features of the Arctic climate, the Sudden Stratospheric Warmings (SSW) have massively improved in ERA5 with respect to ERA-Interim (the previous ECMWF reanalysis; Dee at al. 2011). This improvement is primarily due to changes to the semi-Lagrangian advection scheme, demonstrating that the quality of a forecast model is key for the effective use of observations in data assimilation, and thus highlights the intimate link between modelling and data assimilation.

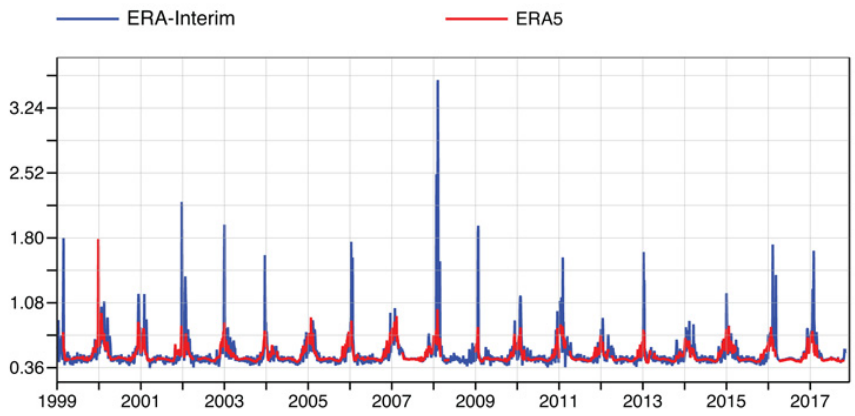

Figure 2 Comparison of microwave radiances measured by the NOAA-15 satellite and equivalent values simulated using the temperature fields of ERA-Interim (blue) and ERA5 (red). The standard deviation of differences is shown for channel 13 (53 GHz) which is primarily sensitive to altitudes around 5 hPa. The large values (spikes) in this time series correspond to Sudden Stratospheric Warming events. It appears the differences between the observed and simulated radiances decreases dramatically in ERA5 compared to ERA-Interim for these events.

Global and regional ocean reanalyses are increasingly used in polar research. Those are typically produced with ocean models, driven by fields and fluxes from atmospheric reanalyses in forced or coupled mode, and assimilating a large variety of data (see e.g. Stammer et al., 2010 or Balmaseda et al., 2015 for a review). In contrast to the atmosphere, game-changing observations in the ocean are relatively recent, e.g. sea-level information from satellite altimetry since 1993 and comprehensive deep profiles from Argo buoys since the late 1990s. In spite of enhanced satellite coverage, the Arctic ocean observational network remains sparse compared to other oceanic regions as e.g. the northern north-Atlantic. For instance, there have been virtually no Argo-buoy deployments north of 70°N, as those buoys cannot operate under perennial sea ice.

Past studies have shown that Arctic sea ice thickness in spring and early summer plays a key role in shaping the end-of-summer Arctic sea ice extent anomalies (e.g. Kauker et al., 2009). A proper initialization of sea ice thickness anomalies (e.g. Chevallier et al., 2013) and sea ice thickness distribution (e.g. Chevallier and Salas-Mélia, 2012) could provide predictive skill up to 6 months ahead, especially for the prediction of the September sea ice extent. In idealized experiments, Day et al. (2014) show that thickness initialization could enhance predictive skill on interannual time scales. For that purpose, operational ocean-sea ice analyses or reanalyses used to initialize sea ice seasonal forecasts do provide the best reconstruction of sea ice thickness, in spite of the sparse thickness data available for assimilation and evaluation.

In a first phase of the Ocean ReAnalyses Intercomparison Project (ORAIP), Chevallier et al. (2017) assessed the quality of Arctic sea ice concentration, thickness, velocity and snow reconstructions in a set of 14 global reanalyses produced by operational centres and academia for several purposes, including initialization of coupled predictions (e.g. GloSea5, GLORYS2v3, ORAP5). Nine of these reanalyses assimilated sea ice concentration (SIC) data while none assimilated sea ice thickness data. The intercomparison of Chevallier et al. (2017) clearly shows that SIC data assimilation does not seem to improve the simulated sea ice thickness, highlighting that the way to constrain sea ice thickness when only SIC measurements are available remains a critical question (Massonnet et al., 2015). Biases in Arctic sea ice thickness – evaluated through comparison with satellite data from ICESAT and CryoSAT2 – are representative of the biases of free-running models, as assessed in the framework of past intercomparison projects (AOMIP, CORE-II).

3. ASSESSMENT OF OPERATIONAL ATMOSPHERIC ANALYSES AND OBSERVATION IMPACT IN THE ARCTIC

The accuracy of the initial conditions is one of the key factors contributing to the skill of weather forecasts over a range of timescales (hours to months). These initial conditions are produced with comprehensive data assimilation systems, which cover a time window (12 hours for the operational ECMWF system) along which they optimally combine the available observations with short-range forecasts to produce a best estimate of the state of the atmosphere (and ocean including sea ice). This is called the analysis. Observations are thus used to correct forecast errors, leading to increments in state variables such as temperature, wind, humidity and geopotential height. This process relies on the observations and their usage (including forward modelling performed with observation operators and an observation bias removal or correction) and a weighting of the observations and short-range forecasts (or background hereafter) by the respective observation and background error covariance matrices. This process also relies on the quality of the forecast model because a good model can exploit the observational information more effectively. In a four-dimensional data assimilation system (4D-Var) such as the one used at ECMWF (Rabier et al., 1998), the assimilation process also relies heavily on the quality of the tangent-linear and adjoint versions of the forecast model (Janiskova and Lopez, 2013).

As mentioned above, Arctic regions pose specific challenges for modelling, observation usage and data assimilation. It is therefore important to understand whether existing observations are used effectively, whether the weighting between observations and background is suitable in polar regions, and how much existing observations contribute to reducing forecast errors in the current NWP systems given imperfect models. To answer these questions, an assessment of the operational ECMWF analyses and short-range forecasts in Arctic regions was carried out at ECMWF in the framework of WP4 (Task 4.1.2) for several winter and summer seasons.

This section summarizes this work, being structured as follows: After outlining a few details regarding the operational ECMWF forecasting system, we present an overview of the usage of atmospheric observations (conventional and satellite) in the Arctic highlighting some known issues with their use in these conditions, and we discuss one of the key elements controlling the weight given to the observations in the data assimilation process (namely the background forecast errors characterized by the spread of the Ensemble of Data Assimilation (EDA); see Sections 3.1 and 3.3). We then employ operationally available statistics and diagnostics from the data assimilation system such as increments, departures, Forecast Sensitivity to Observation Impact (FSOI; Cardinali 2009) to gain insights about the added value of current observing systems, and more particularly about their role in reducing forecast errors in the short-range. The results presented here focus on a few selected recent seasons (June to August 2016 and 2017 and December to February 2017-2018).

3.1 ECMWF operational analyses and medium-range weather forecasts

ECMWF is currently producing global operational deterministic analyses and forecasts up to 10 days ahead at a resolution of ~9 km horizontally (since May 2016) with 137 levels in the vertical. ECMWF also produces ensemble medium-range forecasts (at a resolution of ~18km up to 15 days ahead), monthly and seasonal forecasts (at a resolution of approximately ~32 km), all initialized from the deterministic analyses supported by background error statistics produced by the EDA. In this report, the focus will be on the operational deterministic analyses and short-range forecasts.

The ECMWF analyses and forecasts are produced with the Integrated Forecasting System (https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation), which relies on a 4D-Var data assimilation method (Rabier et al., 1998). Over 40 million atmospheric observations sensitive to temperature, humidity, clouds, wind, and ozone are routinely assimilated globally in each assimilation cycle of this 4D-Var data assimilation system, including both conventional and satellite data (described in more detail in Sect. 3.2).

The weighting of observations and background (short-range forecasts) in the (12 hour) assimilation window in the deterministic analyses is determined by the modelling of the background error together with the assigned observation error. The balance between both errors affect the observation impact since observations cannot properly correct forecast errors if background (observation) errors are too small (large). Furthermore, satellite radiance observations are sensitive to broad atmospheric layers and the background error covariances play an important role in the height distribution of increments due to these observations. A good background error specification is essential for a successful assimilation of observations.

In the ECMWF system flow-dependent background errors are described using a 25-member Ensemble of Data Assimilations (EDA) described in Isaksen et al. (2010). The EDA is an ensemble of independent 4D-Var data assimilations where the main analysis error sources (observation errors, model errors, and boundary condition errors) are represented by perturbing the related quantities (observations, model, sea-surface temperature, etc.). The spread of the EDA at the end of the assimilation window is used to provide the background error information for the following deterministic analysis update. To provide reliable background errors, the spread of the EDA needs to capture both the magnitude and the variation of the errors in the short-range forecasts that are used as background in the deterministic analysis, as well as the correlations of these errors.

3.2 Atmospheric observations assimilated in polar regions and associated challenges

A review of the observational coverage and specific challenges with using these observations in the Arctic regions was carried out at ECMWF for a summer and winter period: June-July-August (JJA) 2016 and December-January-February (DJF) 2017-2018.

The main atmospheric observation types that are assimilated in the Arctic region are:

- conventional observations including radiosondes from land and ships, aircraft data, synoptic and buoy surface pressure observations, and 2m relative humidity observations from some synoptic stations (daytime-only observations),
- microwave (MW) radiances from polar orbiting satellite radiometers,
- infrared (IR) radiances from polar orbiting satellite radiometers,
- GPS Radio Occultation (GPSRO) bending angle observations,
- Atmospheric Motion Vectors (AMVs) derived from polar orbiting satellite image sequences,
- scatterometer 10m equivalent wind (u and v components) over ocean.

Typically, around 2.5 million observations are assimilated per day in the Arctic region, with infrared radiances clearly being the biggest contributor, followed by microwave radiances, conventional observations, GPSRO, AMVs, and scatterometers (Figure 3). The relative contribution of each observation type to the total number of assimilated observations in the Arctic depicted in Figure 3 shows that during winter there is a decrease of the relative number of microwave and infrared observations and an increase in the relative number of conventional observations used compared to the summer season. This is likely associated with the challenges related to the assimilation of satellite observations during the winter season discussed that will be further described in the following.

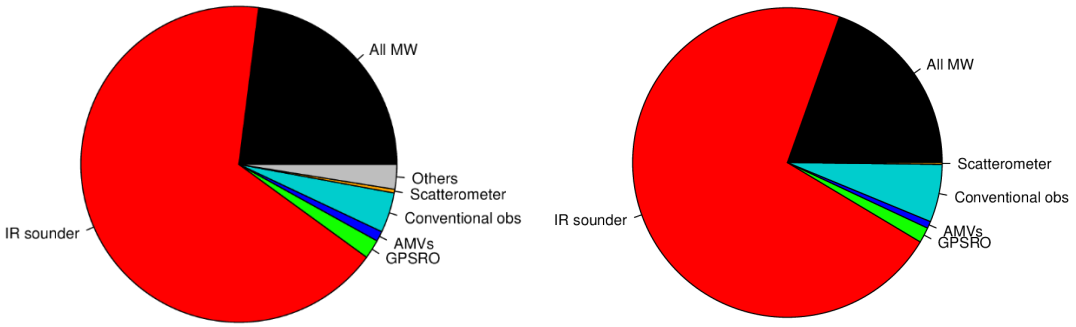

Figure 3: Relative contribution of infrared sounders, microwave sounders, conventional observations, GPSRO, AMVs and other observations to the total number of observations assimilated in the Arctic (above 60N). The relative contributions of each observation type is determined as the ratio between the number of observations of that type that are assimilated in the Arctic and the total number of observations assimilated in the Arctic . Plots are shown for JJA 2016 (left) and DJF 2017/2018, with total observation numbers of around 2.4 Million and 2.7 Million, respectively.

Conventional observations assimilated in the Arctic include surface pressure (ps) observations, temperature (T), wind (u and v) and specific humidity (q) from radiosondes and aircraft, 10-m wind over ocean (u10m and v10m), 2-m relative humidity (rh2m, during day only) and geopotential height (z). Table 1 gives a list of the observations that are assimilated in the Arctic. Note that only a small number of aircraft humidity, and SYNOP geopotential height observations are assimilated in the Arctic. Also note that aircraft data are mainly restricted to 50 – 350 hPa, with most observations being located between 200 - 300hPa, whereas radiosonde data cover all heights from the surface up to around 5 - 20 hPa.

Table 1. Types of conventional observations which are assimilated in the ECMWF system.

Conventional Observation Type	**Observations assimilated**
SYNOP	z, u10m, v10m, rh2m, ps
AIREP (aircraft)	T, u, v, q
BUOY (drifting buoys)	ps, u10m, v10m
TEMP (radiosonde)	T, u, v, q, u10m, v10m
PILOT (wind)	u, v

Conventional observations generally have a better coverage over land than ocean due to most SYNOP, TEMP and PILOT observations being situated over land. The coverage for conventional observations in the Arctic is therefore quite good at latitudes of 60 – 70N, particularly for surface pressure observations (Figure 4) but there are much fewer observations in the Arctic ocean, i.e. above around 70N. In particular, there are very few radiosonde observations above 70N, as shown in Figure 4. Furthermore there are fewer surface pressure and 10-m wind observations over the Arctic ocean in winter than in summer due to less buoy and ship data limited to sea-ice free oceans. Note that there is an increase in the aircraft coverage for the winter period compared to the summer. However, this is due to changes in practice in 2017 where some aircraft began to report more frequently, and it is therefore not indicative of the typical winter/summer difference.

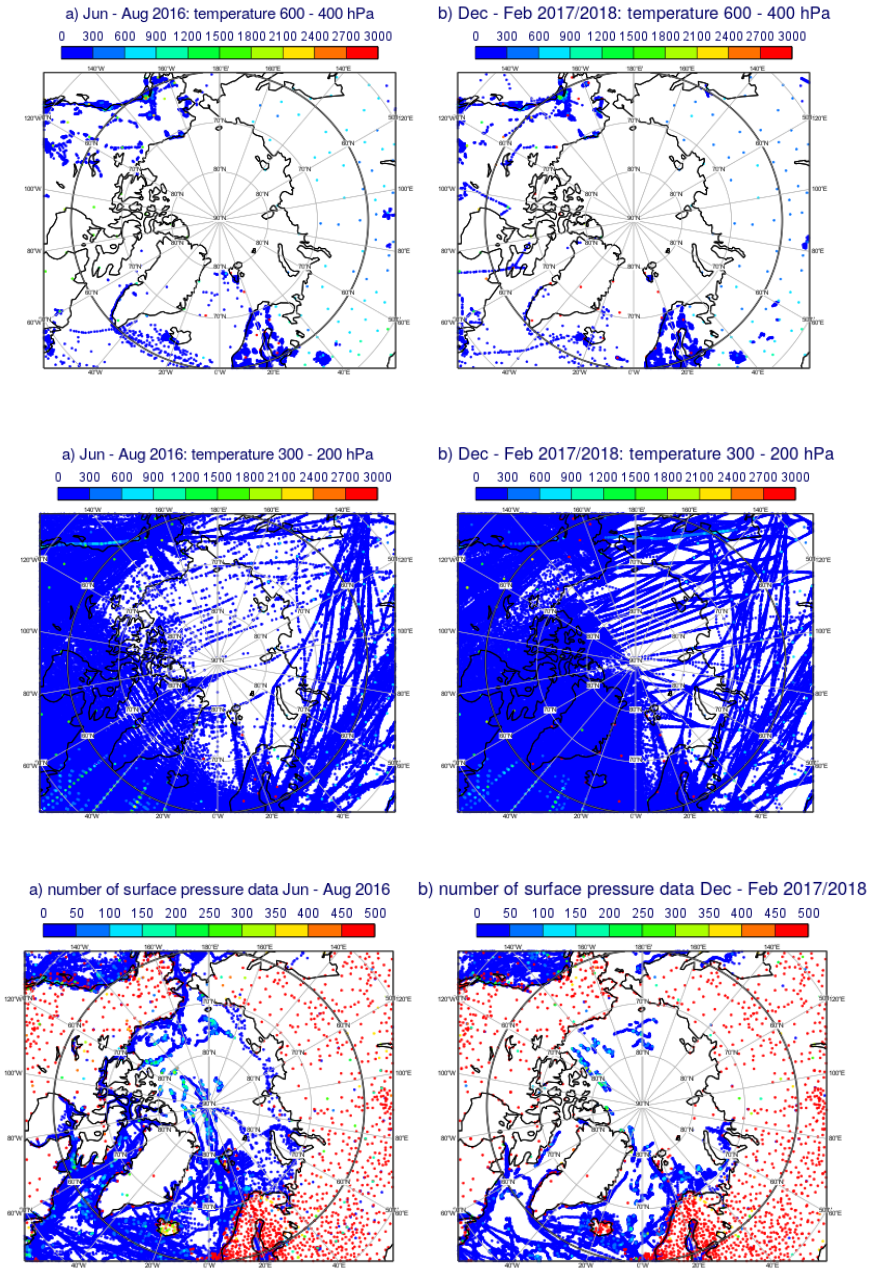

Figure 4: Number of (top) atmospheric conventional temperature observations (radiosondes) at heights of between 400 and 600 hPa, (middle) atmospheric temperature observations (radiosondes and aircraft) at heights of 200 – 300 hPa and (bottom) surface pressure observations (SYNOP land and drifter buoys), assimilated in the operational ECMWF system for the periods (left) JJA 2016 and (right) DJF 2017/ 2018.

The majority of satellite data assimilated in the ECMWF system, both globally and in the Arctic, are in the form of microwave or infrared radiances from polar orbiting satellites. All infrared observations and the temperature-sounding microwave observations are assimilated in clear-sky conditions after applying a cloud-screening and using a clear-sky forward model (RTTOV, Saunders et. al. 2018). The majority of microwave humidity-sensitive observations are assimilated in all-sky conditions, such that the cloud and precipitation affected data can influence the analysis, through use of an all-sky forward model (RTTOV-SCATT, Bauer et al. 2006). The main microwave observations assimilated both globally and in the Arctic from 16 microwave temperature and humidity sounders are: the 50 – 60 GHz temperature sounding channels on the AMSU-A and ATMS instruments, and the 183 GHz water vapour sounding channels on a number of instruments including MHS, ATMS, MWHS-2, MWHS,

GMI[1] and SSMI/S. While the 183 GHz channels are sensitive primarily to humidity they also provide very good wind information in the ECMWF 4D-Var assimilation system. This is due to the tracing effect of humidity and clouds when these observations are assimilated in all-sky conditions, as shown by Geer (2014). Assimilated infrared observations include radiances from four instruments: IASI (2 instruments), CrIS and AIRS, sensitive to atmospheric temperature, humidity and ozone. Infrared observations from these four instruments have a smaller temporal coverage than the microwave observations since they are available from fewer satellites and therefore with fewer local crossing times. However, infrared observations have a greater number of channels and a higher vertical resolution, leading to a greater number of observations (Figure 3) than microwave instruments. Currently 100 - 200 channels are assimilated per instrument, with the majority in the long-wave temperature-sounding band, but also some ozone and humidity sounding channels. Infrared observations are more sensitive to cloud and precipitation than microwave, so that more observations tend to be screened due to cloud than for the microwave temperature sounders.

In the Arctic, the microwave and infrared radiances represent over 90% of the total number of assimilated observations (Figure 3). There are more polar orbiting satellite data assimilated in the Arctic than in other regions, due to the high revisit time over the poles, as each polar orbiting satellite will observe the polar regions approximately every 100 min. The higher number of observations from polar orbiting satellites above 65N than at lower latitudes is illustrated in Figure 5 for the microwave temperature-sounding channel 12 of AMSU-A on MetOp-A (peaking around 10 hPa). Similar observations are provided by instruments from a further 6 microwave temperature sounders, each with comparable coverage. Equally, eleven microwave humidity sounders and four hyperspectral infrared instruments provide additional coverage over the polar regions. The polar areas are therefore the most densely observed regions in terms of passive sounding observations from polar orbiting satellites.

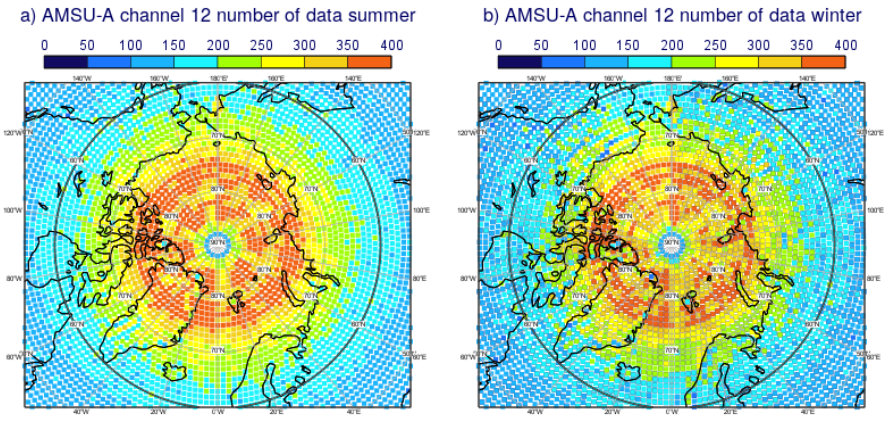

Figure 5: Number of observations assimilated for MetOp-A AMSU-A channel 12 (peaking around 10 hPa) in the ECMWF operational system for a) JJA 2016 and b) DJF 2017/2018.

However, the surface sensitive observations (sensing the lower troposphere) from polar orbiting satellites are particularly difficult to use over snow and sea-ice and in the shallow polar atmosphere. More specifically, we cannot yet make the best use of these observations because model errors in these polar regions are still large and also because observation operators are incomplete or have large errors, limiting the capability to extract information obtained from satellites when we assimilate radiances.

For example, it is more difficult to assimilate surface-sensitive microwave channels over snow and sea-ice due to the difficulties of estimating the surface contribution in the radiative transfer (e.g. in terms of modelling of surface emissivity, skin temperature and reflection properties). This is illustrated by Figure 6 which shows the mean observation minus background (or O – B, also referred to hereafter as background departures) for the lowest peaking microwave temperature sounding channel of

[1] Note that GMI data are only available up to 65N due to the satellite orbit.

AMSU-A, channel 5 (peaking around 500 – 700 hPa), in the Arctic for summer and winter periods. The differences between the observations and the short-range forecasts are larger for the winter period, particularly over areas of sea-ice and snow cover. The mean sea-ice fractions for the summer and winter periods are shown in Figure 7. The larger O - B differences in winter are most likely due to greater errors in the forward model for surface-sensitive channels over snow and sea-ice and because model errors are most likely higher over snow-covered surfaces and in very stably stratified boundary layers which are more frequent in winter. One source of forward model error is that currently a specular reflection is assumed over snow whereas in reality the reflection is more likely to be close to a diffuse (Lambertian) reflection (e.g., Guedj et al. 2010, Bormann et al. 2017). This is an area of ongoing research.

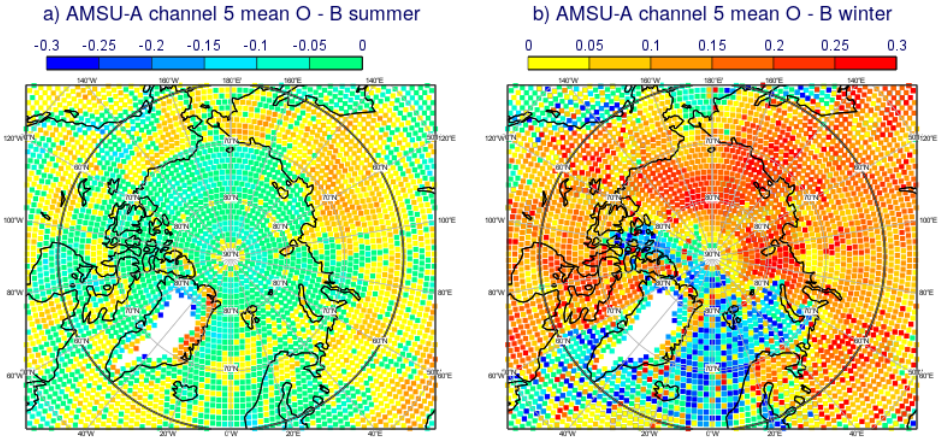

Figure 6: Mean observation minus short-range forecast (background), or O – B, for MetOp-A AMSU-A channel 5 (peaking around 500-700hPa) in the Arctic for a) JJA 2016 and b) DJF 2017/2018. Note that data are shown after bias correction.

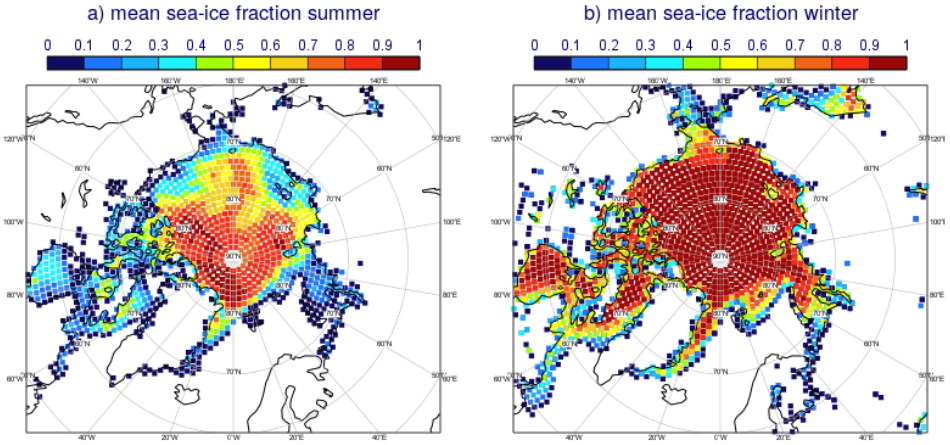

Figure 7: Mean of the model sea-ice fraction for the summer 2016 and winter 2017/2018 periods. The values closest to AMSU-A observation locations are shown.

Since an O – B check is applied in the quality control, the greater mean differences between the observations from the temperature sounding microwave channels and the background lead to more such observations to be rejected in winter than in summer. This is illustrated in Figure 8, which shows the number of assimilated observations from channel 5 of MetOp-A AMSU-A. For this channel 24% more data are used in the polar regions (above 60N) in the summer period than in winter. Note that similarly strong O - B differences can also be seen over sea-ice and snow for the 183 GHz microwave humidity sounding channels (not shown). Improvements in the radiative transfer modelling, the skin temperature treatment or the representation of relevant physical processes may allow more of these observations available in the Arctic region to be used.

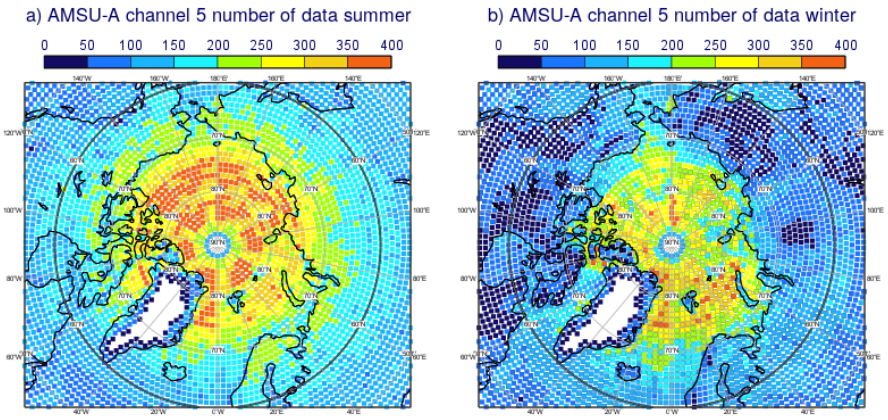

Figure 8: Number of observations assimilated from MetOp-A AMSU-A channel 5 (peaking around 500 – 700 hPa) in the ECMWF operational system in a) JJA 2016 and b) DJF 2017/2018.

There are also difficulties in assimilating infrared observations in polar regions compared to other areas. As for the microwave observations, surface emission is more difficult to model over sea-ice, leading to higher biases for surface sensitive channels. This is illustrated in Figure 9, which shows the mean O -B values for a surface-sensitive infrared temperature sounding channel on MetOp-B IASI, in summer and winter. Note that surface-sensitive infrared observations are not assimilated over land, only over ocean and sea-ice. The high O - B differences over sea-ice in the winter are most likely due to errors in the emissivity estimation or in the model skin temperature, and again due to the fact that model errors are higher.

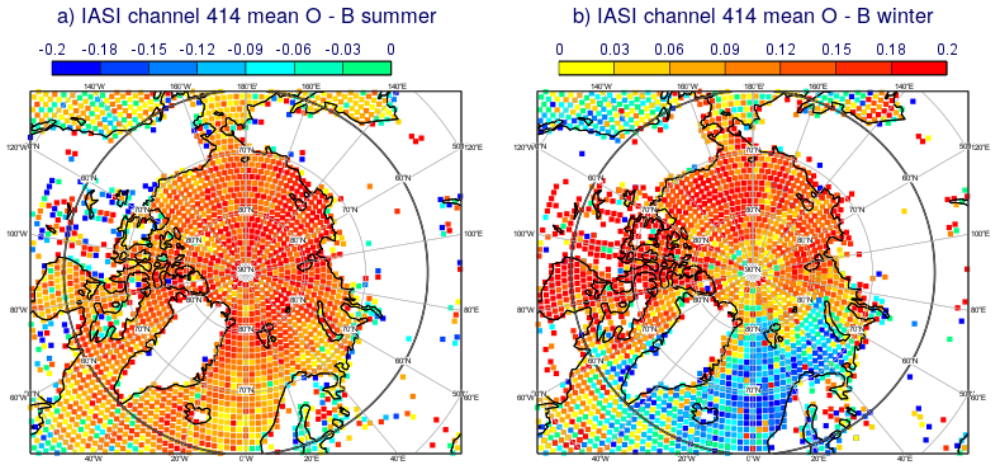

Figure 9: Mean O – B after bias correction for the infrared temperature sounding IASI channel 414 ((748.25 cm^{-1}, peaking around 740 hPa) on METOP-B shown for a) JJA 2016 and b) DJF 2017/2018.

Finally, for infrared observations, there can also be problems with cloud detection, which is more difficult in colder areas (e.g., Eresmaa 2013). Infrared observations are assimilated (in clear sky only), with a clear-sky forward model under the assumption that there are no clouds present. Therefore, if the cloud detection is not accurate this can have a detrimental effect on the analysis. The cloud detection for infrared observations relies primarily on thermal contrast between clouds and the surface. Over warm oceans clouds tend to be visible as a strong cold signal, however in polar latitudes clouds can show up as a weaker warm signal (in temperature inversions) or if the ocean and cloud top temperatures are similar there can be no contrast. Similarly, the cloud detection will rely on an accurate estimate of the model skin temperature, which can be problematic over sea-ice, as highlighted above. These aspects make cloud detection in the infrared part of the spectrum more difficult in polar

regions. The positive mean O – B values shown in Figure 9a for the summer period, in regions with no sea-ice are most likely due to undetected cloud. This will be discussed in Section 3.4.

3.3 Observation-based evaluation of the EDA spread-skill relationship

As explained in Section 3.1, the EDA-spread is key for the reliability of the background errors, and thereby for the quality of the analysis. For the system to be reliable, the spread should match the skill of the EDA. The skill of the EDA is measured by the departures (or errors) of the short-range forecasts against a reference (or truth) which here is chosen to be the observations. In the following, the spread-skill relationship, which is indicative of the reliability of the system, was derived for different regions of the globe and its behaviour in polar regions was compared to that in other regions.

To evaluate the spread-skill relationship, the spread of the EDA was computed in observation space (using conventional observations and satellite radiances) and was compared to the EDA skill, which is measured here by the background departures (observation - background). The methods used to analyse the EDA-spread, and the spread-skill relationship are described in detail by Dahlgren (2018) for conventional observations, and Bormann and Bonavita (2013) for satellite radiances. While there are some differences in the calculation of the statistics, the general approach is as follows: First, the spread of the EDA is calculated as the standard deviation of the values of the EDA analyses for each observation considered. The background departures are computed for the unperturbed control member of the ensemble for the radiance method; and for the ensemble mean for the conventional observations. Next, the observations are binned by the size of the EDA spread, and the root mean square (RMS) of the background departures is calculated for each bin. The RMS of the background departures is a measure of the total EDA error (or skill), including model errors (i.e. of the short-range forecasts) and observation errors. In the figures displayed in the remainder of this section, the contribution from the observation error is removed from the total error for conventional observations, using the observation error assigned in the assimilation. For the satellite radiances this is not done, as the assigned observation errors tend to be inflated and hence not appropriate for this purpose. Instead the raw total error is displayed in this case. When displayed as a function of the EDA spread, the total EDA error (or skill) indicates to what extent the EDA is able to identify regions with larger errors in the short-range forecast. Ideally, the spread should always match the total error, so that the spread-skill curve should lie along the diagonal of each plot. If this curve is above the diagonal the EDA is under-dispersive, while if it is under the diagonal it is over-dispersive. In the following we analyze statistics from ECMWF's operational EDA for two three months periods JJA 2017 and DJF 2017/18.

Spread-skill relationships for surface pressure and 925 hPa temperature from conventional observations indicate that the EDA is more under-dispersive for these parameters during the winter periods over the polar regions than elsewhere over the globe (*Figure* 10). Here, the spread-skill curve is above the diagonal, indicating that the EDA spread is smaller than the short-range forecast error. This may suggest a lack of perturbations for input fields or model parametrisations during winter, possibly related to snow or sea-ice cover. During summer, and for other regions, the relationship is closer to the ideal spread-skill relationship, indicating that the EDA-spread is a good indicator of situations with larger errors.

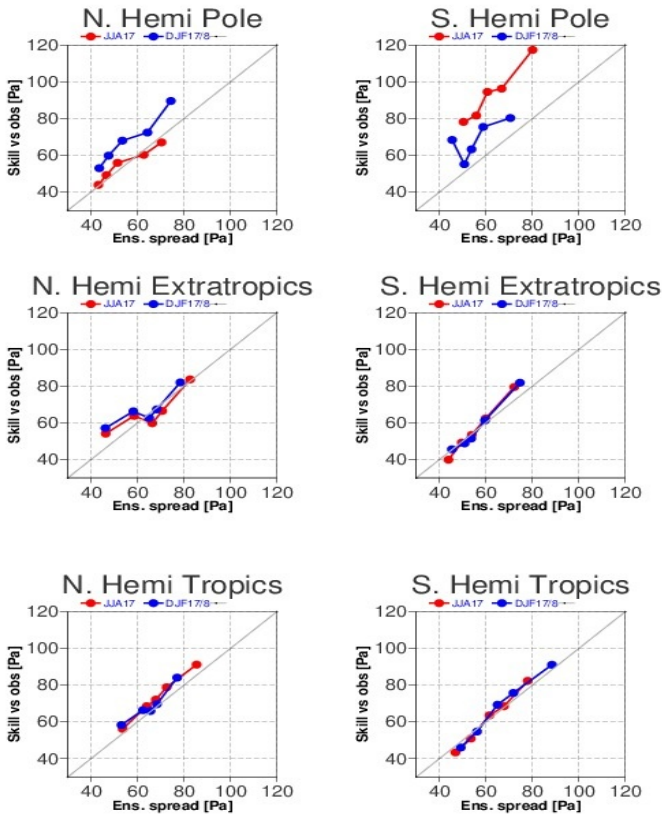

Figure 10*: Surface pressure spread-skill reliability curves for the operational EDA. They have been separated into six zonal bands: 90S-60S (southern hemisphere - pole), 60S-30S (southern hemisphere - extratropics), 30S-0 (southern hemisphere -tropics), 0-30N (northern hemisphere - tropics), 30N-60N (northern hemisphere - extratropics) and 60N-90N (northern hemisphere - pole). JJA 17 is shown in red and DJF 17/8 is shown in blue. The spread-skill relationship looks very good in the tropics and good for the extratropics but for the polar regions the winter pole is very under-dispersive. The statistics include all available assimilated observations from SYNOP, METAR, BUOY, SHIP.*

For mid-tropospheric temperature and humidity, the spread-skill relationships for the polar regions are relatively close to the ideal ones, similarly to the mid-latitudes. For instance, radiosonde temperature observations show only a slightly under-dispersive relationship over the Arctic region for both seasons (*Figure* 11). Similarly, temperature sounding microwave observations from AMSU-A channel 5 (with a weighting-function peak around 500 hPa), also suggest only minor under-dispersiveness of the EDA for the JJA period, with a behaviour for Arctic regions that is similar to the other regions (Figure 12). However, for DJF, the Arctic region shows unusually large departures. These are likely a result of larger radiative transfer errors over snow and sea-ice as mentioned earlier (e.g., Figure 6), rather than an indication of an under-dispersive EDA. Conventional humidity observations also show a reliable spread-skill relationship in the polar regions, and even some indication for an over-dispersive relationship elsewhere (not shown).

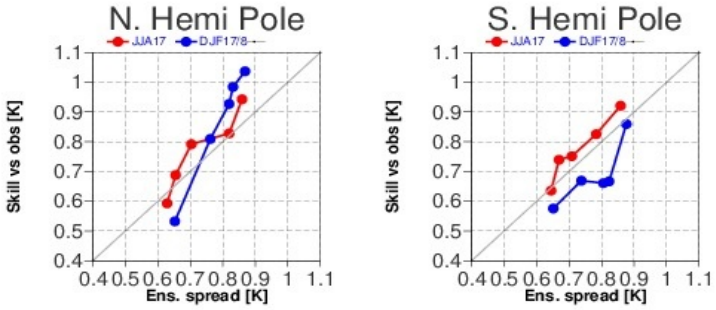

Figure 11*: Spread-skill reliability curves for radiosonde temperature at 500hPa in the operational EDA. Here only showing polar regions. Here the difference in the spread-skill relationship is less pronounced between regions (not shown here). Both periods are far closer to the ideal relationship that channels peaking higher up in the atmosphere or observations sensing further down.*

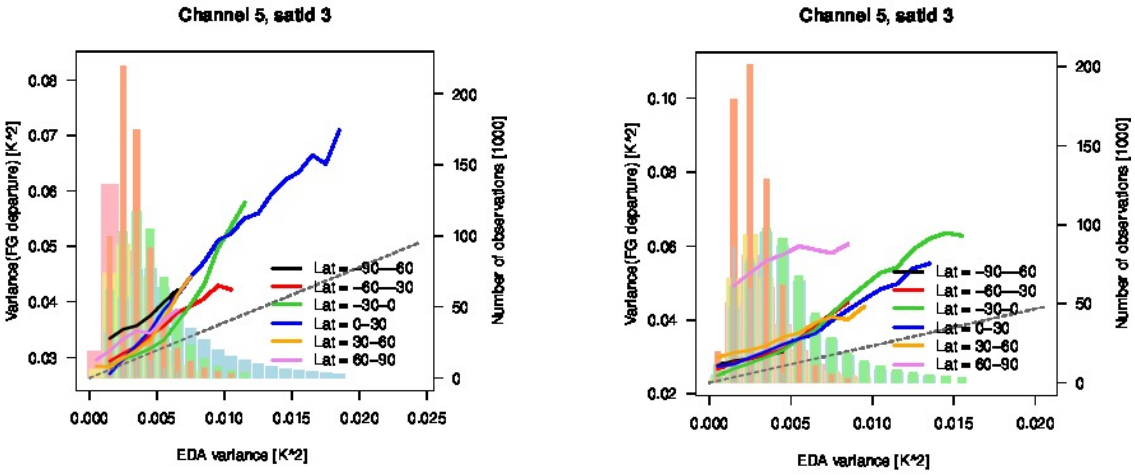

Figure 12*: a) and b) show the variance of background departures for AMSU-A channel 5 on METOP-B as a function of EDA variance (spread squared) for a) JJA 17 and b) DJF 17/8. The statistics are based on data over land and sea after cloud screening and geographical quality control but before the O-B check. They have been separated into six zonal bands: 90S-60S (black), 60S-30S (red), 30S-0 (green), 0-30N (blue), 30N-60N (yellow) and 60N-90N (pink). Also shown is the number of observations in each bin as vertical bars (right y-axis; these use lighter versions of the colours for the six zonal regions); every bin contains at least 3000 observations. In addition, for orientation, a line with the slope 1, indicative of a perfect spread-skill relationship is displayed in grey (dashed). Note that in these plots the observation error has not been subtracted from the total error shown in the x-axis.*

A different behaviour is found for stratospheric temperature. *Figure* 13b shows an under-dispersive relationship during winter for the Arctic and the northern hemisphere extra tropics (30-60N) and a good relationship for the other regions for microwave temperature-sounding observations from AMSU-A channel 10, which peak around 50 hPa. This is also suggested by other AMSU-A channels sensitive to stratospheric temperature and as well as by radiosondes, even though for radiosondes the signal is less clear probably due to more limited observational sampling (not shown). The spread-skill relationship is also under-dispersive in the mid-latitudes and the polar region in the South Hemisphere winter (Figure 13a). This under-dispersive relationship in the mid-to-high latitudes in the winter hemisphere is likely to be due to gravity waves that are not well represented in the background, which causes an increase of the background departures and hence errors in regions where gravity waves are present (not shown). It has been shown in the past that IFS errors in regions with gravity wave activity are related to gravity waves with wavelengths smaller than 100-200 km.

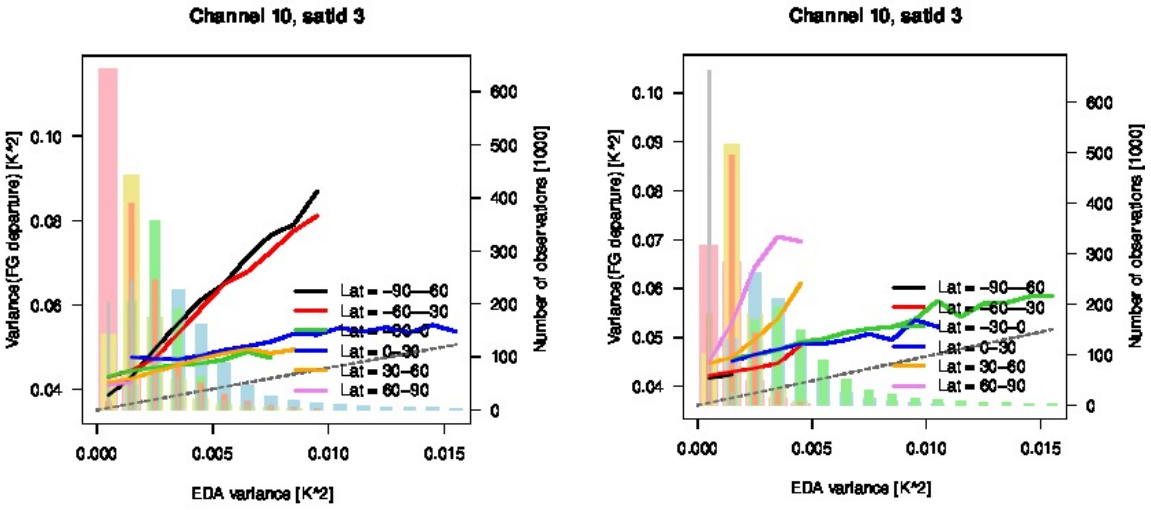

Figure 13*: a) and b) show the variance of background departures for AMSU-A channel 10 on METOP-B as a function of EDA variance for a) JJA 17 and b) DJF 17/8. The statistics are based on data over land and sea after cloud screening and geographical quality control but before the FG check. They have been separated into six zonal bands: 90S-60S (black), 60S-30S (red), 30S-0 (green), 0-30N (blue), 30N-60N (yellow) and 60N-90N (pink). Also shown is the number of observations in each bin as vertical bars (right y-axis; these use lighter versions of the colours for the six zonal regions); every bin contains at least 3000 observations. In addition, for orientation, a line with the slope 1, indicative of a perfect spread-skill relationship is displayed in grey (dashed). Note that in these plots the observation error has not been subtracted from the total error shown in the x-axis.*

3.4 Observation impact on short-range forecasts

The impact of observations on the ECMWF analyses and the short-range forecasts in the Arctic was assessed by examining a range of operationally available diagnostics from the data assimilation system. These diagnostics alone cannot determine the role of each observing system in reducing forecast errors, but they can emphasize the added value of observations in the analyses and short-range forecasts and can suggest which observations have the biggest impact, thus helping to guide the setup of the Observing System Experiments (OSE) in Task 4.2.

We start by analysing the analysis increments, which are indicative of how much the forecast is changed during the analysis process. Increments are the difference between the analysis and the short-range forecast (or background). Analysis increments could be due to a certain observation type or to a combination of different observations. To gain more insights into the contribution of different observing systems to the analysis increments, we compare the increments with the departures with respect to observations of the short-range forecasts and analysis (observation - background or O - B and observations - analysis or O - A). Finally, we assess the forecast sensitivity to observations (FSOI), which provides an estimate of the contribution of each observation type to the reduction of the global short-range forecast error.

Analysis increments

Here we investigate the effect that observations have in the Arctic by comparing analysis increments in the Arctic to those in other regions. To do so we assess the mean and standard deviation of the analysis increments. Mean increments indicate that observations cause a systematic adjustment of the short-range forecast. Data assimilation systems are designed to treat random errors and assume no biases in the observations, forward model, and the background field. Non-zero mean increments indicate, however, that biases are present and that this assumption is not true. Non-zero mean increments can be an indication of observations correcting a systematic model bias or in some cases it could be the result of systematic errors in the observations themselves, or in the forward model. Note

that a bias correction is applied to most observations during assimilation, using a variational bias correction scheme (Dee et al. 2009). However, these corrections tend to be large scale such that local instrument or forward model biases could still affect the assimilation, or the bias correction may not reliably capture specific air-mass characteristics.

There tend to be strong mean increments in the Arctic during the summer, particularly for temperature and geopotential height, when compared to sub-polar regions. During summer, the observations have thus a mean warming effect at low levels (800 – 1000 hPa) of up to 0.4 K, a cooling effect at 400 – 800 hPa (up to 0.2 K) and a strong warming effect at around 200 hPa (0.1 – 0.7 K) (Figure 14). The warming around 200 hPa corrects a known model bias. Indeed, a pronounced cold model bias is present in polar regions in the lower stratosphere, particularly in summer, due to an overestimation of the longwave cooling related to an overestimation of the amount of water vapour penetrating through the tropopause (Hogan et al. 2017). At some levels, these mean increments are largest in specific regions (i.e at 925 hPa, they are predominantly around Greenland and over the sea between North America and Russia (Chukchi sea), Figure 15), while at others they are evenly spread over most of the polar region (200 hPa) (Figure 15).

In winter, the mean increments have similar patterns over the Arctic region (a mean warming/cooling/warming for the lower/mid/upper troposphere, respectively), but the magnitude is generally lower than in summer (Figure 14) and the spatial patterns are quite different in the lower troposphere. While the increments are lower for the winter than for the summer, this does not necessarily mean that errors are smaller in the winter, however. It could simply be due to a lower number of observations assimilated in winter, or due to less atmospheric variability. There is, however, a strong 0.2 – 0.4 K warming effect at 925 hPa located over sea-ice (Figure 15 and Figure 7). This is most likely due to satellite observations since there are no conventional observations at these heights over sea-ice. In both seasons the largest standard deviation of temperature increments is found at low levels, around 800 – 1000 hPa, as shown in Figure 17, with values exceeding 0.3 K. This suggests relatively large adjustments to correct random errors in the Arctic temperature forecasts at low levels.

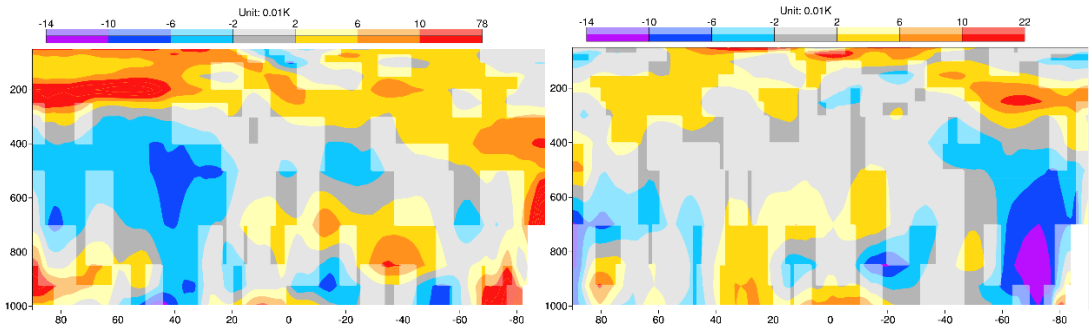

Figure 14: Mean temperature increments for JJA 2016 (left) and DJF 2017/2018, for the ECMWF operational system. Note that the North pole is plotted to the left in this and other latitude-pressure plots

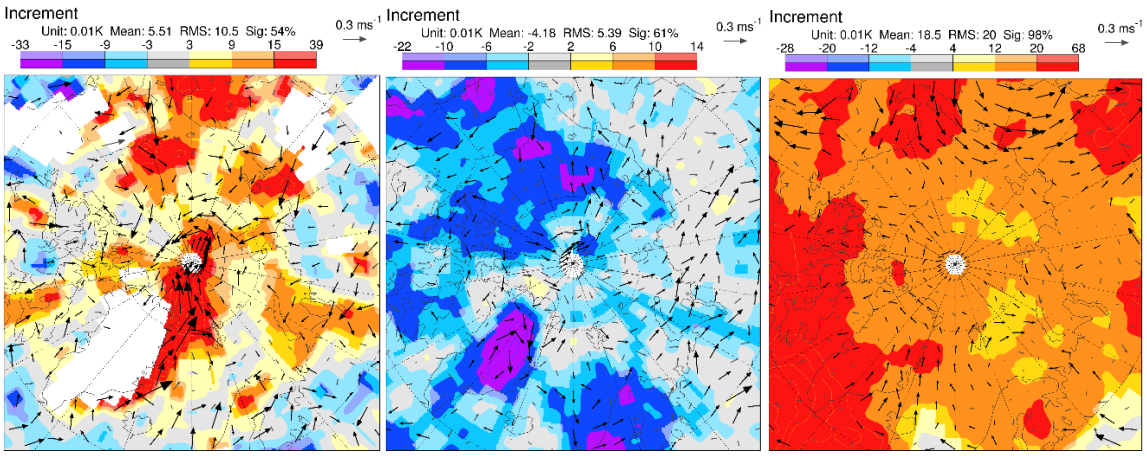

Figure 15: Mean temperature increments at (left-to-right) 925 hPa, 500 hPa and 200 hPa for JJA 2016, for the ECMWF operational system.

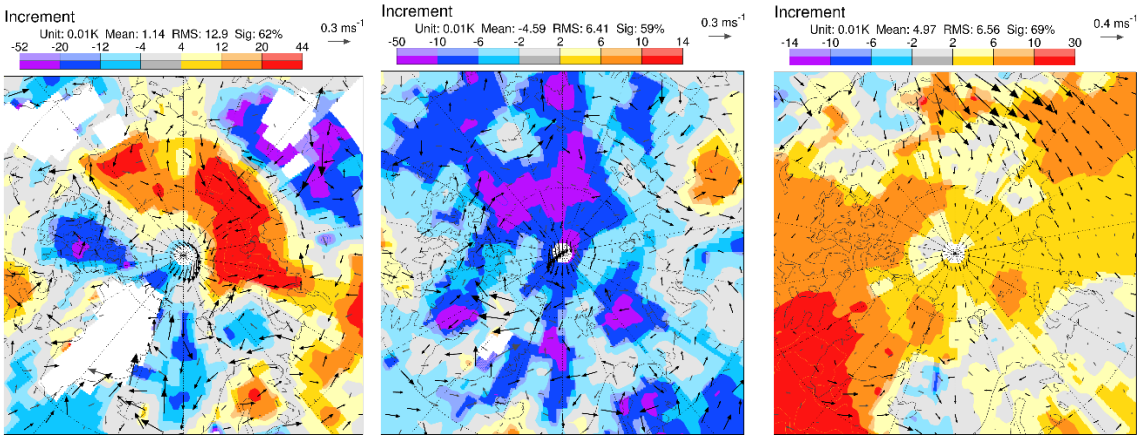

Figure 16: Mean temperature increments at (left-to-right) 925 hPa, 700 hPa and 200 hPa for DFJ 2017/ 2018, for the ECMWF operational system.

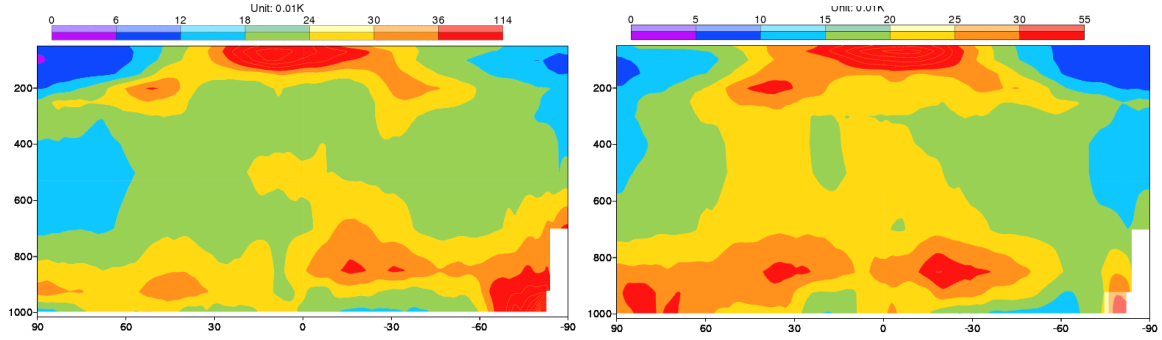

Figure 17: Standard deviation of temperature increments for JJA 2016 (left) and DJF 2017/2018 2018 (right) in the ECMWF operational system.

The observations also introduce mean changes in humidity which have a distinct global pattern, as shown in Figure 18. It is particularly apparent for the DJF period. There is generally a mean moistening effect at 600 – 1000 hPa over the poles, rising to heights of 200 – 400 hPa over the Tropics, and a drying at low levels in the Tropics. In summer, there is also a drying over the North pole at 800 – 1000 hPa, as shown in Figure 18a. There are also strong changes in the mean geopotential height (not shown), which generally have a broad-scale structure. At the winter pole, there is a mean reduction in geopotential height at 200 – 1000 hPa, of around 10 – 20 m^2/s^2, and a mean increase at heights above 200 hPa.

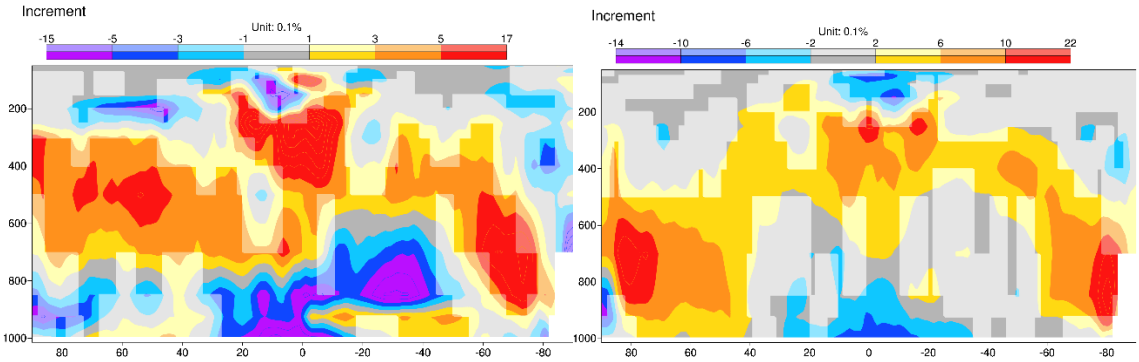

Figure 18: Mean relative humidity increments for JJA 2016 (left) and DJF 2017/ 2018, in operations.

Linking mean increments to observation departures

To better understand the role of observations in reducing short-range forecast errors, it is useful to also analyse departure statistics, such as the observation minus background (O – B) values and observation minus analysis (O – A) values. These are routinely calculated for all observations assimilated in the ECMWF system. By assessing these statistics in the Arctic, we can determine whether the analysis is closer to the observations than the short-range forecasts, at the observation locations and in observation space (i.e. radiances for satellite observations, and temperature, humidity, wind for conventional observations), and further characterise the overall consistency of the analysis.

Here we use observation departure statistics to further analyse the mean temperature increments observed in the Arctic. As mentioned earlier, non-zero mean increments are an indication that uncorrected bias is present in the assimilation, which is introduced by the observations, the observation operator, or the short-range forecasts. As data assimilation systems are designed to correct random errors, it is desirable to identify the origin of this bias, and hopefully correct for it in the future. To further analyse what is causing the mean increments, we provide here an analysis of mean O – B and mean O – A statistics.

We focus on the main temperature sensitive instruments, including AMSU-A, IASI, radiosonde and aircraft temperature data, given the strong temperature increments in the Arctic, particularly during summer. An analysis of mean O – B and mean O – A differences should provide an indication of which observations are responsible for the mean increments discussed above. Furthermore, this provides an indication of whether the increments are correcting a model bias, since if the O – B differences are similar for several types of observations this increases the confidence that they are related to a model bias rather than to an observation-related bias of a certain observation type.

In the following we assess departure statistics after bias correction is applied. Bias correction is applied to most observations, with the exception of RS92 night time radiosonde launches, GPSRO and AMSU-A channel 14 (sensitive to temperature around 1 – 5 hPa). These latter observations act as an anchor for the bias correction, and are therefore most likely to influence the mean state of the model. For satellite radiances and aircraft data, an adaptive, variational bias correction (VarBC) is applied, in which bias corrections are estimated as part of the main analysis (e.g., Dee 2004). Several aspects make a variational bias correction potentially more problematic over the polar regions: fewer anchor observations at the highest latitudes combined with larger model biases mean that VarBC may be prone to also correct model bias in these regions (e.g., Eyre 2016). In addition, AMSU-A observations have geographically varying biases (airmass biases) which tend to be extreme in the polar regions, further adding to a less well-constrained bias correction. These aspects have to be kept in mind when we interpret mean departures after bias correction and link them to mean increments, as the bias correction is likely to be less reliable in polar regions.

The mean O – B and O – A values are shown in Figure 19 for Arctic radiosondes and aircraft temperature data, and in Figure 20 for microwave (from AMSU-A) and infrared (from IASI)

temperature sounding observations for the summer (JJA) 2016 period. Values are calculated for all observations above 60N and are shown as a function of height for radiosonde and aircraft data, channel number for AMSU-A and approximate Jacobian peak height for MetOp-B IASI. The Jacobian peak height is also indicated for AMSU-A channels. While the Jacobian peak height indicates the height at which each channel is most sensitive, the weighting functions tend to be broad such that the channels have sensitivity both above and below their peak heights. Furthermore, the lowest peaking channels have sensitivity to the surface, including channel 5 of AMSU-A which peaks at 500 – 700 hPa.

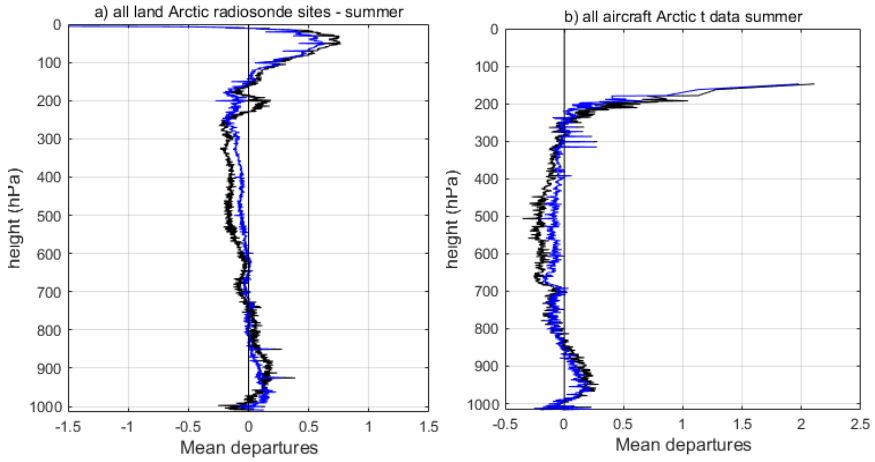

Figure 19: Mean O – B (black) and mean O – A (blue) for a) land radiosonde data assimilated in the ECMWF operational system, b) aircraft temperature observations. Values are averaged for the summer 2016 period (JJA) and shown after bias correction.

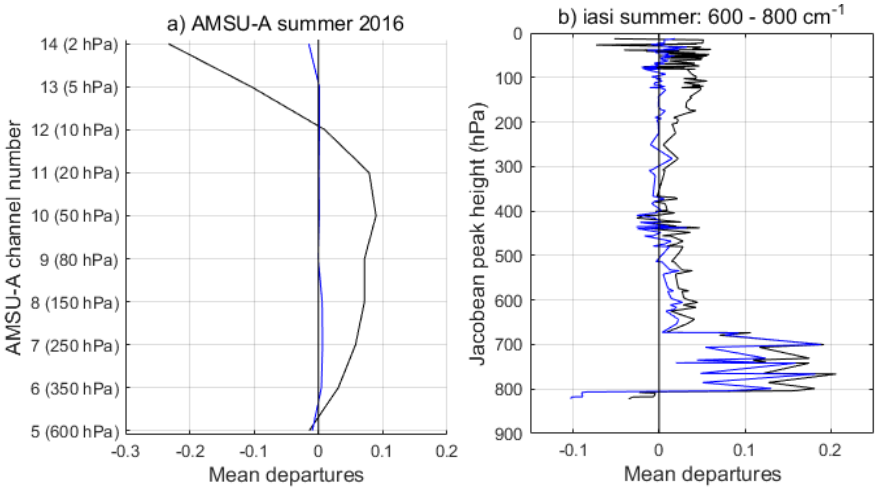

Figure 20: Mean O – B (black) and mean O – A (blue) for a) AMSU-A assimilated data, and b) IASI assimilated data. Values are averaged for the summer 2016 period (JJA) and shown after bias correction.

The mean O – A values are closer to zero than mean O – B for all observation types (Figure 19 and Figure 20), indicating that the analysis is closer to the observations in the Arctic than the short-range forecast, as one would expect of a well-functioning data assimilation system. This is also consistent for the winter period (not shown). Moreover, it appears that the summer temperature increments in the Arctic (Figure 14) are broadly similar to the mean O – B values of the radiosonde and aircraft data between the surface and 300 hPa (Figure 19). That is, the observations act to warm the 800 - 1000 hPa layer and cool the 300 to 800 hPa layer. It is worth noting, however, that surface pressure observations are also likely to introduce temperature increments at 800 – 1000 hPa, and there are more surface pressure observations than radiosonde data in the locations of the 925 hPa temperature increments (Figure 15a). Around 200 hPa, the mean O - A for aircraft temperature observations are approximately

0.3 - 0.7 K, while the mean O - B are around 0.7 - 1.0 K, suggesting that the aircraft observations act to warm the analysis around this height. The radiosonde data also show a mean positive O – B and a very small O - A at 200 hPa, but the signal is smaller than for the aircraft observations. This suggests that the aircraft observations are likely to be one of the main causes of the mean positive temperature increments for the summer season at 200 hPa (Figure 15c).

It is more difficult to compare the satellite (AMSU-A and IASI) summer O – B values to the increments due to the low vertical resolution of the satellite radiances in comparison to the high vertical resolution of the increments. One can compare them, however, to the radiosonde data by first transforming the radiosonde temperature O – B values into radiance space. This has been done using the GRUAN processor, a tool developed at the Met Office during the GAIA-CLIM Horizon-2020 project for use with satellite calibration/validation (cal/val) (Carminati et al, 2018). The GRUAN processor transforms O – B values into radiance space using the radiative transfer forward model RTTOV, transforming also the measurement uncertainties into radiances. The GRUAN station Sodankyla is situated in the Arctic and has O – B temperature values which are representative of the Arctic mean values, as shown in Figure 21a. The mean O – B for Sodankyla was transformed into radiance space, and is shown in Figure 21b and c for AMSU-A and IASI respectively. (Note that higher peaking AMSU-A channels have been excluded since they have sensitivity above the radiosonde top height).

The results show that the mean positive O – B values in the radiosonde data above 250 hPa translate to a bias of around 0.2 - 0.8 K for AMSU-A and IASI channels peaking around 50 - 300 hPa (Figure 21b and c). Qualitatively, this is consistent with the observed positive O - B values for these channels shown in Figure 20a and c. The consistency suggests that the observations act to reduce a model bias here. However, the magnitude of the bias indicated by AMSU-A and IASI is much smaller than what would be consistent with the radiosonde data. This is partly because the variational bias correction inadvertently also corrects for model bias, despite the anchoring provided by radiosondes and GPSRO observations at the relevant levels (e.g., Eyre 2016). Consistent with this interpretation, AMSU-A observations before bias correction show departures that are more in line with the radiosonde differences translated to radiance space (not shown). This suggests that the effect of some systematic model error is likely still present in the analysis.

The main inconsistency between satellite observations and radiosonde data is for the lowest peaking IASI channels which show a positive ‘spike’ in mean O – B for channels peaking around 700 – 800 hPa (Figure 20b) which is not present in the radiosonde statistics translated into IASI radiances (Figure 21c). The positive mean O – B difference for the lowest peaking IASI channels is therefore likely due to undetected cloud. This does not appear to affect the mean increments however since, as shown in Figure 20b, the analysis departures also show this spike.

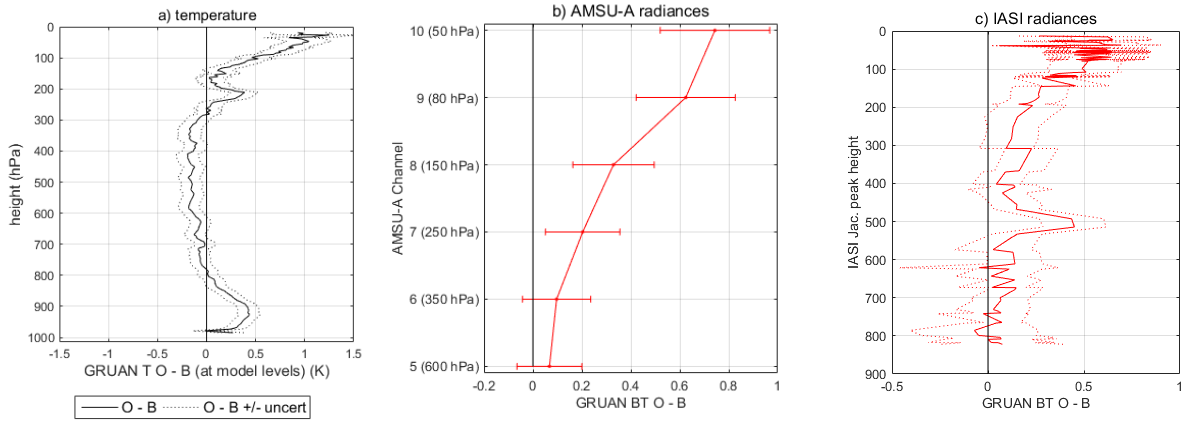

Figure 21: mean O – B for GRUAN temperature profiles at the Sodankyla site, calculated for JJA 2016. Values are shown in temperature (a) and transformed into radiances for AMSU-A (b) and IASI (c). The dashed lines and error bars indicate the GRUAN uncertainties which are in temperature for (a) and radiances for (b) and (c). The transformation into radiance space was performed using the GRUAN processor, which uses the RTTOV radiative transfer forward model.

The mean temperature increments in the winter period are generally smaller than in summer, with the exception of a mean positive increment found over sea-ice at heights of around 800 – 1000 hPa (Figure 16a). There are no conventional data at these heights over sea-ice and so these increments must be due to satellite data. Indeed, positive mean O - B are observed for both the IASI and AMSU-A lower peaking channels, as shown in Figure 6b and Figure 9b. As discussed earlier, these mean O - B values may also be the result of biases in the forward model or the cloud screening, so it is not clear to what extent the analysis correctly identifies and corrects a model bias here. Interestingly, while the positive temperature increments act to draw the analysis closer to the observations, a positive mean analysis departure remains, as would be expected in the presence of an observation-related bias (not shown). It is likely that forecast model biases as well as observation-related biases are higher over sea-ice, as previously discussed, and this makes the partitioning of mean O - B information into background and observation-related biases difficult. Conventional observations with no observational biases would help, but unfortunately these are not available over sea-ice.

Forecast Sensitivity to Observation Impact

One measure of observation impact is the Forecast Sensitivity to Observation Impact diagnostics or FSOI. FSOI uses adjoint methods to estimate the contribution of individual observations to the reduction in the short-range forecast error (24 hours) as measured by a global energy norm when all observations are present (Cardinali, 2009). The measure is therefore different from the observation impact provided by traditional observing system experiments, in which observations are withheld to investigate the influence on subsequent forecasts. The FSOI relies on a number of assumptions, such as a linear error propagation, and uses the analyses of the forecast system as a reference. The FSOI can also be formulated with an observation error norm (Cardinali 2018). However, both 4D-Var, and OSEs based on 4D-Var, rely on the same assumptions, namely a perfect model in data assimilation, an error norm that is either derived from analyses or observations as references, the projection of cost to initial time sensitivities in model space using adjoint models and to observation space using observation operators.

By design, FSOI only evaluates the influence on short-range forecasts, and it hence cannot capture the long-range propagation of forecast errors as typically found in medium-range forecasts. The latter are, for instance, relevant for investigating how better analyses over the Arctic affect forecasts over the mid-latitudes beyond day 3. Despite these limitations, the measure tends to give useful indications of leading contributors to forecast skill.

The FSOI for various observation types assimilated in the Arctic (Figure 3) is shown for JJA 2016 and DJF 2017/2018 in Figure 22. For comparison, Figure 23 shows the FSOI for the same types of observations assimilated at global scale. The relative FSOI values shown in these figures can be interpreted as follows. A relative FSOI value of 3% for microwave Arctic observations indicates for example that the microwave observations assimilated in the Arctic act to reduce by 3% the total global forecast error - as this is the norm FSOI is calculated against. Similarly, a relative FSOI value of 40% for the microwave observations assimilated over the entire globe act to reduce by 40% the total global forecast error.

The three observation types appearing to have most impact, both globally and in the Arctic, are the microwave, conventional and infrared observations. In the Arctic, the microwave observations have the highest impact in the summer period and the conventional observations in the winter period, whereas at global scale microwave observations have the largest impact for both periods. Especially during the winter season, the relative contribution of conventional observations (synoptic stations, buoys, radiosondes) to reducing the global short-range forecast error in polar areas is higher than at global scale. This is likely due to the fact that fewer satellite observations are assimilated in the lower troposphere during winter in these areas, so that conventional observations provide most of the information.

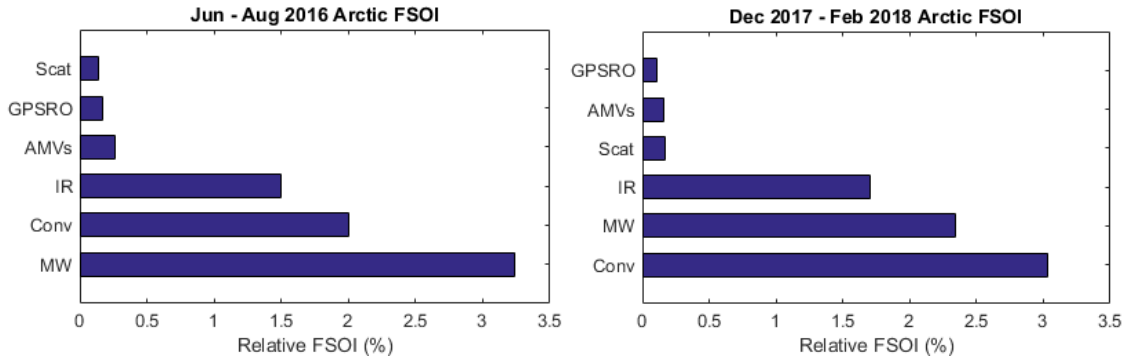

Figure 22: Relative FSOI for different observation types assimilated in the Arctic (latitudes > 60N). These values represent by how much (in %) each observation type contributes to reduce the global short-range forecast error. Values are shown for a) JJA 2016 and b) DJF 2017/ 2018. Values are calculated for experiments for this period using the full ECMWF observing system.

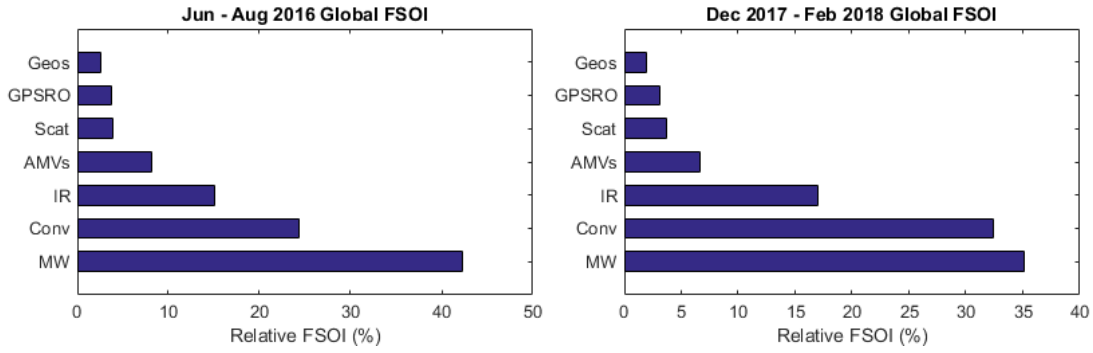

Figure 23: As Figure 22 but for the observations assimilated over the entire globe.

Complementary FSOI analyses in which these observation types were broken down even further (into types of satellite instruments, or various types of conventional observations) showed that the largest contributors in reducing forecast errors in polar regions are microwave temperature sounders, and the SYNOP conventional observations (mainly surface pressure).

4. OCEAN-SEA ICE REANALYSES

Ocean-sea ice reanalyses (ORAs) are crucial for understanding climate variability and predictability in the Arctic Ocean. The evaluation made under task 4.1 is an extension of recent intercomparisons made under the projects ORA-IP (e.g. Balmaseda et al., 2015 ; Chevallier et al., 2017) and PORA-IP (Uotila et al., 2018).

The most widely used ORAs are generally produced using forcing data from atmospheric reanalyses, either in forced or in coupled mode. Since all atmospheric reanalyses have been produced using a prescribed sea ice cover, the atmospheric forcing already contains information on sea ice concentration, which primarily constraints the simulated sea ice cover in the ORAs. The quality of atmospheric reanalyses has been evaluated by Jakobson et al. (2012) against observations obtained in the central Arctic Ocean during the International Polar Year (2007-2008), and by Lindsay et al. (2014) over the whole Arctic region, including land and oceanic areas.

4.1 What observational data is assimilated in current ocean-sea ice reanalyses?

ORA systems assimilate satellite SST, sea level anomaly data from satellite altimetry and in-situ temperature and salinity profiles. In-situ profiles typically include Argo floats, ice-tethered profiles (ITPs) and a large collection of hydrographic cruise data. In the Arctic Ocean, the perennial sea ice cover has long hindered the deployment of Argo-buoys. Thus, most available data for Arctic Ocean hydrography under sea ice comes from ITPs or mooring data (e.g. from the Beaufort Gyre Exploration Project). Due to the sparsity of those data, the constraints in the Arctic Ocean is rather weak. However, data assimilation in data-rich areas like the Northern Atlantic Ocean impacts the Arctic Ocean hydrography due to ocean dynamics and transports (e.g. through Fram Strait and Barents Sea Opening).

Assimilation of SST, as well as the effect of restoring of the sea ice cover toward observed conditions through the forcing applied from atmospheric reanalyses provide a first-order constraint on the sea ice presence in ORAs. However, this restoring is not strong enough to constrain the ice-edge completely (e.g. in the summer) and the sea ice concentration in the pack. Most systems now directly assimilate sea ice concentration (SIC) data derived from satellite passive microwave imagers (SMMR, SSM/I, AMSR-E). Those data are typically processed by NSIDC or the Eumetsat Ocean-Sea Ice Satellite Application Facility (OSI-SAF). So far, the TOPAZ4 regional reanalyses is the lone example of system that routinely assimilates both SIC and sea ice velocity data (Sakov et al., 2012). To date, no ORA system directly assimilates sea ice thickness (SIT) data, and the SIT field is usually not well-constrained during the assimilation process (e.g. Chevallier et al., 2017).

4.2 Arctic sea ice thickness

In the scope of Task 4.1, we have compared fourteen state-of-the-art reanalyses in order to understand the added value of SIC assimilation for the SIT field. Monthly fields of SIT from the following reanalyses are used: **C-GLORS05**, **ECCO-v4**, **ECDA**, **G2V3**, **GECCO2**, **GloSea5**, **GloSea5-GO5**, **MERRA-Ocean**, **MOVE-CORE**, **MOVE-G2**, **ORAP5**, **PIOMAS**, **TOPAZ4** and **UR025-4**. From this ensemble, all reanalyses assimilate SIC data, except **ECDA**, **GECCO2**, **MOVE-CORE** and **MOVE-G2**. Table 2 provides useful information on ORA's providers, horizontal resolution, time span and SIC data assimilation.

Table 2: Ensemble of reanalyses used in Section 4 and their respective providers. Some system configurations are also displayed: horizontal resolution, whether nor not the reanalyses assimilate sea ice concentration data and time span of each product considered in our study. The respective reference(s) for each system is(are) also indicated. For a detailled overview on the reanalyses' specifications and parameters the reader is referred to Chevallier et al. 2017.

Reanalysis	**Provider**	**Horizontal Resolution**	**Sea Ice Data Assimilation**	**Time Period**	**Reference**
C-GLORS05	CMCC	0.5°	Yes	1979–2011	Storto et al., 2014
ECCO-V4	JPL/NASA, MIT, AER	0.4°–1.0°	Yes	1992–2010	Forget et al., 2015
ECDA	GFDL/NOAA	1°	None	1961–2014	Zhang et al., 2013; Chang et al., 2013
G2V3	Mercartor Océan	0.25°	Yes	1993–2011	Ferry et al., 2010
GECCO2	University of Hamburg	1°	None	1948–2011	Köhl, 2015
GloSea5	UK Met Office	0.25°	Yes	1993–2012	Blockley et al., 2014
GloSea5-GO5	UK Met Office	0.25°	Yes	1990–2014	Megann et al., 2014
MERRA-Ocean	GSFC/NASA/ GMAO	0.5°	Yes	1983–2011	Rienecker et al., 2011
MOVE-CORE	MRI/JMA	0.5°x1°	None	1948–2007	Danabasoglu et al., 2014
MOVE-G2	MRI/JMA	0.3°–0.5°x1°	None	1993–2012	Toyoda et al., 2013
ORAP5	ECMWF	0.25°	Yes	1979–2012	Zuo et al., 2015; Tietsche et al., 2017
PIOMAS	APL/PSC	Mean of 22 km - Arctic	Yes	1979–2015	Zhang and Rothrock, 2003
TOPAZ4	ARC MFC	12-16 km	Yes	1993–2013	Sakov et al., 2012
UR025-4	University of Reading	0.25°	Yes	1989–2010	Valdivieso et al., 2014

Our first goal in performing such a comparison was to identify whether or not systems built with assimilation of SIC data are closer to observations in terms of SIT, in contrast to the products built with no sea ice data assimilation. To do so, we have extended the evaluations carried out by Chevallier et al. (2017) and Uotila et al. (2018), and compared the reanalyses against a variety of sources of observations, from moored upward-looking sonars, to submarines, airbornes, satellites and ice boreholes. Fig. 24 shows an example where the SIT difference between reanalyses and ICESat satellite campaigns are plotted. Positive (negative) bias represent an overestimation (underestimation) of the respective reanalysis products regarding the observational product. Overall, most of the reanalyses underestimate the satellite measurements off the Canadian Archipelago and central Arctic. For some of the reanalyses (e.g., **GECCO2, TOPAZ4** and **UR025-4**), the match between modelled and observed values seems to increase across the satellite campaigns (from Laser1 to Laser3i).

The scatter plots shown in Figure 25 combine SIT from each reanalysis and the observational datasets from all sources. The latter are separated into two parameters: sea ice draft (which is the depth of the immersed part of sea ice; black dots) and SIT (green dots). Complementary metrics are employed to evaluate the relationship between measured and modelled values. When directly comparing SIT from reanalyses and observations, we estimate the Root Mean Squared Error (RMSE), the correlation coefficient (R) and the Mean Residual Sum of Squares (MRSS) from the linear fit between both datasets, by having the reanalysis values as predictors and the observational values as predicted variables. When comparing SIT from reanalyses with draft from observations, we estimate R and MRSS. Though SIT and draft are different variables, we consider the linear relationship between both parameters given by the hydrostatic equation, but we do not account for snow variation in order to avoid adding uncertainties.

The results from the different metrics are compiled in Table 3. Numbers highlighted in green and red represent the three cases of best and worst correspondence between reanalyses and observations, respectively. Some reanalyses appear to perform better in terms of the evaluated metrics compared to the others. **C-GLORS05** and **PIOMAS** have 3 out of 5 metrics classified among the best performing systems, while **TOPAZ4** has 4 out of 5 in the top list. The **TOPAZ4** case does not come as surprise, since apart from being a regional product, this is the only reanalyses which assimilates sea ice velocity data. Nevertheless, the SIC data assimilation does not necessarily improve the reproduction of SIT data in all reanalyses (see Table 3). One could speculate that in these cases, the reanalyses do not take the best advantage of the covariances between SIC and SIT, maybe due to the assimilation scheme used to produce the reanalyses.

Table 3: Correlation Coefficient (R) and Mean Residual Sum of Squares (MRSS; in m2) calculated between the SIT from reanalyses and observations, either Draft of SIT. MRSS is estimated from the linear fit having reanalyses as predictors and observations as predict variable. The Root Mean Squared Error (RMSE; in m) is calculated only between SITs. Numbers highlighted in green and red represent the three cases of best and worst correspondence between reanalyses and observations, respectively. The star (*) indicates the reanalyses which do not assimilate sea ice concentration data.

Reanalyses	**Draft**		**Sea Ice Thickness**		
	R	**MRSS (m^2)**	**R**	**MRSS (m^2)**	**RMSE (m)**
C-GLORS05	0.74	0.49	0.58	0.47	0.8
ECCO-V4	0.66	0.51	0.4	0.7	1
ECDA*	0.53	0.95	0.43	0.54	1
G2V3	0.63	0.55	0.55	0.5	1.1
GECCO2*	0.17	1.27	0.64	0.42	0.9
GloSea5	0.54	0.76	0.46	0.54	0.9
GloSea5-GO5	0.63	0.64	0.51	0.52	0.8
MERRA-Ocean	0.1	1.42	0.53	0.5	1
MOVE-CORE*	0.74	0.73	0.47	0.71	0.9
MOVE-G2*	0.59	0.61	0.3	0.62	1.1
ORAP5	0.65	0.74	0.58	0.46	0.8
PIOMAS	0.61	0.75	0.66	0.41	0.7
TOPAZ4	0.76	0.39	0.61	0.41	0.9
UR025-4	0.73	0.42	0.46	0.66	1.1

4.3 Correlation scales of Arctic sea ice thickness

Our second goal was to analyze the impact of SIC assimilation in two important aspects of the SIT variability: the time scale (or persistence) and the length scale of SIT anomalies. The importance of these aspects is reinforced by the fact that the variability and predictability of the SIT field depends on how long the anomalies persist over time and on the spatial scale of the respective anomalies.

In this case, sea ice data assimilation plays a clear role in the referred scales: systems with assimilation of SIC data are characterized by shorter time (Figure 26) and length scales (Figure 27) compared to the systems which do not assimilate sea ice data. The mean time and length scales for reanalyses with data assimilation vary in the range 2.5–5.0 months and 337.0–732.5 km, respectively, while reanalyses with no data assimilation are characterized by values in the range 4.9–7.8 months and 846.7–935.7 km, respectively. Likely, the main reason why this happen is linked to the fact that when a reanalysis assimilates sea ice information, the system is forced towards the assimilated conditions, differently from what occurs with free-running models. It might also be that time and length scales are actually much shorter in the real world than in free-running models, and that assimilation brings these scales towards more realistic values. However, such assessment in the real world is difficult to produce because observational datasets are limited both in time and space.

Finally, data assimilation introduces SIT increments that are not necessarily physical, and so contributes to attenuating the correlation of this variable at a certain grid cell both in time, with their future estimations, and in space, with the neighboring grid points. Which explanation is the correct one is a question to be investigated. In any case and interestingly, for the reanalyses with ice data assimilation, an increase in mean ice thickness seems to lead to an increase in time and length scales, also agreeing with the fact that higher time and length scales are associated with slower sea ice velocities (Figure 28).

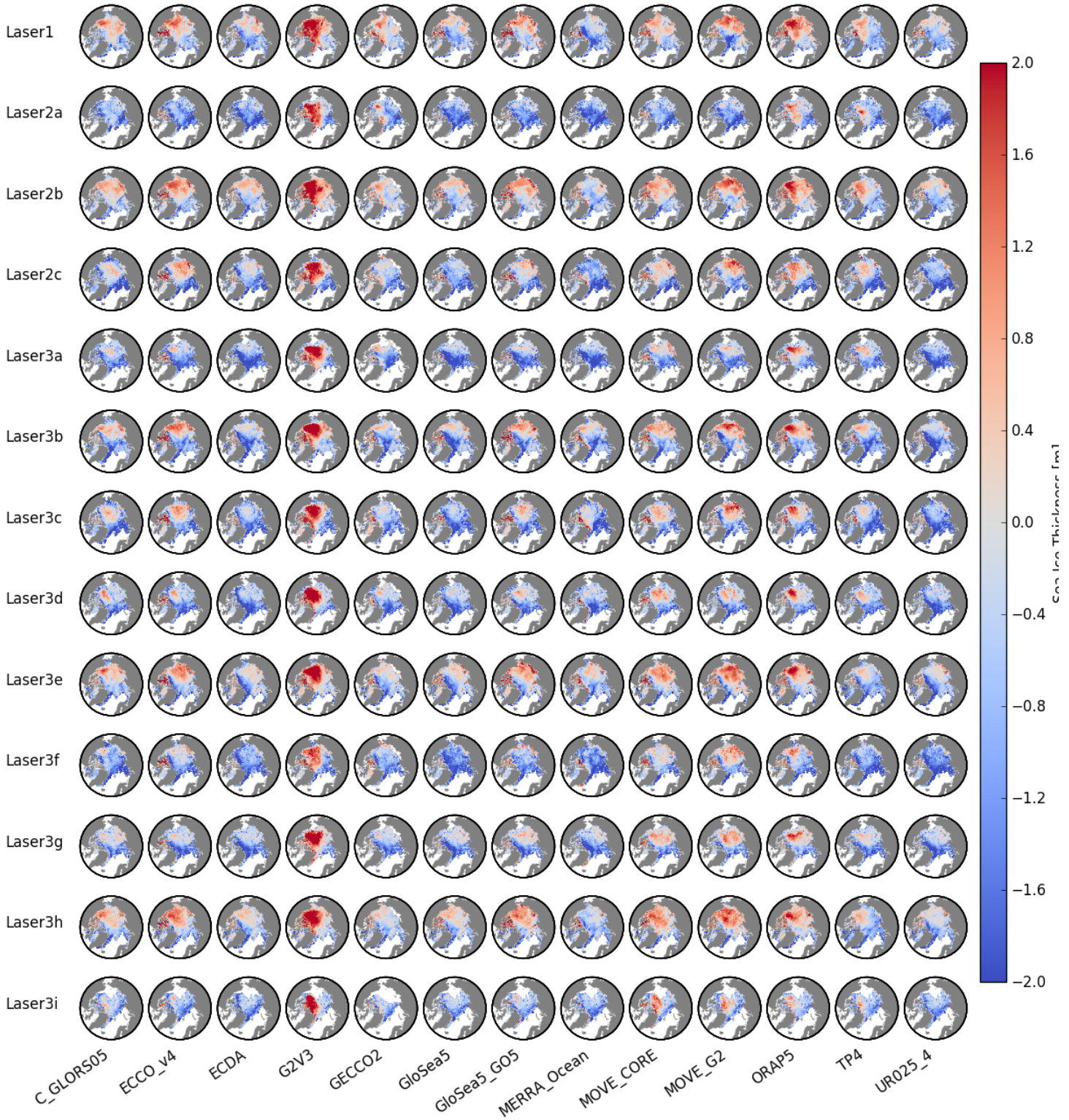

Figure 24: Sea Ice Thickness difference between reanalyses and ICESat satellite campaigns. Positive (negative) bias represent an overestimation (underestimation) of the respective reanalysis products regarding the observational product. All reanalyses assimilate sea ice concentration data, except ECDA, GECCO2, MOVE-CORE and MOVE-G2.

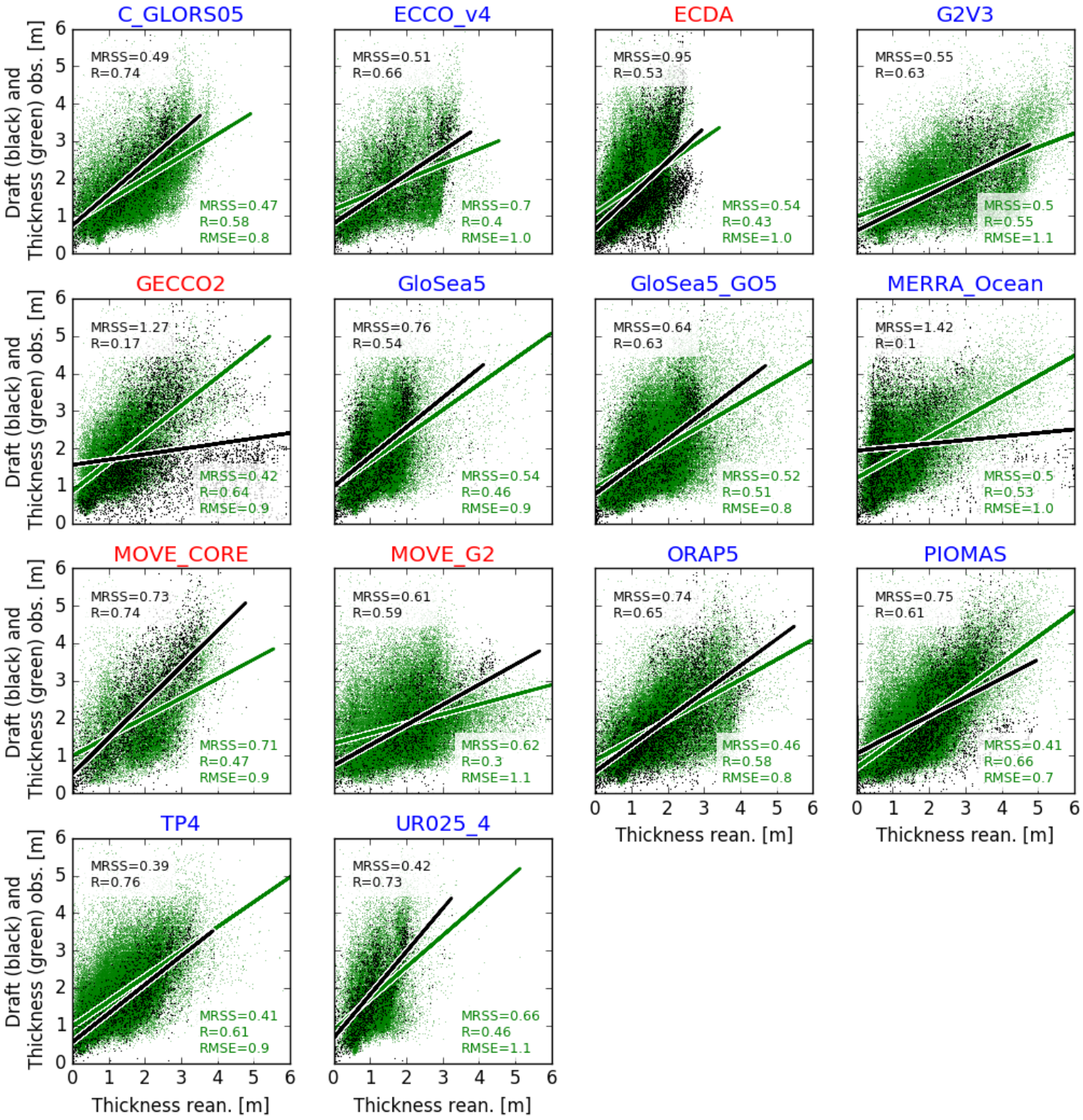

Figure 25: Comparison between sea ice thickness from reanalyses and sea ice thickness (green points) or draft (black points) from observational datasets. The lines represent the linear fits having the reanalysis as the predictor and the observations as predicted variables. The Mean Residual Sum of Squares (MRSS) from the fit, the correlation coefficient (R) and the Root Mean Squared Error (RMSE) are also displayed for each comparison. RMSE is calculated only when comparing SIT from both sources (green), but not when comparing SIT and draft (black). Reanalyses labeled in blue and red highlight whether the datasets were built with or without sea ice data assimilation, respectively.

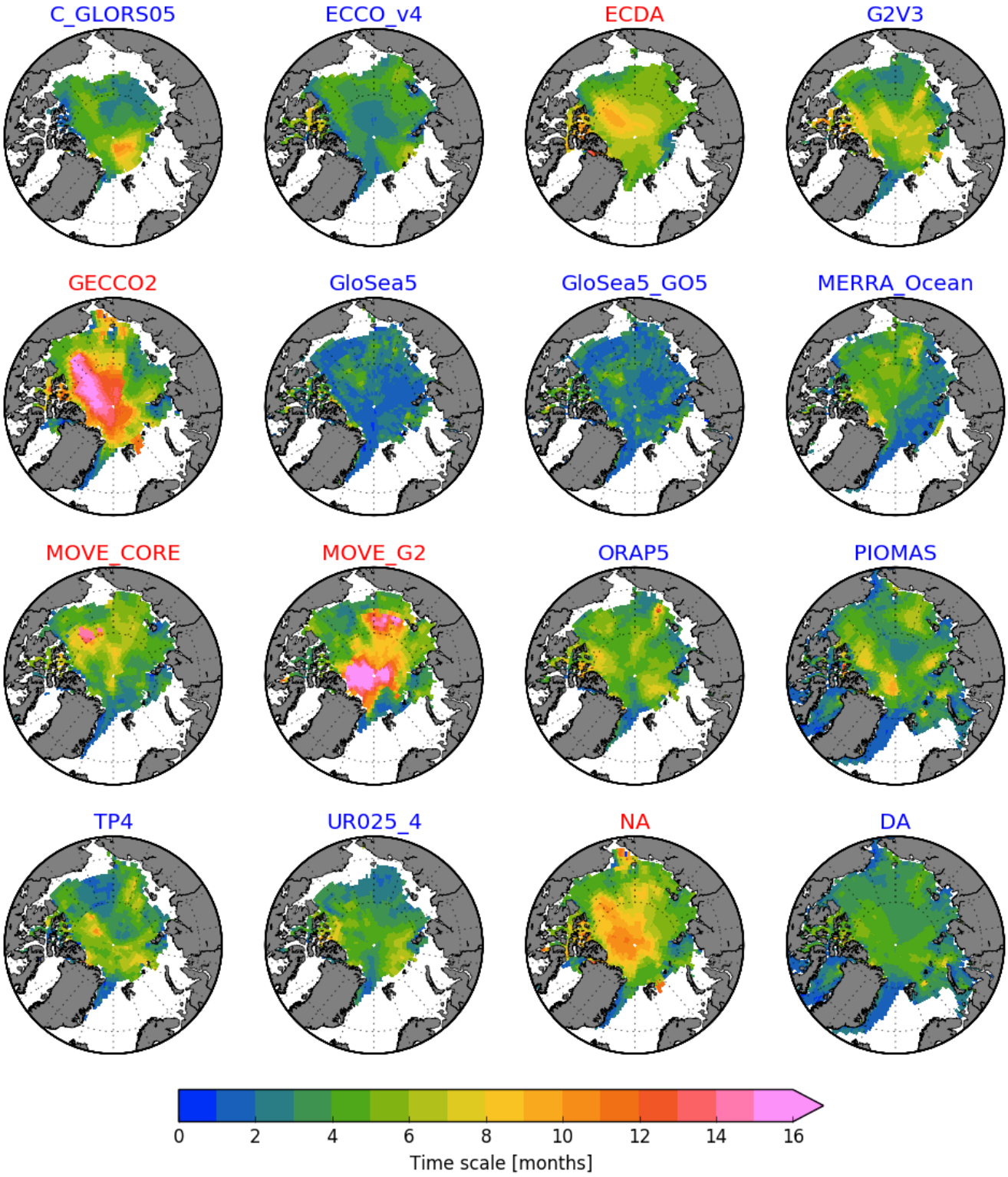

Figure 26: E-folding time scales (or persistence) estimated for the SIT time series. Only grid cells in which the time-mean (for the Jan/1993-Dec/2007 period) SIT at the time of summer minimum is greater than 0.1 m are taken into account for the computations. Averages for the systems with ice Data Assimilation and No data Assimilation are represented by the NA and DA panels. All maps have the 0°-longitude placed at 6-o'clock, while the bounding latitude is 67°N. Reanalyses labelled in blue and red highlight whether the datasets were built with or without ice data assimilation, respectively.

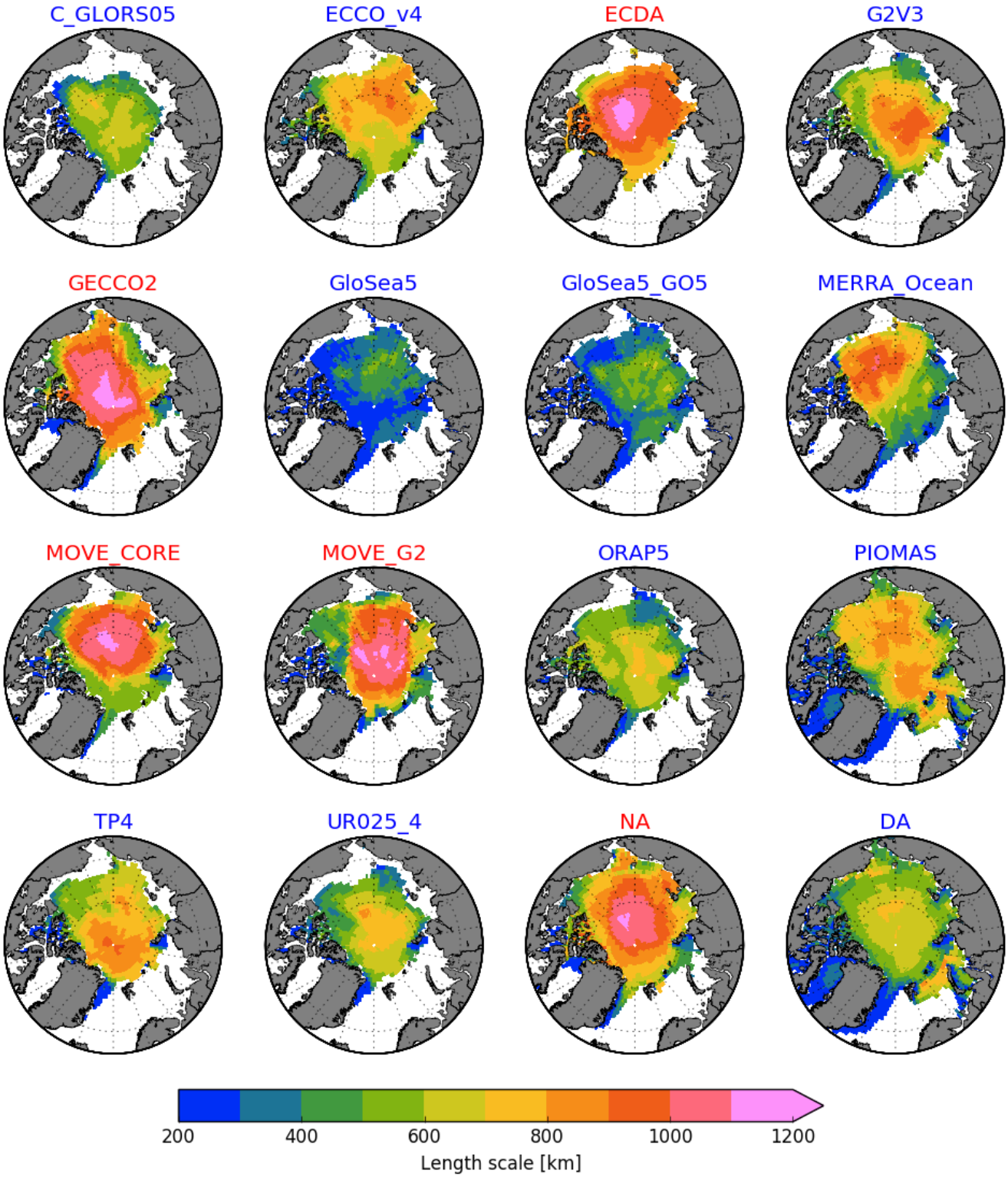

Figure 27: E-folding length scales estimated for the SIT time series. Only grid cells in which the time-mean (for the Jan/1993-Dec/2007 period) SIT at the time of summer minimum is greater than 0.1 m are taken into account for the computations. Averages for the systems with ice Data Assimilation and no Data Assimilation are represented by the NA and DA panels. All maps have the 0◦-longitude placed at 6-o'clock, while the bounding latitude is 67◦N. Reanalyses labelled in blue and red highlight whether the datasets were built with or without ice data assimilation, respectively.

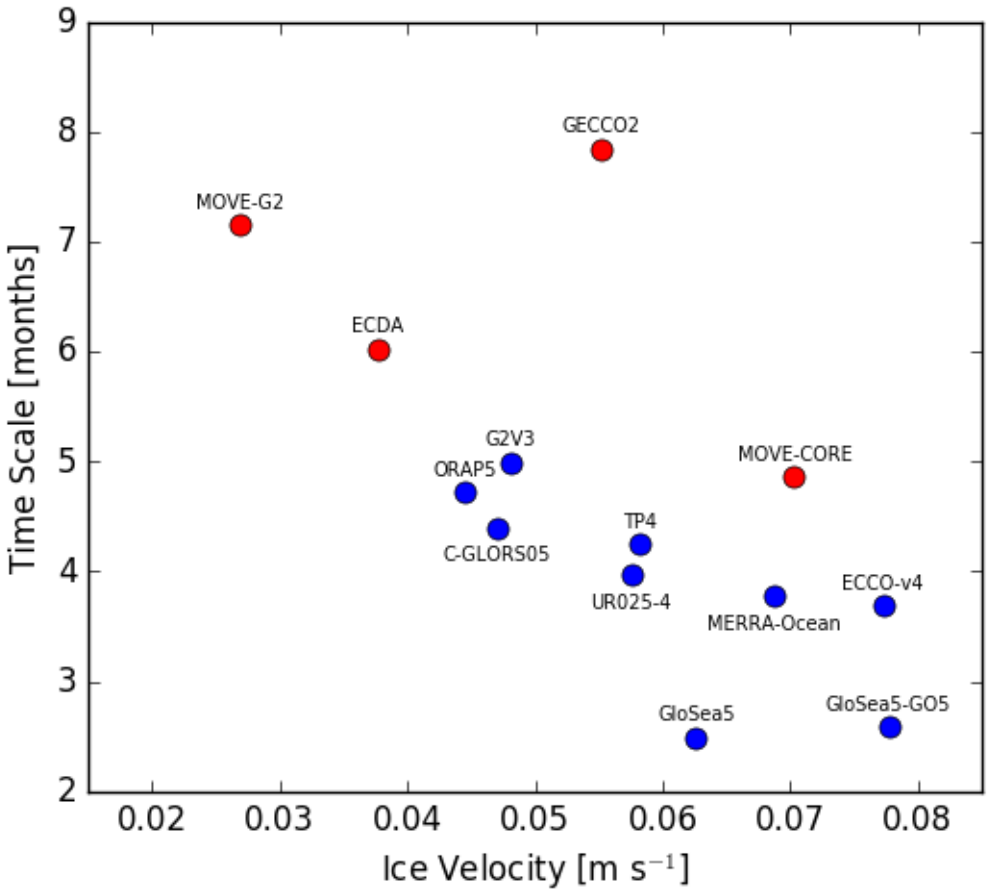

Figure 28: Scatter plot of the Average Weighted Mean (AWM) sea ice velocity versus the AWM time scale. Reanalyses in blue and red highlight whether the datasets were built with or without ice data assimilation, respectively.

4.4 Oceanic properties in the Arctic Ocean in ocean reanalyses

Under the task 4.1, the APPLICATE team has been involved in a first-time-ever intercomparison of hydrographic properties in the Arctic Ocean as reconstructed in global and regional reanalyses.

The study targeted the simulated mixed layer depth, liquid ocean transports through Arctic gateways (Fram Strait, David Strait, Bering Strait and the Barents Sea Opening), ocean heat and salt contents and mean properties of sea temperature and salinity at depth. The emphasis has been put on the performance of the multi-model mean, namely the unweighted mean of all 10 reanalyses considered.

A summary of performance of the set of ORAs considered in the study is provided in Uotila et al. (2018 ; their Table 3), as well as an evaluation of the multi-model mean. We summarize here the key results from this study for the ocean properties.

Comparison with a common reference shows that ORAs overestimate the mixed layer depth in the winter, and underestimate mixed layer depth in the summer. The winter bias is the largest in the Nordic and Barents Seas, where the difference between the reanalyses and the observation-based products locally exceeds 400m. In summer, all ORAs have too shallow mixed layers everywhere in the Arctic Ocean, which is a common bias in coupled and forced models without data assimilation (e.g. Barthélémy et al., 2015).

All ORAs have a colder than observed Atlantic layer in the Arctic Ocean, defined as the warm layer at 300-700m depth. Warm Atlantic Water (AW) penetrates the Arctic Ocean through two pathways : one branch enters through Fram Strait ; another branch through Barents Sea Opening. The AW that enters

the Barents Sea densifies due to loss of heat and release of salt during sea ice formation. Dense shelf water flows into the Arctic interior through submarine valleys (e.g. St. Anna Trough ; Ilicak et al., 2016). Free (ie : without data assimilation) ocean-sea ice models from the CORE-II intercomparison have a large spread in temperature – cold and warm biases – at depth in the Arctic Ocean. Warm bias models have too warm AW inflow entering through Fram Strait, while cold bias models have excessive cold water forming in the Barents Sea entering through the St. Anna Trough (Ilicak et al., 2016).

In the ORAs, ocean heat transport into the Arctic Ocean through Fram Strait is within the observational uncertainty, while on the low side. Moreover, winter heat loss in the Barents Sea is high and many ORAs unrealistically convect to the bottom. Both processes act towards a negative temperature anomaly in the Arctic Ocean interior. One explanation for the too low ocean heat transports through Fram Strait could be the too deep mixed layer depths in the northern North Atlantic, which results in a cooling of the warm AW priori its entrance into the Arctic.

Interestingly, Uotila et al. (2018) show that ocean biases can be related to the mean sea ice state, which acts as a driver of the ocean state which is weakly constrained by ocean observations. For instance, the winter bias of mixed layer depth can be linked to the sea ice biases; in spite of data assimilation, multi-model mean winter sea ice cover is too extensive, while summer sea ice is close to observed. This implies that the simulated sea ice has a larger than observed seasonal amplitude, thus increased formation and melt. Hence during the freezing period more salt is rejected to the upper ocean, thus destabilizing it more than ORAs than in observations. Shallow summer mixed layer in all ORAs could be explained by missing or poorly represented mixing processes under sea ice, rather than by an observationally-based process.

Uotila et al. (2018) also attempted a ranking of the ten ORAs, based on a synthesis of the evaluation of all ORAs for all sea ice and hydrographic diagnostics. Interestingly, for the Arctic Ocean, ORAP5 (ECMWF) and GloSea5-GO5 (Met Office) rank better than the multi-model mean and the other ORAs. Those reanalyses will be used in APPLICATE for the initialization of seasonal forecasting experiments. Note however, that globally – accounting for both Arctic and Antarctic – the multi-model mean ranks better than all individual ORAs.

5. CONCLUSIONS

The work carried out in Tasks 4.1 assessed existing state-of-the-art datasets such as analyses and reanalyses in the Arctic, particularly focusing on the added value of current observational systems. This assessment helps to provide guidance for future reanalyses and numerical experimentation performed in Task 4.2. and in the YOPP consolidation phase.

Operational diagnostics from the ECMWF data assimilation system were used to explore the usage of Arctic observations in the ECMWF IFS and their impact on the analyses and short-range forecasts. This analysis highlighted both the impact of Arctic atmospheric observations and the challenges related to their usage including key areas of model errors. Overcoming these challenges, which are not insurmountable but require a sustained effort, would lead to better analyses and weather forecasts, but also better atmospheric reanalyses. Indeed, atmospheric reanalyses, such as ERA5, are produced with the same NWP systems as the operational analyses and forecasts (ECMWF IFS in this case).

The analysis of the diagnostics from the ECMWF data assimilation system demonstrated that the Arctic is very rich in terms of satellite data, but rather sparse in terms of conventional atmospheric and surface observations. Indeed, there is better polar orbiting satellite data coverage above 65N than in other parts of the globe due to the high revisit time over the poles (AMSU-A, MHS, IASI). At the same time, conventional Arctic atmospheric observations are mainly found at 60 – 70N, with very few conventional observations north, and less ship and buoy data in the winter.

The assimilation of lower tropospheric sounding satellite channels from polar orbiting satellites is however more challenging in the Arctic region, especially during winter. This is due to the fact that model errors are larger and that there are larger errors in the radiative transfer model used to assimilate the satellite radiances over snow and sea-ice. For instance, mean differences between surface-sensitive microwave observations (including AMSU-A channel 5 and all MHS channels) and short-range forecasts are larger over snow and sea-ice. A proportion of this can be linked to errors in the specification of the surface characteristics in the radiative transfer model, even though model biases also contribute. Fewer observations are therefore assimilated from these channels in winter periods, and when these data are assimilated the forward model has a higher uncertainty.

Uncertainties in the specification of surface contributions in the radiative transfer also affect the infrared observations, leading to a more difficult detection of clouds. Reducing these uncertainties in the radiative transfer should be an integral activity for improving the use of observations in the polar region. In the longer term, the multi-layer snow model developed at ECMWF as part of the WP2 of APPLICATE, for instance, offers the possibility to not only reduce errors in the forecast model (e.g., in the specification of surface temperature), but also to allow better forward modelling of the surface characteristics in the radiative transfer for microwave radiances when combined with an appropriate snow emissivity model (such as SMRT, Picard et al. 2018). Addressing these challenges in the use of satellite data in the polar regions will take a sustained effort and the extra observations gathered during the special observing periods of the Year of Polar Prediction will also help by serving as an additional source of independent observational information.

The successful assimilation of atmospheric observations in the ECMWF system also critically depends on a correct specification of the background errors, which control to a large extent the weight given to observations in the data assimilation process. In the ECMWF 4D-Var system the specification of the background errors depends on the spread of the EDA.

The assessment of the EDA-spread performed in observation space using conventional observations and satellite radiances showed a very poor spread in the polar stratosphere, particularly at the winter pole. This is due to gravity wave activity, which is not well represented by the model background. In the mid-troposphere, however, the spread-skill relationships for the polar regions are relatively close to the ideal ones, and are similar to the mid-latitude spread-skill. In the lower troposphere, the EDA in polar regions is generally slightly less dispersive than in the extratropics and tropics. At the surface (i.e. for surface pressure) and near the surface (i.e. for temperature at 925hPa from radiosondes) a lack of spread in the EDA is seen in the winter season. An under-dispersive EDA is not fully capturing the uncertainty in the short-range forecasts.

When the EDA is used to specify background errors for the deterministic 4D-Var system, this means that the background errors for these aspects will be too small, therefore limiting the ability of observations to make corrections during the analysis process. The effect this has in 4D-Var could be studied in more detail through sensitivity experiments that introduce ad-hoc scaling of the background errors over the polar regions, in order to compensate for the under-dispersiveness found. Ultimately, the under-dispersiveness means that some source of uncertainty in the system is not adequately taken into account in the EDA, and this should be addressed directly. For the under-dispersiveness at lower levels, this could be linked to fewer surface-sensitive satellite observations being used in winter, so that the observational uncertainties are not fully captured, and this could be addressed further through a sustained effort to make better use of these observations.

On the other hand, further perturbations of input fields or parameterisations in the EDA may also be needed. Further work is required to identify the reasons for the under-dispersiveness of the EDA in the lower troposphere during winter, and to develop methods to ameliorate it.

The mean analysis temperature and geopotential height increments in atmospheric analyses are larger over the polar regions than elsewhere, further highlighting the challenges associated with these regions. These mean increments suggest the presence of biases in the forecast model, the observation

operator or the observations. For example, the mean temperature increments in the Arctic are strongest during the summer, including most notably a strong warming around 200 hPa. Positive mean departures are also observed for conventional temperature observations at that height, highlighting a known cold model bias. The mean observation departures indicate that the conventional data may be key in reducing model biases in the summer between 1000 and 300hPa as they are used to anchor variable bias corrections applied to satellite data. For example, aircraft observations (where present) contribute to reducing errors around 200hPa. On the other hand, mean temperature increments of around 0.2 - 0.4 K below 800 hPa over sea-ice in winter are most likely due to the assimilation of the satellite microwave and infrared temperature sounding data (AMSU-A, IASI, etc.) which have mean positive departures over sea-ice in the winter. These positive departures are at least partially due to problems in the radiative transfer model (e.g., neglecting Lambertian surface scattering for microwave radiances) or imperfect cloud screening. These effects may therefore lead to a biased analysis. This aspect further highlights the need for a sustained effort to improve the radiative transfer modelling, and the need to use in-situ observations to distinguish between model and observational biases.

This analysis demonstrated the general value of atmospheric observations in operational analyses and short-range forecasts. The impact of various observing types on the quality of the weather forecasts can be estimated by combining diagnostics such as the FSOI, which provide an estimate of the impact of observations on the reduction of the global forecast error (sect 3.4), and Observing System Experiments (OSEs), in which certain observation systems are removed from the data assimilation system. In contrast to FSOI, OSEs measure the actual impact of losing certain observing systems and allow the evaluation of the influence on the medium-range forecast for a range of parameters. The analysis of FSOI statistics from the ECMWF analyses suggested that in the Arctic microwave observations from polar orbiting satellites have the largest impact in summer, while conventional observations have the largest impact in winter. Guided by the FSOI analysis, OSEs are currently being performed at ECMWF (in Task 4.2.2) for the following observing systems: microwave radiances, conventional observations (both all conventional and individual conventional observation types described in Table 1), infrared radiances, GPSRO and AMVs. The OSEs are performed for the same periods discussed in this report and results will be compared to FSOI statistics to further consolidate the observation-impact assessment.

Existing ocean-sea ice reanalyses have also been investigated in order to assess the added value of observational constraints on our understanding of long-term variability of atmosphere, sea ice and ocean states, as well as the quality of ocean-sea ice initial conditions for coupled climate predictions.

Regarding the sea ice, existing observational data sets provide a high constraint on the observable (sea ice concentration) but this constraint fails to be propagated correctly to other, non-observed variables like sea ice thickness. This suggests two avenues: first, a better specification of covariances between observables (here, concentration) and non-observables (here, sea ice thickness) should be undertaken. Previous work has shown that filters that account for the time-varying and space-varying nature of these covariances are able to improve sea ice thickness even if only sea ice concentration information is assimilated (Massonnet et al., 2013). Second, the effective assimilation of sea ice thickness in ocean and sea ice reanalyses would lead to a better representation of that important variable for polar climate prediction. Still, ensuring physical coherency between the assimilated and non assimilated variables is a priority, in order to avoid shocks at the time of initialization.

Our analyses have also shown that the assimilation of sea ice concentration leads to a systematic reduction of time- and space scales of sea ice thickness variability. Verifying if these changes go in the right direction is challenging due to the poor observational coverage of sea ice thickness and the shortness of the observational record.

A first-time-ever systematic assessment of the representation of hydrographic properties in ORAs has been conducted as well for the Arctic Ocean. Few hydrographic observations are assimilated in the Arctic Ocean, thus the ORAs exhibit biases similar as in free models, e.g. from the CORE-II intercomparison. Drivers of the main biases in the mixed layer and Arctic Ocean hydrography can be

related to (i) the poor representation of physical processes (e.g. vertical mixing in ice-infested areas), (ii) biases in the Arctic sea ice covered in the ORAs, in spite of SIC data assimilation and (iii) biases in the North Atlantic Ocean hydrography which propagate through Fram Strait and Barents Sea Opening through advective processes. This assessment must be deepened in future work, through systematic comparison with comparable free simulations, in order to assess the added value of observations.

6. REFERENCES

P. Bauer, E. Moreau, F. Chevallier, U. O'Keefe (2006), "Multiple-Scattering microwave radiative transfer for data assimilation applications," *Q. J. R. Meteorol. Soc.*, vol. 132, pp. 1259 – 1281.

P. Bauer, L. Magnusson, J.- N. Thépaut , T. M. Hamill (2014), "Aspects of ECMWF model performance in polar areas," *Q. J. R. Meteorol. Soc.*, vol. 142, pp. 583-596.

P. Bauer, A. Thorpe, G. Brunet (2015), "The quiet revolution of numerical weather prediction," *Nature*, Vol 525, pp. 47-55.

Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., Mirouze, I., Peterson, K. A., Sellar, A., and Storkey, D.: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts, Geosci. Model Dev., 7, 2613–2638, doi:10.5194/gmd-7-2613-2014, 2014.

N. Bormann and M. Bonavita (2013), "Spread of the ensemble of data assimilation in radiance space," *ECMWF Tech. Memo., vol. 708.*

N. Bormann, C. Lupu, A. Geer, H. Lawrence, P. Weston, S. English (2017), "Assessment of the forecast impact of suface-sensitive microwave radiances over land and sea-ice," *ECMWF Tech. Memo., vol. 804*

C. Cardinali (2009), "Monitoring the observation impact on the short-range forecast," *Q. J. R. Meteorol. Soc.,* vol. 135, pp. 239–250.

Cardinali, C., 2018: Forecast Sensitivity Observation Impact with an observation-only based objective function. Quarterly Journal of the Royal Meteorological Society, in press.

F. Carminati, S. Migliorini, B. Ingleby, W. Bell, H. Lawrence, S. Newman, J. Hocking, A. Smith (2018, submitted), "Using reference radiosondes to characterise NWP model uncertainty for improved satellite calibration and validation*," Atmos. Meas. Tech*.

Chang, Y. S., Zhang, S., Rosati, A., Delworth, T. L., and Stern, W. F.: An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation, Clim. Dyn., 40, 775–803, doi:10.1007/s00382-012-1412-2, 2013.

Chevallier, M., Smith, G., Lemieux, J.-F., Dupont, 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., Wang, X, 2016. Intercomparison of the Arctic sea ice cover in global ocean-sea ice reanalyses from the ORA-IP project. *Climate Dynamics, Special Issue : Ocean Reanalysis*, 49(3), 1107-1136, doi:10.1007/s00382-016-2985-y.

P. Dahlgren (2018), "Evaluation and diagnostics of the CERA-20C climate reanalysis ensemble," *ECMWF Tech. Memo.,* vol. 820.

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., Biastoch, A., Boning, C., Bozec, A., Canuto, V., Cassou, C., Chassignet, E., Coward, A. C., Danilov, S., Diansky, N., Drange, H., Farneti, R., Fernandez, E., Fogli, P. G., Forget, G., Fujii, Y., Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G. K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Clim., 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

D. P. Dee, S. Uppala (2009), "Variational bias correction of satellite radiance data in the ERA-Interim reanalysis," *Q. J. R. Meteorol. Soc*., vol. 35, no. 644, pp.1830–1841.

D. P. Dee S. M. Uppala A. J. Simmons P. Berrisford P. Poli S. Kobayashi U. Andrae M. A. Balmaseda G. Balsamo P. Bauer P. Bechtold A. C. M. Beljaars L. van de Berg J. Bidlot N. Bormann C. Delsol R. Dragani M. Fuentes A. J. Geer L. Haimberger S. B. Healy H. Hersbach E. V. Hólm L. Isaksen P. Kållberg M. Köhler M. Matricardi A. P. McNally B. M. Monge- Sanz J.- J. Morcrette B.- K. Park C. Peubey P. de Rosnay C. Tavolato J.- N. Thépaut F. Vitart (2011), "The ERA- Interim reanalysis: configuration and performance of the data assimilation system," *Q. J. R. Meteorol. Soc.,* vol. 137, pp. 553 – 597.

R. Eresmaa (2013), "Imager- assisted cloud detection for assimilation of Infrared Atmospheric Sounding Interferometer radiances," *Q. J. R. Meteorol. Soc*., vol. 140, pp. 2342 – 2352.

Ferry, N., Parent, L., Garric, G., Barnier, B., and Jourdain, N. C.: Mercator global Eddy permitting ocean reanalysis GLORYS1V1: description and results, Mercator-Ocean Q. Newslett., 36, 15–27, 2010.

Forget, G., Campin, J. M., Heimbach, P., Hill, C. N., Ponte, R. M., and Wunsch, C.: ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation, Geosci. Model Dev., 8, 3071–3104, doi:10.5194/gmd-8-3071-2015, 2015.

A. Geer, F. Baordo, N. Bormann, and S. English (2014), "All-sky assimilation of microwave humidity sounders," *ECMWF Tech. Memo.*, vol. 741.

A. J. Geer, F. Baordo, N. Bormann, P. Chambon, S. J. English, M. Kazumori, H. Lawrence, P. Lean, K. Lonitz, C. Lupu (2017), "The growing impact of satellite observations sensitive to humidity, cloud and precipitation," *Q. J. R. Meteorol. Soc.*, vol. 143, no. 709, pp. 3189-3206.

S. Guedj, F. Karbou, F. Rabier, A. Bouchard (2010), "Toward a Better Modeling of Surface Emissivity to Improve AMSU Data Assimilation Over Antarctica," *IEEE Trans. Geosc. Rem. Sens.*, vol. 48, no. 4, pp. 1976 – 1985.

R. Hogan, M. Ahlgrimm, G. Balsamo, A. Beljaars, P. Berrisford, A. Bozzo, F. Di Giuseppe, R.M. Forbes, T. Haiden, S. Lang, M. Mayer, I. Polichtchouk, I. Sandu, F. Vitart, N. Wedi (2017), "Radiation in numerical weather prediction," *ECMWF Tech. Memo.,* vol. 816

Ilıcak, M., et al. "An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part III: Hydrography and fluxes." *Ocean Modelling* 100 (2016): 141-161.

Isaksen, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher and Laure Raynaud (2010), "Ensemble of Data Assimilations at ECMWF," *ECMWF Tech. Memo., vol. 636.*

M. Janisková, P. Lopez (2013), "Linearized Physics for Data Assimilation at ECMWF" In: Park S., Xu L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II). Springer, Berlin, Heidelberg.

Köhl, A.: Evaluation of the GECCO2 Ocean Synthesis: Transports of Volume, Heat and Freshwater in the Atlantic, Q. J. R. Met. Soc., 141, 166–181, doi:10.1002/qj.2347, 2015.

Lindsay, R., Wensnahan, M., Schweiger, A., & Zhang, J. (2014). Evaluation of seven different atmospheric reanalysis products in the Arctic. *Journal of Climate*, *27*(7), 2588-2606.

Massonnet, F., Mathiot, P., Fichefet, T., Goosse, H., König Beatty, C., Vancoppenolle, M., & Lavergne, T. (2013). A model reconstruction of the Antarctic sea ice thickness and volume changes over 1980–2008 using data assimilation. Ocean Modelling, 64, 67–75.

Massonnet, F., Fichefet, T., & Goosse, H. (2015). Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation. Ocean Modelling, 88, 16–25. .

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., Hyder, P., Siddorn, J., and Sinha, B.: GO5.0: the joint NERC–Met Office NEMO global ocean model for use in coupled and forced applications, Geosci. Model Dev., 7, 1069–1092, doi:10.5194/gmd-7-1069-2014, 2014.

Picard, G., Sandells, M., Löwe, H., (2018). SMRT: an active–passive microwave radiative transfer model for snow with multiple microstructure and scattering formulations (v1.0). Geosci. Model Dev., 11, 2763-2788, 2018, .

F. Rabier, J. Thépaut, P. Courtier (1998), "Extended assimilation and forecast experiments with a four- dimensional variational assimilation system," *Q.J.R. Meteorol. Soc.,* 124: 1861-1887.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K. A., Oke, P. R., & Korablev, A. (2012). TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Science Discussions*, *9*(2).

R. Saunders, J. Hocking, E. Turner, P. Rayer, D. Rundle, P. Brunel, J. Vidot, P. Roquet, M. Matricardi, A. Geer, N. Bormann, C. Lupu (2018) "An update on the RTTOV fast radiative transfer model (currently at version 12)," *Geosci. Model Dev.,* vol. 11, pp. 2717-2737.

Storto, A., Masina, S., and Dobricic, S.: Estimation and Impact of Nonuniform Horizontal Correlation Length Scales for Global Ocean Physical Analyses, J. Atmos. Ocean. Technol., 31, 2330–2349, doi:10.1175/JTECH-D-14-00042.

Tietsche, S., Balmaseda, M. A., Zuo, H., and Mogensen, K.: Arctic sea ice in the global eddy-permitting ocean reanalysis ORAP5, ECMWF technical memorandum 736, 49, 775–789, doi:10.1007/s00382-015-2673-3, 2017.

Toyoda, T., Fujii, Y., Yasuda, T., Usui, N., Iwao, T., Kuragano, T., and Kamachi, M.: Improved Analysis of Seasonal-Interannual Fields Using a Global Ocean Data Assimilation System, Theor. Appl. Mech. Jpn., 61, 31–48, doi:10.11345/nctam.61.31, 2013.

Uotila, P., Goosse, H., Haines, K., Chevallier, M., Barthélémy, A., Bricaud, C., Carton, J., Fuckar, N., Garric, G., Iovino, D., Kauker, F., Korhonen, M., Lien, V., Marnela, M., Massonnet, F., Mignac, D., Peterson, K., Sadikhi, R., Shi, L., Tietsche, S., Toyoda, T., Xie, J. and Zhang, Z., 2018. An assessment of ten ocean reanalyses in the polar regions. *Climate Dynamics*, online, 1-38, doi : 10.1007/s00382-018-4242-z.

Valdivieso, M., Haines, K., Zuo, H., and Lea, D.: Freshwater and heat transports from global ocean synthesis, J. Geophys. Res., 119, 394–409, doi:10.1002/2013JC009357, 2014.

Wang, Q., et al. "An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part I: Sea ice and solid freshwater." *Ocean Modelling* 99 (2016): 110-132.

Wang, Q., et al. "An assessment of the Arctic Ocean in a suite of interannual CORE-II simulations. Part II: Liquid freshwater." *Ocean Modelling* 99 (2016): 86-109.

Zhang, J. L. and Rothrock, D. A.: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates, Mon. Weather Rev., 131, 845–861, 2003.

Zhang, S., Harrison, M. J., Rosati, A., and Wittenberg, A.: System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies, Mon. Weather Rev., 135, 3541–3564, doi:0.1175/MWR3466.1, 2013.

Zuo, H., Balmaseda, M. A., and Mogensen, K.: The ECMWF-MyOcean2 eddy-permitting ocean and sea-ice reanalysis ORAP5. Part 1: implementation, ECMWF technical memorandum 736, 2015.