https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

The Cryosphere Discussions

Statistical predictability of the Arctic sea ice volume anomaly: identifying predictors and optimal sampling locations

Leandro Ponsoni[1], François Massonnet[1], David Docquier[1], Guillian Van Achter[1], and Thierry Fichefet[1]

[1]Georges Lemaître Centre for Earth and Climate Research (TECLIM), Earth and Life Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Correspondence to: Leandro Ponsoni (leandro.ponsoni@uclouvain.be)

Abstract. This work evaluates the statistical predictability of the Arctic sea ice volume (SIV) anomaly – here defined as the detrended and deseasonalized SIV – on the interannual time scale. To do so, we made use of 6 datasets, from 3 different atmosphere-ocean general circulation models, with 2 different horizontal grid resolutions each. Based on these datasets, we have developed a statistical empirical model which in turn was used to test the performance of different predictor variables, as well as to identify optimal locations from where the SIV anomaly could be better reconstructed and/or predicted. We tested the hypothesis that an ideal sampling strategy characterized by only a few optimal sampling locations can provide in situ data for statistically reproducing and/or predicting the SIV interannual variability. The results showed that, apart from the SIV itself, the sea ice thickness is the best predictor variable, although total sea ice area, sea ice concentration, sea surface temperature, and sea ice drift can also contribute to improving the prediction skill. The prediction skill can be enhanced further by combining several predictors into the statistical model. Feeding the statistical model with predictor data from 4 well-placed locations is enough for reconstructing about 70% of the SIV anomaly variance. An improved model horizontal resolution allows a better trained statistical model so that the reconstructed values approach better to the original SIV anomaly. On the other hand, if we look at the interannual variability, the predictors provided by numerical models with lower horizontal resolution perform better when reconstructing the original SIV variability. As per 6 well-placed locations, the statistical predictability does not substantially improve by adding new sites. As suggested by the results, the 4 first best locations are placed at the transition Chukchi Sea–Central Arctic–Beaufort Sea (158.0°W, 79.5°N), near the North Pole (40°E, 88.5°N), at the transition Central Arctic–Laptev Sea (107°E, 81.5°N), and offshore the Canadian Archipelago (109.0°W, 82.5°N), in this respective order. We believe that this study provides recommendations for the ongoing and upcoming observational initiatives, in terms of an Arctic optimal observing design, for studying and predicting not only the SIV values but also its interannual variability.

1 Introduction

The continuous melting of the Arctic sea ice observed in the last decades (e.g., Chapman and Walsh, 1993; Parkinson et al., 1999; Rothrock et al., 1999; Parkinson and Cavalieri, 2002; Zhang and Walsh, 2006; Stroeve et al., 2007, 2012; Notz and Stroeve, 2016; Petty et al., 2018), associated with the respective reduction in total sea ice area (SIA) and volume (SIV), has proven to bring significant impacts at regional and global scales. Regionally, native communities have experienced a

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

disturbance of subsistence activities like fishing, crabbing and hunting (Nuttall et al., 2005), which have also impacted other predators from seabirds (Grémillet et al., 2015; Amélineau et al., 2016) to mammals (Laidre et al., 2008; Lydersen et al., 2017; Wilder et al., 2017; Pagano et al., 2018; Brown et al., 2016). Other pressing local issues are also bringing important implications for the Arctic countries such as the opening of new sailing routes (Lindstad et al., 2016), the development of the tourism industry (Handorf, 2011) and the mineral resource extraction (Gleick, 1989). At global scale, the sea ice depletion is reported to impact some aspects of the weather at low- and mid-latitude regions, by means of both oceanographic (Drijfhout, 2015; Sévellec et al., 2017) and atmospheric teleconnections (Serreze et al., 2007; Overland and Wang, 2010), including the higher occurrence of extreme events (Francis and Vavrus, 2012; Tang et al., 2013; Screen and Simmonds, 2013; Cohen et al., 2014).

Since this meltdown is projected to continue throughout the twenty-first century (e.g., Burgard and Notz, 2017), the interest of the scientific community and policy makers on the sea ice variability and predictability is exponentially increasing, mainly in terms of SIV. The SIV is a primary sea ice diagnostic because it accounts for the total mass of sea ice. In situ measurements of sea ice thickness, that are needed for calculating the SIV, are far too expensive, being spatially sparse and temporally sporadic (Lindsay, 2010; Tilling et al., 2018). In addition, satellite observations present well-known limitations in the warmer seasons and, therefore, sea ice thickness is not made available year-round from the classical satellite campaigns, namely: ICESat (Kwok et al., 2007) and CryoSat-2 (Kwok and Cunningham, 2015).

Due to the lack of long-term observations, the answer to the question of whether or not this negative trend in sea ice affects the interannual variability of the pan-Arctic SIV, and the other way around, is not clear yet, although it has been already shown that trends in the pan-Arctic sea ice extent can be masked by its long-term variability (Swart et al., 2015). Up to now, most of the long-term SIV variability and predictability studies rely on reanalyses and/or model outputs (e.g., Schweiger et al., 2019). Even though atmosphere-ocean general circulation models (AOGCMs), including their sea ice component, are more and more complex nowadays, in situ observations are still required for a more comprehensive model validation and also for assimilation purposes.

In order to respond to the need of having an improved observational system for better understanding the SIV variability, but at the same time minimize the costs required to do so, this work raises the hypothesis that "an ideal sampling strategy characterized by only few optimal sampling locations can provide in situ data for statistically reproducing and/or predicting the SIV interannual variability". To test the hypothesis, this study follows three main directions. First, we propose a statistical empirical model for predicting the SIV. Since we are mainly interested in predicting the interannual variability rather than the trivial seasonal cycle and the long-term trends, we will focus on the SIV without these two components – hereafter defined as SIV anomaly. Second, we aim at inspecting the performance of a set of ocean- and ice-related predictor variables for feeding the empirical model. Third, we intend to localize a reduced number of optimal sampling locations from where the predictor variables could be systematically sampled using oceanographic moorings and/or buoys. Sampling in situ data at optimal locations or, in order words, by collecting data at locations in which most of the pan-Arctic SIV anomaly variability is captured by the predictor variables, makes it much more feasible to sustain a long-term programme of operational oceanography both from logistical and financial points of view.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

To the knowledge of the authors, this study is the first to apply an empirical statistical model for supporting an optimal observing system of the pan-Arctic SIV anomaly, albeit a similar study was conducted by Lindsay and Zhang (2006) a decade ago. However, these authors focused on the predictability of averaged Arctic sea ice thickness, based their results on a single model approach, as well as considered two predetermined sampling locations. Other previous works also applied statistical empirical models for predicting a range of Arctic sea ice properties (e.g., sea ice extent, area and concentration), for lead periods of up to one year, at regional and/or pan-Arctic scales (Walsh, 1980; Barnett, 1980; Johnson et al., 1985; Drobot and Maslanik, 2002; Drobot et al., 2006; Lindsay et al., 2008; Chevallier and Salas-Mélia, 2012; Grunseich and Wang, 2016; Yuan et al., 2016). Unlike the statistical prediction of sea ice extent and area, which have longer and more reliable records of observations allowing the statistical models to be built on this data, the statistical prediction of SIV requires information from models. Thus, even though we claim that in situ observations are crucial for understanding the SIV variability, our study makes use of outputs from 3 AOGCMs. This is the only way to have continuous and broadly distributed data of the predictand and some predictor variables, such as sea ice thickness. Here we assume that the AOGCMs fairly represent the thermodynamic and dynamic processes linking predictors to predictand, while the use of 3 different models attempts to assess the model dependence of our results.

To fully address the three overall goals described above, this study is guided by the following open questions: (i) What are the performance of different pan-Arctic predictors for predicting pan-Arctic SIV anomalies? (ii) What are the best in situ locations for sampling predictor variables to optimize the statistical predictability of SIV anomalies in terms of reproducibility and variability? (iii) How many optimal sites are needed for explaining a large amount, that is to say, at least 70% (an arbitrarily chosen threshold) of the original SIV anomaly variance? (iv) Are the results model dependent, in particular, are they sensitive to horizontal resolution? Following this introduction (Section 1), the manuscript is organized as follows: Section 2 describes the AOGCMs, datasets and the methods (including the development of the statistical empirical model) used in our analyses. Section 3 presents the results which are further discussed in Section 4. This last section also highlights the main conclusions and draw some recommendations for an observing sampling design.

2 Data and methods

2.1 Model outputs

As argued above, this work can not be performed with actual observations and it follows therefore a multi-model approach. It takes advantage of 6 coupled historical runs from 3 different AOGCMs, all conducted within the context of the High Resolution Model Intercomparison Project (HighResMIP; Haarsma et al. (2016)). The HighResMIP is inserted in the framework of the Coupled Model Intercomparison Project 6 (CMIP6; Eyring et al. (2016)) and its main goal is to systematically study the role of horizontal resolution in the simulation of the climate system. In our study, we use 2 different model configurations for each of the 3 models. These configurations differ by their horizontal grid resolution (in both the atmosphere and ocean). The runs start in the early 1950s, spanning for about 65 years until mid-2010s. We extract monthly outputs from these model simulations. Namely, the AOGCMs are: the version 1.1 of the Alfred Wegener Institute Climate Model (AWI-CM; Sidorenko et al. (2015);

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

The Cryosphere Discussions
Open Access
EGU

Rackow et al. (2018)), the European Centre for Medium-Range Weather Forecasts Integrated Forecast System (ECMWF-IFS) cycle 43r1 (Roberts et al., 2018), and the Global Coupled 3.1 configuration of the Hadley Centre Global Environmental Model 3 (HadGEM3-GC3.1; Roberts et al. (2019)). Fig. 1 shows the absolute values and the anomalies (no long-term trend; no seasonal cycle) of the Arctic SIV time series from the 6 model outputs.

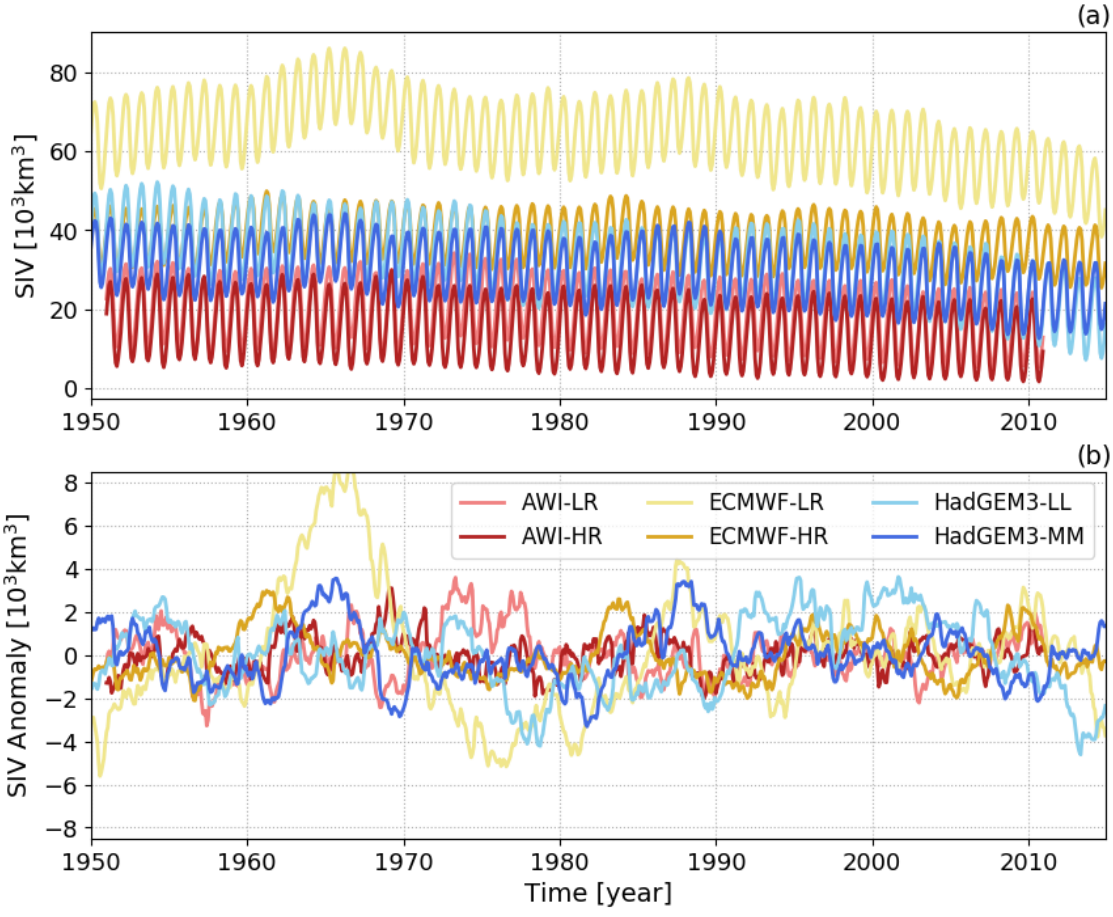


Figure 1. Sea ice volume time series from the 6 model outputs used in this work: (a) absolute values and (b) anomalies (no trend; no seasonal cycle).

A comprehensive comparison including these 3 models and their respective specifications are presented by Docquier et al. (2019). In short, AWI-CM is composed by the European Centre/Hamburg version 6.3 (ECHAM6.3) atmospheric model and by the version 1.4 of the Finite Element Sea ice-Ocean Model (FESOM; Wang et al. (2014); Sein et al. (2016)). ECMWF-IFS is a hydrostatic, semi-Lagrangian, semi-implicit dynamical-core atmospheric model, while the ocean and ice components are composed by the version 3.4 of the Nucleus for European Modelling of the Ocean model (NEMO; Madec (2008)) and version 2 of the Louvain-la-Neuve Sea-Ice Model (LIM2; Fichefet and Maqueda (1997)), respectively. Finally, HadGEM3-GC3.1 is

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

built up with the same ocean model than ECMWF-IFS (NEMO; Madec et al. (2017)), but version 3.6, the atmospheric Unified Model (UM; Cullen (1993)) and the version 5.1 of the Los Alamos sea-ice model (CICE; Hunke et al. (2015)). Hereafter the models are simply referred to as AWI, ECMWF, and HadGEM3.

Overall, the 2 configurations from the same model keep the parameters identical, except for the resolution-dependent parameterizations (Docquier et al., 2019). In terms of ocean–sea ice grid, both AWI versions (data source: Semmler et al. (2017b, a)) use a mesh grid with varying resolution. The low-resolution version (AWI-LR) has a resolution changing from 24 to 110 km, and the high-resolution version (AWI-HR) changes from 10 to 60 km (Sein et al., 2016). Both ECMWF (data source: Roberts et al. (2017b, a)) and HadGEM3 (data source: Roberts (2017a, b)) adopt the tripolar ORCA grid (Madec and Imbard, 1996). The configurations with coarser resolution (ECMWF-LR and HadGEM3-LL) use ORCA1 with a resolution of 1°, while the versions with finer horizontal grid (ECMWF-HR and HadGEM3-MM) use ORCA025 with a resolution of 0.25°.

2.2 Potential predictors

The next step for proceeding with the statistical predictions of the SIV anomalies is to identify potential predictor variables to be used in the empirical statistical model. We recall that the term "anomaly" in this work refers to the detrended and deseasonalized time series. In practical terms, the anomaly is calculated by excluding the individual trend (provided by a second-order polynomial fit) of each individual month. In this section, as a first assessment, we test the performance of different predictors by estimating their correlation against the predictand. This test is performed individually for each model output, which means to say that predictor variables from a certain model configuration are only used for predicting the SIV anomaly from this respective configuration. Overall, two categories of predictors are tested: global variables, intrinsically represented by a single pan-Arctic time series, and local predictors, represented by gridded data. Nevertheless, for this first assessment, the local predictors are considered as their pan-Arctic means. To do so, the gridded values are twice normalized: first, by the grid area of each grid cell and, second, by the correlation maps with the predictand (Drobot et al., 2006). As suggested by Drobot et al. (2006) in the second normalization, the significant correlation coefficients from the different grid cells are used as normalizing factors (as it is the grid-cell area in the first normalization). The idea behind this second normalization is to take the best advantage of the correlations between predictand and predictors since the former is not necessarily correlated to the latter over the entire Arctic domain (Fig. 2). Notice that non–significant correlation coefficients are set to zero so that they do not weight in the normalization. Predictor variables are also used in terms of their anomaly (no trend; no seasonal cycle).

Apart from the condition that all predictor variables could be regularly sampled from observational platforms, we also considered only variables that have the potential to impact the sea ice through dynamic and/or thermodynamic processes. A set of 7 predictors are considered for this preliminary inspection. 3 of them are global variables, that are: pan-Arctic SIV itself, pan-Arctic SIA and Atlantic basin ocean heat transport (OHT) estimated at 60°N. The other 4 predictors are local variables provided by the AOGCMs in a gridded format and reduced to single time series as mentioned above, that are: sea ice thickness (SIT), sea ice concentration (SIC), sea surface temperature (SST) and sea ice drift (Drift). As an example, Fig. 2 compares the time series of predictand against pan-Arctic predictors (Fig. 2a,b,c,e,g,i), and also displays the respective correlation maps

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

used for normalizing the regional predictors (Fig. 2d,f,h,j), for the AWI-LR output. Table 1 shows the correlation coefficient estimated between predictand and predictors for all the models.

Table 1. Correlation coefficient estimated between the predictand (SIV anomaly) and a set of pan-Arctic potential predictors: SIA, OHT, SIT, SIC, SST, and Drift. Regional predictors (SIT, SIC, SST and Drift) are represented by pan-Arctic averages. As for the predictand, all predictors are used with monthly time-resolution and in terms of their anomaly (no seasonal cycle; no long-term trend).

Models	Predictors					
	SIA	OHT	SIT	SIC	SST	Drift
AWI-LR	0.64	-0.08	0.86	0.29	-0.57	0.15
AWI-HR	0.69	–	0.89	0.26	-0.50	0.31
ECMWF-LR	0.20	0.28	0.95	0.31	-0.12	-0.20
ECMWF-HR	0.24	–	0.63	0.37	-0.22	-0.13
HadGEM3-LL	0.63	-0.33	0.91	0.71	-0.54	-0.28
HadGEM3-MM	0.63	–	0.94	0.62	-0.45	-0.31

2.3 Statistical empirical models

The basis of our statistical empirical model (SEM) is a multiple linear regression model where the time series of the dependent variable (y) could be described as a function of the time series of the independent explanatory variables (x_i), as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon, \tag{1}$$

where β_0 is the constant y-intercept, β_k is the slope coefficients for each explanatory variable and ε is the error term (or residual) of the empirical model.

In our case, the reconstructed time series of SIV anomaly (SIV_{rec}) is based on the linear relationship between this variable and the predictors aforementioned in Section 2.2. If the SIV itself is also considered as a predictor, the multiple linear regression in Eq. 1 can be written as:

$$SIV_{rec} = \beta_0 + \beta_1 SIV + \beta_2 SIA + \beta_3 OHT + \beta_4 SIT + \beta_5 SIC + \beta_6 SST + \beta_7 Drift. \tag{2}$$

To bring robustness to the statistical reconstructions, the SEM is applied within a Monte-Carlo loop with 500 repetitions. In every repetition, 70% of the data are randomly selected for training (N_T) the SEM, while the remaining 30% are used for comparing (N_C) the original and the reconstructed SIV. In practical terms, ECMWF and HadGEM3 have 780 data points in time equivalent to the 780 months between Jan-1950 and Dec-2014 (720 for AWI; from Jan-1951 to Dec-2010) so that N_T = 546 monthly values are used for building the SEM and N_C = 234 are used to evaluate how good is the SIV reconstruction (N_T = 504 and N_C = 216 for AWI). Since our main interest lies in the reconstruction of the SIV values, the metric used for comparing the original and reconstructed time series is the root mean squared error (RMSE). In this way, the score (Sc) of the reconstructed SIV can be represented by

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

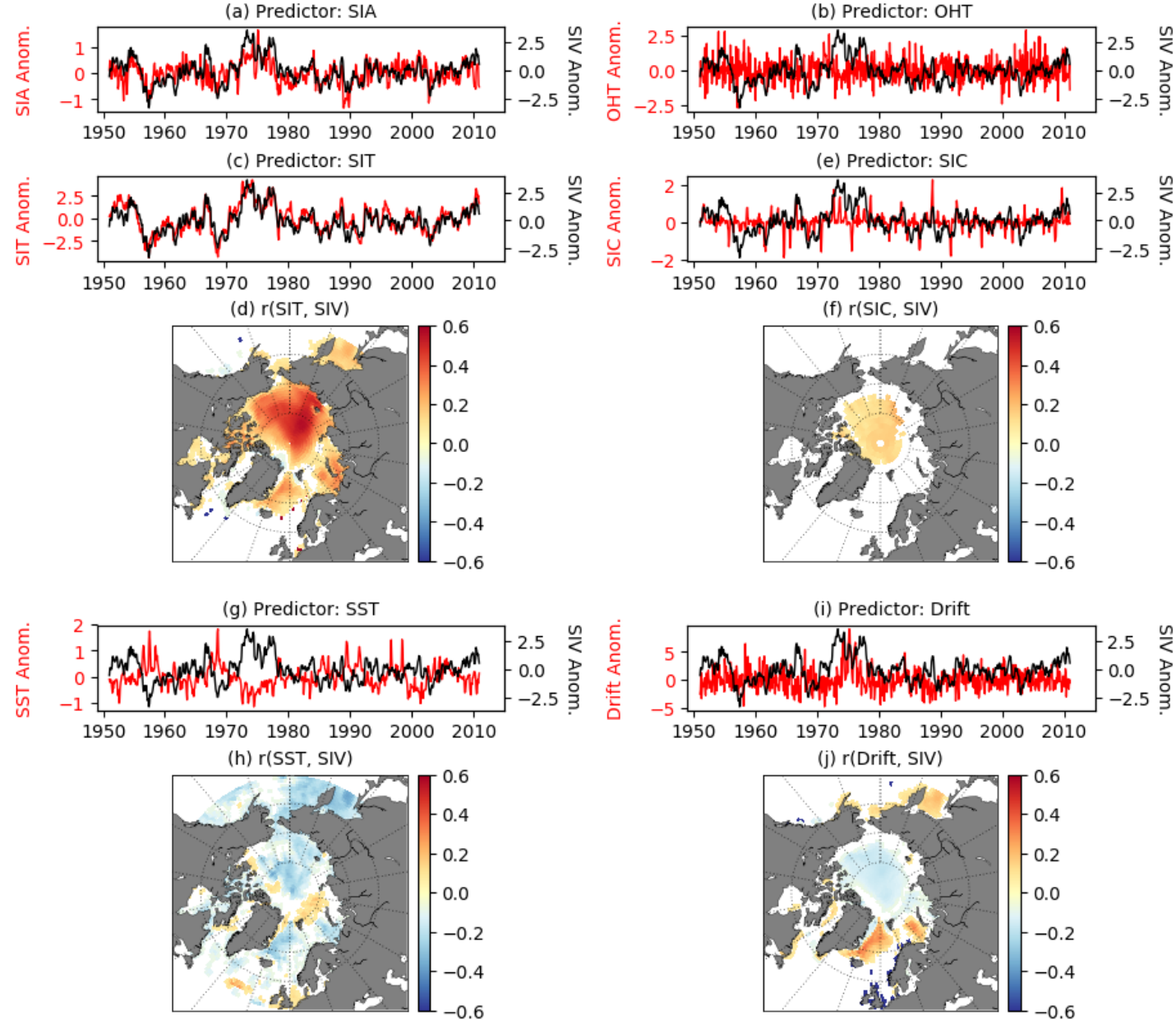


Figure 2. Comparison between the time series from the predictand (SIV [10^3km^3]; black lines) and predictors: (a) SIA [10^6km^2], (b) OHT [PW], (c) SIT [m], (e) SIC [%], (g) SST [°C] and (i) Drift [km day^{-1}] (red lines). The correlation maps used for normalizing the regional predictors are also shown: (d) SIT, (f) SIC, (h) SST and (j) Drift.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

$$Sc = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{\sum_{n=1}^{N_C}(SIV_{rec}(P) - SIV)^2}{N_C}}, \quad (3)$$

where R = 500 indicates the number of interactions in the Monte Carlo loop, P represents the (set of) employed predictor(s) and the index N_C emphasizes that only 30% of the data are used for comparison between original (SIV) and reconstructed SIV (SIV_{rec}) time series. An estimate of the Sc error (Sc_{er}) is given by the standard deviation calculated from the set of RMSEs given at every step of the Monte-Carlo scheme.

Two different approaches for applying the SEM are used in this work. In Section 3.1, we evaluate the individual and combined performances of the pan-Arctic predictors for reconstructing the SIV anomaly at different months of the year (March and September), with a lag of 1 to up to 12 months upfront. In Section 3.2 we make use of the SEM to support an optimal sampling strategy, but using the local predictors in their gridded format rather than their pan-Arctic averages, as the methodology described in Section 2.4.

2.4 Identifying optimal sampling locations

By identifying optimal sampling locations, we intend to spot a reduced number of sites from which predictor variables could offer an optimal representation of the SIV anomaly. To identify the 1st best location, a Score Map (Sc[i,j]) is generated by calculating the Sc at each grid cell [i,j], but now taking into account regional predictors (SIT[i,j], SIC[i,j], SST[i,j], Drift[i,j]) rather than pan-Arctic averages. From the predictors intrinsically represented by single time series (SIV, SIA, OHT), only SIA will be used because in the real world this variable is provided monthly from satellite measurements. SIV is disregarded for an obvious reason since this is the variable that we want to predict while having OHT from observations is a more complex task as it would require oceanic observations broadly distributed both in space and depth. Additionally, OHT is not a good predictor, at least not when it is used with monthly time-resolution, as discussed further in Section 3.1.

This method allows us to build Sc[i,j] where the smaller the score, the better the representation of the pan-Arctic SIV anomaly. Hence, the most optimal location is here defined by the grid point where Sc[i,j] is minimum. In practical terms, the score maps will reveal clusters of grid points defining one region (or more) from where the SIV anomalies would be optimally reconstructed. After determining and fixing the 1st ideal location [i_1,j_1], we can look for a 2nd [i_2,j_2], a 3rd [i_3,j_3], and so on [i_k,j_k], best locations. However, every time that a location is identified, a region of influence surrounding this location is identified to avoid that different stations are placed nearby each other (see details below). In this approach, the regression described in Eq. 2, with k optimal locations, takes the following format:

$$SIV_{rec} = \beta_0 + \beta_2 SIA + \sum \beta_{P1[i_1,j_1]} P1[i_1,j_1] + \sum \beta_{P2[i_2,j_2]} P2[i_2,j_2] + \cdots + \sum \beta_{Pk[i_k,j_k]} Pk[i_k,j_k], \quad (4)$$

where the term $\beta_{Pk[i_k,j_k]} Pk[i_k,j_k]$ represents the product between the valid predictors $Pk[i_k,j_k]$, at the optimal location number k, and their respective slope coefficients $\beta_{Pk[i_k,j_k]}$. It is worthwhile mentioning that only valid predictors, which means only predictors significantly correlated with the predictand, are used to feed the Eq. 4. For instance, for the AWI-LR

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

product, if a grid point placed off the eastern coast of Greenland is one of the N locations, the SEM incorporates SIT, SST and Drift as predictors while SIC is disregarded, as suggested by the correlation maps plotted in Fig. 2d,f,h,j.

For determining the 1st optimal location, this procedure is repeated independently for each of the 6 model outputs. That means that each of the datasets provides its first optimal location (Sc[i_1,j_1]). Not necessarily all the models will suggest the same location as ideal. However, further, in Section 3.2.1, it will be shown that the different model outputs suggest relatively similar clusters of grid-points that can provide a skillful representation of the pan-Arctic SIV anomaly. Subsequently, aiming at spotting a single first optimal location that better represents all datasets, we take the average of the 6 score maps. To give the same weight for all datasets in the averaging, the individual score maps are scaled between 0 and 1 (ScNorm$_{[i,j]}$), as follows (Eq. 5):

$$ScNorm_{[i,j]} = \frac{Sc_{[i,j]} - Sc_{min}}{Sc_{max} - Sc_{min}}, \quad (5)$$

where the indexes min and max indicate the minimum and maximum values in the score map, respectively. Afterward, for having a coherent gridded average, the 6 models are interpolated into a common grid. Finally, the 1st best sampling location is defined as the geographical coordinate where the mean ScNorm map presents its minimum value. The advantage of this approach is to reduce the model dependence of the results by relying on different datasets.

Notwithstanding, before departure for the identification of the 2nd ideal location, we borrow the concept of length scale (Blanchard-Wrigglesworth and Bitz, 2014; Ponsoni et al., 2019) to avoid that 2 optimal sites are placed near each other. The length scale defines a radius where a certain gridded variable is well-correlated to the same variable from the neighboring grid points. In this work we do not use a radius, but a very similar approach: the correlation coefficient of our best local predictor at the selected location (SIT$_{[i1,j1]}$; see Section 3.1) is calculated against the equivalent time series from all the other grid points (SIT$_{[i,j]}$). The region defined by the grid points with a correlation higher than 1/e, a threshold for correlations below which the SIT is assumed to be uncorrelated to the point of interest, is used as a buffer region – hereafter defined as "region of influence". So, all the grid points enclosed by the region of influence are automatically disregarded from being selected as a 2nd location. Figure 3 shows how would be the region of influence for a station arbitrarily placed at the North Pole as defined by the ensemble of datasets. Once the 2nd location is identified for all datasets, we repeat the procedure described above for determining a single 2nd optimal location. This iterative approach is also followed for the identification of the 3rd optimal site, and so on.

3 Results

3.1 Statistical predictability of SIV anomaly: pan-Arctic predictors

In this section, the statistical predictability of the SIV anomaly is quantitatively evaluated by considering leading periods of 1 to 12 months upfront. Also, the predictive performance of 7 pan-Arctic predictors is tested. The predictors are SIV itself, SIA, OHT, SIT, SIC, SST and Drift. Here, we focus on the months with relatively large (March; Section 3.1.1) and reduced (September; Section 3.1.2) SIV at the end of the winter and summer, respectively.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

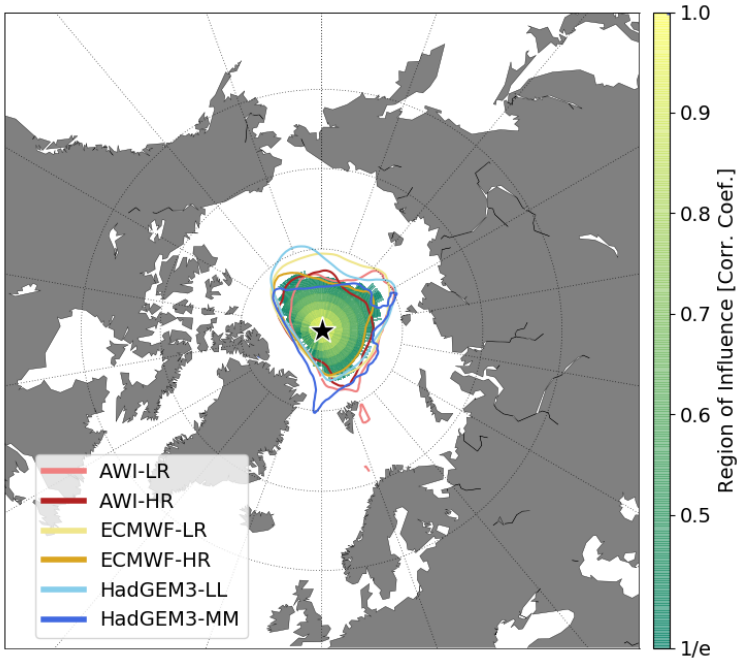


Figure 3. Region of influence for a station arbitrarily placed at the North Pole (black star) as defined by each model (colorful lines) and by the averaged region of influence from the different models (shades of green to yellow).

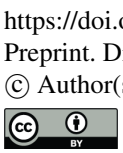
3.1.1 Statistical predictability of March SIV anomaly: pan-Arctic predictors

Figure 4 displays the predictive performance (quantified by the RMSE) of different predictors for estimating March SIV anomalies. The SIV itself is the best predictor variable and its score gradually increases from 12 (Sc = 1.0 $\times 10^3 km^3$) to 4 (Sc = 0.68 $\times 10^3 km^3$) leading months. During this period the mean performance for the ensemble of models increases by about 32%. As per 3 leading months, from December to February, the predictive capacity substantially improves by 43% (Sc = 0.57 $\times 10^3 km^3$), 59% (Sc = 0.41 $\times 10^3 km^3$) and 77% (Sc = 0.23 $\times 10^3 km^3$), respectively (Fig. 4a).

The second best predictor is the SIT, which has performance similar to the SIV predictor from about 12 to 9 leading months (ensemble mean Sc = 1.02 $\times 10^3 km^3$, 1.03 $\times 10^3 km^3$, 1.0 $\times 10^3 km^3$; Fig. 4d). Nevertheless, its score remains relatively stable and improves only by about 25%, from May to February (Sc = 1.0 and 0.75 $\times 10^3 km^3$). SIC (Fig. 4e), SST (Fig. 4f) and Drift (Fig. 4g) have poorer performance compared to SIT, but similar behavior with the score slightly improving over time until 1 leading month.

SIA (Fig. 4b) is a valid predictor for AWI and HadGEM3 models, but it does not seem to be the case for ECMWF versions. Finally, OHT showed to be a poor predictor in terms of monthly predictability. For most of the leading months and models, the statistical reconstruction is not significant when provided by this predictor (Fig. 4c).

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019



The Cryosphere
Discussions
Open Access
EGU

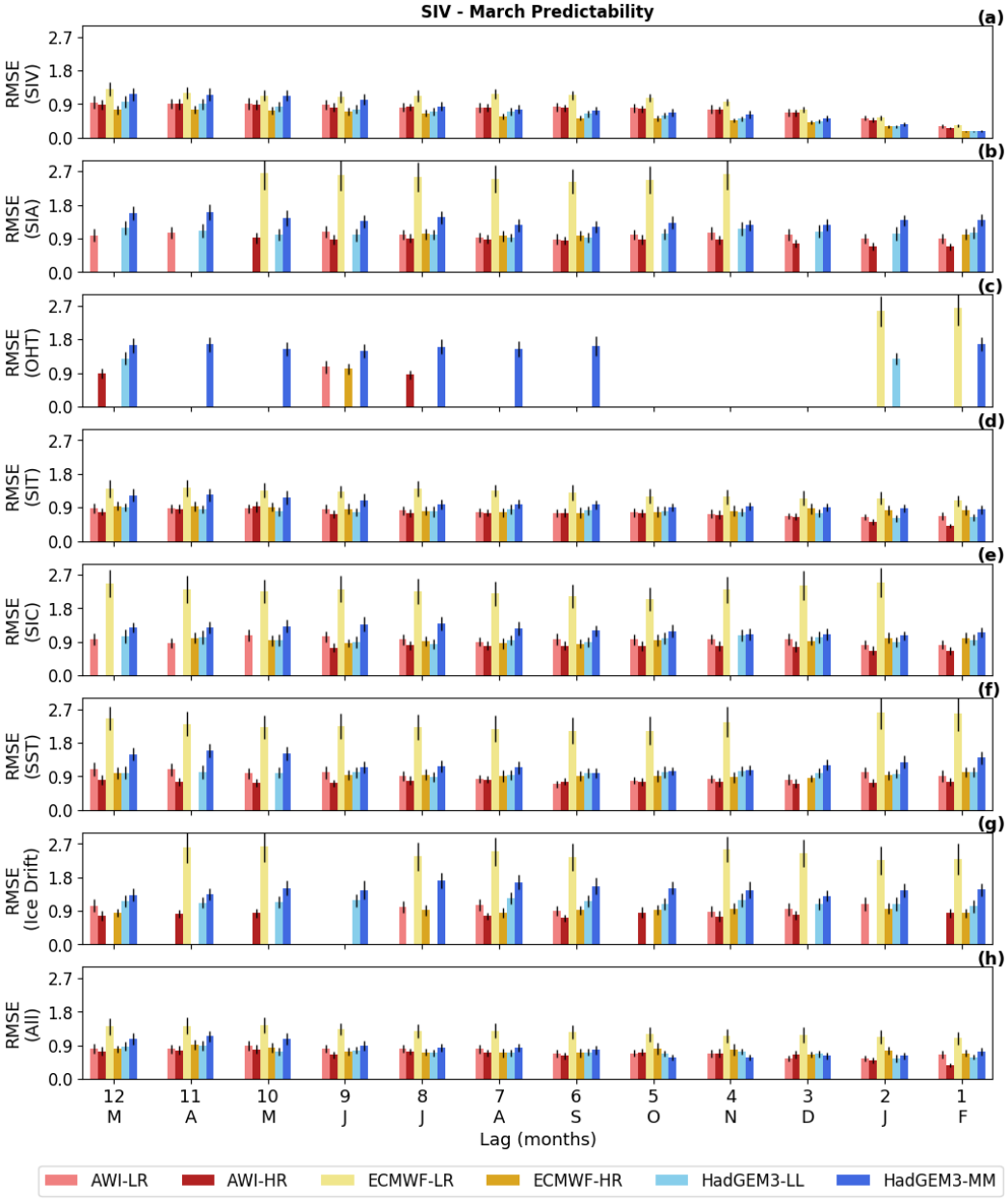


Figure 4. Statistical predictability of the March SIV anomalies, estimated from 12-leading months and quantified by the RMSE (10^3km^3) calculated between the original and reconstructed time series (Sc), as prescribed by 7 predictors: (a) SIV itself, (b) SIA, (c) OHT, (d) SIT, (e) SIC, (f) SST, (g) Drift. The predictions employing ALL predictor variables (except the SIV itself) are displayed in (h). The vertical black lines indicate the error as provided by the 500 Monte Carlo simulations. The statistical predictability follows the methodology introduced in Section 2. Missing vertical bars mean that the statistical reconstruction is not statistically significant. The long-term trend and seasonal cycle are excluded from both predictand and predictors.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

A way of improving further the statistical predictability is to use several predictors at once. Figure 4h shows the case where all the aforementioned predictors (except SIV) are used by the empirical model. For this configuration, the predictive skill is still 10% lower than the case where SIV is standing alone as a predictor, but it is about 10% better than the reconstructions provided only by the SIT.

The inter-model comparison does not show a conclusive answer to the question of whether or not the model resolution plays a role in the statistical predictability of March SIV anomalies. Overall, AWI-HR predictors are more skilled than AWI-LR predictors, though the opposite is observed for HadGEM3. For the ECMWF versions, the SIV anomalies from EMCWF-HR present better reproducibility, while ECMWF-LR presents much larger errors. Note that ECMWF-LR has a mean state characterized by a much thicker sea ice and, consequently, higher variance (see Fig. 1). This is the reason that makes ECMWF-LR an outlier compared to the other 5 model outputs for this and other results found in this manuscript (see further discussion in Section 4).

3.1.2 Statistical predictability of September SIV anomaly: pan-Arctic predictors

A similar scenario compared to March is found for the September SIV anomaly predictability (Fig. 5). The best predictor is the SIV itself (Fig. 5a) for which the predictive skill improves by about 83.6% from June to August (Sc = 1.16 $\times 10^3 km^3$ and 0.19 $\times 10^3 km^3$). This improvement is mainly attributed to the 3 months before September: Sc = 0.71 $\times 10^3 km^3$, 0.44 $\times 10^3 km^3$ and 0.19 $\times 10^3 km^3$ for June, July and August, respectively. The second best predictor is SIT (Fig. 5d), while SIC (Fig. 5e), SST (Fig. 5f) and Drift (Fig. 5g) present an intermediate performance. For the former 4 predictors, the ensemble mean Sc slightly improves from 12 to 1 leading months in about: 28.8% (Sc = 1.04 $\times 10^3 km^3$ and 0.74 $\times 10^3 km^3$), 15% (Sc = 1.40 $\times 10^3 km^3$ and 1.19 $\times 10^3 km^3$), 29% (Sc = 1.26 $\times 10^3 km^3$ and 0.90 $\times 10^3 km^3$) and 24% (Sc = 1.46 $\times 10^3 km^3$ and 1.11 $\times 10^3 km^3$), respectively. Not all tested predictors are statistically significant for reproducing the SIV anomalies. Again, this is the case for OHT (Fig. 5c). SIA also presents poor performance for some models and leading months (Fig. 5b). Another resemblance to March predictability is the relatively poor performance presented by the predictor variables from ECMWF-LR.

3.2 Statistical predictability of SIV anomaly: regional predictors

In this section, the empirical statistical model is used for supporting an optimal sampling strategy by following the methodology described in Section 2.4. To do so, we combine the local predictors at every grid-point rather than use their pan-Arctic averages. The reasoning behind this approach lies in the hypothesis that the statistical empirical model can fairly reproduce and/or predict the SIV anomalies if a few optimal locations provide in situ measurements from the predictor variables. These in situ observations can be applied concomitantly with predictors that are continuously measured by satellites as the pan-Arctic SIA and the local SIC.

Here we assume that numerical models are able to reproduce the main physical processes behind the interactions among predictand and predictors. Practically, we will take into account 4 local predictors that are SIT, SIC, SST and Drift, and 1 pan-Arctic predictor that is SIA, although it is worthwhile reminding that only predictors significantly correlated with the predictand will be incorporated to the statistical model. As per the results of Section 3.1, the OHT will not be included as

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

The Cryosphere
Discussions
Open Access
EGU

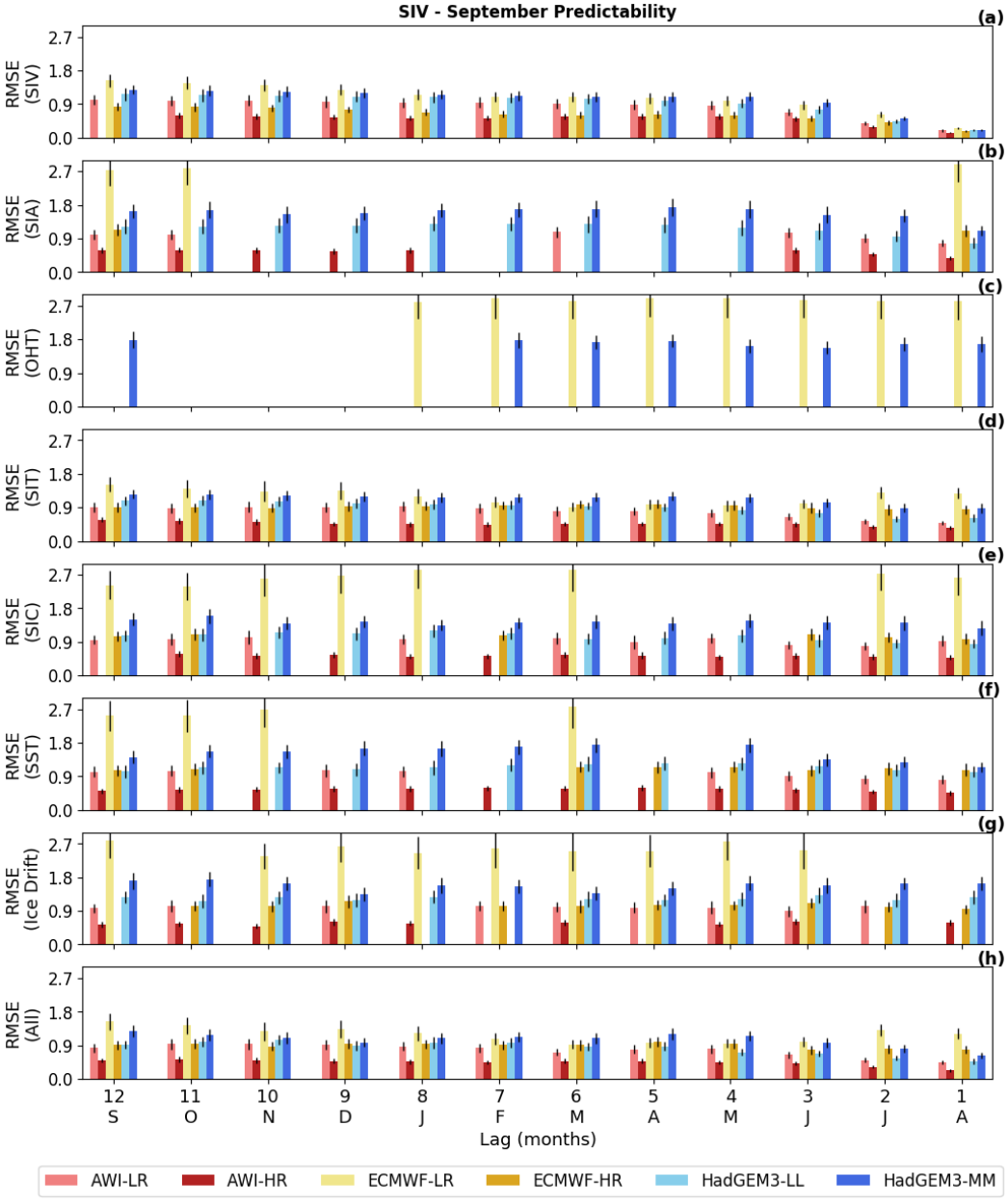


Figure 5. Statistical predictability of the September SIV anomalies, estimated from 12-leading months and quantified by the RMSE (10^3km^3) calculated between the original and reconstructed time series (Sc), as prescribed by 7 predictors: (a) SIV itself, (b) SIA, (c) OHT, (d) SIT, (e) SIC, (f) SST, (g) Drift. The predictions employing ALL predictor variables (except the SIV itself) are displayed in (h). The vertical black lines indicate the error as provided by the 500 Monte Carlo simulations. The statistical predictability follows the methodology introduced in Section 2. Missing vertical bars mean that the statistical reconstruction is not statistically significant. The long-term trend and seasonal cycle are excluded from both predictand and predictors.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

predictor variable due to its poor capacity to provide a skillful prediction, being reinforced by the difficulties associated with the in situ sampling and estimation of this variable.

3.2.1 Optimal sampling locations

For each of the 6 models, score maps (Sc[i,j]; Eq. 3) were determined with the aim of spotting the location that can better reproduce the SIV anomalies as shown in Fig. 6. This location is so defined as the grid point with minimum RMSE calculated between the original and reconstructed time series (Sc[$i1$,$j1$]; black stars in Fig. 6). The spotted ideal location for AWI-LR, AWI-HR, and HadGEM-LL (Fig. 6a,b,e) are relatively close to each other, separated by a maximum of ∼600 km. Even though ECMWF-LR, ECMWF-HR, and HadGEM3-MM (Fig. 6c,d,f) suggest optimal locations that are placed farther from the sites suggested by the other datasets, their score maps still suggest a relatively good skill (low RMSE values) at the common region occupied by the 3 previous referred models. This fact justifies further the multi-model approach used in this work.

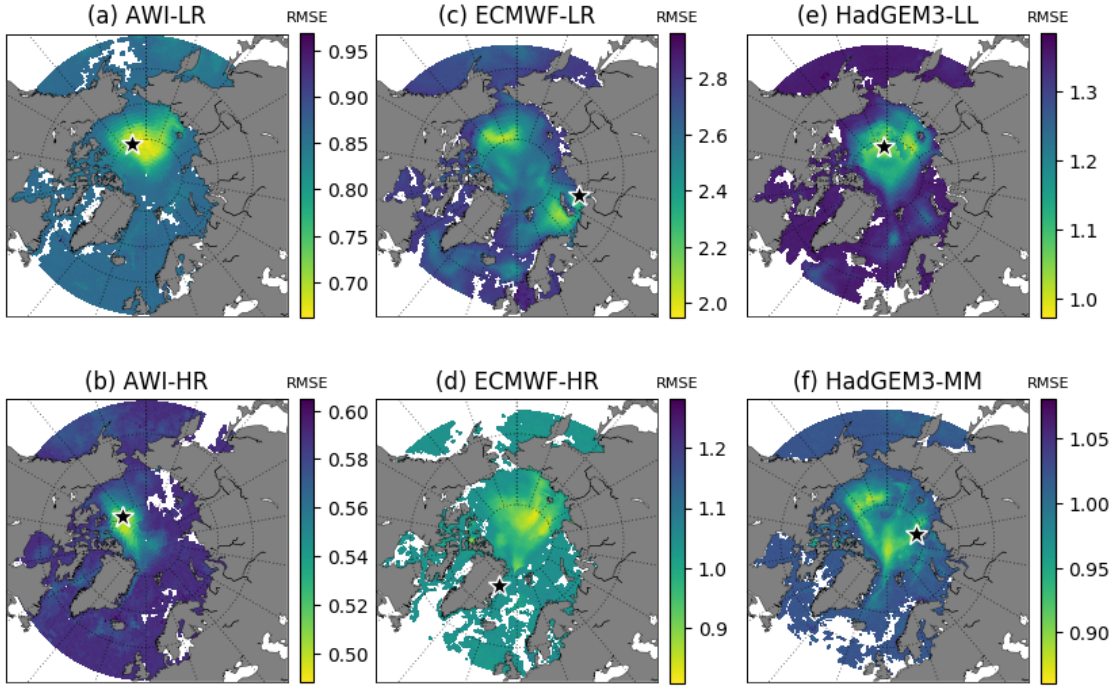


Figure 6. Score Maps (Sc[i,j]) represented by the RMSE (10^3km^3) calculated at every grid cell between the original and the reconstructed SIV anomalies. The smaller the RMSE error (shades of yellow), the higher the performance of the grid point for reconstructing the SIV anomaly. The black star indicates the 1st optimal location for each model Sc[$i1$,$j1$]. Notice that the colormap scale is different for each map.

The RMSEs (and associated STD from the Monte-Carlo scheme) calculated between the original SIV anomalies and the SIV anomalies reconstructed by the ESM, feed with predictor variables from the 1st optimal location (black stars in Fig. 6), are shown in the mid column of Table 2. Based on those values, predictor variables from the AWI systems can better reproduce the

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

Table 2. Mean RMSEs (and associated STDs) from the 500-Monte-Carlo realizations calculated between the original SIV anomalies and the SIV anomalies reconstructed by the ESM. We recall that in each Monte-Carlo realization 70% of the data is randomly used for training the SEM, while 30% is used for calculating the error. The first column shows the values for the case where the predictors are extracted from the individual optimal locations, while the second column shows the values found with predictors from the common optimal location.

Models	RMSE (Error) $\times 10^3 km^3$ 1st Optimal Location Individual location	RMSE (STD) $\times 10^3 km^3$ 1st Optimal Location Common location
AWI-LR	0.66 ($\pm$0.03)	0.67 ($\pm$0.03)
AWI-HR	0.49 ($\pm$0.02)	0.54 ($\pm$0.02)
ECMWF-LR	1.95 ($\pm$0.06)	2.11 ($\pm$0.09)
ECMWF-HR	0.81 ($\pm$0.03)	0.91 ($\pm$0.04)
HadGEM3-LL	0.97 ($\pm$0.04)	1.09 ($\pm$0.05)
HadGEM3-MM	0.86 ($\pm$0.05)	0.95 ($\pm$0.04)

SIV anomalies compared to the predictors from HadGEM and ECMWF. For the 3 models, the high-resolution version provides better statistical predictability.

A common score map, with the indication of a common 1st optimal location placed at the transition Chukchi Sea – Central Arctic – Beaufort Sea (158.0°W, 79.5°N), is shown in Fig. 7a. This common location is found through the ensemble mean of the scaled individual score maps, following the methodology described in Sec. 2.4. If we now come back to the score maps in Fig. 6 and retrieve the RMSE from that common location in Fig. 7a, we find the values displayed in the right column of Table 2. The predictive skill drops by about 10% when the common point is chosen for all models, except for AWI-LR which presents similar results for the two locations. Those values also reinforce that, at least for this 1st location, the predictors from the high-resolution outputs lead to a better predictive skill compared to the low-resolution predictors from their counterpart. Note that this was not the case when using pan-Arctic predictors in Section 3.1.

Once a 1st common optimal site is determined, we fix it for all datasets and so look for the 2nd best location. For that, the neighboring grid points which fell into the region of influence of the 1st best site are not considered as a second option. Fig. 7b shows the 1st location's region of influence.

The procedure followed for identifying the 1st site is so repeated for the nth next locations. Aiming at improving the reconstruction of the SIV anomalies, every time that a new location is set, the valid predictors from this new point add to the predictors from the previous stations into the SEM. Fig. 7c,e,g,i show the 2nd to the 5th optimal sites accompanied by their respective regions of influence (Fig. 7d,f,h,j). The 2nd site is the one closest to the North Pole, from where it is separated by a distance of about 167 km. The 3rd, 4th and 5th points are placed at the offshore domain of the Laptev Sea in the transition with the Central Arctic, in the Central Arctic to the north of the Canadian Islands, and in the central domain of the Beaufort Sea, respectively.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

The Cryosphere
Discussions
Open Access
EGU

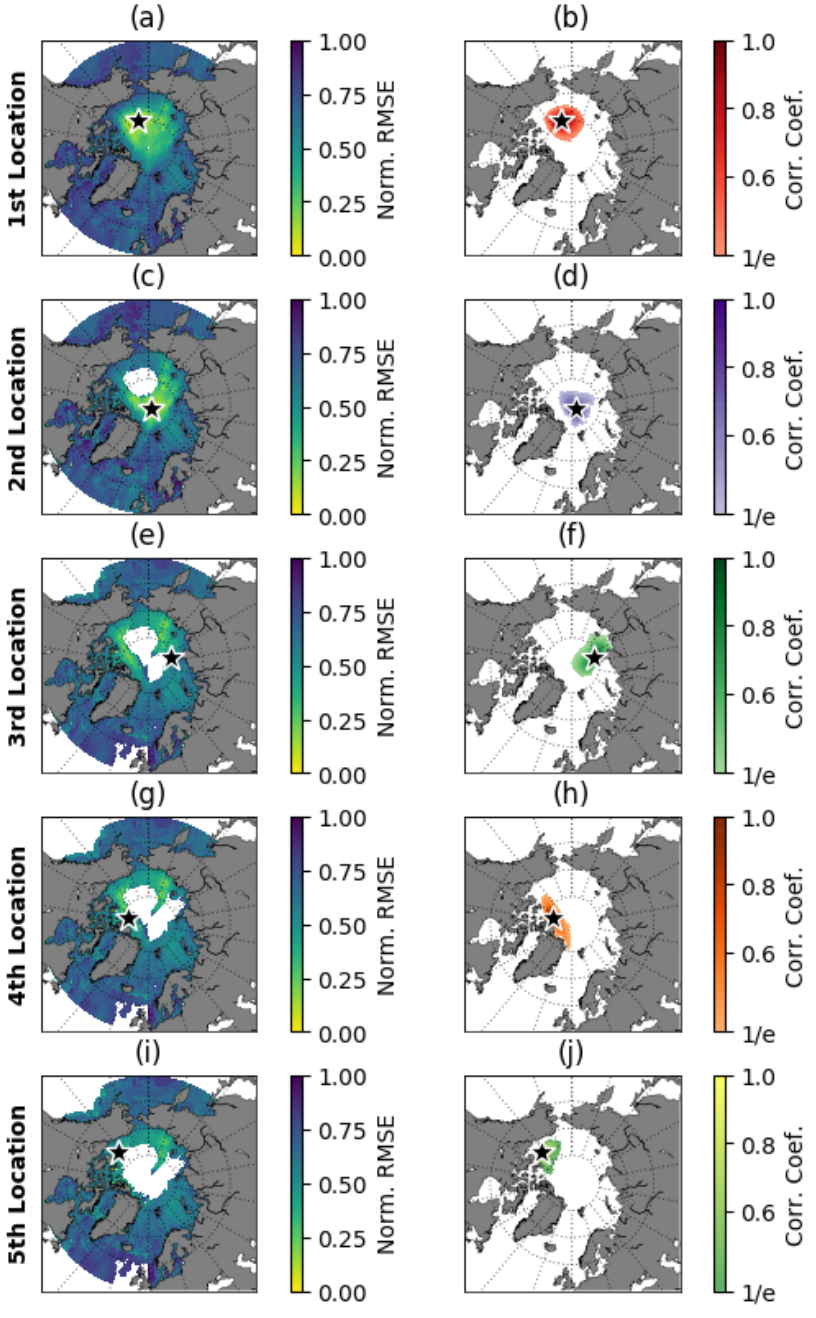


Figure 7. (a) Ensemble mean–normalized score map (ScNorm) for the 1st best sampling location. (b) The region of influence is defined for the 1st best location. The panels (c,d), (e,f), (g,h) and (i,j) represent the same as (a,b) but for the 2nd, 3rd, 4th and 5th best sampling locations, respectively.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

If we think of an optimal observing framework, in which only a few observational platforms are deployed, Fig. 8 represents an idealized scenario with the 10 best locations and their respective regions of influence. In such a context, the selection of points respects the hierarchy of the regions of influence in a way that the 2nd site can not be placed within the region of influence 1 (shades of red), the 3rd point can not be placed within the regions of influence 1 and 2 (shades of red and purple), and so on. Note that with the proposed methodology, the regions of influence from the 10 first locations are covering almost the entire Arctic Ocean and adjacent seas, with exception of the Canadian Archipelago, the Kara Sea, and the Greenland Sea (see Fig. 9). The question of whether or not is indeed required all 10 locations to fairly predict the SIV anomalies, both in terms of anomaly values and variability, will be answered in the next sections.

Table 3 displays the geographical coordinates of the 10 locations as well as the Arctic sub-regions occupied by them, as identified in Fig. 9. The division of the Arctic in sub-regions is based on the classical definition adopted by the broadly used Multisensor Analyzed Sea Ice Extent - Northern Hemisphere (MASIE-NH) product, which is made available by the National Snow & Ice Data Center (NSIDC). Most of the stations are placed within the Central Arctic (2nd, 4th, and 8th), or in the transition of this region with the Chukchi Sea (1st) and Laptev Sea (3rd), where the sea ice tends to be perennial. The 5th location is placed at the central part of the Beaufort Sea, the 6th and 9th stations are located at the offshore and inshore limits of the East Siberian Sea respectively, the 7th site is suggested to be at the Barents Sea off the Severny Island and, finally, the 10th station is occupying the near-coast side of the Laptev Sea.

Table 3. Geographical coordinates for the first 10 optimal sampling locations (second and third columns). The fourth column informs the sub-regions in which each of the points are placed in (see Fig. 9). The limits of the sub-regions are suggested by the National Snow & Ice Data Center (NSIDC).

Optimal Location	Latitude	Longitude	Sub-Region
# 1	79.5°N	158.0°W	Chukchi Sea (CS)
# 2	88.5°N	040.0°E	Central Arctic (CA)
# 3	81.5°N	107.0°E	Central Arctic (CA)
# 4	82.5°N	109.0°W	Central Arctic (CA)
# 5	74.5°N	136.0°W	Beaufort Sea (BeS)
# 6	77.5°N	155.0°E	East Siberian Sea (ESS)
# 7	78.5°N	054.0°E	Barents Sea (BrS)
# 8	83.5°N	001.0°W	Central Arctic (CA)
# 9	72.5°N	176.0°E	East Siberian Sea (ESS)
# 10	74.5°N	134.0°E	Laptev Sea (LS)

3.2.2 Reconstructed SIV anomaly

Once the set of ideal locations are established, these sites are used to effectively reconstruct the entire time series of SIV anomalies from the 6 model outputs, by taking into account only the valid predictors from each location. Again, we will make

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

The Cryosphere Discussions
Open Access
EGU

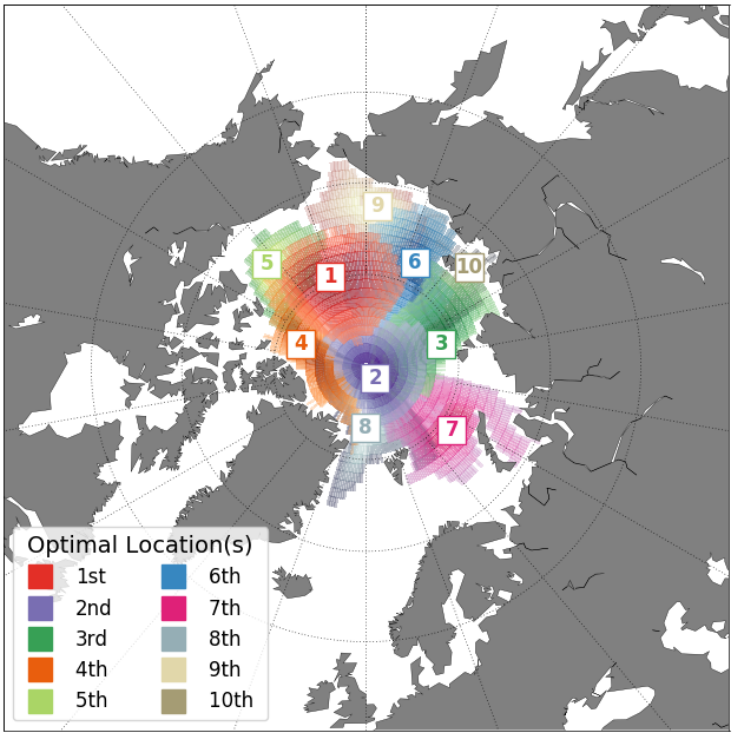


Figure 8. Optimal observing framework, as suggested by the ensemble of model outputs, for sampling predictor variables in order to statistically reconstruct and/or predict the Pan-Arctic SIV anomaly. The numbers indicate the 1st up to the 10th best observing locations in the respective order. The hatched area around each location (same color code) represents their respective region of influence. The selection of points respects the hierarchy of the regions of influence in a way that the 2nd point can not be placed within the region of influence 1 (shades of red), the 3rd point can not be placed within the regions of influence 1 and 2 (shades of red and purple), and so on.

use of the RMSE to evaluate how good is our statistical prediction in terms of absolute values, but here we are also interested in inspecting the ability of the empirical model to reproduce the full variability of the SIV anomalies. For that, apart from the RMSE, we also calculate the coefficient of determination (R^2) between the original and reconstructed time series.

Figure 10 compares the original (black lines) and the reconstructed times series by taking into account the 1st (red lines), the 3 first (green lines) and the 6 first (blue lines) locations. For the first reconstruction, RMSE values are almost identical to the ones shown in the second column of Table 2 (see Fig. 11a; y-axis=1). Again, for all 3 models, the predictor variables from the higher resolution versions present better performance in reproducing the SIV anomaly values. The relatively poor skill of the ECMWF-LR predictors compared to the other 5 systems is remarkable (Figure 10c).

Figure 11a summarizes the RMSE values for the reconstructions conducted with data from the only the 1st up to all 10 locations. The pattern of better prediction skill for the models with higher grid resolution revealed by the 1st location remains when more sites are incorporated into the SEM. From the ensemble means the RMSE ($\times 10^3 km^3$) values are, respectively,

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

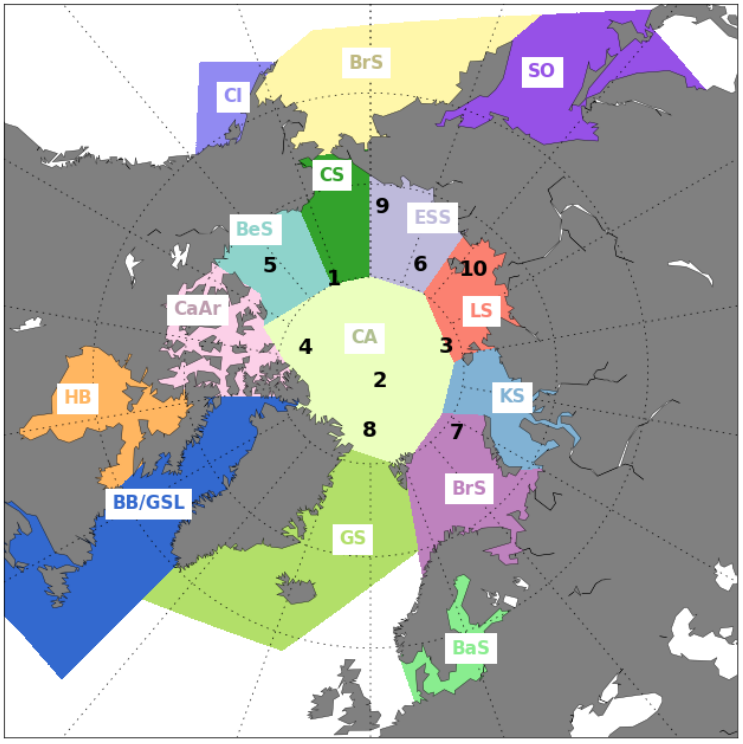


Figure 9. Optimal observing framework for sampling predictor variables in order to statistically reconstruct and/or predict the Pan-Arctic SIV anomaly. The numbers indicate the 1st up to the 10th optimal sites. Each of the colored areas represent an Arctic sub-region according to the Arctic subdivision suggested by the National Snow & Ice Data Center (NSIDC).

1.06, 0.95, 0.90, 0.81, 0.78, 0.70, 0.65, 0.63, 0.60 and 0.59 for the reconstruction with 1 to 10 locations (black curve/points in Fig. 11a). By excluding the outliers from ECMWF-LR, the previous RMSEs reduce to about 20% as shown by the gray curve-points in Fig. 11a). For most of the datasets, the statistical reconstruction seems to improve better until the incorporation of the 5th to 6th locations, from when on the improvement seems to attenuate (Figure 11a).

Figure 11b introduces a similar analysis but quantified by the R^2. Interestingly, for this metric, the ECMWF-LR is not outstanding from the others, and its predictors present a similar performance for reproducing the SIV anomaly variability. By account the reconstructions with 1 to 10 optimal sites, the ensemble means of R^2 values are: 0.53, 0.63, 0.67, 0.73, 0.75, 0.80, 0.81, 0.83, 0.84 and 0.84, respectively. These ensemble means suggest that the statistical empirical model could reproduce more than 60% of the SIV variability by using predictors from only the 3 first optimal locations. AWI and HadGEM datasets indicate that 4 locations are enough for reproducing more than 70% of the variability. With 6 well-positioned sites, about 80% of the SIV anomaly could be explained as suggested by the ensemble mean (Fig. 11b). As per the 6th station, the gain from adding new locations seems to be minimal (∼1%). Also interestingly is the fact that for the R^2 metric, the opposite from the RMSE is observed since the best performing predictors are the ones coming from the model's version with lower grid resolution.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

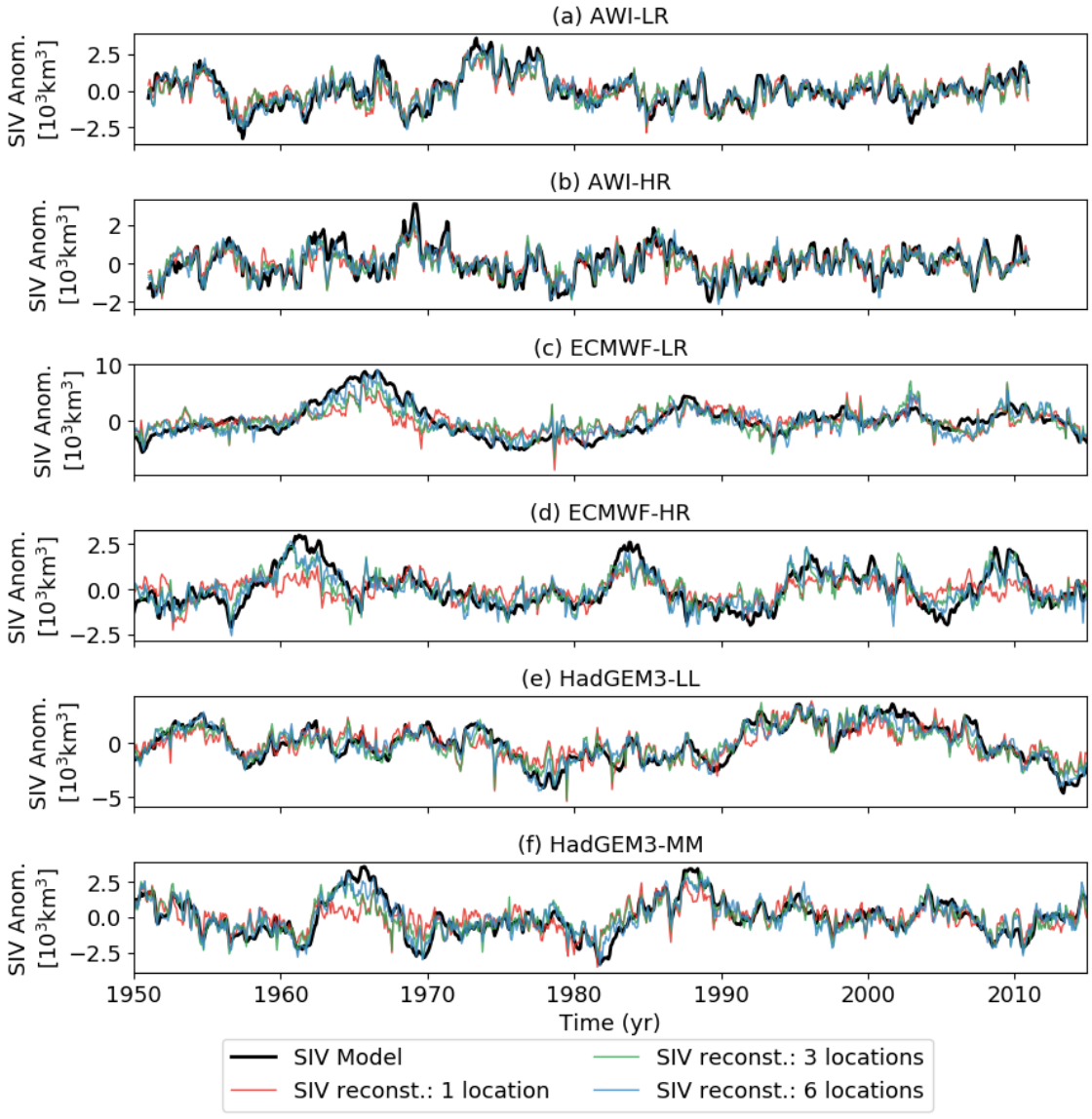


Figure 10. Original (black) and statistically reconstructed SIV anomalies. The reconstruction takes into account the 1st (red), the 3 first (1st–3rd; green) and the 6 first (1st–6th; blue) optimal locations: (a) AWI-LR, (b) AWI-HR, (c) ECMWF-LR, (d) ECMWF-HR, (e) HadGEM-LL and (f) HadGEM-MM. Notice the different scales in the y-axes.

In terms of used predictor variables, Figure 11c reiterates that SIT is the most skillful of the local predictors. From the 60 cases that the SEM was applied (6 datasets, 10 locations), SIT was used 59 times. SIT was not a valid predictor only for the 9th location in ECMWF-HR. SST and Drift were used in about two thirds (20 and 22 times, respectively), while SIC was used only in half (31 times) of the cases. If we look at the individual model outputs, HadGEM (the 2 resolutions comprised) is the

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

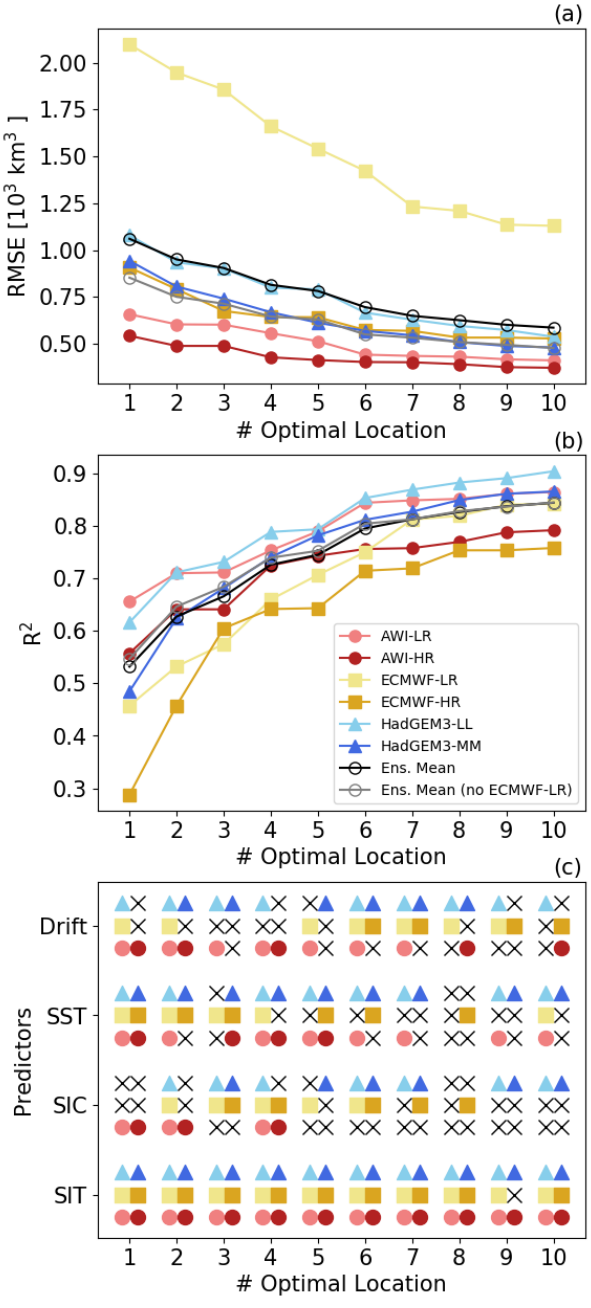


Figure 11. (a) RMSE (y-axis) estimated between the original and reconstructed time series by taking into account predictor variables from 1 up to 10 optimally selected locations (x-axis). (b) Same as (a) but using R^2 (y-axis) to compare original and reconstructed time series. (c) Valid predictors, as determined by the correlation maps, retrieved from each targeted location. If a predictor is valid (y-axis), its respective symbol, as defined in the inset legend from (b), is plotted. A black cross indicates that the predictor is not valid at the respective location.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

one in which the empirical model takes the best advantage of the available gridded predictors, having neglected one of them in only 15 out of 80 cases, while ECMWF and AWI have ignored predictors in 29 and 30 out of 80 cases, respectively.

4 Discussion and conclusions

In this work, we have introduced a statistical empirical model for predicting the Arctic SIV anomaly (no trend; no seasonal cycle). on the interannual time scale. The model was built and tested with data from 3 AOGCMs (AWI-CM, ECMWF-IFS, and HadGEM3-GC3.1), each of which provided with 2 horizontal resolutions, performing a total of 6 datasets. We have first inspected the predictive skill of 7 different pan-Arctic predictors, namely: SIV, SIA, OHT, SIT, SIC, SST, and Drift. These predictors were tested since they have dynamical and/or thermodynamical influence on the SIV. The 3 first are intrinsically represented by single time series, while the remaining are gridded variables that were reduced to mean pan-Arctic time series. From this first assessment, performed for the months of March and September, the results (Section 3.1) show that the best predictors are the SIV itself and the SIT, whilst SST, Drift, SIC and SIA provide some intermediate-skill predictions. In general, such results are valid for predictions performed from 1 back to 12 leading months. For the SIV predictor, the skill substantially increases in the last 3 leading months. For the remaining aforementioned predictors, the skill slightly improves from 12 to 1 leading month. OHT provided a very poor predictive skill. Docquier et al. (2019) recently showed (their Fig. 12) a relatively good correlation between OHT and the SIV. However, these authors correlated annual averages of OHT against monthly values of SIV, but here we are considering monthly means for all predictors. Based on that, the results from both manuscripts suggest that the OHT has a cumulative impact on the sea ice throughout the year, which is not so remarkable when looking at individual months, even if several leading months are considered. From Section 3.1's results is also noticeable that the ECMWF-LR predictors present a relatively poor skill compared to the others. This is explained by the fact that this model has a mean state characterized by a much thicker sea ice (see Fig. 1), impacting the RMSE used as a metric for evaluating the prediction skill. That being said, we can recapitulate and objectively answer the first open question posed in the introduction of this manuscript:

(i) What are the performance of different pan-Arctic predictors for predicting pan-Arctic SIV anomalies?

If we take into account the ensemble mean, and use the average RMSE calculated between original and reconstructed SIV time series (Section 3.1; Figs. 4 and 5) for the last 3 leading months as score, the best predictors for March are sorted in the following order: SIV (0.41×10^3km^3), SIT (0.78×10^3km^3), SIA (1.01×10^3km^3), SIC (1.10×10^3km^3), SST (1.15×10^3km^3), Drift (1.32×10^3km^3) and OHT (2.05×10^3km^3). The best predictors for September are sorted as: SIV (0.45×10^3km^3), SIT (0.76×10^3km^3), SST (0.96×10^3km^3), SIA (1.07×10^3km^3), SIC (1.12×10^3km^3), Drift (1.22×10^3km^3) and OHT (2.24×10^3 km^3). If ALL predictors are used (except SIV itself), the averaged scores for 3 leading months are 0.70×10^3km^3 for both March and September, respectively.

Once the statistical empirical model is developed and the potential predictor variables are identified, we made use of this information for recommending an optimal observing system. Such observations could eventually be performed in the framework

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

of an operational oceanography program to continuously provide predictor data for the statistical model. So, we considered parameters that could be locally sampled by observing platforms (e.g., oceanographic moorings and/or buoys) as SIT, SST and Drift. It is fair also to consider the SIC and the pan-Arctic SIA since this information is regularly provided from satellite measurements. The OHT and the SIV are here disregarded as predictors. The first did not turn out to be a skillful predictor (at least not when using monthly means). The second is the variable that we supposedly do not have and the one we want to predict. From a realistic point of view, our analyses were restricted to a maximum of 10 optimal locations, although a reduced number of stations would be already enough to fairly reproduce the SIV anomaly, and so to explain a large amount of its variance (see below). The results from Section3.2 provide us with elements to answer the other 3 open questions of this study, as follows:

(ii) What are the best in situ locations for sampling predictor variables to optimize the statistical predictability of SIV anomalies in terms of reproducibility and variability?

We have here identified 10 optimal locations. The exact coordinates of these locations are provided in Table 3 and also plotted in Figs. 8 and 9. As suggested by the ensemble of model outputs, the 1st optimal location is placed at the transition Chukchi Sea–Central Arctic–Beaufort Sea (158.0°W, 79.5°N). The 2nd, 3rd and 4th best locations are placed near the North Pole (40°E, 88.5°N), at the transition Central Arctic–Laptev Sea (107°E, 81.5°N) and offshore the Canadian Archipelago (109.0°W, 82.5°N).

(iii) How many optimal sites are needed for explaining a large amount, that is to say, at least 70% of the original SIV anomaly variance?

By considering an arbitrary threshold of 70%, the systems AWI-LR (75%), AWI-HR (73%), HadGEM3-LL (79%) and HadGEM3-MM (74%) suggest that only 4 stations are enough to overpass this threshold, what is also confirmed by the ensemble mean (73%). Even though the ECMWF predictors have slightly low skill, they are still not far from the threshold: ECMWF-LR (66%) and ECMWF-HR (64%). The ensemble mean indicates that 5 and 6 well-placed stations could explain about 75% and 80% of the SIV anomaly variance, respectively. As per these numbers, the statistical predictability does not substantially improve by adding new sites, taking into account that 10 locations explain about 84% of the variance. However, as suggested by Fig. 8, even though the SEM seems to fairly reproduce the SIV anomaly variance and, therefore, the long-term variability, it found more difficulties to reproduce the short-term variabilities.

(iv) Are the results model dependent, in particular, are they sensitive to horizontal resolution?

The results suggest that statistical predictability is affected by model resolution. Notwithstanding, the question of whether or not a finer horizontal resolution provides better statistical predictability depends on the metric used to evaluate the predictions (Section 3.2.2 and Fig. 11). That is the case for RMSE, where the main target is to evaluate the reproducibility of the reconstructed values. It seems that an improved horizontal resolution allows a better trained statistical model so that the reconstructed values approach better to the original SIV anomaly (Fig. 11a). On the other hand, if we look at the interannual variability, the predictors provided by numerical models with lower resolution are more able to approach the reconstructed time series to the

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

original SIV anomaly (Fig. 11b). In this case, it is possible to argue that the low-resolution versions provide smoother time series, with less amount of short-term variability, making it easier for the statistical model to represent the long-term variation of SIV anomaly over time. Along the same lines, it might be harder to achieve skillful predictions in the real world employing statistical tools because the actual SIV variability is likely noisier than the one described by AOGCM outputs. Some caution should be taken since our findings could be slightly different for other AOGCMs. A good perspective for addressing this issue is to reapply the methodology developed in this manuscript, but using all models that will be made available through the CMIP6. Also, with the sea ice depletion, some of the optimal sampling locations here suggested might be in a free-ice region in the future.

Finally, it is worthwhile mentioning the recent effort from the scientific community to enhance the Arctic observational system. This effort takes place through recent observational programs such as the Year Of Polar Prediction (YOPP) (Jung et al., 2016) and the MOSAiC International Arctic Drift Expedition (https://www.mosaic-expedition.org/; last access: 23 July 2019). Within this context, we expect that this manuscript will provide recommendations for the ongoing and upcoming initiatives towards an Arctic optimal observing design.

Competing interests. No competing interests are present.

Acknowledgements. The work presented in this paper has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement no. 727862: APPLICATE project (Advanced prediction in Polar regions and beyond). François Massonnet and Leandro Ponsoni are F.R.S.-FNRS PostDoc and Research Associate, respectively. David Docquier is funded by the EU Horizon 2020 PRIMAVERA project, grant agreement no. 641727. Guillian Van Achter is founded by PARAMOUR project which is supported by the Excellence Of Science programme (EOS), also founded by FNRS. The datasets used in this study were made available through PRIMAVERA project.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

References

Amélineau, F., Grémillet, D., Bonnet, D., Le Bot, T., and Fort, J.: Where to Forage in the Absence of Sea Ice? Bathymetry As a Key Factor for an Arctic Seabird, PLoS ONE, 11(7), doi:10.1371/journal.pone.0157764, 2016.

Barnett, D. G.: Empirical orthogonal functions and the statistical predictability of sea ice extent, in Sea Ice Processes and Models, R. S. Pritchard, ed., Univ. Wash. Press, Seattle, 1980.

Blanchard-Wrigglesworth, E. and Bitz, C.: Characteristics of Arctic Sea-Ice Thickness Variability in GCMs, J. Clim., 27, 8244–8258, doi:10.1175/JCLI-D-14-00345.1, 2014.

Brown, T. A., Galicia, M. P., Thiemann, G. W., Belt, S. T., Yurkowski, D. J., and Dyck, M. G.: High contributions of sea ice derived carbon in polar bear (Ursus maritimus) tissue, PLoS ONE, 13(1), e0191 631, doi:10.1371/journal.pone.0191631, 2016.

Burgard, C. and Notz, D.: Drivers of Arctic Ocean warming in CMIP5 models, Geophys. Res. Let., 44, 4263–4271, doi:10.1002/2016GL072342, 2017.

Chapman, W. L. and Walsh, J. E.: Recent variations of sea ice and air temperature in high latitudes, Bull. Amer. Meteor. Soc., 74, 33–47, doi:10.1175/1520-0477(1993)074<0033:RVOSIA>2.0.CO;2, 1993.

Chevallier, M. and Salas-Mélia, D.: The role of sea ice thickness distribution in the Arctic sea ice potential predictability: a diagnostic approach with a coupled GCM, J. Clim., 25, 3025–3038, doi:10.1175/JCLI-D-11-00209.1, 2012.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Overland, D. E. J., and Jones, J.: Recent Arctic amplification and extreme mid-latitude weather, Nat. Geosci., 7, 627–637, doi:10.1038/NGEO2234, 2014.

Cullen, M. J. P.: The unified forecast/climate model, Meteorological Magazine, 112, 81–94, 1993.

Docquier, D., Grist, J. P., Roberts, M. J., Roberts, C. D., Semmler, T., Ponsoni, L., Massonnet, F., Sidorenko, D., Sein, D. V., Iovino, D., Bellucci, A., and Fichefet, T.: Impact of model resolution on Arctic sea ice and North Atlantic Ocean heat transport, Clim. Dyn., 2019.

Drijfhout, S.: Competition between global warming and an abrupt collapse of the AMOC in Earth's energy imbalance, Sci. Rep., 5, 1–12, doi:10.1038/srep14877, 2015.

Drobot, S. D. and Maslanik, J. A.: A practical method for long-range forecasting of sea ice severity in the Beaufort Sea, Geophys. Res. Lett., 29, 54–1–54–4, doi:10.1029/2001GL014173, 2002.

Drobot, S. D., Maslanik, J. A., and Fowler, C. F.: A long-range forecast of Arctic Summer sea-ice minimum extent, Geophys. Res. Lett., 33, L10 501, doi:10.1029/2006GL026616, 2006.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., E., R. J. S. K., and Taylor: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, doi:10.5194/gmd-9-1937-2016, 2016.

Fichefet, T. and Maqueda, M. A. M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, J. Geophys. Res., 102, 12 609–12 646, doi:10.1029/97JC00480, 1997.

Francis, J. A. and Vavrus, S. J.: Evidence linking Arctic amplification to extreme weather in mid-latitudes, Geophys. Res. Lett., 39, L06 801, doi:10.1029/2012GL051000, 2012.

Gleick, P. H.: The implications of global climatic changes for international security, Climatic Change, 15, 309–325, doi:10.1007/BF00138857, 1989.

Grémillet, D., Fort, J., Amélineau, F., Zakharova, E., Le Bot, T., Sala, E., and Gavrilo, M.: Arctic warming: nonlinear impacts of seaice and glacier melt on seabird foraging, 21, 1116–1123, doi:10.1111/gcb.12811, 2015.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

Grunseich, G. and Wang, B.: Predictability of Arctic Annual Minimum Sea Ice Patterns, J. Climate, 29, 7065–7088, doi:10.1175/JCLI-D-16-0102.1, 2016.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., J, S., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, doi:10.5194/gmd-9-4185-2016, 2016.

Handorf, U.: Tourism booms as the Arctic melts. A critical approach of polar tourism, GRIN Verlag, Munich, 2011.

Hunke, E., Turner, W. L. A., Jeffery, N., and Elliott, S.: CICE: the Los Alamos sea ice model, documentation and software user's manual, Version 5.1. LA-CC-06-012, Tech. rep., Los Alamos National Laboratory, p. 116pp, 2015.

Johnson, C. M., Lemke, P., and Barnett, T. P.: Linear prediction of sea ice anomalies, J. Geophys. Res., 90, 5665–5675, doi:10.1029/JD090iD03p05665, 1985.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., amd M. Holland, H. F. G., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M., and Yang, Q.: Advancing polar prediction capabilities on daily to seasonal time scales, Bull. Am. Meteorol. Soc., 97, 1631–1647, doi:10.1175/BAMS-D-14-00246.1, 2016.

Kwok, R. and Cunningham, G. F.: Variability of Arctic sea ice thickness and volume from CryoSat-2, Phil.Trans.R.Soc.A, 73: 20140157, doi:10.1098/rsta.2014.0157, 2015.

Kwok, R., Cunningham, G. F., Zwally, H. J., and Yi, D.: Ice, Cloud, and land Elevation Satellite (ICESat) over Arctic sea ice: Retrieval of freeboard, J. Geophys. Res., 112, C12 013, doi:10.1029/2006JC003978, 2007.

Laidre, K. L., Stirling, I., Lowry, L. F., Wiig, Ø., Heide-Jørgensen, M. P., and Ferguson, S. H.: Quantifying the sensitivity of Arctic marine mammals to climate-induced habitat change, Ecol. Appl., 18(2), S97–S125, doi:10.1890/06-0546.1, 2008.

Lindsay, R. W.: A new sea ice thickness climate data record, EOS, 91, 405–406, doi:10.1029/2010EO440001, 2010.

Lindsay, R. W. and Zhang, J.: Arctic Ocean Ice Thickness: Modes of Variability and the Best Locations from Which to Monitor Them, J. Phys. Oceanogr., 36, 496–506, doi:10.1175/JPO2861.1, 2006.

Lindsay, R. W., Zhang, J., Schweiger, A. J., and Steele, M. A.: Seasonal predictions of ice extent in the Arctic Ocean, J. Geophys. Res., 113, C02 023, doi:10.1029/2007JC004259, 2008.

Lindstad, H., Bright, R. M., and Strømmanb, A. H.: Economic savings linked to future Arctic shipping trade are at odds with climate change mitigation, Transp. Policy, 45, 24–34, doi:10.1016/j.tranpol.2015.09.002, 2016.

Lydersen, C., Vaquie-Garcia, J., Lydersen, E., Christensen, G. N., and Kovacs, K. M.: Novel terrestrial haul-out behaviour by ringed seals (Pusa hispida) in Svalbard, in association with harbour seals (Phoca vitulina), Polar Research, 36:1, 1374 124, doi:10.1080/17518369.2017.1374124, 2017.

Madec, G.: NEMO ocean engine, Tech. rep., Institut Pierre-Simon Laplace (IPSL), 2008.

Madec, G. and Imbard, M.: A global ocean mesh to overcome the north pole singularity, Clim. Dyn., 12, 381–388, doi:10.1007/BF00211684, 1996.

Madec, G., Bourdallé-Badie, R., Bouttier, P., Bricaud, C., Bruciaferri, D., Calvert, D., Chanut, J., Clementi, E., Coward, A., Delrosso, D., Ethé, C., Flavoni, S., Graham, T., Harle, J., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., Masson, S., Mocavero, S., Paul, J., Rousset, C., Storkey, D., Storto, A., and Vancoppenolle, M.: NEMO ocean engine (Version v3.6), Notes Du Pôle De Modélisation De L'institut Pierre-simon Laplace (IPSL), doi:10.5281/zenodo.1472492, 2017.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

Notz, D. and Stroeve, J.: Observed Arctic sea-ice loss directly follows anthropogenic CO_2 emission, Science, 354 (6313), 747–750, doi:10.1126/science.aag2345, 2016.

Nuttall, M., Berkes, F., Forbes, B., Kofinas, G., Vlassova, T., and Wenzel, G.: Arctic Climate Impact Assessment, Cambridge University Press, Cambridge, 2005.

Overland, J. E. and Wang, M.: Large-scale atmospheric circulation changes are associated with the recent loss of Arctic sea ice, Tellus, 62A, 1–9, doi:10.1111/j.1600-0870.2009.00421.x, 2010.

Pagano, A. M., Durner, G. M., Rode, K. D., Atwood, T. C., Atkinson, S. N., Peacock, E., Costa, D. P., Owen, M. A., and Williams, T. M.: High-energy, high-fat lifestyle challenges an Arctic apex predator, the polar bear, Science, 359 (6375), 568–572, doi:10.1126/science.aan8677, 2018.

Parkinson, C. L. and Cavalieri, D. L.: A 21-year record of Arctic sea-ice extents and their regional, seasonal and monthly variability and trends, Ann. Glaciol., 34, 441–446, doi:10.3189/172756402781817725, 2002.

Parkinson, C. L., Cavalieri, D. L., Gloersen, P., Zwally, H. J., and Comiso, J. C.: Arctic sea ice extents, areas, and trends, 1978–1996, J. Geophys. Res., 104, 20 837–20 856, doi:10.1029/1999JC900082, 1999.

Petty, A. A., Stroeve, J. C., Holland, P. R., Boisvert, L. N., Bliss, A. C., Kimura, N., and Meier, W. N.: The Arctic sea ice cover of 2016: a year of record-low highs andhigher-than-expected lows, The Cryosphere, 13, 433–452, doi:10.5194/tc-12-433-2018, 2018.

Ponsoni, L., Massonnet, F., Fichefet, T., Chevallier, M., and Docquier, D.: On the timescales and length scales of the Arctic sea ice thickness anomalies: a study based on 14 reanalyses, The Cryosphere, 13, 521–543, doi:10.5194/tc-13-521-2019, 2019.

Rackow, T., Goessling, H. F., Jung, T., Sidorenko, D., Semmler, T., Barbi, D., and Handorf, D.: Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part II: climate variability, Clim. Dyn., 50, 2369–2394, doi:10.1007/s00382-016-3192-6, 2018.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., and Keeley, S.: ECMWF ECMWF-IFS-HR model output prepared for CMIP6 HighResMIP, doi:10.22033/ESGF/CMIP6.2461, https://doi.org/10.22033/ESGF/CMIP6.2461, 2017a.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., and Keeley, S.: ECMWF ECMWF-IFS-LR model output prepared for CMIP6 HighResMIP, doi:10.22033/ESGF/CMIP6.2463, https://doi.org/10.22033/ESGF/CMIP6.2463, 2017b.

Roberts, C. D., Senan, R., Molteni, F., Boussetta, S., Mayer, M., and Keeley, S. P. E.: Climate model configurations of the ECMWF Integrated Forecast System (ECMWF-IFS cycle 43r1) for HighResMIP, Geosci. Model Dev., 11, 3681–3712, doi:10.5194/gmd-11-3681-2018, 2018.

Roberts, M.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 HighResMIP, doi:10.22033/ESGF/CMIP6.1901, https://doi.org/10.22033/ESGF/CMIP6.1901, 2017a.

Roberts, M.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 HighResMIP, doi:10.22033/ESGF/CMIP6.1902, https://doi.org/10.22033/ESGF/CMIP6.1902, 2017b.

Roberts, M. J., Baker, A., Blockley, E. W., Calvert, D., Coward, A., Hewitt, H. T., Jackson, L. C., Kuhlbrodt, T., Mathiot, P., Roberts, C. D., Schiemann, R., Seddon, J., Vannière, B., and Vidale, P. L.: Description of the resolution hierarchy of the global coupled HadGEM3-GC3.1 model as used in CMIP6 HighResMIP experiments, Geosci. Model Dev., doi:10.5194/gmd-2019-148, 2019.

Rothrock, D. A., Yu, Y., and Maykut, G. A.: Thinning of the arctic sea ice cover, Geophys. Res. Let., 26, 3469–3472, doi:10.1029/1999GL010863, 1999.

Schweiger, A. J., Wood, K. R., and Zhang, J.: Arctic Sea Ice Volume Variability over 1901–2010: A Model-Based Reconstruction, J. Climate, 32, 4731–4752, doi:10.1175/JCLI-D-19-0008.1, 2019.

Screen, J. A. and Simmonds, I.: Exploring links between Arctic amplification and mid-latitude weather, Geophys. Res. Lett., 40, 959–964, doi:10.1002/GRL.50174, 2013.

https://doi.org/10.5194/tc-2019-257
Preprint. Discussion started: 22 November 2019

Sein, D. V., Danilov, S., Biastoch, A., Durgadoo, J. V., Sidorenko, D., Harig, S., and Wang, Q.: Designing variable ocean model resolution based on the observed ocean variability, J. Adv. Model Earth Sy., 8(2), 904–916, doi:10.1002/2016MS000650, 2016.

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Hegewald, J., Sein, D., Wang, Q., and Jung, T.: AWI AWI-CM 1.1 HR model output prepared for CMIP6 HighResMIP, http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.AWI.AWI-CM-1-1-HR, 2017a.

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Hegewald, J., Sein, D., Wang, Q., and Jung, T.: AWI AWI-CM 1.1 LR model output prepared for CMIP6 HighResMIP, http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.HighResMIP.AWI.AWI-CM-1-1-LR, 2017b.

Serreze, M. C., Holland, M. M., and Stroeve, J.: Perspectives on the Arctic's Shrinking Sea-Ice Cover, Science, 315, 1533–1536, doi:10.1126/science.1139426, 2007.

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H. F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D. V., and Wang, Q.: Towards multi-resolution global climate modeling with ECHAM6-FESOM. Part I: model formulation and mean climate, Clim. Dyn., 44, 757–780, doi:10.1007/s00382-014-2290-6, 2015.

Stroeve, J. C., Holland, M. M., Meier, W. N., Scambos, T., and Serreze, M.: Arctic sea ice decline: Faster than forecast, Geophys. Res. Let., 34, L09 501, doi:10.1029/2007GL029703, 2007.

Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M. M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, Geophys. Res. Let., 39, L16 502, doi:10.1029/2012GL052676, 2012.

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E., and Jahn, A.: Influence of internal variability on Arctic sea-ice trends, Nat. Clim. Change, 5, 86–89, doi:10.1038/nclimate2483, 2015.

Sévellec, F., Fedorov, A. V., and Liu, W.: Arctic sea-ice decline weakens the Atlantic Meridional Overturning Circulation, Nat. Clim. Change, 7, 604–610, doi:10.1038/nclimate3353, 2017.

Tang, Q., Zhang, X., Yang, X., and Francis, J. A.: Cold winter extremes in northern continents linked to Arctic sea ice loss, Environ. Res. Lett., 8, 014 036, doi:10.1088/1748-9326/8/1/014036, 2013.

Tilling, R. L., Ridout, A., and Shepherd, A.: Estimating Arctic sea ice thickness and volume using CryoSat-2 radar altimeter data, Adv Space Res., 62, 1203–1225, doi:10.1016/j.asr.2017.10.051, 2018.

Walsh, L. E.: A long-range ice forecast method for the north coastof Alaska, in Sea Ice Processes and Models, R. S. Pritchard, ed., Univ. Wash. Press, Seattle, 1980.

Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., and Schröter, J.: The Finite Element Sea Ice-Ocean Model (FESOM) v.1.4: formulation of an ocean general circulation model, Geosci. Model Dev., 7, 663–693, doi:10.5194/gmd-7-663-2014, 2014.

Wilder, J. M., Vongraven, D., Atwood, T., Hansen, B., Jessen, A., Kochnev, A., York, G., Vallender, R., Hedman, D., and Gibbons, M.: Polar bear attacks on humans: Implications of a changing climate, Wildl. Soc. Bull., 41(3), 537–547, doi:10.1002/wsb.783, 2017.

Yuan, X., Chen, D., Li, C., Wang, L., and Wang, W.: Arctic Sea Ice Seasonal Prediction by a Linear Markov Model, J. Climate, 29, 8151–8173, doi:10.1175/JCLI-D-15-0858.1, 2016.

Zhang, X. and Walsh, J. E.: Toward a Seasonally Ice-Covered Arctic Ocean: Scenarios from the IPCC AR4Model Simulations, J. Clim., 19, 1730–1747, doi:10.1175/JCLI3767.1, 2006.