Universität Rostock Traditio et Innovatio

The potential of the Baltic Sea for seasonal to decadal (S2D) predictability

Master's thesis
written at the Institute of Physics
of the Faculty of Mathematics and Natural Sciences
of the University of Rostock

Marti Wolff, born 22nd October 1999 in Magdeburg

Supervisor and 1. Examiner : Prof. Dr. H. E. M. Meier, Leibniz Institute for Baltic Sea Research Warnemünde, Rostock
2. Examiner: Dr. F. Börgel, Leibniz Institute for Baltic Sea Research Warnemünde, Rostock

Rostock, 21.10.2025

https://doi.org/10.18453/rosdok_id00005038

Abstract

Seasonal to decadal (S2D) predictions form a research subject of growing importance providing valuable information for numerous stakeholders. The Baltic Sea's predictability on S2D timescales was investigated by analyzing an ensemble run of the Modular Ocean Model 6 (MOM6) in regard to the temporal variability, memory capacity and response to external forcing of the sea. Potential for skillful decadal salinity predictions of the entire Baltic Sea, as well as the Gotland Sea and the Gulf of Bothnia, was found due to the salinity memory and partial predictability of freshwater forcing and thereby the low-frequent salinity variability of the Baltic Sea. Further results indicate seasonal predictability of the Baltic Sea water temperature and Baltic Sea, Gotland Sea and Gulf of Bothnia sea ice, based on the temperature memory of the Baltic Sea. Generally, future predictions of any variable should aim to leverage the combination of memory capacity and accurate estimations of future forcing to maximize their skill.

Zusammenfassung

Saisonale bis dekadische (S2D) Vorhersagen sind ein Forschungsgebiet von wachsender Relevanz, welches zahlreichen Interessengruppen wertvolle Informationen zur Verfügung stellt. Die Vorhersagbarkeit der Ostsee auf S2D Zeitskalen wurde untersucht, indem ein Ensemblelauf des Modular Ocean Models 6 (MOM6) hinsichtlich zeitlicher Variabilität, Vorhersagegedächtnis und Reaktion der Ostsee auf äußeren Antrieb analysiert wurde. Potenzial für dekadische Vorhersagen des Salzgehaltes der gesamten Ostsee, sowie der Gotlandsee und des Bottnischen Meerbusens, wurde durch das Vorhersagegedächtnis für Salzgehalt und die teilweise Vorhersagbarkeit des Frischwasserantriebs und damit verbunden der niederfrequenten Salzgehalt-Variabilität der Ostsee gefunden. Weitere Ergebnisse deuten auf saisonale Vorhersagbarkeit der Wassertemperatur der Ostsee und des Seeeises von Ostsee, Gotlandsee und Bottnischem Meerbusen hin, basierend auf dem Vorhersagegedächtnis für die Temperatur der Ostsee. Allgemein sollten zukünftige Vorhersagen von jeder Variable versuchen, die Kombination von Erinnerungsvermögen und präzisen Einschätzungen des zukünftigen Antrieb für maximalen Vorhersageskill zu nutzen.

List of Figures

List of Tables

List of Abbreviations

AB Arkona Basin.

ACC Anomaly correlation coefficient.

AMV Atlantic Multidecadal Variability.

ARP Annually resolved predictions.

BS Baltic Sea.

CMIP Coupled Model Intercomparison Project.

DCPP Decadal Climate Prediction Project.

GB Gulf of Bothnia.

GS Gotland Sea.

Ice CN Ice concentration.

Ice HI Ice thickness.

MBI Major Baltic Inflow.

MOM6 Modular Ocean Model 6.

MRP Monthly resolved predictions.

MSSS Mean square skill score.

NAO North Atlantic Oscillation.

NOAA National Oceanic and Atmospheric Administration.

S2D Seasonal to decadal.

SSH Sea surface height.

SST Sea surface temperature.

Contents

Chapter 1.

Introduction

Seasonal to decadal (S2D) predictions bridge the gap between long-term climate projections like IPCC [2023] and Meier et al. [2022] and short-term weather forecasts [Boer et al., 2016]. They deliver crucial information for decision making and planning of the next 10 years and help society anticipate, prepare for, and adapt to climate variability and change over the next decade [O'Kane et al., 2023]. More precisely, decadal predictions facilitate the development of climate-resilient farming practices [Solaraju-Murali et al., 2022], the adaption of the energy sector to modern day challenges [Hutchins et al., 2025], better understanding of fish stock evolution under changing ocean conditions [Payne et al., 2022] and comprehension of climate change impact on future tourism [Gössling and Scott, 2025], among many more applications.

Mathematically, S2D predictions lie at the intersection of initial-value and boundary-value problem [Meehl et al., 2021, Boer et al., 2016, Stocker et al., 2013]. Understanding and modeling both aspects is crucial [Kushnir et al., 2019], particularly considering the influence of model bias and drift [Meehl et al., 2022, 2021]. The Decadal Climate Prediction Project (DCPP) aims to enhance the understanding of decadal climate predictability and variability [Boer et al., 2016] in order to make a contribution to the Coupled Model Intercomparison Project 6 (CMIP6) [Eyring et al., 2016]. A key challenge, especially on decadal timescales, forms the so-called signal-to-noise paradox, where the predictable signal is underestimated and masked by internal variability [Scaife and Smith, 2018, Smith et al., 2020]. Recent studies tackle this obstacle by using large ensembles to obtain meaningful prediction skill [Smith et al., 2019, Yeager et al., 2018, Eade et al., 2014, Dobrynin et al., 2022]. Furthermore the utilization of the slowly varying major modes of the climate system presents potential for skillful S2D predictability [Smith et al., 2019,

Meehl et al., 2021, Mariotti et al., 2020, Smith et al., 2020].

While the Baltic Sea hasn't been specifically investigated regarding its S2D predictability, the proximate North Atlantic ocean already shows potential for skillful predictions on decadal time frames. Borchert et al. [2018, 2019] found decadal hindcast skill for North Atlantic sea surface temperature if northward ocean heat transport during initialization is significantly weaker or stronger than on average. Lipfert et al. [2022] concluded that North Atlantic variability in marine heat wave days and frequency is predictable for up to 8 years, and, likewise, Fan et al. [2023] demonstrated skillful salinity predictions over the same time horizon by accounting for downstream effects in the North Atlantic subpolar gyre. This raises the question of whether the Baltic Sea also holds potential for skillful decadal prediction and how significant the North Atlantic's role could be in that regard.

It is known that the North Atlantic Oscillation (NAO) and the Atlantic Multidecadal Variability (AMV), two large-scale modes located in the North Atlantic, are key drivers of climate variability of Northern Europe and the Baltic Sea region in particular [Börgel et al., 2020, 2022]. The former describes changes in the normalized sea-level pressure difference between the Azores High and the Icelandic Low [Hurrell, 1995]. It influences Northern Europe's climate by modulating the strength and path of westerly winds across the Atlantic [Hurrell et al., 2003, Börgel et al., 2020] with energies on 4–10 year timescales [Meier et al., 2023, Visbeck et al., 2001]. The positive (negative) phase of the oscillation typically brings mild (cold), wet (dry) winters [Hurrell, 1995, Hurrell et al., 2003]. Skillful NAO predictions on decadal timescales used large ensembles [Patrizio et al., 2025, Smith et al., 2014, Athanasiadis et al., 2020, Strommen et al., 2023] and methods like variance matching [Smith et al., 2020]. The AMV characterizes low-frequent fluctuations of North Atlantic sea surface temperature, oscillating on 60–90 year timescales [Knight et al., 2006, Meier et al., 2023]. It influences Northern European precipitation [Börgel et al., 2022, 2018] and thereby Baltic Sea salinity [Meier et al., 2023], and furthermore holds an influence on Baltic Sea temperatures [Börgel et al., 2023, Barghorn et al., 2025]. Decadal predictions of the future AMV state are improving [Yeager, 2020, Buckley et al., 2019] with the use of tools like neural networks [Liu et al., 2023] or a linear inverse model [Huddart et al., 2016]. NAO and AMV influence each other [Klavans et al., 2019, Börgel et al., 2020, Smith et al., 2020] and together shape the variability of the Baltic Sea region temperature, precipitation and ice among others [Börgel et al., 2023, 2022].

The Baltic Sea possesses unique salinity dynamics as the result of inflowing saltwater from the North Sea mixed with precipitation and river runoff [Mohrholz, 2018, Lehmann et al., 2022]. These saline inflows occur primarily through large, irregular, so-called Major Baltic Inflows (MBIs) [Mohrholz, 2018, Lehmann et al., 2022], which generate horizontal and vertical salinity gradients as the dense saline water propagates along the seabed through the Arkona Basin and Bornholm Basin into the Gotland Sea and the Gulf of Finland [Liblik et al., 2018]. Baltic Sea salinity exhibits a pronounced multidecadal variability with a periodicity of around 30 years [Winsor et al., 2001, Meier and Kauker, 2003], caused by the AMV and NAO modulating the freshwater supply and amplified by a positive feedback mechanism between freshwater input and saline inflows [Meier et al., 2023]. More precisely, the AMV and the zonal displacement of the NAO centers of action force the low-frequent variability of atmospheric flow over Northern Europe and hence Baltic Sea catchment area precipitation variability on timescales corresponding to the squared AMV harmonic of about 30 years [Meier et al., 2023]. That precipitation variability dictates changes in the freshwater supply of the Baltic Sea on timescales matching the residency time of water masses in the Baltic Sea [Döös et al., 2004, Meier, 2007], yielding the multidecadal salinity variability of the Baltic Sea of about 30 years, which is further amplified by a positive feedback of the Baltic Sea freshwater supply influencing the salinity of saline inflows into the sea [Meier et al., 2023]. The variabilities of Baltic Sea temperature, especially SST [Kniebusch et al., 2019], ice cover [Omstedt and Hansson, 2006, Karpechko et al., 2015] and sea-level [Novotny et al., 2006, Karimi et al., 2022] are primarily atmospherically forced. A low-frequent variability component, as that of salinity, is not known for these properties. Baltic Sea memory capacities are governed by its water and heat balance. The water balance influences salinity and operates on timescales around 30y years [Döös et al., 2004, Meier, 2007, Omstedt and Hansson, 2006]. In contrast, the heat balance, which controls temperature and ice cover, displays e-folding timescales ranging between 4-7 months [Meier, 2006] and 1 year [Omstedt and Hansson, 2006].

Different approaches were used to evaluate Baltic Sea predictability timescales regarding average and surface temperature, sea-level, ice thickness and concentration as well as average and surface salinity. Thereby all calculations were performed for the entire Baltic Sea as well as for the Arkona Basin, Gotland Sea and Gulf of Bothnia.
At first, the sub-ensemble averages over all forcings were investigated. Information

about memory timescales of the different variables was obtained through the development of their respective initializations. Further, the influence of the external forcing, comprising river runoff, atmospheric conditions and open boundary conditions to the North Sea, was studied by comparing reference run and sub-ensemble average slices that start from the same initial field. The regard of the external forcing was continued with a subsampling technique [Dobrynin et al., 2018, 2022]. Predictions on S2D timescales were attempted through the estimation of future forcing and evaluated for their skill.
After an explanation of the methodology of the study the results are presented and discussed. Subsequently, the findings are summarized and an outlook on future research possibilities is given.

Chapter 2.

Methods

2.1. Model setup and ensemble construction

All model runs of this study were performed with the ocean model MOM6 coupled with the ice model SIS2 using a horizontal resolution of 8 nautical miles and 100 vertical layers. The model domain spanned the whole Baltic Sea including Kattegat and Skagerrak up to a longitude of 8.12 °E (see Fig. 2.1). Propagating in time, the model was forced by a set of boundary conditions. The atmospheric forcing, consisting of mean sea level pressure, rain fall amount, snow fall amount, 2m specific humidity, short- and longwave radiation, 2m temperature and U- and V-components of 10m wind, was based on coastDat-3 COSMO-CLM reconstruction data. Open boundary conditions at the domain border to the North Sea were implemented with a statistical model based on westerlies and the river runoff into the sea followed the general runoff data of the HELCOM PLC-6 assessment [Sonesten et al., 2018] comprising all significant rivers that flow into the model domain. Overall, the model setup was oriented at the work of Neumann et al. [2022]. All computations were performed at the high performance compute cluster HAUMEA of the University of Rostock.

At first, a reference run was calculated for the 72 year period 1948-2019 (see Appendix A.1 for an overview), initialized with a 50 year spin-up and forced by the boundary conditions explained above. Afterwards, a large ensemble was computed pairing each of 13 initial fields with each of 63 forcing time series (consisting of atmospheric forcing, open boundary conditions and river runoff), as shown in Fig. 2.2. For initial fields the reference run states on January 1st, every 5 years from 1951 to 2011, were used. The 63 forcing time series were obtained by cutting 10 year long slices out of the reference run forcing (which is 72 years long in total),

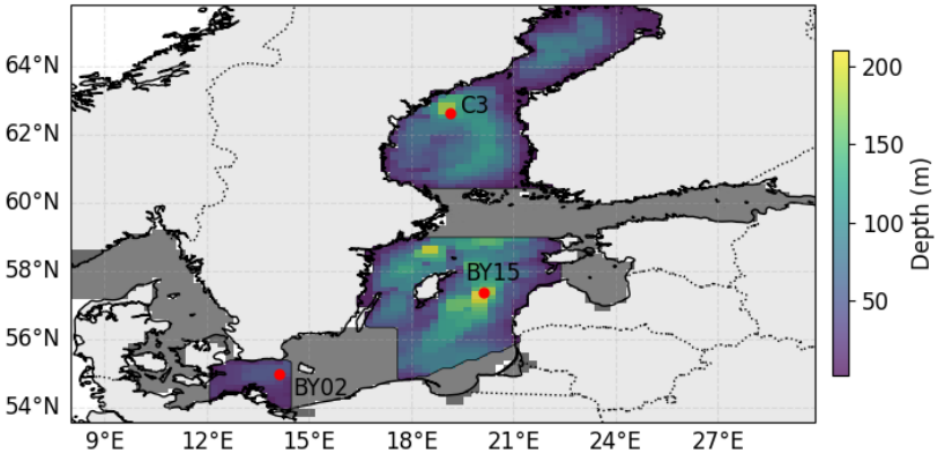


Figure 2.1: Analyzed basins of the Baltic Sea (colored areas) with vertical model output stations located at their deepest points: BY02 - Arkona Basin, BY15 - Gotland Sea, C3 - Gulf of Bothnia. The full model domain comprises the basins together with the dark grey area.

with the first time series starting 1948 and all following time series starting 1 year later than the previous one, until the last one starting in 2010. Accordingly, the whole ensemble consists of 13 sets of 63 members. These sets will be referred to as sub-ensembles. Each member spans 10 years, and the sub-ensembles start every 5 years, in accordance with the CMIP5 protocol regarding decadal predictions [Taylor et al., 2012]. The construction of the entire ensemble provides a trade off between accessible computational resources and temporal sampling.

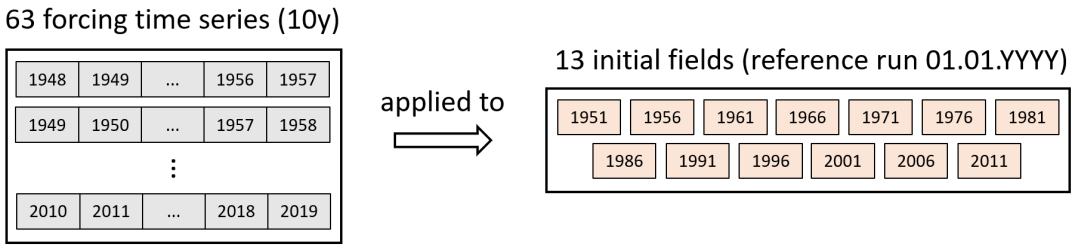


Figure 2.2: Sketch of the construction of the ensemble model run. All 13 initial fields were combined with each of the 63 forcing time series.

2.2. Postprocessing

2.2.1. Analyzed variables

This section aims to give an overview on the model output configuration and the resulting data of the analyzed variables: temperature and salinity of the whole water body and of the sea surface, sea-level, ice thickness and ice concentration.

Not only the Baltic Sea (BS) as a whole was investigated, but also the Arkona Basin (AB), the Gotland Sea (GS), and the Gulf of Bothnia (GB). Sea surface salinity (SSS), sea surface temperature (SST), sea surface height (SSH), ice thickness (Ice HI) and ice concentration (Ice CN) were computed by the model using their respective MOM6 and SIS2 diagnostics and afterwards averaged over the surface of the according areas (entire model domain for BS or single basin surface), with consideration of the grid cell surface areas. The Ice CN output distinguished between different category bounds, which were all summed up. The description of average temperature and salinity differs between the entire BS and its sub-basins. Basin-specific values were obtained by multiplying temperature and salinity of the basins deepest water column (at the station BY02 of the AB, BY15 of the GS and C3 of the GB, as shown in Fig. 2.1) with the basins respective hypsographic function. The values for the whole BS were taken from the control file ocean.stats.nc, which contained heat content and salt mass of each grid cell integrated over the whole model domain (diagnostics: Heat, Salt). Since that domain also includes the Kattegat and Skagerrak, the salt and heat of these two areas were included in the diagnostics of the Baltic Sea, but not of the single basins. In order to maintain consistency in the results regarding the Baltic Sea, surface and ice values were averaged over the entire model domain, involving Kattegat and Skagerrak as well. Total heat content and salt mass of the BS are to a good degree directly proportional to average temperature and average salinity, so these terms were used synonymously. All variables have a temporal resolution of monthly values.
As a result, the whole data has the following five degrees of freedom:

- 1) Variable: SST, SSS, SSH, average temperature, average salinity, ice concentration, ice thickness
- 2) Basin: BS, AB, GS, GB
- 3) Initialization year: 1951, 1956, 1961, 1966, 1971, 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2011
- 4) Forcing time interval: 10 year interval starting each year from 1948 to 2010
- 5) Lead time: Months 0–120

2.2.2. Prediction generation through subsampling

For the subsampling analysis in this study many subsamples were created out of the large ensemble, by applying a set of different predictors to all of the 13

sub-ensemble initializations. Each sub-ensemble initial field was classified into a threshold (strong/moderate positive/negative or neutral) in regard to each predictor. Members of the same sub-ensemble with a forcing that matches the initial field's threshold regarding a specific predictor were grouped into a subsample. Consequently each subsample is classified by its initial field and the applied predictor. The subsample averages were treated as predictions. Table 2.1 gives an overview of all used predictors. The thresholds were defined based on the standard deviation of the time series, as shown in Fig. 2.3.
As an example of the subsampling, the initial field of 1980, like all others, was forced by the 63 different forcing time series, thus creating 63 members starting from the same 1980 initial field. That initial field of 1980 was evaluated for all predictors and assigned a threshold for each predictor. For example the future 3y AMV index (average AMV index over the next three years) regarded from 1980 is 'moderate negative', as it is for the years 1963–1969, 1978–1981, 1986, 1989–1990, 1993 (see Fig. 2.3). Now, out of the 63 members of the 1980 initial field, all members created by the forcing starting at these years were added to the subsample "future 3y AMV moderate negative" for the initial field 1980. The mean of that subsample was interpreted as prediction starting 1980 based on knowledge of the future 3y AMV index.

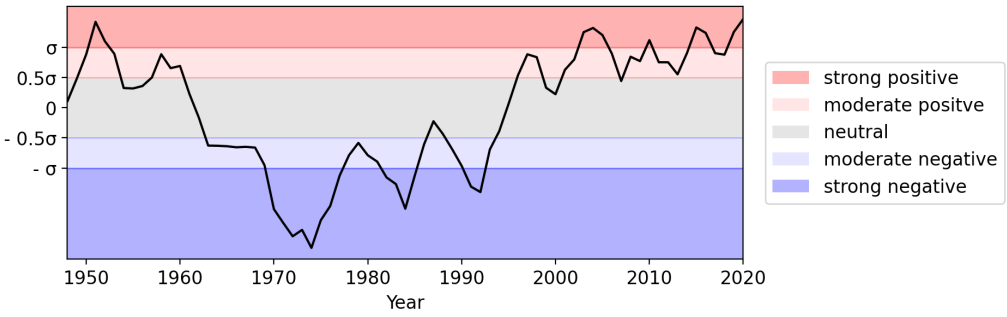


Figure 2.3: Sketch of the general threshold classification, applied to the future 3y AMV index as an example. The data is grouped into five windows based on the time series standard deviation.

Predictors

The used AMV index was provided by the National Oceanic and Atmospheric Administration (NOAA). It is derived from North Atlantic SST anomaly data [Kaplan et al., 1998] through averaging and detrending, spanning from 1938 until 2019. Runoff and precipitation data was extracted directly from the model forcing. Past

BS salinity values were taken from the reference run. All predictor time series $P(t)$ were calculated separately through the application of various running means to the above explained metrics. Subsequently they were standardized according to:

$$P_{\text{standardized}}(t) = \frac{P(t) - P_{\text{mean}}}{P_{\text{std}}} \tag{2.1}$$

$P_{\text{standardized}}(t)$ – standardized predictor time series; P_{mean} – mean of $P(t)$; P_{std} – standard deviation of $P(t)$. As a result, past and future AMV index, past and future runoff plus precipitation, past salinity of the BS, averaged over multiple time spans and standardized, as well as the initial Baltic Sea SST, temperature, SSH, SSS and salinity were obtained as predictors. Tab. 2.1 gives a full overview.

Predictor	**Time Direction**	**Time Span**
Baltic Sea average salinity	Past	5y 10y 15y
AMV index	Past	3y 5y 10y
Sum of runoff and precipitation	Past	3y 5y 10y
AMV index	Future	3y 5y 10y
Sum of runoff and precipitation	Future	3y 5y 10y
Initial Baltic Sea SST, SSH, SSS, average temperature, average salinity	-	-

Table 2.1.: List of all the predictors that are used for subsampling with their corresponding time direction and time span.

Quantification of prediction skill

Subsampling prediction skill was analyzed for monthly resolved predictions (MRP) and annually resolved predictions (ARP). For the MRP, effects of the seasonal cycle of the Baltic Sea were removed by subtracting the respective climatology from the prediction and the according 10 year reference run slice. Afterwards, for MRP and ARP and only if SST or average temperature were predicted, prediction and reference run slice were additionally detrended by subtracting their respective linear trend. Further reasoning of the methodology is given in Appendix B.1.
Prediction skill in regard to the reference run was quantified through the Anomaly Correlation Coefficient (ACC) and the Mean Square Skill Score (MSSS). Each MRP/ARP spans ten years of monthly/annual values. For MRP the skill metrics were calculated for the timescales first year, first 5 years, years 6–10 and whole 10 years. Skill metrics of the ARP were calculated for the first and last 5, 6, 7, 8 and 9 years of the prediction and for the whole 10 year prediction.
The ACC, calculated as Pearson correlation coefficient, describes how well the predictions phase aligns with the phase of the reference run. Its values span from -1, which corresponds to a perfect inverse correlation, to 1, corresponding to perfect agreement in phase.

$$\mathrm{ACC} = \frac{\Sigma_i \left[(r_i - \bar{r}) \, (p_i - \bar{p}) \right]}{\sqrt{\Sigma_i \left(r_i - \bar{r} \right)^2 \; \Sigma_i \left(p_i - \bar{p} \right)^2}} \tag{2.2}$$

r - reference run, p - prediction, $\bar{r}$ & $\bar{p}$ - mean over the prediction time span.
In addition to the ACC, the MSSS is a common metric to assess the closeness of predicted values to a reference [Ho et al., 2013, Smith et al., 2020, 2013]. While that reference can be measurement or reconstruction data, in this study it is given by the reference run.

$$\mathrm{MSSS} = 1 - \frac{\Sigma_i \left(p_i - r_i \right)^2}{\Sigma_i \left(r_i - \bar{r} \right)^2} \tag{2.3}$$

The MSSS is calculated based on the mean square error between prediction and reference run scaled by the variance of the reference [Wilks, 2011]. The maximum value of 1 indicates a perfect prediction matching the reference run at all times. A minimum value does not exist. Both metrics are sensitive to outliers motivating the investigation of smaller time windows.

Blocked Bootstrap significance testing

This study uses a blocked bootstrapping test to quantify the statistical significance of the determined MSSS or ACC between a subsampled prediction and the reference run. The method allows to account for the autocorrelation of the analyzed variables through an adjustable blocksize [Wilks, 2011]. The following scheme describes the calculation of the p-value of each calculated MSSS/ACC for a given pair of prediction and reference run.

- Divide both time series equally into blocks of the same length/size, depending on the analyzed time window and the autocorrelation of the analyzed variable.
- Identically resample both time series with replacement.
- Compute MSSS/ACC of the two resampled versions.
- Repeat the last two steps 999 times.
- Examine the resulting distribution and perform a one-sided hypothesis test for a positive MSSS/ACC with the null hypothesis H_0: of $\mathrm{MSSS} = 0$ or $\mathrm{ACC} = 0$ respectively.

The block size is required to preserve the autocorrelation signal while avoiding excessive averaging of the internal variability [Kunsch, 1989, Wilks, 2011]. This was achieved by using blocksizes similar to the autocorrelation length of the analyzed variable, and, in the case of excessive autocorrelation lengths, the blocksize which maximizes the standard error of the bootstrapped distribution was utilized [Hall et al., 1995, Politis and White, 2004, Mignani and Rosa, 1995]. The full blocksize computation process is explained in Appendix B.2.

Chapter 3.

Results and Discussion

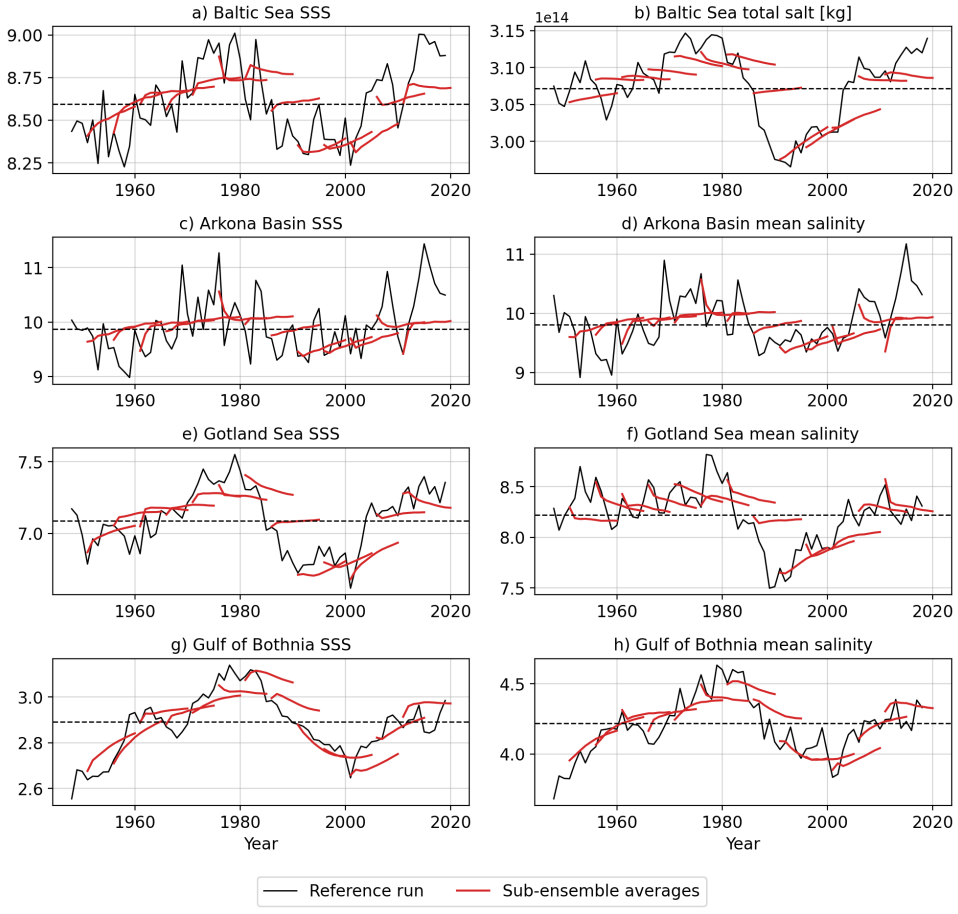


Figure 3.1: Temporal salinity dynamics of the reference run and the sub-ensembles. Solid black line: Annually averaged reference run values. Solid red line: Sub-ensembles (characterized by their initial states) averaged over all forcings and annually in time. Dashed black line: Average value of the whole ensemble.

Reference run SSS and average salinity of the Baltic Sea (BS) and its subbasins exhibit low-frequent variability on multidecadal timescales, as shown in Fig. 3.1a-h.

The variability is most pronounced in the Gulf of Bothnia (GB), followed by the Gotland Sea (GS), and weakest in the Arkona Basin (AB). This pattern is consistent with the two dominant processes governing salinity across the Baltic Sea: inflowing saltwater from the North Sea and low-frequency variability in freshwater input, on around 30 year timescales, amplified by a positive feedback influencing saline inflows. These saline inflows don't reach the GB directly, meaning its salinity dynamics primarily follow the freshwater input variability. Major Baltic Inflows (MBIs) propagate along the seafloor and can reach the GS, whose salinity is primarily influenced by MBIs and freshwater input. This is reflected in the salinity decline during the period from 1980-1990, a known stagnation period. The AB is the first basin of the Baltic Sea affected by saline inflows, including smaller events. While the low-frequency salinity variability is still visible there, it is superimposed by higher-frequent fluctuations caused by both large and small inflows.
Averaging the computed ensemble over all forcings (atmospheric forcing, open boundary condition to the North Sea, river runoff) yields 13 sub-ensemble averages starting every five years (1951, 1956, ..., 2011) from January 1st of the reference run, that are shown in Fig. 3.1 for surface and average salinity. During year one the sub-ensemble salinity adapts to the different forcings that mostly don't match the initialization, followed by a decay to the mean for the rest of the 10 year window. The latter trend confirms that the forcing is effectively averaged out to a good degree and doesn't drive the sub-ensemble averages. Generally the decay to the mean salinity is happening slowly in the BS, GS and GB, illustrating their salinity memory. At the same time salinity initializations of the AB clearly dissipate faster the initializations of the other basins which indicates a smaller memory. The low-frequent salinity variability of the entire BS is slightly blurred in the model results because Kattegat and Skagerrak, which don't feature that variability, are included as well adding more noise on top of the signal. A figure similar to Fig. 3.1, that additionally contains the sub-ensemble initial values can be found in the appendix under Fig. A.4.

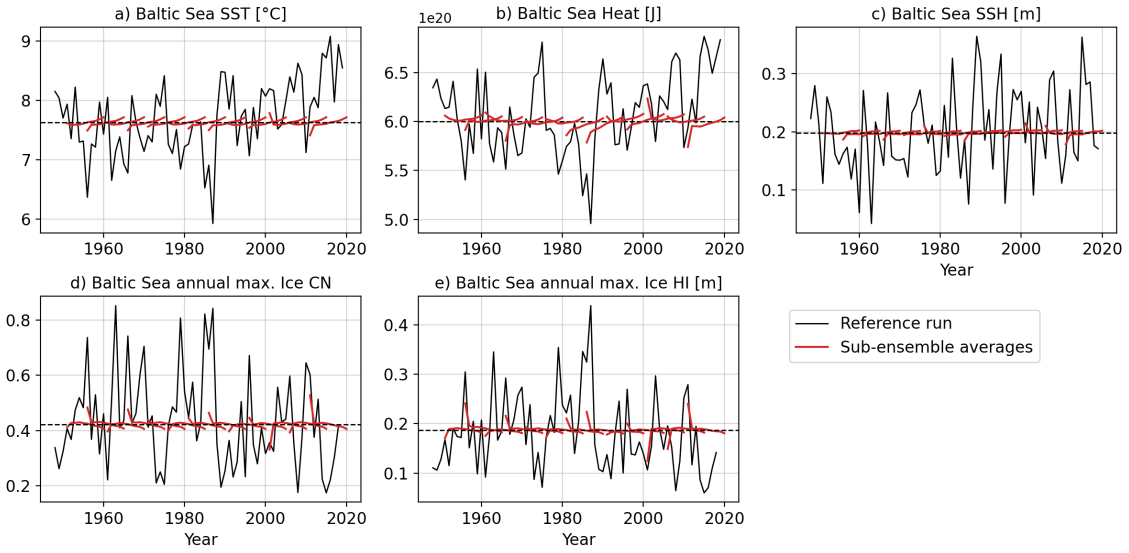


Figure 3.2: Temporal Baltic Sea temperature, sea-level and ice dynamics of the reference run and the sub-ensembles. Solid black line: Annually averaged reference run values. Solid red line: Sub-ensembles (characterized by their initial states) averaged over all forcings and annually in time. Dashed black line: Average value of the whole ensemble.

In contrast to salinity, the reference run sea surface and average temperature, sea-level as well as ice concentration and thickness are governed by higher-frequent climate variability and long-term trends caused by global climate change: warming of the Baltic Sea causing less sea ice, and sea-level rise, as shown in Fig. 3.2a-e for the whole Baltic Sea. The sub-ensemble averages of these variables decay to the mean considerably quicker than those of salinity indicating a smaller memory (see Fig. 3.2a-e). Nonetheless, that decay is not instantaneous and the initial temperature signal can be found in the year one average of SST, average temperature and maximum ice concentration and thickness. Especially initializations of very low (1956,1966,1986,2011)/high (1961,1976,2001) temperatures yield lower/higher year one sea surface and average temperatures and a higher/lower maximum ice concentration and thickness. This provides evidence for a distinct temperature memory of the Baltic Sea on timescales of several months starting January 1st. From there the sub-ensemble averages follow the general trends of the reference run, as they are forced by the average reference run forcing. Baltic Sea SST (average reference run trend per decade: $+0.15°\mathrm{C}$), heat content ($+4.7 \cdot 10^{18}\mathrm{J}$), and SSH ($+0.005\mathrm{m}$) increase while its annual maximum ice concentration (-0.0145) and annual maximum ice thickness ($-0.0056\mathrm{m}$) decrease. The dynamics of the AB, GS and GB look similar to the BS dynamics displayed in Fig. 3.2. A similar figure, that additionally contains the sub-ensemble initial values can be found in

the appendix under Fig.A.5.

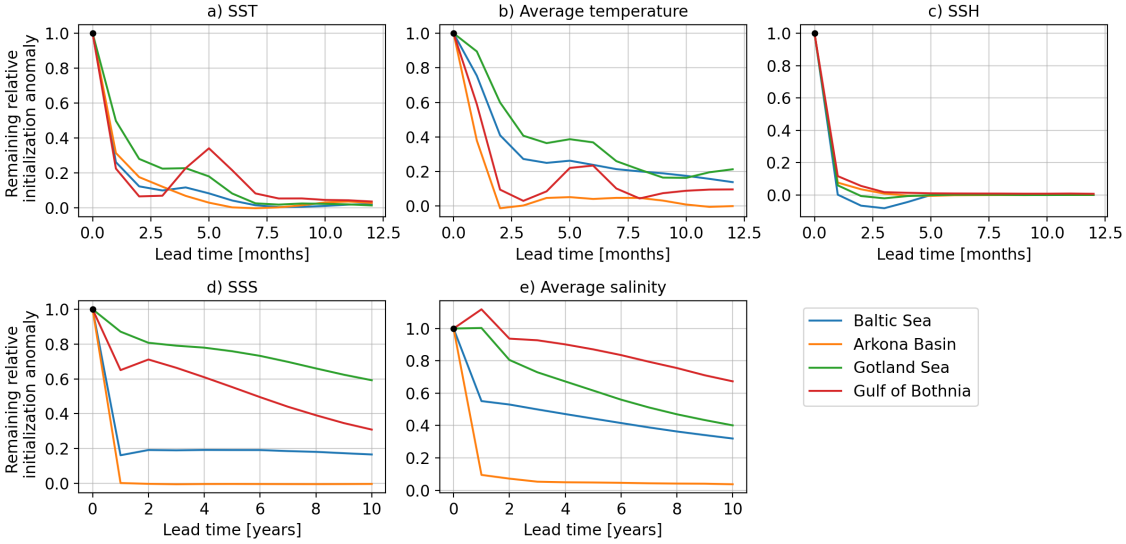


Figure 3.3: Development of SST, mean temperature, SSH, SSS and mean salinity initialization anomalies averaged over all sub-ensembles. Anomalies were computed in regard to the ensemble average over all initializations and forcings. Only initialization anomalies larger/smaller than $\pm$ 0.25 times the initialization standard deviation were considered. Note, that the upper row displays the first year of model calculations, while the lower row shows the whole 10 years.

As a different method to study the memory of the Baltic Sea, the temporal development of the SST, average temperature, SSH, SSS and average salinity initializations (always on January 1st) anomalies was analyzed (see Fig. 3.3). To calculate the anomalies, the average 10-year time series of the whole ensemble was subtracted from each sub-ensemble average. This method also effectively removes the climatology and isolates the initialization development. Only sub-ensemble initialization anomalies larger/smaller than $\pm$ 0.25 times the standard deviation of all initializations were considered. Otherwise small absolute changes of the smaller initializations causing large relative changes, potentially even with a change of sign, overshadow the underlying signal.
SST initialization anomalies decay to zero faster than average temperature initialization anomalies across all basins. This behavior is physically consistent, as a whole water body has a higher temperature memory than only its surface. GS temperature initialization anomalies remain highest over the first 12 months, because it is the largest sub-basin and consequently possesses the largest temperature memory, and because the BS values are influenced by shallower basins with less memory. Around 20% of GS and BS initialization anomalies remain after the first 12 months, implying a distinct temperature memory. When the sea ice starts melting

during the spring months cold meltwater is added to the sea water, decreasing its temperature. This effect leads to an artificial increase of the remaining surface and average temperature initialization anomaly of the GB and also to a smaller degree of the GS, as seen in Fig. 3.3a,b. SSH initializations decay close to the mean already during the first month across all basins indicating that Baltic Sea SSH memory timescales are shorter than one month (see Fig. 3.3c). The salinity initialization anomalies, of the sea surface and averaged, decay much slower compared to the other variables, apart from the AB for which no SSS signal and only a small mean salinity signal remain in the year one average values. In contrast to the AB, the relative remaining initialization anomalies of sea surface and mean salinity of the GS and the GB never fall below 30% over the whole ten year model period. Of the initialization anomalies 59% of GS SSS and 40% of GS mean salinity, as well as 31% of GB SSS and 67% of GB mean salinity anomalies remain in year ten, compared to a remainder of 17% of the BS SSS and 32% of the BS mean salinity initialization anomaly. Values for the BS without Kattegat and Skagerrak included in the data would likely be higher as these two areas are well connected with the open North Sea so that their initialization anomaly decreases quicker than that of the actual Baltic Sea. Generally, the quantitative results have to be treated with caution, because the $\pm$ 0.25 standard deviation threshold is rather arbitrarily and using higher, lower or no restriction at all influences the outcome. Nonetheless qualitative conclusions can be drawn.

SSH memory was not detectable for the BS or any sub-basin on monthly resolution, which is in agreement with the sea-level being primarily forced by surface winds and anomalies evening out quickly without external forcing. Basin-averaged temperature is memorized longer than SST on timescales of several months up to a year for the Baltic Sea and Gotland Sea. This pattern can be explained by relating the temperature memory of a water body to its volume and thus the total heat it contains. Of the analyzed basins the AB has the lowest salinity memory, in accordance to it being the smallest one and most influenced by saline inflows from the North Sea. BS, GS and GB display salinity memory on interannual to decadal timescales, as the initialization anomalies of surface and average salinity remain distinct over the whole model period.

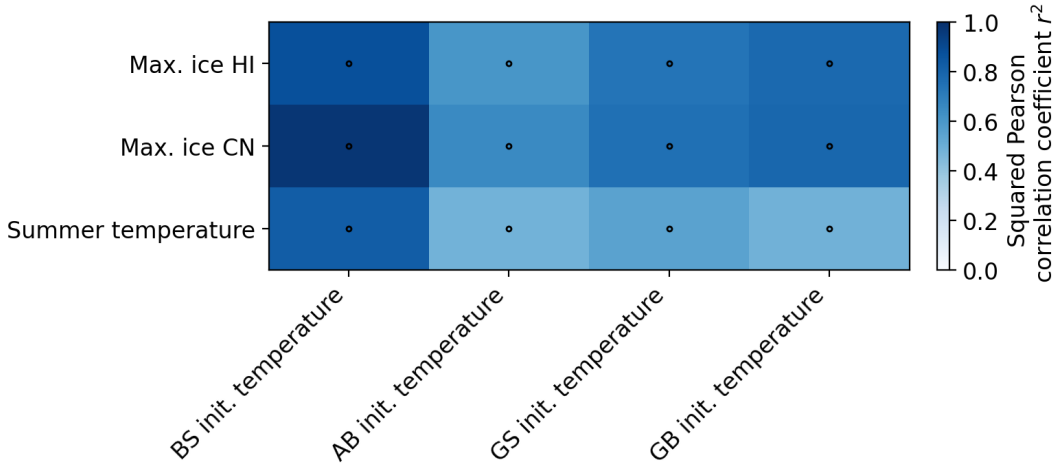


Figure 3.4: Pearson correlations between the initial temperature of sub-ensemble averages and their year one winter ice and summer temperature for each basin. Winter ice is quantified as maximum ice thickness (ice HI) and maximum ice concentration (ice CN). Summer temperature is calculated as JJA mean temperature. Correlations are shown for the whole Baltic Sea (BS), Arkona Basin (AB), Gotland Sea (GS) and Gulf of Bothnia (GB). A panel is dotted if p<0.05.

The temperature memory of the Baltic Sea and its basins was further assessed by studying the sub-ensembles averaged over all forcings. Correlating the initial temperature of a sub-ensemble average with its year one maximum ice thickness, year one maximal ice concentration, as well as year one summer (JJA mean) temperature allowed an investigation of the development of the temperature initialization from a different framework. Since this studies ensemble comprises 13 sub-ensembles, the sample size for each computed correlation was 2x13.
As already seen in Fig. 3.3a,b, temperature initializations generally decay to the mean over time. Nonetheless, this process does not happen instantaneously and across all basins a correlation signal was found between initial temperature and year one maximum ice thickness and concentration (between lead months 2-4) of the sub-ensemble averages, as well as between initial temperature and summer mean temperature (average of lead months 6-8) (see Fig. 3.4) of the sub-ensemble averages. The correlations regarding ice cover are generally higher than those for summer temperature, likely because the initial signal has decayed less until months 2-4 than until months 6-8. Furthermore the BS correlations are highest among the analyzed basins with $\mathrm{r}^2 = 0.83\ (\mathrm{p} < 0.001)$ for summer temperature, $\mathrm{r}^2 = 0.97\ (\mathrm{p} < 0.001)$ for ice concentration and $\mathrm{r}^2 = 0.88\ (\mathrm{p} < 0.001)$ for ice thickness, followed by GB ice concentration $\mathrm{r}^2 = 0.79\ (\mathrm{p} < 0.001)$ and ice thickness $\mathrm{r}^2 = 0.78\ (\mathrm{p} < 0.001)$ as well as GS ice concentration $\mathrm{r}^2 = 0.75\ (\mathrm{p} < 0.001)$ and ice thickness $\mathrm{r}^2 = 0.73\ (\mathrm{p} < 0.001)$.

These significant correlations between initial temperature and year one winter ice and summer temperature were only achieved because the sub-ensembles were averaged over their forcings. If this studies ensemble is large enough, which it seems to be to a good degree, the high-frequent external forcing on the temperature is averaged out in the sub-ensemble averages, decreasing the perturbation of the initialization signal, hence resulting in a decay to the mean to a good degree. After a few months of computation the amplitude of the remaining initialization anomaly of the sub-ensemble averages has decreased, but a large part of the phase remains, which is what the Pearson correlation quantifies. This enables the high correlation coefficients. Figure 3.4 demonstrates, that the signal of January 1st temperature is memorized across all analyzed basins at least until the summer months, with a stronger memory for larger basins. The degree to which the atmospheric forcing influences the detectability of that signal is analyzed below.

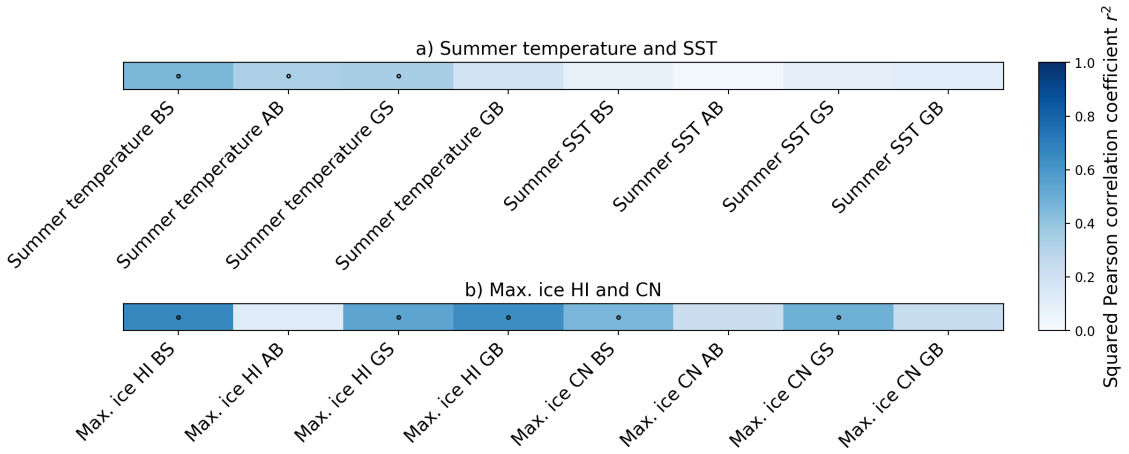


Figure 3.5: Pearson correlations between reference run and sub-ensemble average year one winter ice as well as summer temperature. Reference run slices and sub-ensemble-averages start from the same initialization. Winter ice is quantified as maximum ice thickness (ice HI) and maximum ice concentration (ice CN). Summer temperature and SST are calculated as JJA mean. Correlations are shown for the whole Baltic Sea (BS), Arkona Basin (AB), Gotland Sea (GS) and Gulf of Bothnia (GB). A panel is dotted if $p<0.05$.

In order to investigate the actual predictability of winter ice and summer temperatures based on temperature memory of the Baltic Sea, the model forcing has to be taken into account. Therefore the year one winter ice as well as the year one summer temperatures of the sub-ensemble averages were correlated with the according reference run values (see Fig. 3.5). Once again all Pearson correlations base on samples with 13 values each.
Statistically significant correlations between the basin-averaged year one summer

(JJA mean) temperatures of sub-ensemble averages and reference run were found for the BS ($r^2 = 0.46,\ p = 0.01$), AB ($r^2 = 0.34,\ p = 0.04$) and GS ($r^2 = 0.35,\ p = 0.03$), but not for the GB, and also not at all if the temperature of the sea surface was analyzed instead of basin-averaged. Calculated the same way, $r^2 > 0.45$ correlations between sub-ensemble averages and reference run were detected for the year one maximum ice thickness of the BS ($r^2 = 0.66,\ p < 0.01$), GS ($r^2 = 0.54, p < 0.01$) and GB ($r^2 = 0.64, p < 0.01$), as well as for the ice concentration of the BS ($r^2 = 0.46, p = 0.01$) and GS ($r^2 = 0.49, p < 0.01$). All correlations are shown in Fig. 3.5.
The sub-ensemble year one averages are unforced to a good degree. Hence, this method compares the forced and unforced development of temperature initializations. The appearance of correlations can be attributed to the temperature memory of the Baltic Sea, while the height of the correlations is dictated by the reference run forcing, mainly the surface air temperature perturbing the initial temperature signal via sensible and latent heat flux. These heat exchange processes between atmosphere and sea surface influence the SST strongest, masking any predictable summer SST signal after an initialization on January 1st. The correlations for BS and GS year one summer temperature can be explained by the temperature memory of the two large water bodies that sustains parts of the initial temperature signal through the influence of the experienced atmospheric forcing. However this doesn't explain the correlation found for the year one summer temperature of the very small AB. The generally higher correlations for year one maximum winter ice than for summer temperatures are likely caused by the shorter time span of forcing influence (2-4 vs. 6-8 months). AB ice cover happens only during very cold winters, thus it depends on atmospheric forcing. Consequently no correlations between forced and unforced initialization developments were found for AB ice thickness or concentration. Higher correlations for ice thickness than for concentration are likely due to different sensitivities of the two variables to atmospheric forcing, surface air temperature in particular. Warm surface air can melt larger areas of thin ice quickly, while the thicker ice loses only some of its thickness, causing larger changes in ice concentration than thickness (similar for freezing due to cold surface air). Furthermore, the GB is the northernmost part of the Baltic Sea and most likely to develop an ice cover. During cold winters the GB surface can freeze completely. Because ice concentration is capped at 1, the variable cannot distinguish between differently cold winters if the GB freezes fully. As a result there was no correlation regarding GB ice concentration found between sub-ensemble averages and the reference run.

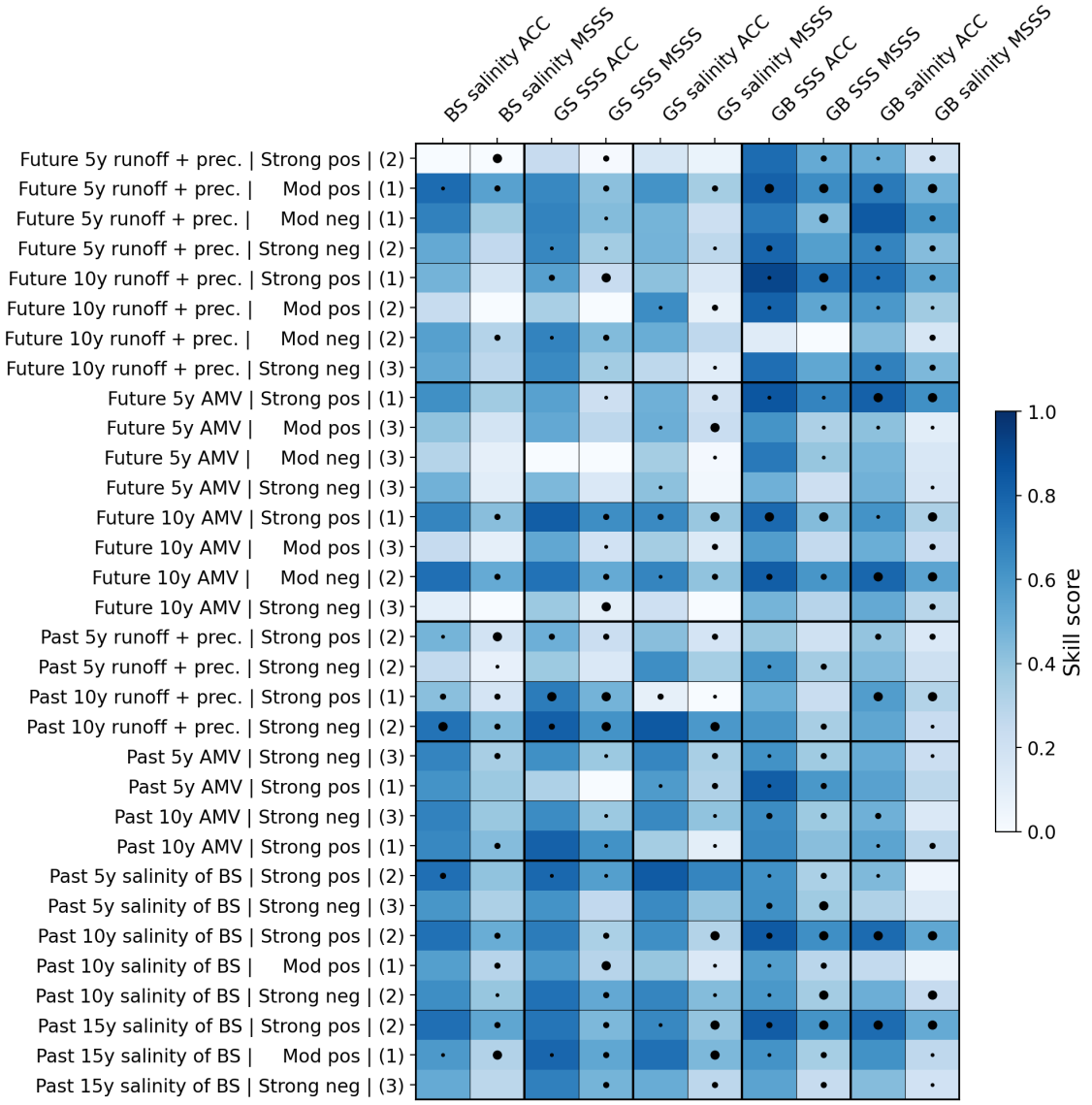


Figure 3.6: Anomaly correlation coefficient (ACC) and mean square skill score (MSSS) for monthly resolved subsampling predictions of ten years length. Baltic Sea (BS), Gotland Sea (GS) and Gulf of Bothnia (GB) salinity as well as GS SSS and GB SSS were predicted utilizing the sum of past or future runoff plus precipitation (runoff + prec.), the past or future AMV index (AMV) and past BS salinity as predictors. Therefore runoff plus precipitation and AMV index were each averaged over the last 5 or 10 years prior to the prediction (Past 5y/10y ...) or over the first 5 or 10 years of the prediction (Future 5y/10y ...) and past BS salinity was averaged over the last 5, 10, or 15 years (Past 5/10/15y salinity of BS). All predictions base on the categorization of the predictors into the thresholds strong or moderate positive or negative (Strong/Mod pos/neg), as explained in 2.2.2. Numbers in parentheses behind the thresholds indicate the number of initial fields of the ensemble matching that threshold. Each initial field yields its own prediction and the shown skill scores are averaged per threshold. The results are dotted based on their significance level: large dot for $p<0.05$, medium dot for $0.05<p<0.10$ and small dot for $0.10<p<0.15$.

In an attempt to generate predictions in the framework of this study and thereby to further investigate predictability, a subsampling method was applied to the ensemble data. A detailed description of the methodology is given in section 2.2.2. Basically the members of a sub-ensemble were grouped into subsamples based on different predictors. The subsample average over all members was then treated as a prediction. These predictions were evaluated for their skill in comparison to the reference run through the anomaly correlation coefficient (ACC) and the mean square skill score (MSSS).

The analysis incorporated annually resolved predictions (ARP) as well as monthly resolved predictions (MRP). Resulting skill scores of the ARP were not statistically significant on the basis of an applied blocked bootstrap significance testing (see section 2.2.2 for the methodology). Also no significant skill was found for predictions of the first year, the first five years, or years six to ten. As a consequence all shown and discussed results base on MRP on prediction time scales of ten years only. Further, no significant skill was found for predictions of the AB in general, as well as for predictions of SST, mean temperature, SSH, ice thickness, ice concentration of the BS, GS and GB. Nonetheless skillful MRP of BS mean salinity and GS and GB SSS and mean salinity over a ten year time frame were achieved with different predictors, as shown in Fig. 3.6. These predictors comprise the past salinity of the Baltic Sea, the AMV index and the sum of river runoff and precipitation into the Baltic Sea. All three of them were regarded as average values over different time windows relative to the initialization time of the model runs. Past salinity was averaged over the last 5, 10 and 15 years before initialization, while AMV index and runoff plus precipitation were averaged over the last 5 and 10 years prior to and and the first 5 and 10 years after initialization. Found prediction skill of BS salinity based on strong negative past 10 year runoff plus precipitation amounted to $\mathrm{ACC} = 0.74$ ($\mathrm{p} = 0.03$), $\mathrm{MSSS} = 0.44$ ($\mathrm{p} = 0.06$). The same predictor yielded $\mathrm{ACC} = 0.82$ ($\mathrm{p} = 0.08$), $\mathrm{MSSS} = 0.61$ ($\mathrm{p} = 0.04$) for predictions of GS SSS. Subsampling for strong positive past 15 year Baltic Sea salinity resulted in GB salinity predictions with $\mathrm{ACC} = 0.77$ ($\mathrm{p} = 0.04$), $\mathrm{MSSS} = 0.52$ ($\mathrm{p} = 0.01$).

All the significant results (shown in Fig, 3.6) are average values for 1-3 (out of 13 total) initializations fitting the same predictor threshold, i.e. strong negative. Thus the sampling error of this subsampling method is very high due to the limited ensemble size. Computing the ensemble with more initial fields would increase the number of initializations per predictor threshold enabling more robust results, that might be significant on shorter prediction time scales. Nonetheless this studies ensemble size allows for qualitative takeaways. The subsampling approach achieves

prediction skill, in the form of ACC or MSSS, by predicting the future forcing of the Baltic Sea, composed of open boundary conditions, atmospheric forcing and river runoff in this study. Variability of surface and average temperature as well as of ice concentration and thickness are primarily forced by variations in SAT which were not predictable on interannual to decadal timescales by any of this studies predictors. Hence no significant prediction skill was found for these variables on a multi-annual timescale. The same rationale applies for the sea-level, which is mainly forced by wind fields. Prediction skill for salinity, of the surface and depth-averaged, is accomplished by gaining information about the current phase of the Baltic Sea in its low-frequent salinity variability. The predictor past salt of the Baltic Sea achieves just that. Since the Baltic Sea salinity variability is caused by freshwater forcing, which in turn is partly modulated by the AMV, runoff plus precipitation and AMV index are also suitable predictors for future salinity. Predictability of AB salinity is generally low because the basins salinity is perturbed more frequently by saline inflows resulting in a less pronounced multidecadal salinity variability compared to the BS, GS and GB, as shown in Fig. 3.1. If BS salinity values were not incorporating Kattegat and Skagerrak salt that clouds the low-frequent salinity variability signal, higher skill scores could have possibly been achieved in BS salinity predictions and predictions utilizing past BS salinity as predictor. Besides the already mentioned predictors, the initial fields of the sub-ensembles were also used to generate subsamples. But due to their sensitivity to higher-frequent natural variability they were less suited as predictors than the multi-year averages of past Baltic Sea salinity, AMV index and runoff plus precipitation. Generally the prediction skill of the latter three predictors was highest if they were averaged over at least a 10 year time span, where large parts of higher-frequent variability are averaged out, so that the signal of the low-frequency salinity variability is unmasked further. For the same reason the longest prediction time scale, ten years, was the one providing significant results. Furthermore strong positive or negative predictor thresholds generally allowed for most of the significant results and the highest skill scores as they provide the clearest phase information regarding the predictors variability and in turn the Baltic Sea salinity variability.

Chapter 4.

Conclusion and Outlook

This studies goal was to obtain an overview of the Baltic Sea's potential for seasonal to decadal predictability. Therefor an ensemble was computed pairing 13 initial fields with 63 forcing time series, each run lasting ten years (see Fig. 2.2). Two different approaches were used in the analysis. First, temperature, sea-level, and salinity memory timescales were investigated through the development of the initial fields. Second, predictions based on the future forcing were attempted by using a subsampling method.

As expected, a pronounced salinity memory on a decadal scale was found for the entire Baltic Sea (BS), the Gotland Sea (GS) and the Gulf of Bothnia (GB). The Arkona Basin (AB) exhibited the smallest salinity memory spanning only seasonal timescales. Generally, the memory regarding depth-averaged salinity was higher than that of surface salinity (see Fig. 3.3 d,e). Same applies for temperature where sea surface temperature initializations decay to the mean in only a few months, while the relative initialization anomalies of BS and GS depth-averaged temperature remain at around 20% after lead year one (see Fig. 3.3 a,b). The temperature memory was further reflected in significant correlations between the initial temperature and year one winter ice and summer temperature of sub-ensemble averages (see Fig. 3.4). As each of the sub-ensembles of this study incorporates 63 different forcing time series, the mean over all sub-ensemble members averages out the influence of external forcing, especially of surface air temperature, to a good degree and the sub-ensemble averages reflect the unforced development of the initializations. Consequently, the signal of initial temperature does not fully vanish in year one and is mainly masked by the external forcing influence. That external forcing was accounted for by correlating forced reference run and quasi-unforced sub-ensemble averages regarding their winter ice and summer temperature (see Fig. 3.5). Significant correlations with $\mathrm{r}^2 > 0.45$ were found for BS, GS and GB

winter ice and BS summer temperature, indicating a distinct detectability of the initial temperature signal at least throughout the first eight lead months of external forcing. All these findings imply the predictability of heat-related variables on seasonal timescales.
The conducted subsampling approach yielded skillful results through the prediction of future forcing on the Baltic Sea. This was achieved for salinity predictions (see Fig. 3.6), where the phase of the low-frequent salinity variability of the BS, GS and GB was estimated utilizing multi-year averages of past BS salinity as well as of past or future AMV index or runoff plus precipitation. The latter two predictors are well suited because they describe the freshwater forcing on the Baltic Sea causing its low-frequent salinity variability.

To sum up, predictability of Baltic Sea temperature and ice cover was found on seasonal timescales, based on the seas temperature memory. Further understanding of that memory, possibly in combination with the leveraging of the NAO to predict future winter surface air temperature variability, could possibly enable interannual temperature and ice cover prediction skill. No evidence for sea-level predictability on seasonal or longer timescales was detected. Salinity of the BS, GS and GB was identified as the best predictand among the analyzed variables displaying decadal predictability potential. Skillful decadal salinity predictions are accomplishable because of a high salinity memory of the Baltic Sea paired with the predictability of freshwater forcing causing the seas distinct low-frequent salinity variability.

Future improvements of this studies findings could be achieved by computing a larger ensemble that incorporates more initial fields. Thereby the development of the initial anomalies, especially of larger ones, could be analyzed more thoroughly, and the subsampling could yield more meaningful results if it uses more than 1-3 initializations per predictor threshold. Also the predictor runoff plus precipitation could be improved to describe the actual freshwater input into the sea, by additionally accounting for evaporation. Furthermore, model runs starting at times other than January 1st are needed to increase the understanding of the seasonal dependency of the Baltic Sea's temperature memory. Lastly, excluding the Kattegat and Skagerrak from the Baltic Sea domain would raise the accuracy of the results regarding the entire Baltic Sea, especially its salinity.

Appendix A.

Additional Plots

A.1. Reference run overview

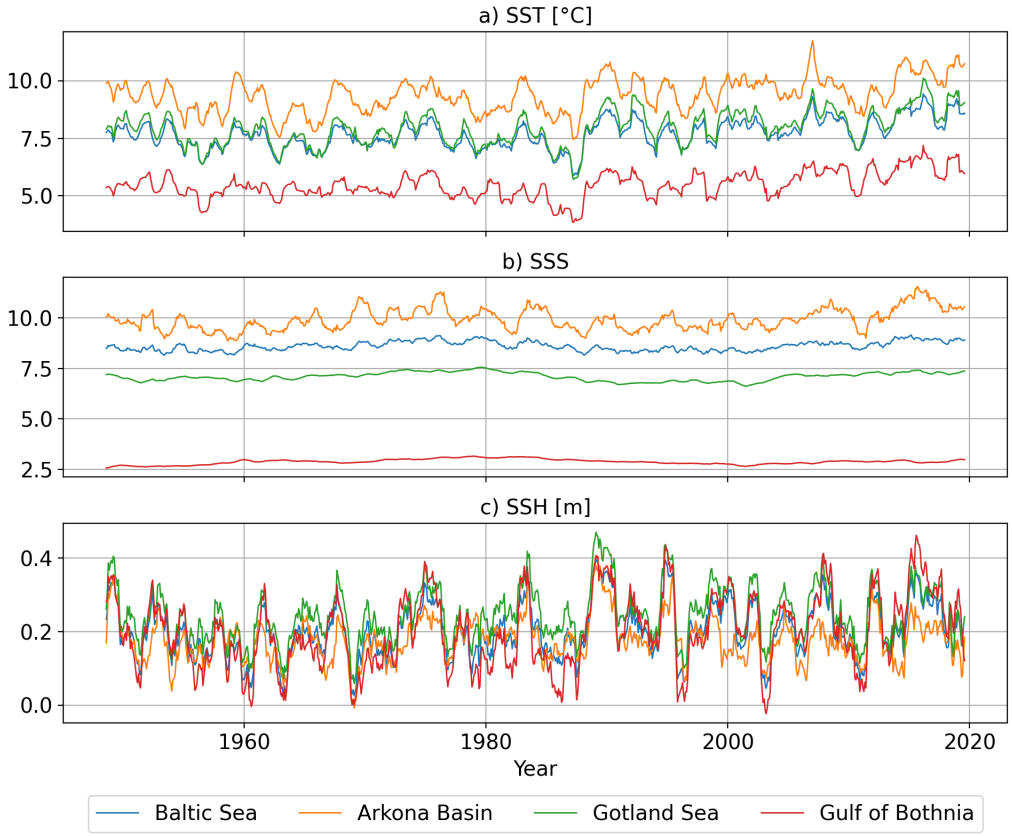


Figure A.1: Reference run overview: SST (a), SSS (b) and SSH (c) of the Baltic Sea and its subbasins. A 12 month smooth was applied to all shown variables.

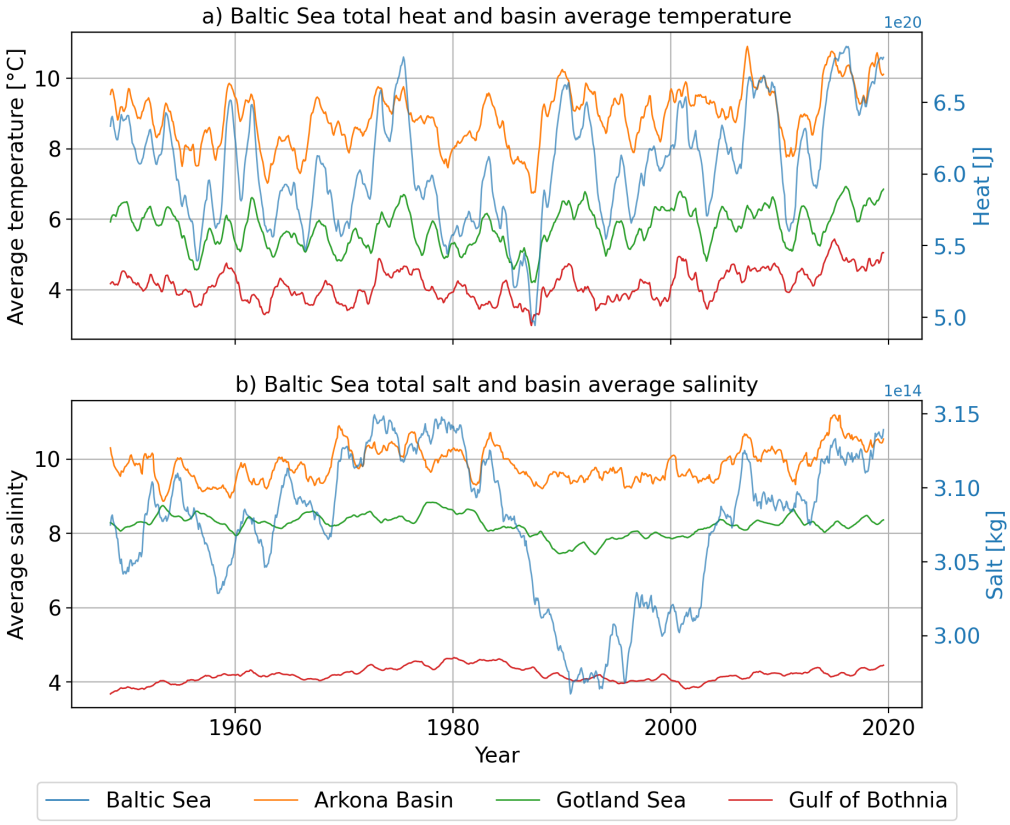


Figure A.2: Reference run overview: Heat content (a) and salt mass (b) integrated over the whole Baltic Sea and average temperature (a) and average salinity (b) of its subbasins. A 12 month smooth was applied to all shown variables.

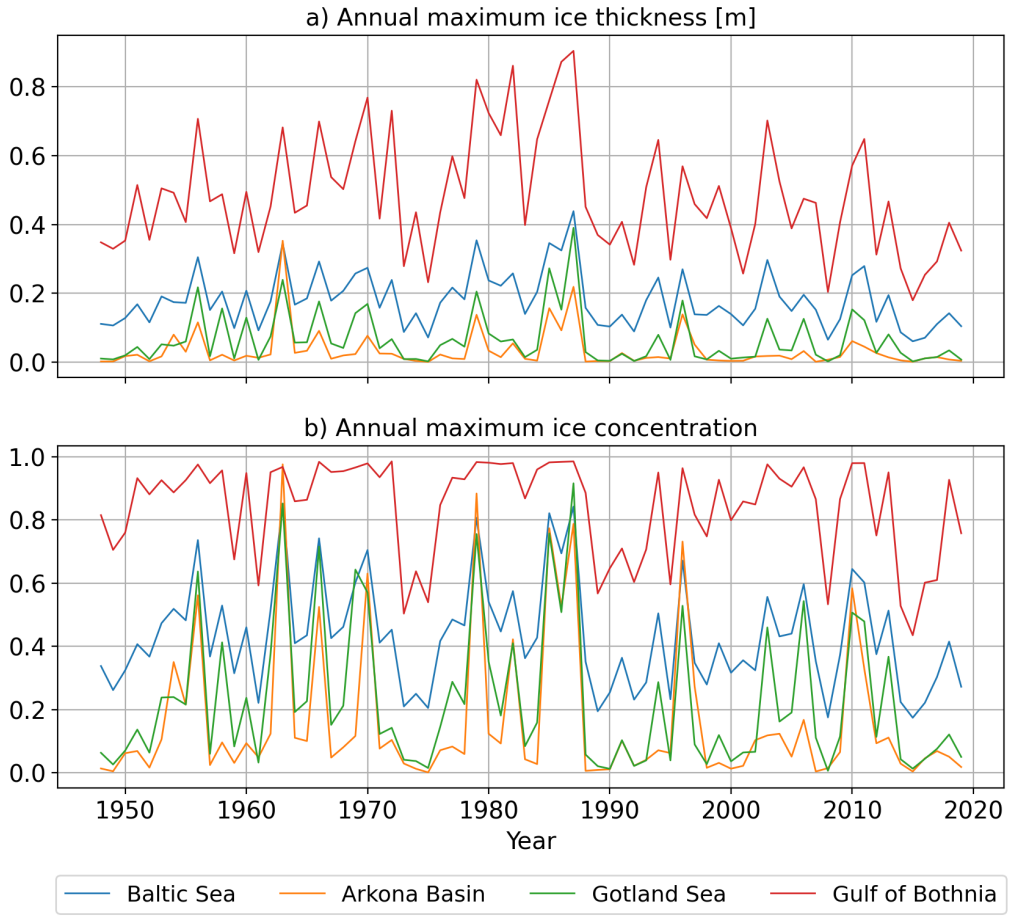


Figure A.3: Reference run overview: Annual maximum ice thickness (a) and annual maximum ice concentration (b), averaged over the Baltic Sea and its subbasins.

A.2. Results

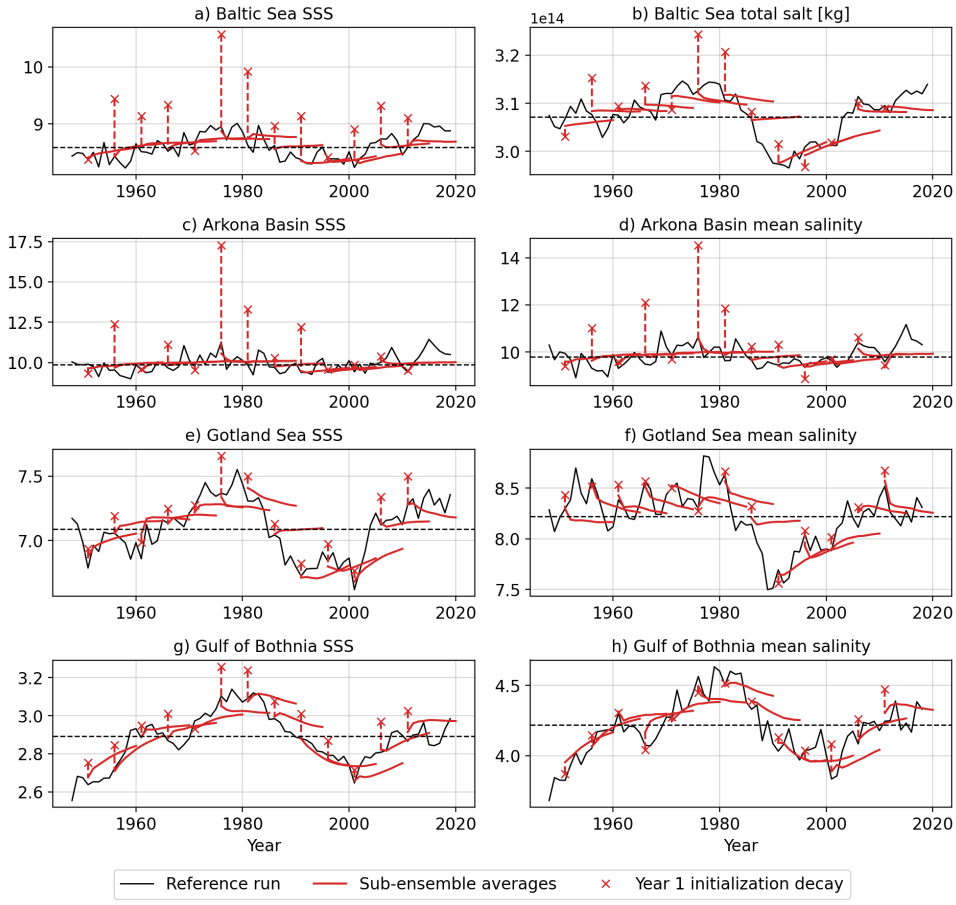


Figure A.4: Temporal salinity dynamics of the reference run and the sub-ensembles. Solid black line: Annually averaged reference run values. Solid red line: Sub-ensembles (characterized by their initial states) averaged over all forcings and annually in time. Dashed black line: Average value of the whole ensemble. Red X: Sub-ensemble initial values on January 1st of year 1. Red dashed line: initialization decay to the year 1 mean, where the sub-ensemble averages start. For visualization purposes this figure shows both annual average values of the reference run and sub-ensemble averages as well as the initialization snapshot on January 1st.

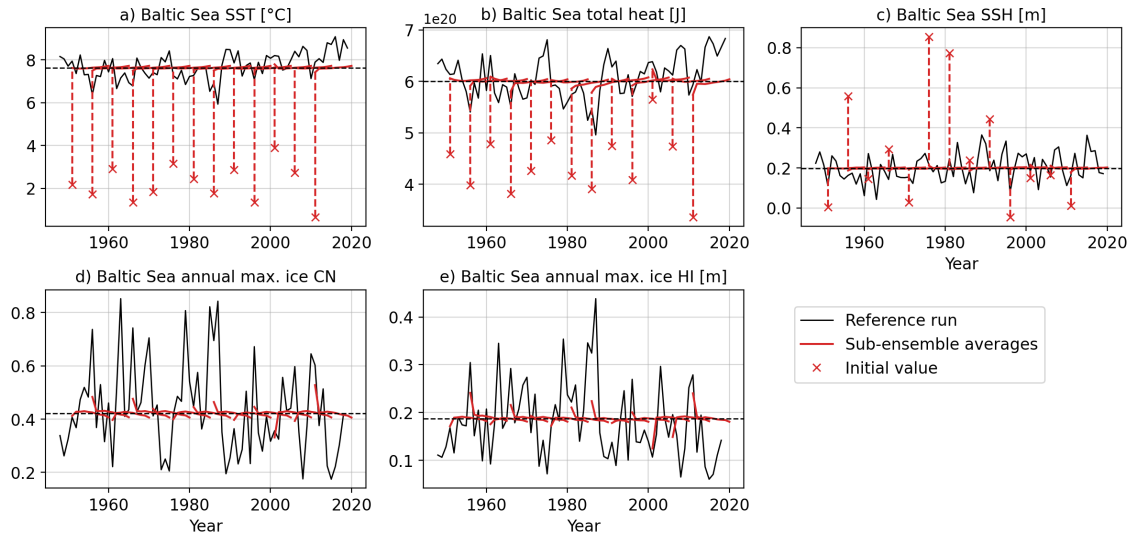


Figure A.5: Temporal Baltic Sea temperature, sea-level and ice dynamics of the reference run and the sub-ensembles. Solid black line: Annually averaged reference run values. Solid red line: Sub-ensembles (characterized by their initial states) averaged over all forcings and annually in time. Dashed black line: Average value of the whole ensemble. Red X: Sub-ensemble initial values on January 1st of year 1. Red dashed line: initialization decay to the year 1 mean, where the sub-ensemble averages start. For visualization purposes this figure shows both annual average values of the reference run and sub-ensemble averages as well as the initialization snapshot on January 1st.

Appendix B.

Methodological Considerations

B.1. Reasoning of deseasonalizing and detrending methods

In order to deseasonalize the MRP, the climatology of either the reference run or the prediction could be subtracted. The former method induces a new seasonal cycle, as shown in Fig. B.1a, since the climatologies of reference run and prediction differ (see Fig. B.1b). Thus the MRP were deseasonalized through subtraction of their own climatology. In order to prevent a falsification caused by the strong initialization anomaly, the first year of the prediction was left out in the climatology calculations.

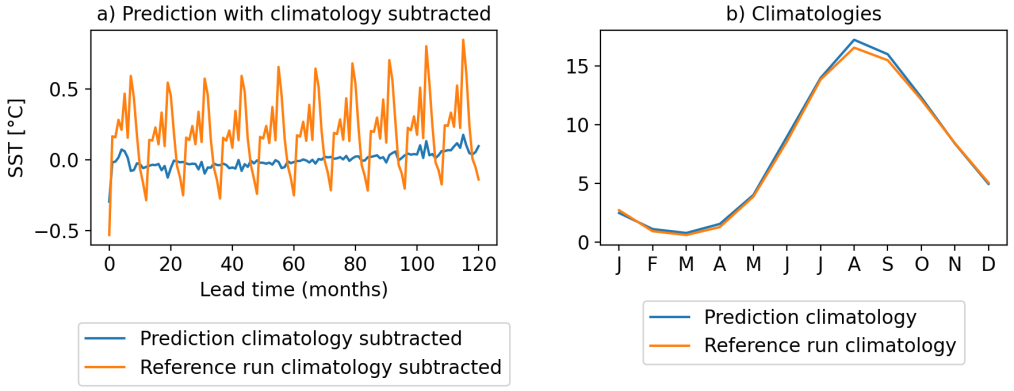


Figure B.1: Deasonalization illustration. Left: An SST MRP deseasonalized through subtraction of its own and the reference run's climatology. Right: The subtracted climatologies of prediction and reference run. Note the difference between the climatologies especially in later summer months.

The warming of the Baltic Sea masks the internal variability this study aims to analyze. Consequently, linear warming trends were removed from SST and temperature MRP and ARP and the according reference run parts. Both, the prediction and the reference run (10 year slice), had their own linear trends subtracted. Subtracting any trend from the prediction, other than its own, didn't reliably remove the warming trend. Also, the calculation of the trend excluded the first prediction year, similar to the deseasonalizing process.
Predictions of SSS, average salinity, SSH were not detrended as there are no known long-term salinity or sea level trends of the Baltic Sea. For salinity, detrending would actually remove the internal variability signal. Salinity oscillates on multi-decadal timescales and looking at a 10 year window could suggest a trend (see Fig. A.2b), i.e. from 1980-1990). Furthermore, Ice CN and HI predictions were not detrended, because this study uses monthly values and the subtraction of a trend would induce negative values during summer months, which aren't physical.

B.2. Bootstrapping blocksize determination

Monthly Resolved Predictions (MRP)

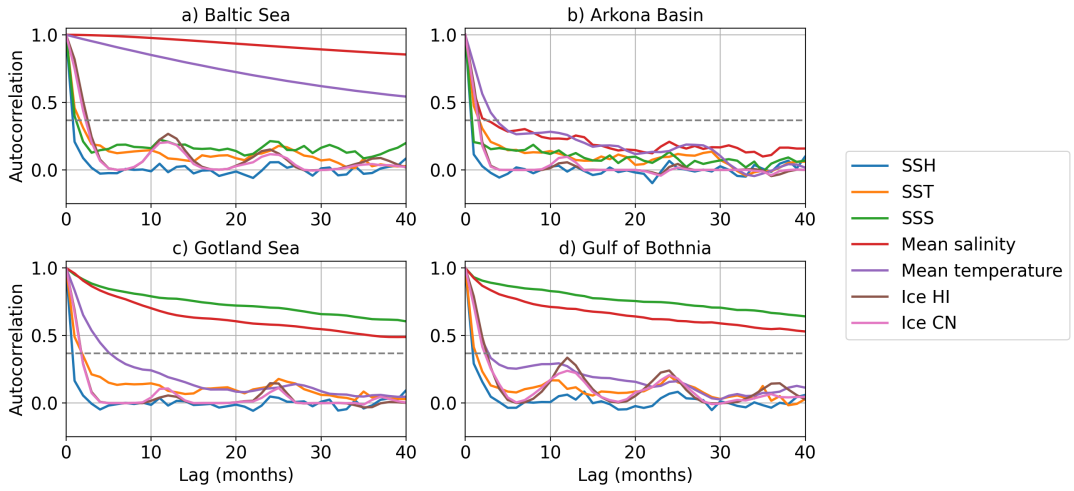


Figure B.2: Reference run autocorrelation functions of monthly resolution for all analyzed regions and variables. Dashed lines mark the 1/e threshold.

Defining the autocorrelation length as the time span after which the region-specific autocorrelation function of a variable doesn't exceeds 1/e anymore, most of the basin-specific variables analyzed in this study display autocorrelation lengths between 1-4 months (see Fig. B.2). These values were directly used as blocksizes for MRP on all timescales (see Table B.1). Additionally there are 7 instances with

autocorrelation lengths over 4 months, as shown in Fig. B.2: Temperature and salinity of the whole Baltic Sea, Gotland Sea SSS, temperature and salinity and Gulf of Bothnia SSS and salinity (all > 60 months, except Gotland Sea temperature 6 months). In these cases the blocksizes were determined separately for each analysis timescale. Therefore the standard errors of the bootstrapped distribution were calculated for varying blocksizes and each timescale, as shown in Fig. B.3 - B.9, for one exemplary pair of reference run and prediction timeseries with relatively high MSSS and ACC. The blocksizes were picked around the plateau of the standard error, under consideration that they don't exceed the autocorrelation length they base on, and that the timescale length can be divided by them, i.e. blocksizes of 1, 2, 3, 4 or 6 for the timescale 1 year. Table B.1 shows all results. Note that the MRP analysis uses monthly values, so the blocksizes describe the number of months per block.

Table B.1.: Blocked bootstrapping blocksizes for MRP. Single digit entries indicate the same block length for all timescales. Four digit entries display the blocksize for prediction timescales of 1y/5y/6-10y/10y.

Region	SST	SSS	SSH
Baltic Sea	2	2	1
Arkona Basin	2	1	1
Gotland Sea	2	4/20/10/40	1
Gulf of Bothnia	2	4/6/20/40	1

Region	Salinity	Temperature	Ice CN	Ice HI
Baltic Sea	6/10/20/40	3/10/10/10	3	3
Arkona Basin	3	4	2	2
Gotland Sea	4/20/15/40	3/6/6/6	2	2
Gulf of Bothnia	4/10/20/40	3	3	3

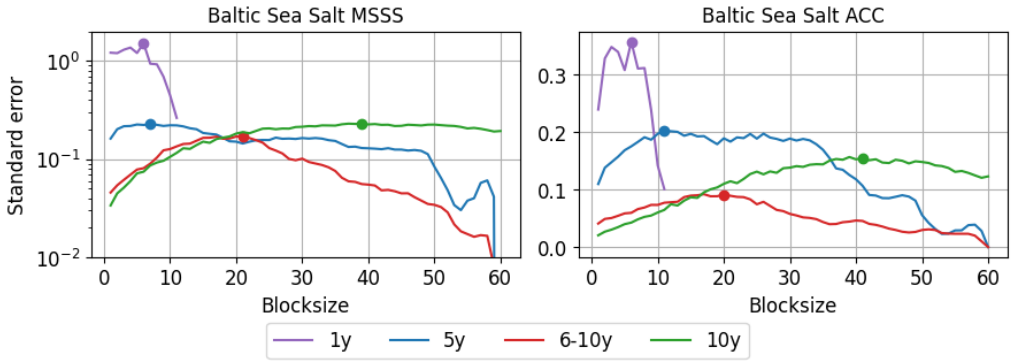


Figure B.3: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Baltic Sea salinity as a function of the blocksize (predictor: 1981 past 10y BS salinity).

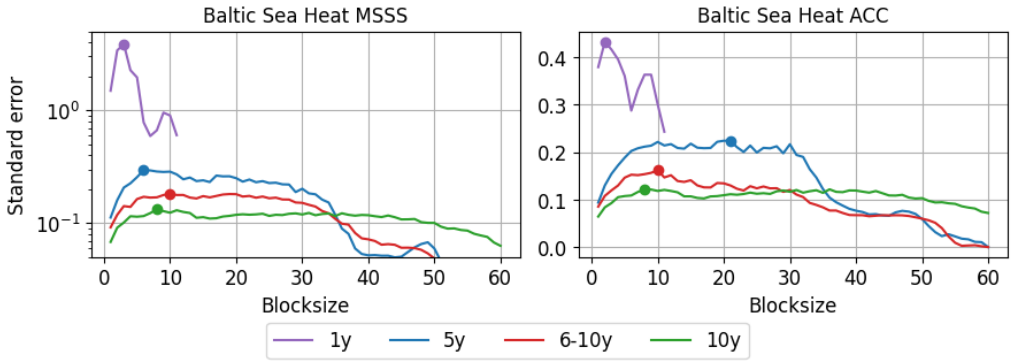
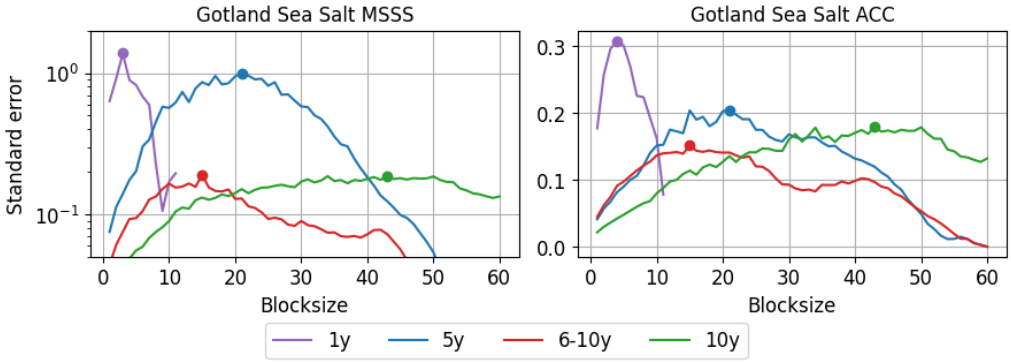


Figure B.4: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Baltic Sea temperature as a function of the blocksize (predictor: 1961 initial BS temperature).



[h]

Figure B.5: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Gotland Sea salinity as a function of the blocksize. (predictor: 1981 past 10y AMV).

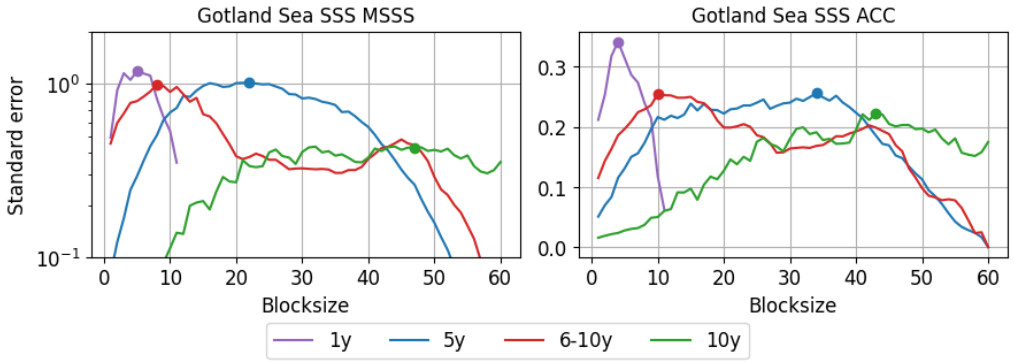


Figure B.6: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Gotland Sea SSS as a function of the blocksize (predictor: 2001 past 3y AMV).

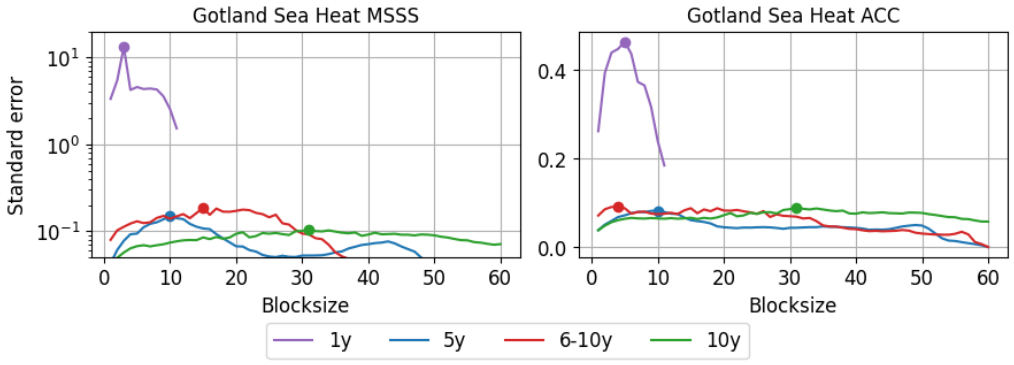


Figure B.7: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Gotland Sea temperature as a function of the blocksize.

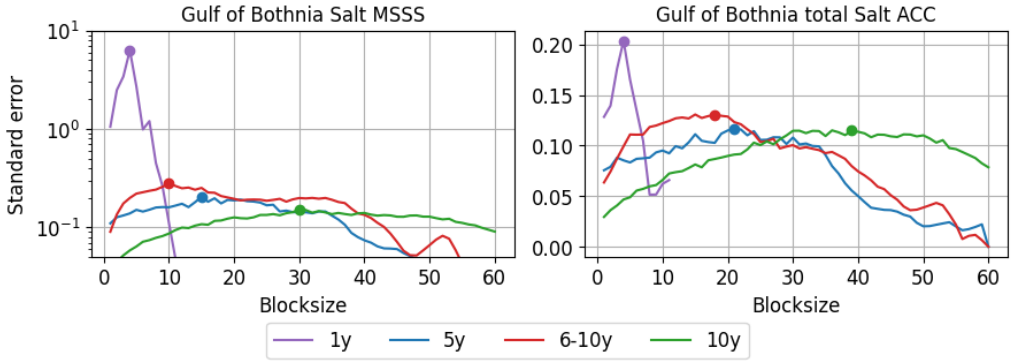


Figure B.8: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Gulf of Bothnia salinity as a function of the blocksize (predictor: 1951 initial BS salinity).

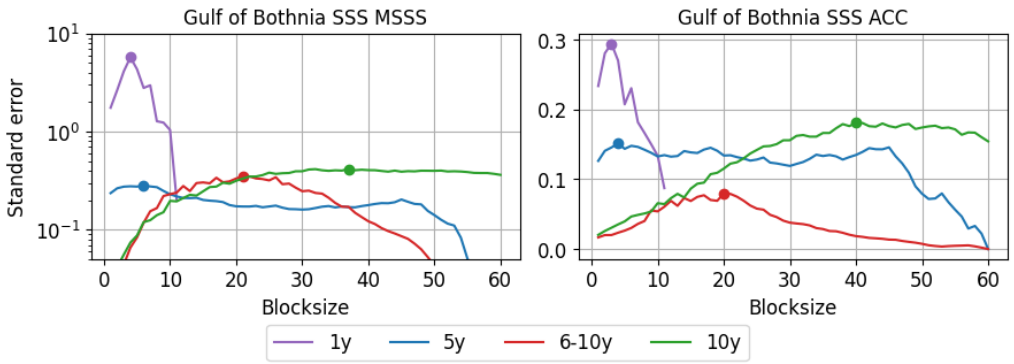


Figure B.9: Standard error of bootstrapped MSSS (left) and ACC (right) distributions for Gulf of Bothnia SSS as a function of the blocksize (predictor: 1951 initial BS salinity).

Annually resolved predictions (ARP)

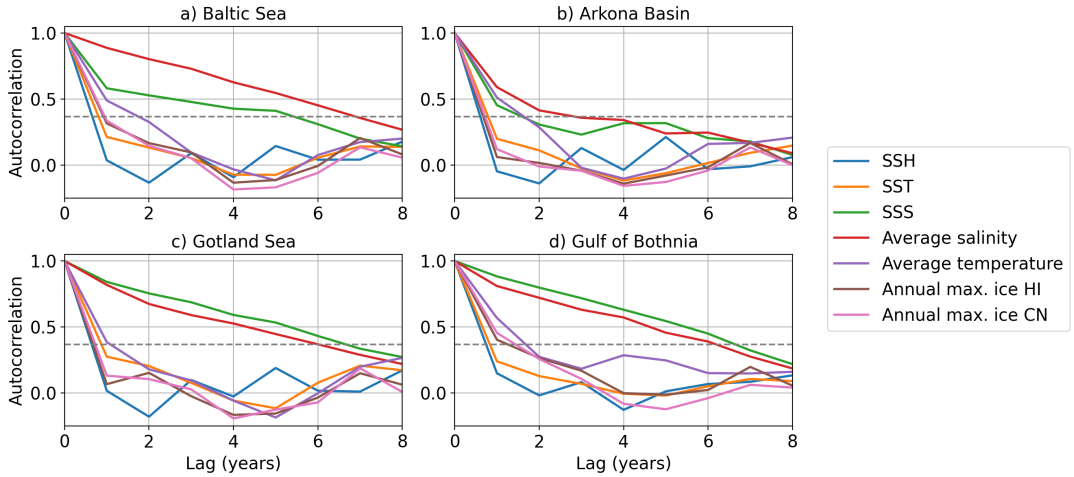


Figure B.10: Reference run autocorrelation functions of annual resolution for all analyzed regions and variables. Dashed lines mark the 1/e threshold.

Based on the annual autocorrelation functions shown in Fig. B.10 the blocksizes of the ARP were determined. In contrast to the MRP, where each block length unit corresponds to 1 month, here it corresponds to 1 year. Except for SSS and salinity all autocorrelation lengths lay at 1 or 2 years for all basins and are taken as blocksizes directly. The blocksizes for SSS and salinity were chosen so that each resample consists of at least three blocks. All ARP blocksizes are shown in Tab. B.2.

Table B.2.: Blocked bootstrapping blocksizes for ARP. Single digit entries indicate the same block length for all timescales. Six digit entries display the blocksize for prediction lengths of 5/6/7/8/9/10 years.

Region	SST	SSS	SSH
Baltic Sea	1	2/2/3/3/4/4	1
Gotland Sea	1	2/2/3/3/4/4	1
Gulf of Bothnia	1	2/2/3/3/4/4	1
Arkona Basin	1	2	1

Region	Salinity	Temperature	Ice CN	Ice HI
Baltic Sea	2/2/3/3/4/4	2	1	1
Gotland Sea	2/2/3/3/4/4	2	1	1
Gulf of Bothnia	2/2/3/3/4/4	2	1	1
Arkona Basin	2/2/3/3/3/3	2	1	1

Bibliography

IPCC. Climate change 2023: Synthesis report. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change, 2023. https://doi.org/10.59327/IPCC/AR6-9789291691647.

H. E. M. Meier, c. Dieterich, M. Gröger, et al. Oceanographic regional climate projections for the baltic sea until 2100. *Earth Sys. Dyn.*, 13:159–199, 2022. https://doi.org/10.5194/esd-13-159-2022.

G. J. Boer, D. M. Smith, C. Cassou, et al. The decadal climate prediction project (dcpp) contribution to cmip6. *Geosci. Mod. Dev.*, 9(10):3751–3777, 2016. https://doi.org/10.5194/gmd-9-3751-2016.

T. J. O'Kane, A. A. Scaife, Y. Kushnir, et al. Recent applications and potential of near-term (interannual to decadal) climate predictions. *Front. Clim.*, 5, 2023. https://doi.org/10.3389/fclim.2023.1121626.

B. Solaraju-Murali, D. Bojovic, N. Gonzalez-Reviriego, et al. How decadal predictions entered the climate services arena: an example from the agriculture sector. *Clim. Serv.*, 27:100303, 2022. https://doi.org/10.1016/j.cliser.2022.100303.

B. W. Hutchins, D. J. Brayshaw, L. C. Shaffrey, et al. Decadal prediction for the european energy sector. *Meteorol. Appl.*, 32(3):e70054, 2025. https://doi.org/10.1002/met.70054.

M. R. Payne, G. Danabasoglu, N. Keenlyside, et al. Skilful decadal-scale prediction of fish habitat and distribution shifts. *Nat. Commun.*, 13:2660, 2022. https://doi.org/10.1038/s41467-022-30280-0.

S. Gössling and D. Scott. Tourist demand and destination development under climate change: complexities and perspectives. *J. Sustain. Tour.*, pages 1–32, 2025. https://doi.org/10.1080/09669582.2025.2543953.

G. A. Meehl, J. A. Richter, H. Teng, et al. Initialized earth system prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.*, 2:340–356, 2021. https://doi.org/10.1038/s43017-021-00155-x.

T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, editors. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013. ISBN 978-1-107-05799-1. doi: 10.1017/CBO9781107415324.

Y. Kushnir, A. A. Scaife, R. Arritt, et al. Towards operational predictions of the near-term climate. *Nat. Clim. Change*, 9(2):94–101, 2019. https://doi.org/10.1038/s41558-018-0359-7.

G. A. Meehl, H. Teng, D. G. Smith, et al. The effects of bias, drift, and trends in calculating anomalies for evaluating skill of seasonal-to-decadal initialized climate predictions. *Clim. Dyn.*, 59:3373–3389, 2022. https://doi.org/10.1007/s00382-022-06272-7.

V. Eyring, S. Bony, G. A. Meehl, et al. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geosci. Mod. Dev.*, 9(5):1937–1958, 2016. https://doi.org/10.5194/gmd-9-1937-2016.

A. A. Scaife and M. Smith. A signal-to-noise paradox in climate science. *npj Clim. Atmos.*, 1:28, 2018. https://doi.org/10.1038/s41612-018-0038-4.

D. M. Smith, A. A. Scaife, R. Eade, et al. North atlantic climate far more predictable than models imply. *Nature*, 583(7818):796–800, 2020. https://doi.org/10.1038/s41586-020-2525-0.

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

S. Yeager, G. Danabasoglu, N. Rosenbloom, et al. Predicting near-term changes in the earth system: A large ensemble of initialized decadal prediction simulations using the community earth system model. *Bull. Amer. Meteor. Soc.*, 99, 2018. https://doi.org/10.1175/BAMS-D-17-0098.1.

R. Eade, D. M. Smith, and A. A. Scaife. Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, 41(15):5620–5628, 2014. https://doi.org/10.1002/2014GL061146.

M. Dobrynin, A. Düsterhus, K. Fröhlich, et al. Hidden potential in predicting wintertime temperature anomalies in the northern hemisphere. *Geophys. Res. Lett.*, 49(20):e2021GL095063, 2022. https://doi.org/10.1029/2021GL095063.

A. Mariotti, C. Baggett, E. A. Barnes, et al. Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, 11(5): E608–E625, 2020. https://doi.org/10.1175/BAMS-D-18-0326.1.

L. F. Borchert, W. A. Müller, and J. Baehr. Atlantic ocean heat transport influences interannual-to-decadal surface temperature predictability in the north atlantic. *J. Clim.*, 31(17):6763–6782, 2018. https://doi.org/10.1175/JCLI-D-17-0734.1.

L. F. Borchert, A. Düsterhus, S. Brune, et al. Forecast-oriented assessment of decadal hindcast skill for north atlantic sst. *Geophy. Res. Lett.*, 46(20):11444–11454, 2019. https://doi.org/10.1029/2019GL084758.

L. Lipfert, S. Brune, and J. Baehr. Decadal prediction of marine heatwaves in mpi-esm. *Geophys. Res. Lett*, 49(15):e2022GL099347, 2022. https://doi.org/10.1029/2022GL099347.

H. Fan, L. F. Borchert, S. Brune, et al. North atlantic subpolar gyre provides downstream ocean predictability. *npj Clim. Atmos.*, 6:145, 2023. https://doi.org/10.1038/s41612-023-00469-1.

F. Börgel, C. Frauen, T. Neumann, and Meier H. E. M. The atlantic multidecadal oscillation controls the impact of the north atlantic oscillation on north european climate. *Environ. Res. Lett.*, 15(10):104025, 2020. https://doi.org/10.1088/1748-9326/aba925.

F. Börgel, Meier H. E. M., M. Gröger, et al. Atlantic multidecadal variability and the implications for north european precipitation. *Environ. Res. Lett.*, 17(4):044040, 2022. https://doi.org/10.1088/1748-9326/ac5ca1.

J.W Hurrell. Decadal trends in the north atlantic oscillation. *Science*, 269:676–679, 1995. https://doi.org/10.1126/science.269.5224.676.

J.W Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck. An overview of the north atlantic oscillation. pages 1–35, 2003. https://doi.org/10.1029/134GM01.

H. E. M. Meier, L. Barghorn, F. Börgel, M. Gröger, L. Naumov, and H. Radtke. Multidecadal climate variability dominated past trends in the water balance of the baltic sea watershed. *npj Clim. Atmos.*, 6:58, 2023. https://doi.org/10.

1038/s41612-023-00380-9.

M. Visbeck, J. M. Hurrell, L. Polvani, and H. M. Cullen. The north atlantic oscillation: Past, present, and future. *Proc. Natl. Acad. Sci.*, 98(23):12876–12877, 2001. https://doi.org/10.1073/pnas.231391598.

C. R. Patrizio, P. J. Athanasiadis, D. M. Smith, and D. Nicoli. Ocean-atmosphere feedbacks key to nao decadal predictability. *npj Clim. Atmos.*, 8:146, 2025. https://doi.org/10.1038/s41612-025-01027-7.

D. M. Smith, A. A. Scaife, R. Eade, and J. R. Knight. Seasonal to decadal prediction of the winter north atlantic oscillation: emerging capability and future prospects. *Q. J. R. Meteorol. Soc.*, 142(695):611–617, 2014. https://doi.org/10.1002/qj.2479.

P. J. Athanasiadis, S. Yeager, Y.-O. Kwon, et al. Decadal predictability of north atlantic blocking and the nao. *npj Clim. Atmos.*, 3:20, 2020. https://doi.org/10.1038/s41612-020-0120-6.

K. Strommen, T. Woolings, P. Davini, et al. Predictable decadal forcing of the north atlantic jet speed by sub-polar north atlantic sea surface temperatures. *Weather Clim. Dyn.*, 4(4):853–874, 2023. https://doi.org/10.5194/wcd-4-853-2023.

J. R. Knight, C. F. Folland, and A. A. Scaife. Climate impacts of the atlantic multidecadal oscillation. *Geophys. Res. Lett.*, 33(17), 2006. https://doi.org/10.1029/2006GL026242.

F. Börgel, C. Frauen, T. Neumann, et al. Impact of the atlantic multidecadal oscillation on baltic sea variability. *Geophy. Res. Lett*, 45(18):9880–9888, 2018. https://doi.org/10.1029/2018GL078943.

F. Börgel, M. Gröger, Meier H. E. M., et al. The impact of atlantic multidecadal variability on baltic sea temperatures limited to winter. *npj Clim. Atmos.*, 6(4):64, 2023. https://doi.org/10.1038/s41612-023-00373-8.

L. Barghorn, F. Börgel, M. Gröger, and H. E. M. Meier. Atlantic multidecadal variability control on european sea surface temperatures is mainly externally forced. *Environ. Res. Lett.*, 20(3):034044, 2025. https://doi.org/10.1088/1748-9326/adb6bf.

S. Yeager. The abyssal origins of north atlantic decadal predictability. *Clim. Dyn.*, 55:2253–2271, 2020. https://doi.org/10.1007/s00382-020-05382-4.

M. W. Buckley, T. DelSole, M. S. Lozier, and L. Li. Predictability of north atlantic sea surface temperature and upper-ocean heat content. *J. Climate*, 32:3005–3023, 2019. https://doi.org/10.1175/JCLI-D-18-0509.1.

G. Liu, P. Wang, and Y.-O. Kwon. Physical insights from the multidecadal prediction of north atlantic sea surface temperature variability using explainable neural networks. *Geophys. Res. Lett.*, 50(24):e2023GL106278, 2023. https://doi.org/10.1029/2023GL106278.

B. Huddart, A. Subramanian, L. Zanna, and T. Palmer. Seasonal and decadal forecasts of atlantic sea surface temperatures using a linear inverse model. *Clim. Dyn.*, 49:1833–1845, 2016. https://doi.org/10.1007/s00382-016-3375-1.

J. M. Klavans, A. C. Clement, and M. A. Cane. Variable external forcing obscures the weak relationship between the nao and north atlantic multidecadal sst variability. *J. Climate*, 32:3847–3864, 2019. https://doi.org/10.1175/JCLI-D-18-0409.1.

V. Mohrholz. Major baltic inflow statistics – revised. *Front. Mar. Sci.*, 5, 2018. https://doi.org/10.3389/fmars.2018.00384.

A. Lehmann, K. Myrberg, P. Post, et al. Salinity dynamics of the baltic sea. *Earth Sys. Dyn.*, 13:373–392, 2022. https://doi.org/10.5194/esd-13-373-2022.

T. Liblik, M. Naumann, P. Alenius, et al. Propagation of impact of the recent major baltic inflows from the eastern gotland basin to the gulf of finland. *Front. Mar. Sci.*, 5, 2018. https://doi.org/10.3389/fmars.2018.00222.

P. Winsor, J. Rodhe, and A. Omstedt. Baltic sea ocean climate: an analysis of 100 yr of hydrographic data with focus on the freshwater budget. *Clim. Res.*, 18:5–15, 2001. https://doi.org/10.3354/cr018005.

H. E. M. Meier and F. Kauker. Modeling decadal variability of the baltic sea: 2. role of freshwater inflow and large-scale atmospheric circulation for salinity. *J. Geophys. Res.*, 108:3368, 2003. https://doi.org/10.1029/2003JC001799.

K. Döös, H. E. M. Meier, and R. Döscher. The baltic haline conveyor belt or the overturning circulation and mixing in the baltic. *Ambio*, 33:261–266, 2004. https://doi.org/10.1639/0044-7447(2004)033[0261:TBHCBO]2.0.CO;2.

H. E. M Meier. Modeling the pathways and ages of inflowing salt- and freshwater in the baltic sea. *Estuar. Coast. Shelf Sci.*, 74:610–627, 2007. https://doi.org/10.1016/j.ecss.2007.05.019.

M. Kniebusch, H. E. M. Meier, T. Neumann, and F. Börgel. Temperature variability of the baltic sea since 1850 and attribution to atmospheric forcing variables. *J. Geophys. Res.: Oceans*, 124(6):4168–4187, 2019. https://doi.org/10.1029/2018JC013948.

A. Omstedt and D. Hansson. The baltic sea ocean climate system memory and response to changes in the water and heat balance components. *Cont. Shelf Res.*, 26:236–251, 2006. https://doi.org/10.1016/j.csr.2005.11.003.

Y. Karpechko, K. A. Peterson, A. A. Scaife, J. Vainio, and H. Gregow. Skilful seasonal predictions of baltic sea ice cover. *Env. Res. Lett.*, 10(4):044007, 2015. https://doi.org/10.1088/1748-9326/10/4/044007.

K. Novotny, G. Liebsch, Lehmann A., and R. Dietrich. Variability of sea surface heights in the baltic sea: An intercomparison of observations and model simulations. *Mar. Geod.*, 29(2):113–134, 2006. https://doi.org/10.1080/01490410600738054.

A. A. Karimi, K. Ghobadi-Far, and M. Passaro. Barystatic and steric sea level variations in the baltic sea and implications of water exchange with the north sea in the satellite era. *Front. Mar. Sci.*, 9, 2022. https://doi.org/10.3389/fmars.2022.963564.

H. E. M. Meier. The baltic sea ocean climate system memory and response to changes in the water and heat balance components. *Clim. Dyn.*, 27(1):39–68, 2006. https://doi.org/10.1007/s00382-006-0124-x.

M. Dobrynin, D. I. V. Domeisen, W. A. Müller, et al. Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophys. Res. Lett.*, 45(8):3605–3614, 2018. https://doi.org/10.1002/2018GL077209.

L. Sonesten, Svendsen L. M., H. Tornbjerg, B. Gustafsson, D. Frank-Kamenetsky, and J. Haapaniemi. *Sources and pathways of nutrients to the Baltic Sea*, volume 153 of *Baltic Sea Environment Proceedings*. HELCOM, 2018. https://helcom.fi/wp-content/uploads/2019/08/BSEP153.pdf.

T. Neumann, H. Radtke, B. Cahill, M. Schmidt, and G. Rehder. Non-redfieldian carbon model for the baltic sea (ergom version 1.2) – implementation and budget estimates. *Geosci. Mod. Dev.*, 15(22):8473–8540, 2022. https://doi.org/10.5194/gmd-15-8473-2022.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, 93:485–498, 2012. https://doi.org/10.1175/BAMS-D-11-00094.1.

A. Kaplan, M. A. Cane, Y Kushnir, et al. Analyses of global sea surface temperature 1856-1991. *J. Geophys. Res.*, 103(C9):18567–18589, 1998. https://doi.org/10.1029/97JC01736.

C. K. Ho, E. Hawkins, L. Shaffrey, and F. M. Underwood. Statistical decadal predictions for sea surface temperatures: a benchmark for dynamical gcm predictions. *Clim. Dyn.*, 41(3):917–935, 2013. https://doi.org/10.1007/s00382-012-1531-9.

D. M. Smith, E. Rosie, and H. Pohlmann. A comparison of full-field and anomaly initialization for seasonal to decadal climate. *Clim. Dyn.*, 41(11):3325–3338, 2013. https://doi.org/10.1007/s00382-013-1683-2.

D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 100 of *International Geophysics Series*. Academic Press, Elsevier, 3rd edition, 2011.

H. R. Kunsch. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.*, 17(3):1217–1241, 1989. https://doi.org/10.1214/aos/1176347265.

P. Hall, J. L. Horowitz, and B.-Y. Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574, 1995. https://doi.org/10.1093/biomet/82.3.561.

D. N. Politis and H. White. Automatic block-length selection for the dependent bootstrap. *Econom. Rev.*, 23(1):53–70, 2004. https://doi.org/10.1081/ETC-120028836.

S. Mignani and R. Rosa. The moving block bootstrap to assess the accuracy of statistical estimates in ising model simulations. *Comput. Phys. Commun.*, 92(2):203–213, 1995. https://doi.org/10.1016/0010-4655(95)00114-7.

Code availability

Reference run and ensemble were both computed with the MOM6 coupled with the SIS2. The models source code can be found under `http://github.com/mom-ocean/MOM6`. Specific model calibration code is available from the author upon request.

All code used in the data evaluation is uploaded under: `https://github.com/martiwolff/master-thesis`.

Data availability

Unsmoothed AMV index data, based on North Atlantic SST, was taken from `https://www.psl.noaa.gov/data/timeseries/AMO/`, with the last access on May 12th 2025.

Reference run and ensemble data are gladly available from the author upon request.