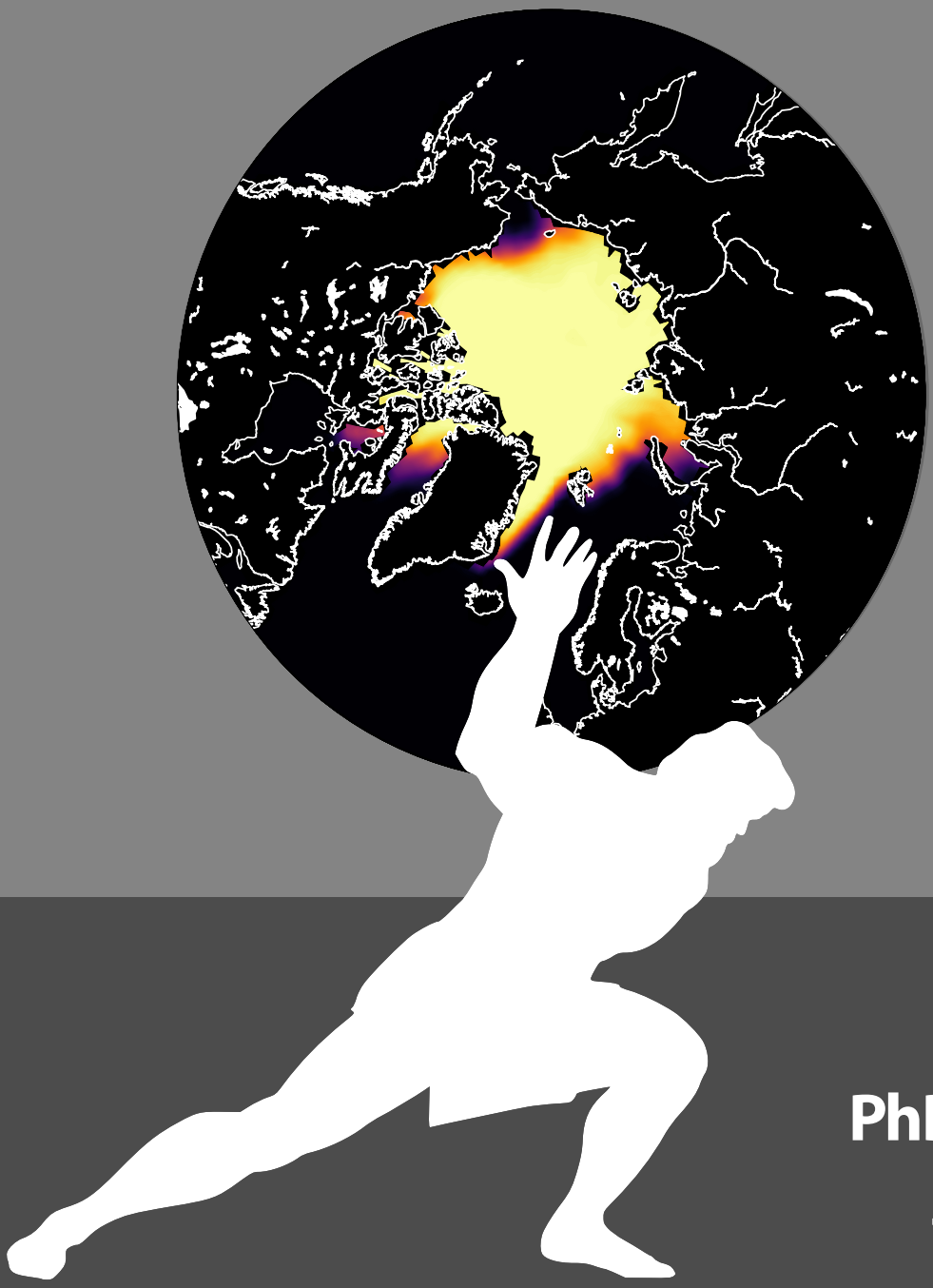
Seasonal Arctic sea ice predictability and prediction

Rubén Cruz García



PhD Thesis
2020

Seasonal Arctic sea ice predictability and prediction

Predictibilidad y predicción estacional del hielo marino del Ártico

Memoria que presenta

Rubén Cruz García

para optar al grado de

Doctor en Ingeniería Ambiental

Directores:

Dr. Pablo Ortega Montilla

Dra. Virginie Guemas

Tutor:

Dra. María Gonçalves Ageitos

Universitat Politècnica de Catalunya
Barcelona, 2020

Dedicado a mis
abuelas Lola y Catalina
y a mi abuelo Paco

Agradecimientos

Aunque una tesis la lleve a cabo una sola persona, suele tratarse de un gran proyecto que necesita de la colaboración de muchas otras personas, y sin las cuales el proyecto no sería posible. Seguramente estas palabras no sean suficiente para agradecer la ayuda recibida durante la elaboración de esta tesis, pero por intentarlo que no quede:

Igual que el arco de Tito se sustenta sobre dos grandes pilares, este trabajo reposa sobre dos personas fundamentales, sin las cuales esta tesis no sería como es: Pablo y Virginie. Pablo, gracias por todo el tiempo que has dedicado, por todos los buenos consejos y echarme una mano en los momentos más difíciles. Virginie, merci de m'avoir tant appris, de ta patience, de m'inspirer par ton exemple, ainsi que de m'avoir donné l'opportunité de démarrer ce projet. Gracias a los dos por ser exigentes, a la par que flexibles cuando lo necesitaba.

Matthieu, merci pour ton soutien, et de m'avoir guidé lors de mes débuts lorsque je sentais encore un peu perdu. Merci également de m'avoir appris à aborder les problèmes sous différents angles. J'aurais aimé t'avoir à mes côtés jusqu'à la fin de ce marathon; mais peut-être que de cette façon j'apprécierai davantage l'aide précieuse que tu m'as donné au début.

Quiero dar las gracias a Juan por estar siempre disponible para cualquier duda, pregunta o reflexión. También a Valentina, por ejercer de profesora particular para lanzar mis primeras simulaciones. A Nicolau por enseñarme tanto de *R* y resolverme tantos quebraderos de cabeza. A Pierre Antoine por ayudarme con problemas de distinta índole, desde *bash* hasta todo lo que tenga que ver con descargar, organizar, manipular y formatear datos. A Javi Vegas por estar disponible cuando lo necesitaba para cualquier diagnóstico. A Francesco por responder todas mis dudas de *python*. A Edu y Roberto por responder mis mini-preguntas varias. A Javi García por las discusiones iniciales. A Kim por todo el soporte técnico. Gracias a todos los integrantes del *Climate Prediction Group*, que me han hecho sentir como parte de una gran familia. Gracias también a todas las personas que han pasado por

el departamento y que me han ayudado. Hay muchas que me dejo, pero que igualmente merecen mi agradecimiento.

Je remercie aussi François pour toute son aide dès les premiers instants, tant sur le plan scientifique que technique, ansi que de m'avoir accueilli à Louvain-la-Neuve où je me sentais si bien. Gracias a Paco por su ecuanimidad y su compromiso. Por interesarse por mi trabajo y por discutir resultados en momentos clave. Quiero agradecer a María su esfuerzo por ayudarme con la burocracia y por gestionar eficientemente mis clases en la universidad.

Gracias a Danila por ser tan atenta y por responder a cualquier tipo de duda, además de ejercer de "hermana" de investigación. Gracias a todos los que me hicieron sentir como en casa en Toulouse: Robin, Michael, Sofia, Francesca y Claudia (y los que seguramente me deje). Gracias a Aude por transmitirme su alegría. Igualmente, mil gracias a toda la gente de Louvain-la-Neuve por acogerme: Sylvain, Jean, Leandro, Koffi, Guillian, Steve, Lilian y Quentin.

Thanks to Lauriane and Steffen for reviewing the thesis with so little time and for their valuable suggestions. I think their comments have improved the final manuscript.

Gracias a Lluís por compartir las penas, pero sobre todo las alegrías. También por ayudarme a adaptarme al entorno de trabajo y a Barcelona. No puede faltar mi agradecimiento a los *PhDs et al.*, por el apoyo recibido y tantos buenos momentos: Dani, Bala, Bianca, Jaime, Vincenzo, Cristina, Elisa, Froila, Ferrán y Carlos. Empezamos siendo compañeros y acabamos como amigos. Gracias por tantas risas.

Gracias a mis amigos de *La Panda*, que tanto en Cartagena como en sus visitas a Barcelona me han traido las risas y la alegría que me han mantenenido cuerdo. Gracias a Mikel por ser un amigo a 600km y por estar siempre disponible.

Esto no hubiese sido posible sin el apoyo de mi familia. Gracias por todo vuestro sacrificio para dármelo todo, por animarme a hacer lo que me gustaba y por enseñarme a perseverar.

Por último, gracias a la persona que me soporta todos los días, Agata, merecedora de una medalla. Por alentarme a seguir en los momentos más difíciles.

Contents

Abstract

Arctic sea ice plays a central role in the Earth's climate. Changes in the sea ice on seasonal-to-interannual timescales impact ecosystems, populations and a growing number of stakeholders. A prerequisite for achieving better sea ice predictions is a better understanding of the underlying mechanisms of sea ice predictability.

In the first part of the thesis, we investigate the seasonal-to-interannual Arctic sea ice predictability in so-called "perfect-model" experiments performed with six different global climate models, exploring the model ability to predict itself up to 3 years in advance if the initial conditions were exactly known. Robust mechanisms for reemergence of sea ice skill are highlighted, i.e. increases in predictability in sea ice properties after an initial loss. Similar pan-Arctic winter sea ice extent reemergence is found for HadGEM1.2, GFDL-CM3 and E6F, while a sea ice volume persistence from 1 to 3 years is confirmed for all models. Robust similarities in winter sea ice extent predictability reemergence in the GIN seas and Baffin Bay are found even though models have distinct background sea ice states. A robust summer sea ice volume skill reemergence is also found in the Barents, Kara and Chukchi seas. An in-depth analysis of the regional sea ice predictability in EC-Earth2.3 suggests that Arctic basins can be classified according to three distinct regimes. The central Arctic drives most of the pan-Arctic sea ice volume persistence. In peripheral seas, we find predictability for the sea ice extent in winter associated with ocean thermal anomalies persistence, but low predictability throughout the rest of the year. The Labrador Sea stands out as the region with sea ice predictability extending the longest, i.e. up to 1.5 years, the reemergence of predictability in winter being driven by the advection of heat content anomalies along the subpolar gyre.

The potential predictability experiments provide a reference of the upper level of skill that could be achieved in real predictions if the model represented perfectly the real world, and

perfect and complete observations were able to initialize the predictions. In real predictions, however, forecast errors appear due to inconsistencies between the initial states of the different model components on one hand and due to the development of the inherent model biases on the other hand. This thesis thus identifies and quantifies the contribution of initial condition inconsistencies and systematic model errors to the forecast model errors in two sets of seasonal forecasts (May and November initialized, respectively) produced with EC-Earth3.2 during the first forecast month. After 24 (19) days, the inherent model biases become the largest contributor to the forecast error for the May (November) initialized forecasts, while the initial inconsistency dominates in the previous days. This initial inconsistency is mostly associated to a mismatch between the sea ice and ocean initial conditions, with a marginal role associated to differences with the atmosphere, and its effect can be seen in the Greenland Sea and the Baffin Bay, in particular in November. The development of both types of errors is sensitive to the month of initialization: the initial shock is more pronounced in November than in May because the initial ocean is warmer and less consistent with the initial sea ice cover. Furthermore, in both cases, the initial shock leads to sea ice melting, but, unlike in May, in November it happens while sea ice is expanding. These findings highlight the importance of looking at high frequency data to disentangle the evolution of errors within the first forecast month.

To study how to minimize the initial shocks and the model drift while maximizing the observational information provided at initialization to the model, the last part of the thesis aims at comparing three seasonal forecast systems. These systems are based on EC-Earth and initialized through three different strategies for the sea ice and ocean components: (1) using both the sea ice and ocean initial conditions from a native reconstruction that assimilates ORAS4 temperature and salinity with a weak surface restoring coefficient, (2) taking the sea ice initial conditions from the same reconstruction as in *1* and the oceanic initial conditions from ORAS4 and (3) the same as in *1* but using a stronger restoring coefficient towards ORAS4. The focus is set on evaluating the impact of these methodological choices on the sea ice biases and skill. Initialization leads to reduced sea ice biases for the three systems which have not reached yet the model attractor after seven months. Initialization strategy *2* induces an initial shock because of a too warm polar surface ocean in ORAS4 for the reconstructed sea ice initial conditions, the forecast system recovering from the shock in less than 20 days. For strategy *3*, a strong ocean nudging towards ORAS4 produces a too warm ocean and a sea ice deficit, which pushes the forecast to have initial biases of opposite signs from the inherent model state. The forecast rapidly drifts and the sea ice shortage is reverted while a cold bias develops in less than one month in agreement with the model attractor.

For the prediction initialized using method *1*, no shocks are observed, and it converges faster towards the systematic model error, since it is initially closer to the attractor. The sea ice errors of strategies *2* and *3* converge in less than a month, demonstrating the role of the ocean initial conditions (which are very close one to another) in driving the sea ice biases. The lowest sea ice concentration bias in both DJF and MAM was found in the Atlantic sector for the forecast using strategy *3*, while in the Pacific sector for strategy *2*. Regarding the sea ice prediction skill, no significant differences appear between the three forecast systems. However all three show significantly higher skill than the historical experiment for predicting the pan-Arctic sea ice extent during the first two forecast months. For the individual regions, the Labrador Sea stands out as the only basin with a significant added value of initialization during the whole prediction (7 months), whereas only about 3 months of added-value are detected in the peripheral basins.

Resumen

El hielo marino del Ártico juega un papel central en el clima de la Tierra. Los cambios en el hielo marino en escalas temporales estacionales a interanuales afectan a los ecosistemas, las poblaciones y un número creciente de sectores socio-económicos (como el transporte marino, o el turismo). Para lograr mejores predicciones de hielo marino se necesita una mejor comprensión de los mecanismos subyacentes de su predictibilidad.

En la primera parte de la tesis, investigamos la predictibilidad del hielo marino del Ártico en escalas de tiempo estacional a interanual en los llamados experimentos de "modelo perfecto" realizados con seis modelos climáticos globales diferentes, en los que se explora la capacidad de cada modelo para predecirse a sí mismo (hasta 3 años en el futuro) si se tuvieran unas condiciones iniciales perfectas. El análisis multimodelo ha permitido identificar distintos mecanismos de reemergencia de la predictibilidad del hielo marino, es decir, un aumento de la predictibilidad en las propiedades del hielo marino después de una pérdida inicial. Por ejemplo, se ha encontrado una reemergencia similar para la predictibilidad de la extensión del hielo marino panártico de invierno en HadGEM1.2, GFDL-CM3 y E6F, mientras que para el volumen de hielo marino todos los modelos muestran una clara persistencia de 1 a 3 años. A escala regional, hemos encontrado similitudes sólidas en el resurgimiento de la predicción de la extensión del hielo marino invernal en los mares GIN y la Bahía de Baffin a pesar de que los modelos tienen estados climatológicos de hielo marino distintos. En los mares de Barents, Kara y Chukchi también se encuentra un resurgimiento robusto del volumen del hielo marino en verano. Un análisis en profundidad de la predictibilidad regional del hielo marino en EC-Earth2.3 sugiere además que las cuencas árticas se pueden clasificar de acuerdo con tres regímenes distintos. El Ártico central es la región que contribuye en mayor parte a la persistencia del volumen de hielo marino panártico. En los mares periféricos, encontramos predictibilidad para la extensión del hielo marino en invierno asociada con la persistencia de

anomalías térmicas oceánicas, pero poca predictibilidad durante el resto del año. El Mar de Labrador destaca como la región con la predictibilidad del hielo marino que se extiende por más tiempo, es decir, hasta 1.5 años, y el resurgimiento de la predictibilidad en invierno se debe a la advección de anomalías en el contenido de calor procedentes del giro subpolar del Atlántico Norte. Los experimentos de predictibilidad potencial proporcionan una referencia del nivel superior de la habilidad de predicción que se podría lograr en predicciones reales si el modelo representara perfectamente el mundo real, y se dispusiera de observaciones perfectas y completas para inicializar las predicciones. Sin embargo, en las predicciones reales las condiciones iniciales incluyen grandes incertidumbres que pueden limitar la capacidad predictiva. Además, errores de predicción también pueden aparecer, por ejemplo, debido a inconsistencias entre los estados iniciales de las diferentes componentes del modelo, y también debido al desarrollo de sesgos inherentes al modelo.

La segunda parte de esta tesis identifica y cuantifica la contribución de las inconsistencias de las condiciones iniciales y los errores sistemáticos del modelo a los errores en el primer mes de predicción en dos sistemas de predicciones estacionales (inicializados en mayo y noviembre, respectivamente) producidos con EC-Earth3.2. Después de 4 (19) días, el error sistemático del modelo se convierte en el mayor contribuyente al error de predicción para las predicciones inicializadas en mayo (noviembre), mientras que la inconsistencia inicial domina en los días anteriores. Esta inconsistencia inicial se asocia principalmente a un desajuste entre las condiciones iniciales del hielo marino y del océano, con un papel marginal asociado a las diferencias con la atmósfera. Su efecto se manifiesta principalmente en el Mar de Groenlandia y la Bahía de Baffin, y en particular en noviembre. El desarrollo de ambos tipos de errores es sensible al mes de inicialización: el choque inicial es más pronunciado en noviembre que en mayo porque el océano inicial es más cálido y menos consistente con la cubierta inicial de hielo marino. Además, en ambos casos, el choque inicial conduce al derretimiento del hielo marino, pero, a diferencia de mayo, en noviembre ocurre mientras el hielo marino se expande, lo que acelera su desarrollo. Este análisis ha sido posible gracias a la disponibilidad de productos observacionales con resolución diaria con los que caracterizar los errores y su evolución a lo largo del primer mes de predicción. Para estudiar cómo minimizar los choques iniciales y la deriva del modelo mientras se maximiza la información observacional proporcionada en la inicialización del modelo, la última parte de la tesis tiene como objetivo comparar tres sistemas de predicción estacionales con el modelo EC-Earth.

Los tres sistemas se diferencian en las estrategias utilizadas para la inicialización de sus componentes de hielo marino y de océano: (1) utilizando para ambas componentes condiciones iniciales de una reconstrucción nativa que asimila la temperatura y la salinidad de

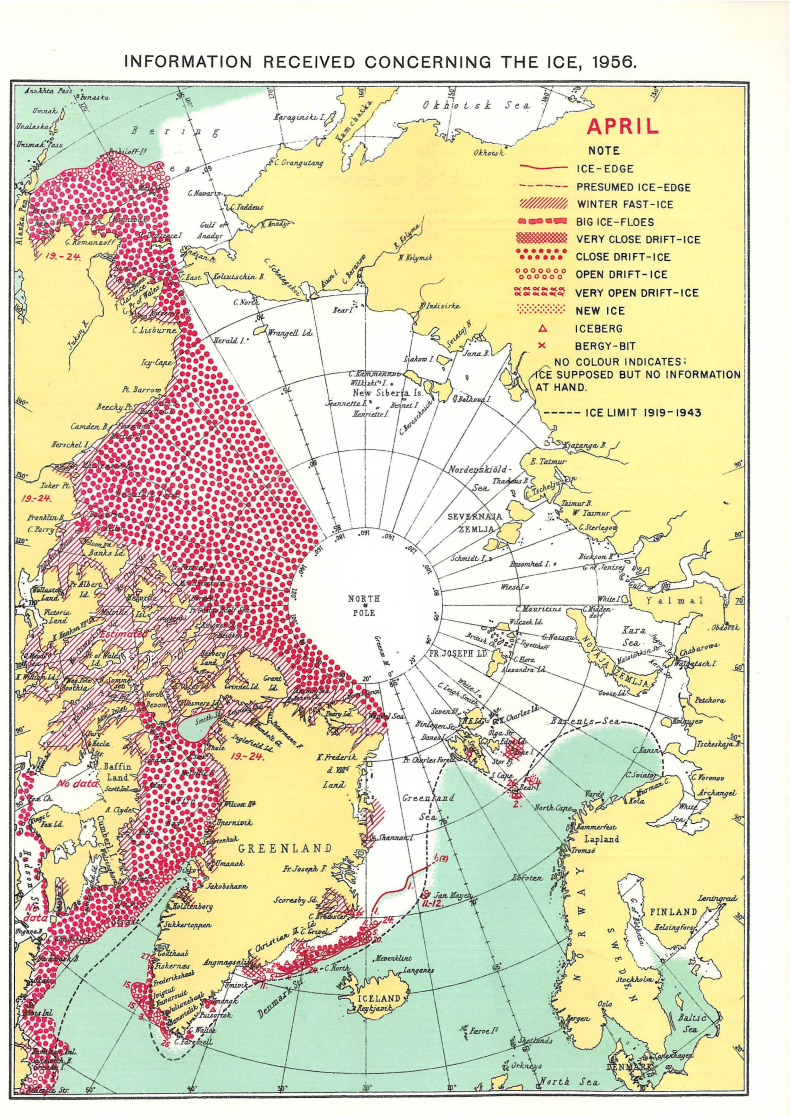
ORAS4 con un coeficiente de restauración superficial débil, (2) tomando las condiciones iniciales de hielo marino de la misma reconstrucción que en 1 y las condiciones iniciales oceánicas de ORAS4, y (3) lo mismo que en 1 pero usando un coeficiente de restauración más fuerte hacia ORAS4. El objetivo principal es evaluar el impacto de estas 3 metodologías en los errores de predicción y la capacidad predictiva del hielo marino. Comparados con simulaciones históricas no inicializadas, los tres sistemas de predicción muestran menores errores en el hielo marino durante los siete meses de predicción, lo que indica que el modelo no ha alcanzado aún su atractor. La estrategia de inicialización 2 provoca un choque inicial debido a que el océano inicial (ORAS4) está demasiado cálido en el Ártico para mantener las condiciones iniciales de hielo marino en la reconstrucción, aunque los efectos del choque desaparecen en menos de 20 días. Para la estrategia 3, la asimilación fuerte del océano hacia ORAS4 produce también un océano sensiblemente cálido, en este caso acompañado de condiciones también bajas de hielo marino. Las predicciones revierten rápidamente este estado de bajo hielo marino debido al efecto del error sistemático del modelo, que es de signo opuesto y domina el error de la predicción en menos de un mes. Para la predicción inicializada usando el método 1 no se observan choques, y converge más rápidamente hacia el error sistemático del modelo, ya que inicialmente está más cerca del atractor. Los errores de hielo marino de las estrategias 2 y 3 convergen en menos de un mes, lo que demuestra el papel de las condiciones iniciales del océano (que están muy cerca para las dos estrategias) en la evolución de los errores de predicción de hielo marino. El error más bajo de concentración de hielo marino en Diciembre-Enero-Febrero y Marzo-Abril-Mayo se encontró en el sector Atlántico para la predicción que utiliza la estrategia 3, mientras que en el sector del Pacífico para la estrategia 2. Con respecto a la capacidad de predicción del hielo marino, no aparecen diferencias significativas entre los tres sistemas de predicción. Sin embargo, los tres muestran una capacidad significativamente mayor que el experimento histórico para predecir la extensión del hielo marino panártico durante los primeros dos meses de predicción. Para las regiones individuales, el Mar de Labrador destaca como la única cuenca con un valor añadido significativo por la inicialización durante toda la predicción (7 meses), mientras que solo se detectan aproximadamente 3 meses de valor añadido en las cuencas periféricas.

Septentrionalium Terrarum, the first map of the Arctic. Version of 1623 of the illustration of Gerardus Mercator.

The ancient exploration of new lands share some similarities with the research that scientists carry out nowadays. Unlike scientists, the navigators' motivations to explore the Arctic were however to gain wealth and reputation. Sea ice observations exist since the man started to sail the Arctic margins and exploit its resources: from commercial whaling or walrus hunting in the 1600s to coal mining on Spitsbergen Island (1899) and the most recent oil drilling. Although the earliest data collected from ship logbooks and diaries correspond to 1750 (Divine and Dick, 2007), continuous records of the position of the Arctic sea ice margins have been gathered only since approximately 1850 (Walsh, 1978; Walsh et al., 2017). Records of sea ice edge, for example, have been produced interpreting sea ice charts from the Danish Meteorological Institute (Underhill et al., 2014). Sea ice information from 47 million years ago to the last millennium is mostly obtained through proxy records (Polyak et al., 2010), such as the sea salt in the ice cores (e.g. Grumet et al., 2001; Kinnard et al., 2006), the quartz content in the sediment (Eiríksson et al., 2000) or a combination of ice cores and tree ring records (Fauria et al., 2010). However, proxies tend to be insufficient and discontinuous in time and space. Historical records, such as the observations of drift ice by the Iceland inhabitants in the last 1200 years have been used to build a sea ice index (Ogilvie, 1996). Widespread continuous Arctic sea ice measurements only began in 1978 with the launching of the Scanning Multichannel Microwave Radiometer (SMMR) flown on Nimbus 7 (Cavalieri et al., 2003; Gloersen and Campbell, 1988). Since then, our understanding of the climate system has amazingly evolved (Yang et al., 2013). These observations have allowed to document a drastic shrinking of the Arctic sea ice cover.

The fast decline of the Arctic sea ice in the last few decades (Stroeve and Notz, 2018) has fostered human activity in the North Pole in sectors such as the maritime transport, the fishery, oil and mineral exploitation or the eco-tourism. Maritime cargo transportation companies are aware of the huge saving in both time and money that the Northwest Passage and the Northern Sea route represent as alternatives to the Panama Canal and the Suez Canal routes, correspondingly. Recent studies have projected an increase in the number of maritime routes and the ships going through them over the next decades in parallel with the significant summer sea ice reduction (Melia et al., 2016; Smith and Stephenson, 2013; Stephenson et al., 2011). Access to new sites will allow the extraction and transportation of oil and minerals (Monitoring et al., 2011), while climate change will certainly affect the distribution of fisheries (Lam et al., 2016). Arctic eco-tourism has received a lot of attention in the last 20 years due to the advances of transport means, but associated risks for these trips have also increased (Hall and Saarinen, 2010). The number of people going to the Arctic has steadily grown, and it is projected to continue increasing (Maher, 2016). Indigenous popu-

lations are also affected by the changes in sea ice cover, mainly in relation to transportation and shoreline exposure, the latter leading to hazards for the waterfront infrastructure (Meier et al., 2014). All these new activities in the Arctic will also cause new impacts on its climate (Huntington et al., 2007), such as the permafrost degradation or the change in drainage patterns (derived from new human constructions). These emerging socio-economic activities and the new public risks caused by the changing climate call for actionable and trustworthy climate predictions in the Arctic region, including reliable predictions of the Arctic sea ice edge.



Example of a sea ice chart compiled by the Danish Meteorological Institute. Chart is for April, 1956. Further information: https://nsidc.org/data/g10007.

1 Introduction

1.1 A changing Arctic climate

The global surface air temperature (SAT) increase of the last decades is mostly a response to the increasing atmospheric greenhouse gas (GHG) concentrations (e.g., Mitchell et al., 1995; Washington and Meehl, 1996). This temperature increase has been faster and more pronounced in the Arctic than in the rest of the planet (e.g., Bekryaev et al., 2010; Hinzman et al., 2013; Overland et al., 2004; Serreze and Francis, 2006). The ratio of Arctic warming to tropical warming is known as Arctic amplification (AA; Huang et al., 2017; Manabe and Stouffer, 1980; Pithan and Mauritsen, 2014; Serreze et al., 2009; Serreze and Barry, 2011). AA results mostly from two major contributions. First, the positive surface albedo feedback: when the sea ice and snow on top of it melt, the surface albedo decreases, which in turn increases the amount of energy absorbed by the surface, further enhancing sea ice melt (Hall, 2004; Kashiwase et al., 2017). The second contributor is the lapse-rate feedback, which is positive in the Arctic, and that is associated to an uneven warming of the lower and upper stratosphere under stable stratification conditions, which acts to reduce the outgoing longwave radiation and thus to enhance the surface warming (e.g., Goosse et al., 2018; Stuecker et al., 2018). These are, however not the only factors behind AA. Other important factors are the increase in downwelling longwave radiation following the increase in GHGs (Kapsch et al., 2016; Notz and Stroeve, 2016), the increase in cloud cover (Jun et al., 2016) or the changes in oceanic heat content (Carmack et al., 2015; Ivanov et al., 2016). Further factors are still debated (Stuecker et al., 2018).

The rapid Arctic warming has contributed to substantial melting of sea ice (Comiso et al., 2008; Stroeve et al., 2007; Stroeve and Notz, 2018; Stroeve et al., 2012; Vaughan et al., 2013). On September 18, 2019, the Arctic sea ice extent dropped to 4.15 million square kilometres (Fig. 1.1), the second lowest summer minimum in the satellite record since measurements

started in 1979. Likewise, on March 7, 2017, the Arctic sea ice extent dropped to 14.41 million square kilometres, a new record low winter maximum extent. These record lows follow a long list of record winter (2015, 2016) and summer (2007, 2012, 2016) Arctic sea ice lows. An increasing attention is being drawn towards the sea ice melting processes since the debate on the timing of an ice-free Arctic is still open (e.g., Jahn, 2018; Notz and Marotzke, 2012; Notz and Stroeve, 2018; Perovich et al., 2007; Screen and Deser, 2019; Sigmond et al., 2018; Tsamados et al., 2015). Given the record low sea ice conditions of the recent winters, the discussion on the reduced winter sea ice growth has also increased (e.g., Petty et al., 2018; Stroeve et al., 2018). Arctic sea ice loss is not exclusively linked to the effect of radiative forcing, with internal climate variability playing also a paramount role (Day et al., 2012; Zhang, 2015), accounting for 30 to 50% of the Arctic sea ice decline in the period 1979-2014 (Ding et al., 2017). Likewise, internal variability will be a key factor in the near-term future of the Arctic, influencing the timing of the first ice-free summer (Jahn, 2018).

Predicting this, and other local dramatic reductions in sea ice cover is of paramount importance not only for the region, but also beyond the Arctic, via several teleconnection mechanisms, which are climate variability links between two distant regions.

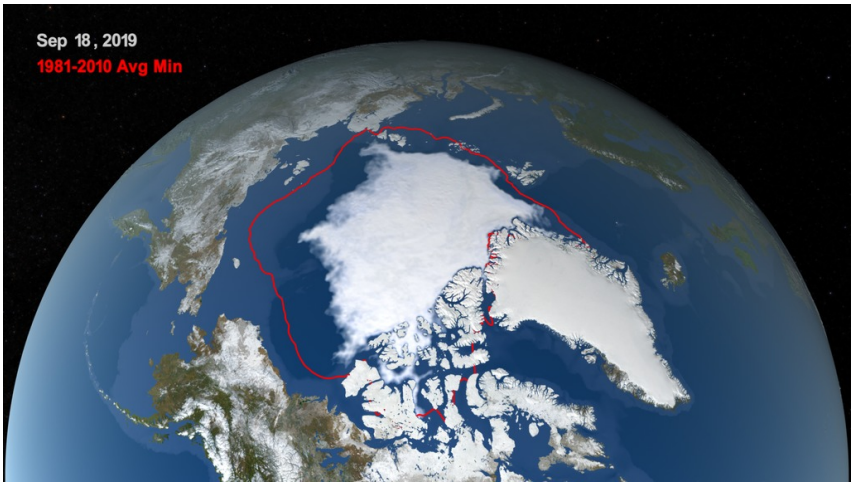


Figure 1.1: Arctic sea ice on September 18, 2019. Source: NASA (https://svs.gsfc.nasa.gov/4757).

1.2 Impact on the mid-latitude climate and weather

Sea ice is a major component of the climate system, which acts as an insulating layer, damping heat, mass and momentum exchanges between atmosphere and ocean, and affecting substantially the Earth's albedo and therefore the amount of solar radiation absorbed by the

Earth. It also affects the Atlantic thermohaline circulation/Atlantic meridional overturning circulation (AMOC; e.g. Mikolajewicz et al., 2005) through the effect of buoyancy fluxes on the North Atlantic deep-water-formation rate. Sea ice is, on the one hand, sensitive to small external perturbations (e.g. changes in surface temperatures or radiative forcing) but also, on the other hand, associated with a series of important feedbacks (e.g., Ebert and Curry, 1993; Goosse et al., 2018), such as the aforementioned positive ice-albedo feedback, which acts to amplify initial perturbations of the sea ice conditions (Hansen et al., 1984) or the negative ice-growth thickness feedback (Bitz and Roe, 2004). Therefore, sea ice acts not only as an early indicator of climate change but also as an amplifier of climatic perturbations (e.g., Serreze and Barry, 2011; Vihma, 2014).

At seasonal-to-interannual timescales, observations, reanalysis and model experiments have shown that sea ice may influence the climate of mid and high latitude regions (e.g., Deser et al., 2010; Francis and Vavrus, 2012; Liu et al., 2012; Yang and Christensen, 2012). The simplest case of a remote effect (or linkage) would be when positive temperature anomalies generated in the Arctic due to sea ice melting are propagated horizontally by the atmospheric circulation (Serreze et al., 2009). Other proposed remote impacts are mediated via the stratosphere, through an effect on planetary-scale waves (Nakamura et al., 2016), or even by causing a reorganization in the tropical convection that triggers an anticyclonic response over the North Pacific (Cvijanovic et al., 2017). It is important to note that the underlying mechanisms are not purely linear, and the processes involved do not necessarily follow a cause-and-effect chain (Overland et al., 2016). Future changes in the mid-latitude climate and weather could also be linked to the tropics (e.g. Peings et al., 2018; Zappa and Shepherd, 2017).

The most studied linkages/effects associated to the Arctic sea ice decline are: (1) a potential increase in snowfall over Canada, Alaska and Siberia (Cohen et al., 2012; Deser et al., 2010; Liu et al., 2012); (2) a weakening of the mid-latitude westerlies (Outten and Esau, 2012) (which help maintaining a warm European climate by transporting heat from the Atlantic); (3) a potential increase in winter cold spells (Bellprat et al., 2016; Cohen et al., 2012; Honda et al., 2009; Liu et al., 2012; Petoukhov and Semenov, 2010; Yang and Christensen, 2012); (4) a potential cooling of the polar vortex (Kim et al., 2014); (5) a poleward shift and strengthening of the subtropical jet (Fu and Lin, 2011); (6) a decline of the AMOC (Sévellec et al., 2017), which might lead to a generalized decrease in temperature over Europe (Jackson et al., 2015). Fig. 1.2 illustrates some of the most important potential responses to sea ice loss. Although most of the recent studies have focused on winter and autumn sea ice reduction impact, this affects all seasons, including summer. For example,

Arctic sea ice changes might drive European summer precipitation (Screen, 2013). Although not necessarily as a consequence of the Arctic sea ice decline, AA has been seen to influence substantially the mid-latitude summer circulation by weakening the storm tracks (Coumou et al., 2018). Polar to mid-latitudes linkages have been proven to be useful even to understand the origin of some extreme climate events (e.g. Acosta Navarro et al., 2019; Cohen et al., 2014). The potential increase in frequency of these extreme mid-latitude climate and weather events might affect millions of people in coming decades.

The atmospheric response to sea ice changes strongly depends on the background state of the atmosphere (Balmaseda et al., 2010). However, the observational record is relatively short and the atmospheric noise relatively large to be able to detect such an impact of sea ice. Besides that, the remote responses to Arctic sea ice decline are uncertain and hard to be confirmed (Screen, 2013), with a lack of consistency between modelling studies (Screen et al., 2018). Simulating faithfully the sea ice processes realistically is the first step to guarantee that the teleconnection mechanisms represented in the models are also realistic.

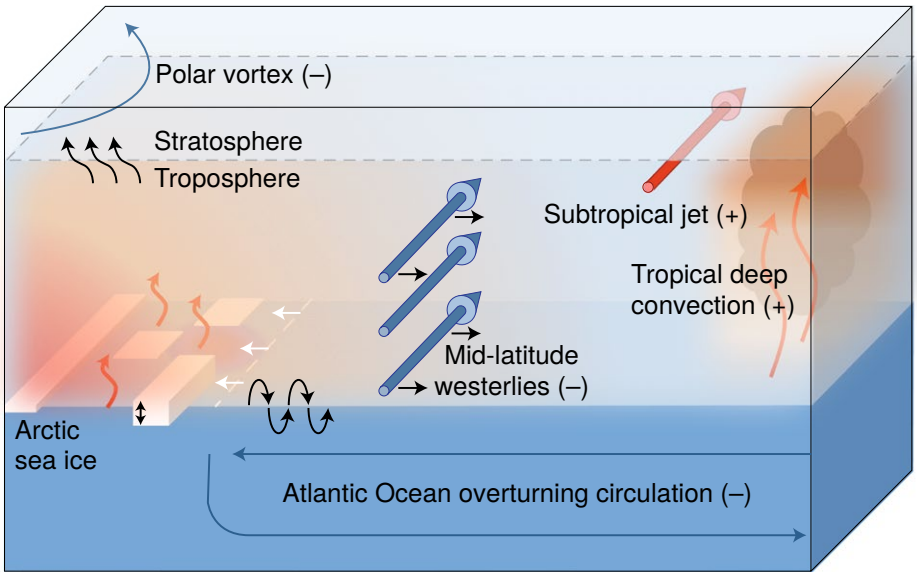


Figure 1.2: Schematic representation of the potential climate response to Arctic sea-ice loss. An illustrative cross-section from the North Pole to the Equator. Major atmospheric and oceanic circulation features that are weakened by Arctic sea-ice loss are shown by blue arrows and labelled with minus signs, and those that are strengthened by Arctic sea-ice loss are shown by red arrows and labelled with plus signs. Red/orange shading indicates regions of greatest warming in response to sea-ice loss. Source: Screen et al. (2018).

1.3 Sea ice processes

The processes acting on sea ice can be divided into two categories: thermodynamical processes, which involve the transfer of heat and direct effects from the radiative fluxes, and dynamical processes, which can move and deform the sea ice.

Within the first category, the thermodynamical processes, the most important is the reflection of incoming solar radiation, which is controlled by the sea ice albedo. Sea ice not only has an albedo (or reflective capacity) that is much higher than for open water, it also provides a surface on which snow can accumulate, which in most cases further increases the albedo and therefore the amount of reflected solar radiation toward the atmosphere. During the melting season, melt ponds appear, which are pools of water that form on the sea ice surface. As melting progresses, the ponds deepen and extend and much of the thin ice melts completely, exposing a bigger area of the dark ocean and thus reducing effectively the total sea ice albedo. This is additionally relevant because of the heat capacity of the ocean is much larger than that of the sea ice, and can therefore potentially absorb larger amounts of energy. A substantial part of the ocean-atmosphere heat exchanges include latent heat fluxes from evaporation, whereas latent heat fluxes from ice sublimation are negligible in the ice-atmosphere heat exchanges. Finally, among the sea ice thermodynamical processes we also have to account for the conduction of heat within the sea ice layer.

Salinity is a thermodynamic state variable. Together with temperature, it controls the relative brine volume encased in the ice, which in turn affects all ice properties (thermal, mechanical and radiative), thus influencing, e.g., the seasonal cycle of ice thickness (Vancoppenolle et al., 2009), by affecting the growth and melt rates, e.g. a vertically-varying salinity profile has thinner ice compared to an isosaline one (Vancoppenolle et al., 2006).

The dynamic processes are governed by different factors, like the momentum, whose equations determine the ice motion; the rheology, which refers to the relationship between sea ice stress (caused by floe–floe or floe–lead interactions) with the resulting large-scale deformation of the ice cover; and the mechanical redistribution, arising from ridging and rafting of sea ice when densely packed ice converges (Hunke et al., 2010).

A good representation of all of these processes in climate models becomes essential to properly simulate the real climate. And what is more important in our context, it provides the key to obtain certain predictability mechanisms.

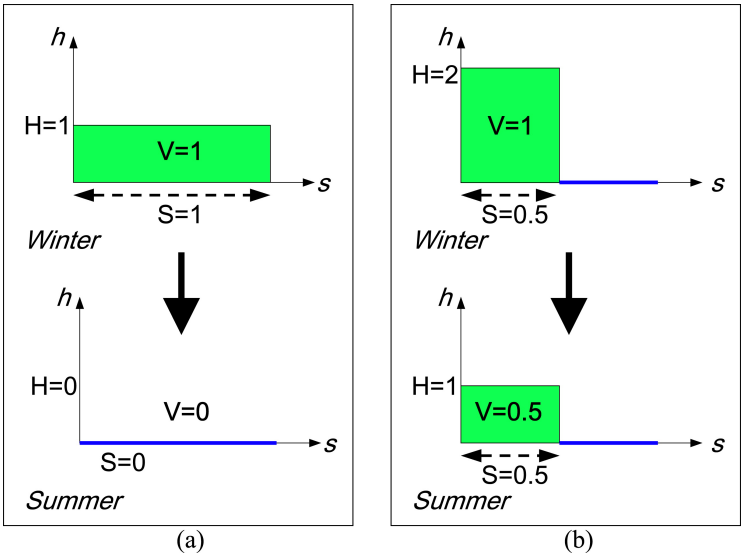


Figure 1.3: Idealized examples to illustrate the role played by the winter ice thickness distribution vs winter ice volume in the determination of the summer ice area. Each graph presents a grid cell of surface 1 m^2, covered by ice over a surface S (m^2, x axis), with thickness H (in m, y axis). Open water surface is $1 - S$, and the grid cell ice volume is $V = S \times H$ (m^3). Each panel exhibits (top) initial (winter) ice coverage of the grid cell and (bottom) the ice cover after a vertical melting of 1 m. Source: Chevallier and Salas y Mélia (2012).

1.4 Arctic sea ice predictability mechanisms

There exist several potential sources of Arctic sea ice predictability on seasonal to decadal time-scales:

1. The most important is *persistence*. The persistence, or capacity of a given variable or property to remain in (or close to) a given state, can be measured by the time-scale until which its autocorrelation stays significant. It is usually interpreted as the memory of that variable. The characteristic persistence of sea ice depends on the variable considered. For the sea ice concentration (SIC) it is of 3-7 weeks at the local scale while it ranges from a few months to a few years for the local sea ice thickness (SIT); for the sea ice area (SIA) the longest time scales, from 2 to 5 months, occur in the central Arctic, with even longer values of 4 to 10 months being associated to the central Arctic sea ice volume (SIV; Guemas et al., 2014b). An important source of summer sea ice predictability is what Chevallier and Salas y Mélia (2012) named *memory regime*. Thanks to the memory regime, the September SIA is potentially predictable up to 6 months in advance, using as a predictor the area covered by thick ice. This is better a predictor than simple persistence, for instance. Indeed, only the thickest ice present in May survives the melt season. Fig. 1.3 illustrates sea ice covers

with similar volumes but different thickness which are subjected to the same forcing and experience very different changes. More information on the potential predictability of these idealized sea ice covers is contained in their thickness rather than in their total SIV or total SIA.

2. *Reemergence* is a phenomenon in which anomalies of a given variable during one season (or month) are positively correlated with anomalies of the same variable during another season (or month), despite a loss of correlation in-between. Blanchard-Wrigglesworth et al. (2011a) highlighted a melt-to-freeze season reemergence which occurs between pairs of months from different seasons but where the ice edge is in a similar position (e.g., May and December; Fig. 1.4) and is due to persistence of Sea Surface Temperature (SST) anomalies. When the sea ice edge returns to the same vicinity as the preceding spring, the SST anomaly "inherited" from the original sea ice area anomaly has an influence on the rate at which sea ice forms again in that vicinity. Note that this mechanism could be taken as a predictability mechanism coming from ocean, but we have preferred to differentiate. They also highlighted a freeze-to-melt reemergence mechanism which also occurs between pairs of months where the sea ice edge has close positions and is due to the persistence of sea ice thickness anomalies. In this mechanism, a positive (negative) SIA anomaly in the growth season is associated with an early (late) date of freeze up, locally creating a positive (negative) SIT anomaly at the sea ice edge that slows down (accelerates) the sea ice retreat during the next spring when it reaches the same vicinity and is therefore associated with a local positive (negative) SIA anomaly. Day et al. (2014) confirmed this mechanism in the HadGEM1.2 model.

3. *Advection* of sea-ice anomalies by the mean Arctic circulation can also provide additional predictability over simple persistence. The mean Arctic sea-ice circulation is characterized by an anticyclonic gyre in the Beaufort Sea and a cyclonic gyre of smaller extent in the Laptev Sea. At their frontiers appears the Transpolar Drift Stream (Sokolov, 1962). Gudkovich (1961a,b) described two types of circulation. In the anticyclonic regime, the area of the Beaufort gyre increases and the area of the Laptev gyre shrinks; the Transpolar Drift Stream originates from the Laptev, East Siberian and Chukchi Seas and transports ice toward the Greenland Sea. In the cyclonic regime, a contraction of the Beaufort gyre occurs simultaneously with an expansion of the Laptev gyre; the Transpolar Drift Stream slows down and its entrance shifts toward the Beaufort Sea. The cyclonic regime is associated with a larger sea-ice export through Fram Strait because the exported ice is thicker than in the anticyclonic regime (Polyakov

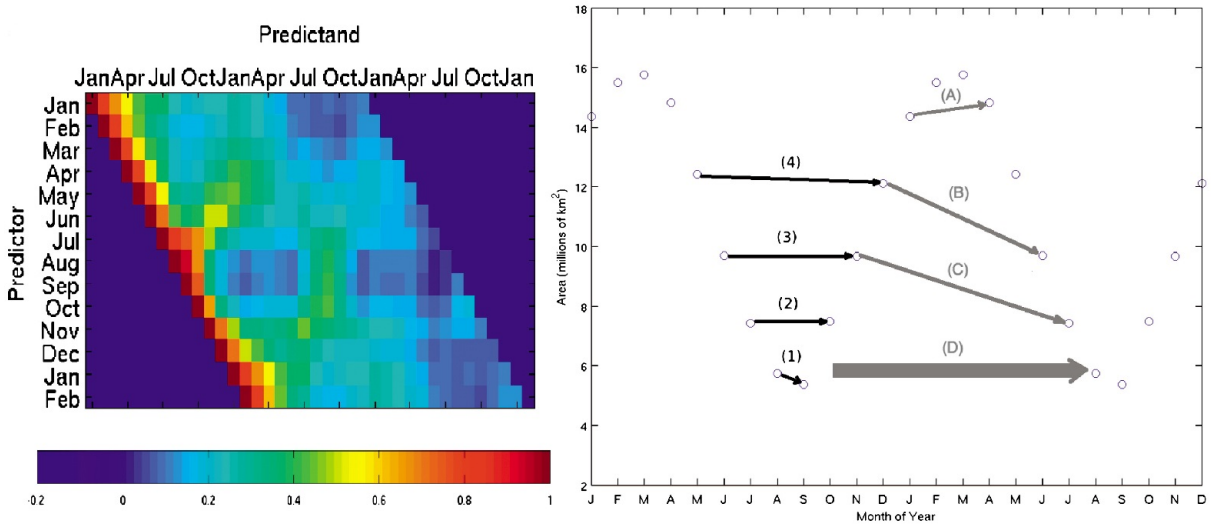


Figure 1.4: Left panel: Correlations values for all months and all lags for the sea ice area of CCSM3 model. Months along the x axis indicate the month whose anomaly (predictand) is correlated with the month's anomaly along the y axis (predictor), from lags of 0 months (thus values of 1 along the diagonal) to 23 months. January and February data have been duplicated for ease of following structures through winter. **Right panel**: Mean annual cycle of sea ice area in millions of square kilometres in the CCSM ensemble run for 2001–30. The arrows are used to illustrate pairs of months that show high correlation values in their sea ice area anomalies. The black arrows show pairs of months during the summer limb of memory seen in left panel: (1) August–September, (2) July–October, (3) June–November, and (4) May–December. The gray arrows represent pairs of months with memory reemergence in the winter limb: (A), January with April; (B), December with June; (C), November with July; and (D) late summer (July–October) with the following September. Source: Blanchard-Wrigglesworth et al. (2011a).

et al., 1999). Karklin (1977) proposed a periodicity of 6–7 years in the ice drift and the circulation regimes, which could then provide predictability in the sea-ice conditions on decadal time-scales.

4. Sea ice predictability can also arise from the *ocean*, which has a long memory, involving seasonal, interannual, decadal and multidecadal time scales, which makes it an important source of climate variability (e.g. (Boer, 2004)). Heat advection along the North Atlantic Current strongly influences the Arctic sea ice conditions, in particular in the Atlantic sector, with a delay of a few years (e.g., Furevik, 2001; Vinje, 2001). Another example would be the warm water inflow through the Bering Strait (Fig. 1.5), which, e.g., accounted for about one-third of the sea-ice melt in 2007 (Woodgate et al., 2010).

5. The *atmosphere* can also exert an important influence on Arctic sea ice variability. However, because of its low predictive capacity beyond 1 or 2 weeks, its potential as a source of sea ice predictability is usually constrained to that short horizon (Guemas

et al., 2014b).

All these mechanisms, when well reproduced in models, can provide predictability. Capabilities of current dynamical prediction systems will be discussed in the following section.

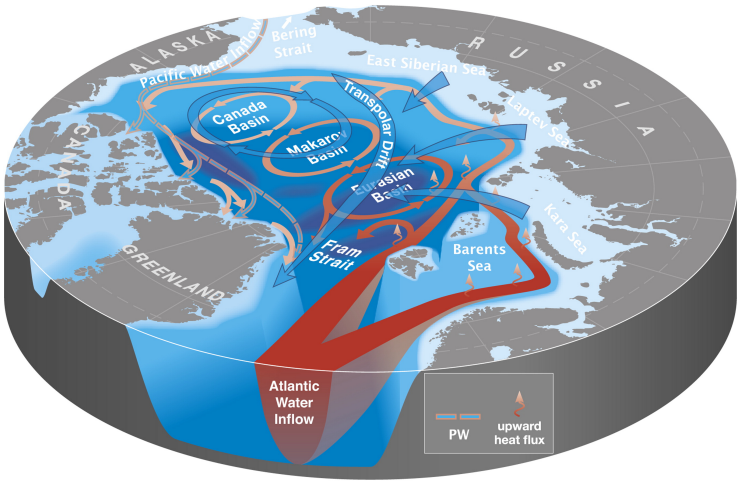


Figure 1.5: Circulation of the surface water (blue), intermediate Pacific Water (blue arrow surrounded by pink), and Atlantic Water (red) of the Arctic Ocean. Source: Carmack et al. (2015).

1.5 State-of-the-art of climate prediction

Due to the chaotic nature of the atmosphere, it is not possible to predict the weather at a set location more than 2 weeks ahead. But useful information can be achieved beyond that horizon. As opposed to weather forecasts, climate predictions inform us about the likelihood of shifts from the normal climatic conditions of a certain region for periods ranging between a month and several years (e.g., Doblas-Reyes et al., 2013). The large-scale circulation modes such as the El Niño–Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO) are important sources of seasonal forecast skill in the mid-latitudes (Doblas-Reyes et al., 2003; Jin et al., 2008; Scaife et al., 2014; Wu et al., 2009), while the ocean variability, in particular in the North Atlantic is a major source for multi-year predictions (also referred to as decadal predictions (García-Serrano et al., 2015; Meehl et al., 2014; Yeager et al., 2018)). However, the skill achieved in these prediction systems over the continents is still rather modest. For example, skill arising from the ENSO region barely translates into predictive capacity over other mid-latitude regions, although it shows some potential for predicting specific events in certain regions, like the occurrence of hot days over Eastern North America (McKinnon et al., 2016). These limitations might reflect model difficulties in representing the teleconnection

mechanisms that enable the impacts over land, which some studies suggest that may have weakened over the past 30 years (e.g., Kam et al., 2014) and will continue doing so in a warmer future climate (Meehl et al., 2006). Other studies point to an absence of stationarity in the ENSO response on European and Mediterranean rainfall (Ineson and Scaife, 2009; Knippertz et al., 2003), something that might also affect other modes of variability and regions, including the Arctic, which is changing substantially since the early 2000s (Meier et al., 2014).

Brief history of seasonal forecasting of sea ice Seasonal predictions as we know them nowadays, which focus on time horizons from 1 month to slightly over than 1 year, started in the 70s, with scientists aiming at predicting the ENSO variability (Quinn, 1974; Wyrtki et al., 1976). In the 80s, the first statistical models were applied to simulate specifically the sea ice and to investigate its predictability (e.g. Johnson et al., 1985; Lemke et al., 1980; Walsh, 1980). Sea ice modelling improved significantly between 1979 and the early 90s (e.g. Barry et al., 1993; Fleming and Semtner Jr, 1991; Hakkinen et al., 1992; Hibler III, 1979, 1980; Hibler III and Bryan, 1987; Walsh et al., 1985), which fructified in 1987 in the first sea ice predictions performed by a non-statistical model (Polar Ice Prediction System, Preller and Posey, 1989). These predictions were related to operational activities, with clear economic purposes, and aligned with the latest computational advances at the time. Models continued improving, but it was not until the beginning of the 2000s that the attention on seasonal Arctic sea ice predictions increased again. The availability of two decades of good quality satellite observations, together with the development of the first global circulation models, allowed scientists to identify a pronounced decline in Arctic sea ice, and to associate it to the accelerated increase in GHG concentrations, respectively. The study of observations also allowed to investigate the link between atmospheric circulation structures and sea ice variability, and to identify significant correlations between the Arctic Oscillation and the summer Arctic sea ice (Rigor et al., 2002), or the NAO and the Barents sea ice extent in winter (Sorteberg and Kvingedal, 2006). However, the fast changes in the Arctic climate became a limitation for the statistical assumptions required in the observational studies, almost definitely giving way to the use of dynamic models (Holland and Stroeve, 2011). The first seasonal predictions produced with physical models were made with ocean-sea ice models forced by atmospheric reanalyses (e.g. Kauker et al., 2009; Zhang et al., 2008). In response to the Arctic SIE record minimum of 2007, the Sea Ice Outlook (SIO) was established in 2008 inside the program Study of Environmental Arctic Change (SEARCH). This initiative collects and compares seasonal predictions of September Arctic SIE using a variety of modelling, statistical and heuristic approaches (Stroeve et al., 2014). Currently, the

production of trustworthy operational climate predictions relies on demonstrating the ability to forecast the past. These retrospective predictions, generally referred to as *hindcasts*, are evaluated against observations to assess the performance of the prediction systems, and have shown encouraging improvements in the recent years. And further improvements can be expected. The latest initialization techniques, including data assimilation of different ocean and sea ice observations (e.g. Blockley and Peterson, 2018; Massonnet et al., 2015), together with the increasing efforts in model intercomparison (e.g. ENSEMBLES[1], SIPN2[2]), augur a promising future in the seasonal forecasting of Arctic sea ice.

1.5.1 Overview of initialization methods

Seasonal prediction skill largely relies on a good initialization of the forecast systems. The increasing density and quality of observations for the various components of the climate system has allowed for more accurate initial conditions (ICs) for seasonal predictions. The recent improvements in predictive capacity of the Arctic sea ice are also related to advances in forecast initialization techniques to phase the models with the observed climate evolution (e.g. Bushuk et al., 2017; Collow et al., 2015; Wang et al., 2013). In particular, efforts have been made to improve the initialization of sea ice (e.g., Blanchard-Wrigglesworth et al., 2011b; Blockley and Peterson, 2018; Dirkson et al., 2017), ocean (in particular the ocean's top thermal structure; e.g., Balmaseda and Anderson, 2009; Balmaseda et al., 2009), atmosphere (Infanti and Kirtman, 2016) and land surface (Koster et al., 2010). The classic approach is to use full-field initialization, in which the model begins from an actual observed or reanalyzed state. This method, however, is affected by the presence of important model biases which cause a drift in the predictions as the model transitions from observed climate towards its own attractor (e.g. Magnusson et al., 2013; Meehl et al., 2014; Sanchez-Gomez et al., 2016). This disagreement between model and "observed" climatology can produce initialization shocks (Balmaseda and Anderson, 2009), and thus introduce large errors in the prediction at the short and even the long lead times (from days to years) that need to be corrected a posteriori under certain assumptions (e.g. that the model drift is independent of the start date; ICPO, 2011). To minimize this drift, another method usually employed is anomaly initialization, in which the model is started from a synthetic state in which the observed anomalies are added on top of the the mean model climate. Some studies suggest that anomaly initialization outperforms the full-field initialization (Kröger et al., 2017), in particular over the Arctic (Volpi et al., 2016). Others disagree and state that anomaly initialization methods might

[1] http://ensembles-eu.metoffice.com/
[2] https://atmos.uw.edu/sipn/

not be the best options to reduce the initial shock (He et al., 2017) and that full-field initialization provides more skilful predictions at seasonal timescales (Smith et al., 2013). The advantages of each approach still remain unclear, and are a current matter of debate. Regardless of the choice of full field or anomaly initialization, the particular method used to produce the ICs can also have an impact in the development of the drift. This should be higher (in particular for full field initialization) when initializing directly from interpolated products from operational reanalyses, like those from the European Center for Medium-Range Weather Forecasts (ECMWF), as the forecasts start from a climate state completely independent from the model attractor.

A large variety of alternative techniques has been developed to assimilate observations and/or reanalyzed products to produce in-house reconstructions (typically for the sea ice, since the only publicly distributed sea ice reanalysis is PIOMAS) with the same models (or components of the models) used to perform the forecasts. These go from simple approaches, like Newtonian relaxation or nudging (Lindsay and Zhang, 2006; Tietsche et al., 2013), to more sophisticated methods like the Ensemble Optimal Interpolation (EnOI, e.g. Dulière and Fichefet, 2007; Stark et al., 2008) or the Ensemble Kalman Filter (EnKF, e.g. Evensen, 2003; Massonnet et al., 2013). One of the key advantages of the latter is its multivariate nature, as it is designed to update consistently all ocean/atmosphere/sea ice variables after every forecast step (as long as they are related to the assimilated variable) using model covariances. For example, EnKF assimilation of SIC has been proven to have a positive impact in the representation of sea ice thickness (Massonnet et al., 2013; Mathiot et al., 2012). This technique is appropriate for nonlinear models (like the sea ice ones) and to minimize the expected forecast drifts, although it is computationally more expensive than, e.g., the EnOI (Blyverket et al., 2019) and the nudging. In Chapter 5, three different full-field initialization approaches and their impact on the sea ice biases and skill will be investigated.

1.5.2 Establishing the limits in seasonal predictability

As discussed above, model initialization has still room for improvement. On top of that, it is hard to determine whether the imperfect forecast skill obtained in most of studies derives from the quality of initial conditions, the overall model performance, if there are simply certain events which are themselves unpredictable because of the climate chaos or a mixture of all of these factors. These uncertain questions can be potentially answered thanks to perfect model predictions, in which the model is used to predict itself, instead of predicting the real world. For this, model simulations are used as a surrogate for the real climate. Ensemble predictions are typically initialized from a control run by introducing

small perturbations. Potential predictability is a measure of the lack of amplification of those perturbations, i.e. the fraction of the signal which is inherently predictable in spite of the climate noise. Such experiments using state-of-the-art models also provide an indication of the level of skill that could be achieved in real predictions if perfect initial conditions were available, if all processes were perfectly represented by the models and no initialization shocks occurred. As such, predictability estimated from perfect model approaches tends to be larger than the one obtained in realistic retrospective predictions (e.g. Bushuk et al., 2019). Eade et al. (2014) even suggest that current potential predictability studies may be underestimating the true predictability for variables such the near-surface temperature, sea level pressure and precipitation in seasonal to decadal predictions. However, Kumar et al. (2014) argue that there is not necessarily a relationship between potential predictability and actual predictability due to model biases (in which the potential predictability would always represent the true upper limit of prediction skill). The initial idea triggering the use of perfect model experiments was to estimate the inherent predictability of climate phenomena, but this information cannot exactly be obtained: the predictability that is estimated through a model is necessarily affected by model errors. Some climate phenomena could be more predictable that what perfect model experiments tend to show (e.g. Scaife and Smith, 2018), but our models are not good enough.

1.5.3 Ensemble size: a key step for more reliable climate forecasts

One key factor to better constrain forecast skill is the generation of ensemble predictions. An ensemble prediction is a set of predictions that present the range of future climate possibilities. Each single prediction is called "member", and each member starts from a slightly different set of initial conditions. Ideally, the difference in the initial conditions is representative of the observational uncertainties, which are usually included as stochastic perturbations. Most studies agree that increasing the number of ensemble members reduce the forecast uncertainties and the excessive noise (Atger, 1999; Eade et al., 2014; Murphy et al., 2004), avoiding thus to infer misleading conclusions (e.g., Ménégoz et al., 2018). This is because for each single member only the predictable part of the signal is phased, while the rest of the climate signal behaves like chaotic noise. Then, when averaging the different members, this noise is removed and only the predictable signal emerges. The larger the number of members, the more efficient the averaging process is. The limited computational resources (e.g. Leutbecher, 2018; Palmer, 2014; Palmer et al., 2005; Tintó-Prims et al., 2018) constrains the number of ensemble members and start years to be run though. Something to take also into account when designing a forecast is that, although increasing the ensemble size normally

improves the forecast skill, the skill gain is smaller for large ensemble sizes, as it depends on the inherent predictability of the given variable (Chen et al., 2013). For these two reasons, Hawkins et al. (2016) discussed the benefits of increasing the ensemble size: for both the SIE and SIV, the ACC and RMSE saturate with around eight ensemble members (with the same number of start dates; Fig. 1.6), meaning that the improvement in skill and decrease in error are negligible after reaching a certain ensemble size.

However, the various ensemble prediction techniques do not take into account the systematic errors, that are largely model-dependent, either for the internal variability or for the mean state. Consequently, ensemble forecasts are usually over-confident (Slingo and Palmer, 2011). Multimodel predictions tend to provide a more adequate approach to deal with these uncertainties and are able to outperform the best single model (Dirkson et al., 2019; Hagedorn et al., 2005; Weigel et al., 2008). Note that some processes are modeled similarly in all models though, which sometimes are also initialized from similar initial conditions to forecast the same event.

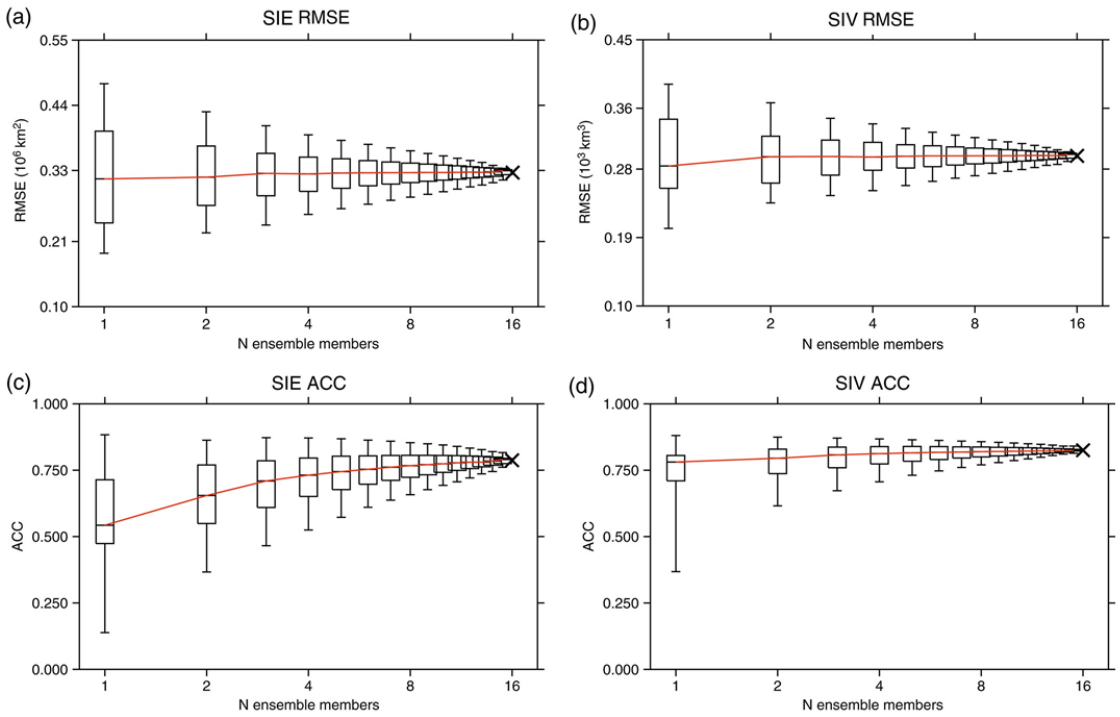


Figure 1.6: Box and whisker plots showing quantiles (5, 25, 50, 75 and 95%) of RMSE in Arctic (a) sea ice extent and (b) volume for September, for all possible choices of N (out of 16) ensemble members. (c) Same as (a,b) for the anomaly correlation coefficient of Arctic (c) sea ice extent and (d) volume for May as start month (September is at a lead time of 5 months). Source: Hawkins et al. (2016).

1.6 Climate Prediction Shortcomings

1.6.1 Sea ice models and their limitations

Climate models, and their individual components, are limited to represent the reality at certain scales due to their too coarse resolution or their inability to reproduce important phenomena (i.e. because certain processes are not included in the model). For example, there are some regions as the Odden ice tongue in the Greenland Sea that are misrepresented by NEMO-LIM3 (the main sea ice model used in this thesis) because none of the hypothesized mechanisms for ice formation (e.g., cold air outbursts, instabilities of the East Greenland Current) are represented in enough detail (Vancoppenolle et al., 2009). Another good example would be the Northern Baffin and Hudson Bays, where there are numerous mesoscale features, as the sea breeze, or some important processes (e.g., tides or river runoff) which cannot be resolved or are misrepresented by NEMO-LIM3 due as well to its coarse resolution (Vancoppenolle et al., 2009). Recently, Docquier et al. (2019) showed how increasing the ocean resolution can improve the location of the Arctic sea ice edge. Regarding the biases, Prodhomme et al. (2016) showed, however, that resolution does not necessarily have a noticeable effect on the long-standing Arctic sea ice biases. The representation of physical processes and characteristics like the lateral sea ice melting, the ice-atmosphere interactions, the rheology or the melt ponds are in the spotlight of climate scientists and sea ice modellers (Hunke et al., 2013, 2010; Notz, 2012), and improvements due to increasing resolution are becoming real (Docquier et al., 2019; Jung et al., 2012).

Currently, typical sea ice models do not have enough resolution to represent the complexity and the various types of ice which are present within a grid cell. The ice thickness distribution (ITD) is a probability density function [usually written $g(h)$] that describes the probability that the ice cover over some region R has thickness h (Bitz et al., 2001). Sea ice models nowadays rely on a discretized form of $g(h)$, i.e. a small number of ice categories are defined according to their ice thickness. The processes controlling $g(h)$ are ice growth, melt, divergence, ridging and advection, but only ice growth, melt and ridging can lead to a transfer of ice between thickness categories (Bitz et al., 2001). Those sea ice models which use an ITD solve thermodynamic and dynamic equations for each sea ice category. Bitz et al. (2001) demonstrated that thin ice grows quicker than thick ice in fall. They found that in order to capture the effect of an ITD on the evolution of SIV, between five and ten sea-ice thickness categories are needed.

Salinity is a crucial variable for the determination of the total heat fluxes at the ice–ocean interface (Vancoppenolle et al., 2005, 2006). There are different ways for the sea ice models

to represent the salinity: some of them, as the version 1 of Louvain-la-Neuve Sea Ice Model (LIM1; LIM is embedded in the ocean modelling system NEMO) (Vancoppenolle et al., 2010), consider the ice salinity as a constant, while others, such as version 2 (LIM2) (Bouillon et al., 2009), include steady and simplified sea ice salinity profiles, only used while computing the salt/freshwater flux at the ice bottom interface (Fichefet and Maqueda, 1997a), or use a vertical linear profile evolving throughout the time, such as LIM3 (Vancoppenolle et al., 2009).

1.6.2 Impact of the limited observational coverage

Most of our understanding of Arctic sea ice comes from satellite measurements using passive microwave sensors which, since October 1978 (Cavalieri et al., 1996; Rayner et al., 2003), monitor the SIC and the SIE. Before the advent of satellite imagery, only in-situ observations and aerial snapshots were available, but these do not allow for a proper monitoring of the continuous evolution of the pan-Arctic and regional sea ice cover. On top of that, daily SIC measurements took ten more years to arrive (in 1988). SIT, which has more potential for seasonal prediction (Blockley and Peterson, 2018), is even less well sampled, with gridded observations only available for some specific periods of the year since 2010. For this reason, model-assimilated products or reanalysis of SIT have been traditionally used for climate studies covering the SIC observational period (e.g. Labe et al., 2018). Luckily, in the past few years, many efforts have been devoted to enhance the observational network over the poles (Jung et al., 2016).

The problem of shortage of satellite data adds to others such as the short operational life span of individual satellites and their instruments (typically below a decade). The inhomogeneity of satellite products (which need to be combined to cover longer periods) is a problem that also affects ocean measurements (e.g., the Argo data only start in the early 2000s, Riser et al., 2016).

These limitations are important for climate predictions because satellite observations serve for three main purposes: (1) they expand our knowledge of the climate system and its processes (e.g. Overpeck et al., 2011); (2) they allow us to validate our retrospective predictions (e.g. Onarheim et al., 2015), improving thus our future predictive capacity. However, the evaluation of skill can be affected by the uncertainty of the observational products (Bunzel et al., 2016; Massonnet et al., 2015) against which the forecasts are evaluated, which in the particular case of sea ice can be related, among different factors, to the specific satellite characteristics (resolution, spectral bands, precision) or to the calibration/retrieval algorithms (Fig. 1.7; Ivanova et al., 2014). And (3) they are a key source of information to

be assimilated into climate models (e.g. Ghent et al., 2011) for forecast initialization. Thus, satellite data and climate models stand as complementary tools.

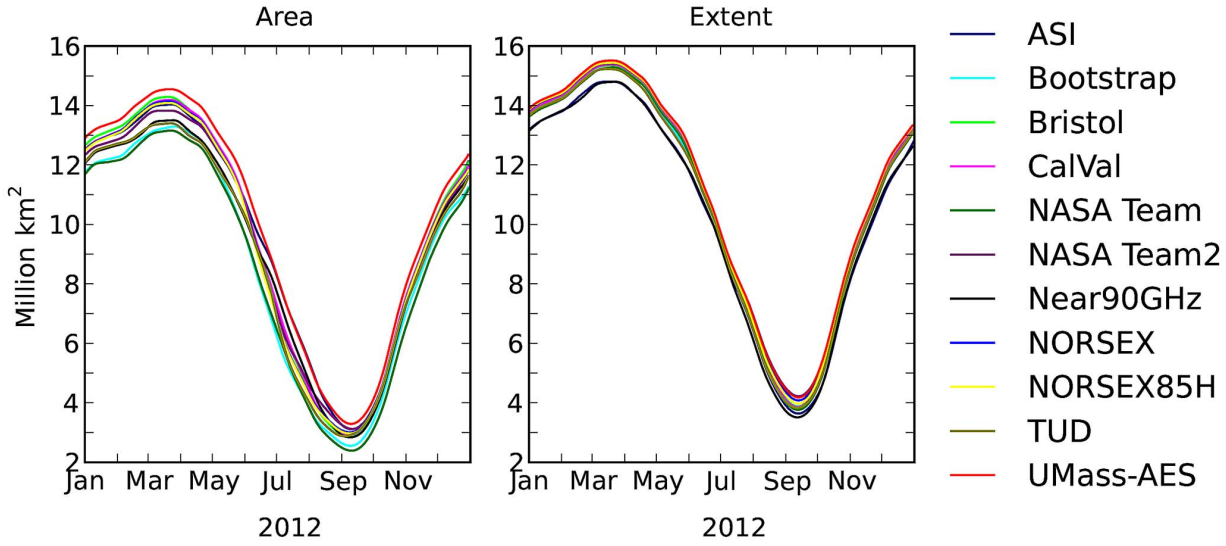


Figure 1.7: Daily Arctic sea-ice area (left) and extent (right), 2012, obtained by the 11 algorithms. Source: Ivanova et al. (2014).

1.7 Questions with incomplete answers

This introduction has given a broad overview of the physical basis and the current status of research in seasonal Arctic sea ice prediction. From the questions that are still under debate or unanswered in the sea ice forecasting community, those of particular interest for this PhD thesis are:

- Are the pan-Arctic sea ice predictability mechanisms applicable on a regional perspective? Is the representation of these mechanisms model-dependent? Which are the factors that can lead to different predictability in different models?

- How does the uncertainty in climate satellite observations affect initial conditions and the quality of Arctic sea ice predictions? Is their effect comparable to the one of initialization shocks? What is the nature of these shocks? Which are the regions in which they manifest more intensely? And for how long their effect persist?

- What strategies are most efficient in minimizing initialization shocks? To what extent these reduced initialization shocks improve the forecast sea ice biases? And do they also lead to improved forecast skill?

Please note that this is not an exhaustive list of all the open questions regarding Arctic sea ice prediction, but gathers the main ones that this thesis aims to answer, at least partially.

1.8 Objectives

Sea ice prediction is a relatively novel field, in which the knowledge of the sources and mechanisms of predictability of the Arctic sea ice conditions is still incomplete. To understand these sources as well as the development of forecast biases, a regional approach is needed because their different driving mechanisms can largely depend on the location. Through a better understanding of the different forecast biases at play and how they unfold, this thesis aims to contribute to the development of improved forecast systems. This thesis has three general objectives, each tackling more specific scientific sub-objectives:

1. Identifying and quantifying the sources and mechanisms of sea ice predictability both at pan-Arctic and regional scales in a perfect model analysis.

   (a) Assessing the predictability mechanisms leading to successful sea ice extent and volume predictions for various Arctic basins in EC-Earth.

   (b) Understanding how predictability at the pan-Arctic scale could be explained by regional-scale mechanisms.

   (c) Extending the evaluation of the potential predictability of Arctic sea ice to 5 additional forecast systems, based on different General Circulation Models (GCMs).

2. Disentangling and quantifying different sources of Arctic sea ice forecast biases in a set of retrospective seasonal predictions produced with EC-Earth3.2.

   (a) Assessing the inconsistencies between the different initialization products (i.e. sea ice, ocean and atmosphere) and how they affect the development of Arctic sea ice biases during the prediction. Quantifying the potential impact of the sea ice-ocean initial inconsistencies at the regional level.

(b) Documenting the development and relative importance of the systematic model drift as a function of forecast time. Studying its competing effect with initialization shocks during the first forecast month.

(c) Characterizing the forecast biases at daily resolution. The manifestation of certain fast initial shocks may be hidden when monthly averages are used, and daily means can help to better diagnose the potential sources of forecast biases.

3. Comparing three seasonal forecast systems based on EC-Earth3.2 initialized through different strategies and quantifying their Arctic sea ice forecast biases and prediction skill.

(a) Assessing the benefits/disadvantages of strong versus weak restoring coefficients in the ocean when assimilating an observational dataset during the production of the initial conditions.

(b) Comparing the use of non-native (i.e. built with an alternative ocean model) versus native (produced with the same ocean model used in forecast mode) ocean initial conditions, and how these impact the forecast performance.

(c) Assessing whether strategies preventing or minimizing the initial shocks can also achieve improved predictive capacity.

1.9 Thesis structure

The thesis is organized in six chapters, as described in the following: **Chapter 1** has offered an overview of the state-of-the-art in Arctic sea ice prediction, including the current main limitations; **Chapter 2** provides a description of the different methodological aspects concerning the thesis and a summary of key concepts; **Chapter 3** evaluates and compares the potential predictability of pan-Arctic and regional sea ice in 6 GCMs. It also concentrates on EC-Earth to deepen in the understanding of potential predictability regimes in some specific regions; **Chapter 4** analyzes the development of the different forecast biases during the first forecast month in a set of retrospective seasonal predictions with EC-Earth; **Chapter 5** provides an assessment of the Arctic sea ice forecast biases and predictive skill in retrospective seasonal predictions based on three different sea ice-ocean initialization strategies; finally, an overview of the main conclusions is given in **Chapter 6**.

2 Methods

This chapter starts by defining important sea ice definitions (section 2.1), concepts (section 2.2) and metrics (section 2.3) that have been used throughout the thesis. The two versions of the EC-Earth model employed in the thesis are briefly described in section 2.4. Finally, in section 2.5 the observational references used for this study are presented. Specific details will be given in the methodology of each chapter.

2.1 Sea ice definitions

Before giving detailed account of the most important methodological aspects considered for this thesis, we provide a list of the key definitions used to describe and evaluate the different aspects of the sea ice state, as well as the main Arctic regions considered:

1. *Sea Ice Concentration* (SIC) is the fraction of a grid cell that is ice-covered. SIC is expressed as a percentage.

2. *Sea Ice Thickness* (SIT) is the average thickness of sea ice over the ice-covered portion of the grid cell. It is measured in metres.

3. *Sea Ice Area* (SIA) is the sum of grid cell areas weighted by their SIC in a predefined domain (e.g. pan-Arctic, Barents Sea, Baffin Bay, etc.). It is given in square metres.

4. *Sea Ice Extent* (SIE) is the integrated area of all grid cells with SIC above 15%. SIE is thought to be more accurately observed from satellites than SIA (Parkinson and Cavalieri, 2008), for two main reasons: (1) SIA estimates are affected by the presence

of random ice floes which are equatorward of the main ice pack; (2) atmospheric interferences with the microwave signal due to the water vapour and clouds produce errors that affect predominantly areas well inside or outside the ice pack. In particular, the first kind produces many regions well outside of the ice pack where the ice concentration calculations erroneously suggest a small amount of ice, predominantly under 15%. Choosing a 15% threshold for ice coverage removes most of the atmospheric interferences while still retaining the majority of the ice cover (Parkinson and Cavalieri, 2008).

5. *Sea Ice Volume* (SIV) is the integral of sea ice thickness multiplied by the area of each grid cell covered by sea ice. It is measured in cubic kilometres.

6. The *Sea Ice Edge* is the boundary between open water and sea ice, and it is typically defined as the SIC threshold of 15%.

7. The *marginal ice zone* (MIZ) is the transition zone ($0 < \text{SIC} < 1$) between the interior ice pack ($\text{SIC} = 1$) and the open ocean ($\text{SIC} = 0$). The MIZ captures most of the seasonal and longer term SIC variability (Germe et al., 2014). The *Arctic regions* can be categorized according to the processes at stake in their sea ice variability. The basins can be classified into two groups: the internal basins, which are entirely ice covered during winter (central Arctic, Canadian Archipelago, Beaufort Sea, East Siberian Sea and Laptev Sea) and the peripheral regions, which include summer ice-free regions (Barents Sea, Kara Sea, Greenland-Icelandic-Norwegian (GIN) seas, Irminger Sea, Baffin Bay, Labrador Sea, Hudson Bay, Bering Sea, Sea of Okhotsk and Chukchi Sea). A map illustrating the boundaries of the different Arctic basins used in this thesis is shown in Fig. 2.1.

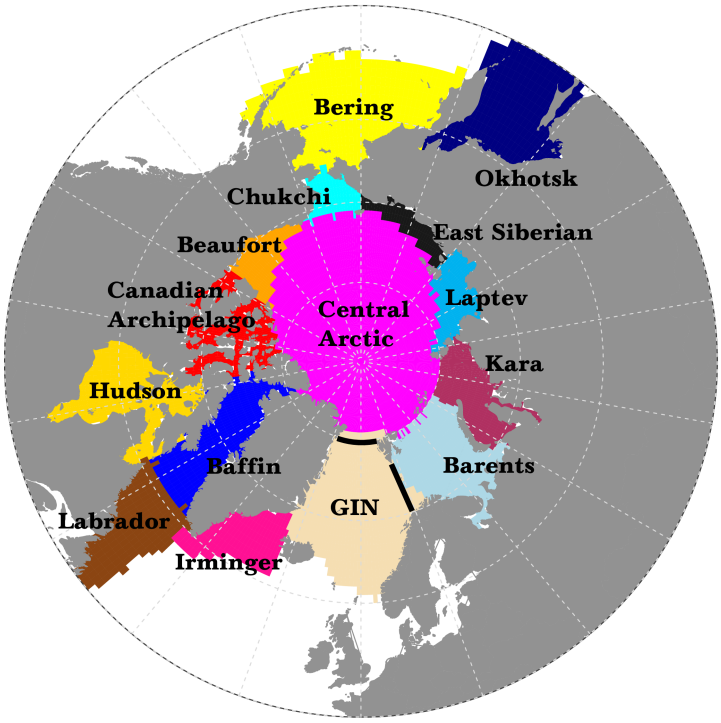


Figure 2.1: Map of the Arctic seas. The black lines indicate the sections used for the calculation of the Atlantic heat transport into the Arctic (Fram Strait plus Barents Sea Opening). The GIN region is formed by the Greenland, Icelandic and Norwegian seas.

2.2 Forecast evaluation concepts

We now describe the key concepts that need to be present when producing and evaluating a forecast:

- A *climatology* is the mean state of a given climate variable over a specific period, and is generally used as a reference (e.g. to compute anomaly; see below). It can be computed in a model simulation, in a forecast system, as well as in observations, although the exact methods vary in each case.

- The *model climatology* is calculated as the temporal average (typically as an average annual cycle) of the variable over a long reference simulation (typically in a present day fixed forcing or a historical simulation), or a specific period in it (e.g. the overlap period with the observations). This reference simulation or period has to be long enough to smooth out the effect of internal climate variability (optimally 30 years

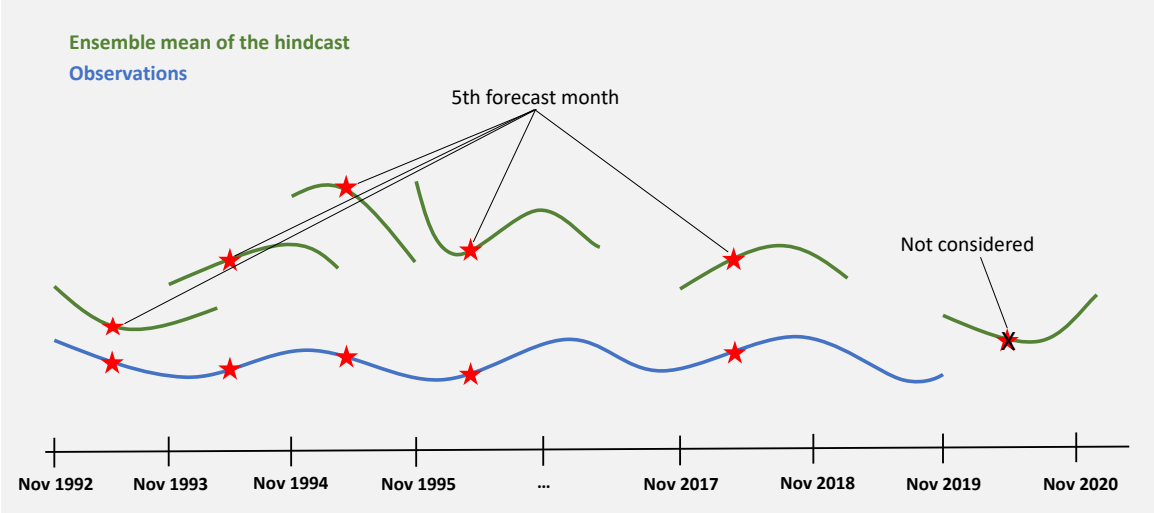


Figure 2.2: Illustration of the per-pair method used to select forecast and reference data for computing the forecast climatology for the month 5 of the hindcast, for an arbitrary climate variable. In green are shown the model hindcasts and in blue the observations. The 5th month of the last hindcast is not considered as there is no observation available.

or more). Changing the period over which we compute the climatology slightly, e.g. by one year, would lead to a slightly different climatology, although representative of the same inherent model climatology. The difference between the two resultant climatologies gives an idea of the uncertainty there is in estimating this climatology from available data. This type of uncertainty is called 'sampling uncertainty' because it arises from the size of the sample used to estimate the climatology.

- The *observed climatology* is calculated as the model climatology, but using observations instead.

- In climate prediction, the model is initialized from a state as close as possible from observations so that its average state (in a statistical sense) evolves during the prediction from being close to an observed climatology to being close to a model climatology. The *forecast climatology* is a function of the forecast time and, as described by the per-pair method (García-Serrano and Doblas-Reyes, 2012), it is estimated by averaging the hindcast variable across all members and start dates. Also, it is typically computed only over dates for which observations are available (Fig. 2.2). A forecast climatology has one value per forecast time, while a model climatology has one value per calendar

day or month, depending on the temporal resolution in turn.

- The *forecast bias* is the systematic error in the forecast as a function of the forecast time (f). It is computed, for each forecast time, by subtracting the observed climatology (o) from the forecast climatology.

$$b_f = y_f - o_f \tag{2.1}$$

where y_f is the forecast climatology and o_f is the observational reference at the forecast time f.

- In climate simulations non-initialized from observations but from a random state in the range of possible model state for a given climate, the *model bias* (also referred to as systematic model error in this thesis) is the systematic error in the model when compared to the observations. It is computed by subtracting the observed climatology from the model climatology. When based on climatologies previously computed per calendar day or month, the model bias provides an annual cycle of model inherent errors (e.g., too much or too little sea ice during the boreal summer).

$$b_m = y\prime_c - o_c \tag{2.2}$$

where $y\prime_c$ is the model climatology and o_c is the observed climatology for a specific calendar day or month c over the same period.

- The term *anomaly* refers to the difference between the contemporaneous state of a variable and its long-term average or climatology, calculated over a predefined period of time.

- The *forecast anomalies* are computed by subtracting to the raw predictions the forecast climatology for each forecast time. This is done for every member and start date:

$$y\prime_{f,s,m} = y_{f,s,m} - \overline{y_f} \tag{2.3}$$

where $y\prime$ is a predicted anomaly, y is the raw forecast, $\overline{y}$ is the forecast climatology for the same forecast time f as $y\prime$, s is the start date and m the member.

- The *observed anomalies* are computed as the difference between the observed values and a given observed climatology::

$$o\prime_{f,s} = o_{f,s} - \overline{o_f} \tag{2.4}$$

where $o\prime$ is an observed anomaly, o is the raw observed signal for the same forecast time f as $o\prime$ and the same date s as the forecast start date, and $\overline{o}$ is the observed climatology.

2.3 Forecast verification

Forecast skill can be assessed through deterministic or probabilistic verification metrics. Deterministic verification focuses on assessing how close the ensemble mean prediction (the average over all the members) is from the observation, whereas probabilistic verification focuses on assessing to which extent the distribution of the ensemble members of the prediction is representative of the likelihood of the observed event to occur. In this thesis, we have only used deterministic metrics to estimate the Arctic sea ice prediction skill.

2.3.1 Verification metrics

Anomaly correlation coefficient (ACC) and *root mean-square error* (RMSE) are examples of deterministic scores and the most commonly used for prediction skill assessment. The ACC is defined as:

$$ACC(f) = \frac{\sum_{s=1}^{S} y\prime\,(f,s) \cdot x\prime_o\,(f,s)}{\sqrt{\sum_{s=1}^{S} y\prime\,(f,s)^2 \cdot \sum_{s=1}^{S} x\prime_o\,(f,s)^2}} \tag{2.5}$$

where S is the total number of start dates, s is the start date, f is the forecast time, $y\prime$ represents the ensemble mean of the predicted anomalies and $x\prime_o$ the observed anomalies.

The ACC is particularly sensitive to errors in the sign and to outliers (Kim et al., 2015). It ranges between 1 and -1. If ACC = 1, there is a perfect association between the predicted and observed anomalies, while ACC $\leq$ 0 reveals that there is no skill in the prediction. Since the true ACC of a system cannot be perfectly estimated because its computed over a reduced sample of start dates, it is important to use statistical significance tests to evaluate if the estimated ACC value is different than 0, and the system is therefore skilful. These significance tests that are introduced in section 2.3.2.

The SIE is an integrated diagnostic, and as such, its bias can only inform if the total SIE is overestimated or underestimated over a selected region. This overestimation or underestimation could be small on average over the region, but it is also possible that large areas of overestimation compensate large areas of underestimation, a forecast deficiency that the SIE bias would overlook. The *Integrated Ice Edge Error* (IIEE, Goessling et al., 2016) is defined to circumvent this limitation as the area where the forecast and the target reference disagree on the sea ice concentration being above or below 15%. It equals the sum of all areas where the SIE is overestimated (O) or underestimated (U). The IIEE is, therefore, a more informative error estimation metric than the SIE bias. Its formal definition is:

$$IIEE = O + U \tag{2.6}$$

with

$$O = \int_a max(c_f - c_o, 0) dA \tag{2.7}$$

and

$$U = \int_a max(c_o - c_f, 0) dA \tag{2.8}$$

where A is the area of interest over the Arctic, c = 1 where the SIC is above 15% and c = 0 elsewhere, and subscripts f and o denote the forecast time and the target (Figure 2.3). Thus, the definition of the IIEE is equivalent to the difference between the areas enclosed by the forecasted and the target ice edge.

The IIEE can be decomposed into the *absolute extent error* (AEE) and the *misplacement error* (ME) components:

$$IIEE = AEE + ME \tag{2.9}$$

with

$$AEE = |O - U| \tag{2.10}$$

and

$$ME = 2 \cdot min(O, U) \tag{2.11}$$

The AEE represents the absolute difference in SIE between the forecast and observations, while the ME integrates sea ice that has been predicted at an incorrect location.

Figure 2.3: The sea ice edges (15% sea ice concentration contours) of the AWI-CM idealized forecast ensemble on 15 September of an arbitrary year. Interpreting the blue contour as forecast and the red contour as truth (observations), the IIEE is the sum of all light blue (sea ice extent overestimated; O) and light red (sea ice extent underestimated; U) areas. Source: Goessling et al. (2016).

2.3.2 Statistical hypothesis testing

Statistical hypothesis testing allows to make an informed decision of whether a result is real or only the product of sample uncertainty. This is typically done by choosing between a null hypothesis (H_0) of statistical relationship between two data sets and an alternative hypothesis (H_1) which typically describes the range of possibilities that may be true when H_0 is false (i.e., Fig. 2.4). The test can only have two outcomes: either H_0 is rejected or it is not rejected. To decide whether or not to reject the null hypothesis, the sample data is used to estimate the probability of this null hypothesis to be true. The significance level (α) of the test is the probability threshold below which the null hypothesis will be rejected. Serve as an example: we want to know if the current global warming is provoked by human activity. Then:

H_0 = Global warming is caused by natural factors.

H_1 = Global warming has an anthropogenic origin.

Imagine we have 2 datasets covering the last 100 years, one with the global mean temperatures and another with the global CO_2 emissions (for simplicity we will suppose that

there is an immediate and direct impact of CO_2 on temperatures). We obtain a correlation value (R score) of 0.25 between the two samples. How likely is that the correlation between global temperatures and CO_2 emissions is actually zero, 0.25 being only the random result of sampling uncertainty? We compute the probability of obtaining a correlation at least as high as the one in our sample data, under the assumption of no true correlation, i.e. the null hypothesis. This probability is known as 'p-value'.

H_0: correlation = 0

H_1: correlation $\neq$ 0

In our case the p-value is 0.012. This means that if H_0 is true, there is a 1.2% probability of obtaining a sample correlation at least as high as what we computed. To estimate this probability, the following steps can be applied: (1) we generate two random samples of data with no correlation between them and estimate their correlation, which will not be exactly 0 because of sampling uncertainty; (2) we repeat this operation a large number of times to be able to estimate a probability distribution function (pdf) of the estimated correlation of two samples with no real correlation; (3) we locate our measured correlation (here 0.25) in the pdf to obtain the probability of occurrence of a correlation at least as high when two samples are not correlated. These steps describe a non-parametric test. An alternative procedure is to assume that the sample data follows a given distribution, e.g. a Gaussian distribution, and deduce which type of distribution should follow the computed metrics, here the correlation, which would follow a t-distribution, as in our case. The pdf is readily obtained from this assumed distribution instead of being built as in a non-parametric test. The next ingredient to our test is the choice of a probability threshold to accept or reject H_0. This probability threshold is known as the significance criterion for the test α. Typically, significance criterion are chosen to be 1%, 5% or 10%, with 5% being the most widely used in climate sciences. This significance criterion is used as follows:

p-value $< \alpha \implies reject \quad \mathrm{H}_0 \implies accept \quad \mathrm{H}_1$

p-value $\geq \alpha \implies fail \quad to \quad reject \quad \mathrm{H}_0$

Let's use 5% for this test. The probability of obtaining the measured correlation under H_0 is only 1.2%, hence below the threshold of 5% and H_0 is rejected, H_1 being considered as the most likely option: global warming has actually an anthropogenic origin. Therefore, a null hypothesis is simply the default assumption the study sets out to be disproved by statistical inference. Statistical hypothesis testing has two limitations: (1) there is a probability equal to the significance criterion of the test that a true H_0 will be rejected- this is the error of the

first kind; (2) H_1 can be rejected being actually true, with unknown probability - this is the error of the second kind. A decrease in the error of the first kind goes along with an increase in error of the second kind. It is a trade-off. It is important to note that smaller the sample size, the more likely an error of the second kind is to occur. Therefore, in seasonal prediction, the typically limited sample size available (regarding start dates, ensemble members and forecast time) hinders our ability to detect a significant level of prediction skill.

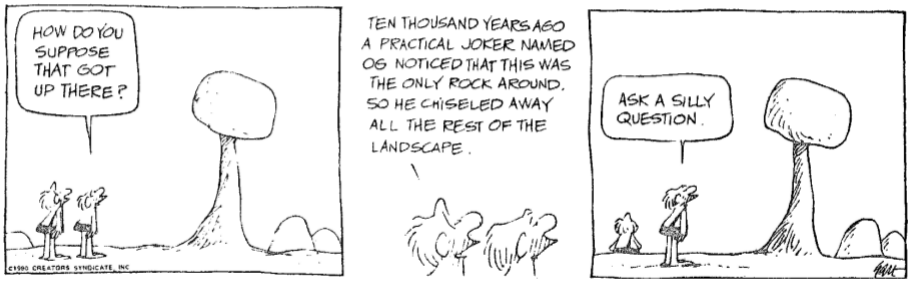


Figure 2.4: Creation of the Mexican Hat rock of Utah: Null hypothesis correctly rejected. Source: Von Storch and Zwiers (2001).

Significance tests Most of the significance analyses on this thesis have been based on the *T-test* (through a one-sided student-T distribution), but for some specific cases the *F-test* has also been considered.

The T-test is a parametric test that exploits the t-distribution by computing a variable that is expected to follow a t-distribution when H_0 is true. It has been used to estimate if correlations are significant or not, the differences in means and the difference in correlations. The F-test computes ratio of the forecast and observed variances which follows a f-distribution.

The validity of these two tests relies on three assumptions: (1) all sample values entering the tests are considered to be independent, an hypothesis that is violated when there is temporal correlation between the start dates driven by climate phenomena operating on long timescales. This problem can be dealt with by using an effective sample size that takes into account this temporal autocorrelation (Guemas et al., 2014a); (2) the variables entering the tests are assumed to be normally distributed and to have equal variance - the normality assumption is very often valid in climate (with exceptions like precipitation) but the observed and model variances often differ; (3) the distribution that generates realizations of X (or Y) is the same for each observation in the X (or Y) sample. Decremer et al. (2014) compared the autocorrelation of precipitation over 20 continuous-year simulations (validating with simulations of up to 1000 years) using five different statistical techniques, concluding that the Student's t-test stands out as the best method overall. More information about how

these tests are performed can be found in Von Storch and Zwiers (2001).

2.4 The model: EC-Earth

The EC-Earth GCM has been used, throughout this thesis, to produce retrospective predictions, historical simulations and the reconstructions used as initial conditions. Two different versions of EC-Earth were used: EC-Earth2.3 and EC-Earth3.2, both of them with a horizontal resolution of 1° (Doblas-Reyes et al., 2018; http://www.ec-earth.org/). Information about those experiments and how they were produced is provided in the respective chapters.

- **EC-Earth version 2.3** (Hazeleger et al., 2012): The ocean component is the second version of NEMO (Nucleus for European Modelling of the Ocean) (Ethé et al., 2006; Madec et al., 2015) with ORCA1 configuration (about 1 degree with enhanced tropical resolution) and 42 vertical levels. The sea ice component is LIM2 (Fichefet and Maqueda, 1997b) embedded into NEMO. The atmospheric component is based on the ECMWF integrated forecasting system (IFS), with 62 vertical levels and a TL159 horizontal resolution. The atmospheric and ocean components are coupled via OASIS3 (Valcke, 2006; Fig. 2.5). The vegetation component is LPJ-GUESS (Smith, 2001), which uses a present-day climatology. The information on the atmospheric chemistry and aerosols is prescribed from observations. Finally, the land model uses H-TESSEL (Balsamo et al., 2009).

- **EC-Earth version 3.2**: The ocean component is the third version of NEMO with ORCA1 configuration and 75 vertical levels. The sea ice component is LIM3 (Vancoppenolle et al., 2009)) embedded into NEMO. For the atmosphere, IFS is employed, in a configuration with 91 vertical levels and a T255 horizontal resolution. All components are coupled via OASIS3. The vegetation component is a present-day climatology provided by LPJ-GUESS. The atmospheric chemistry and aerosols are prescribed and the land model is H-TESSEL.

Ensemble generation As discussed in Chapter 1, the production of a large ensemble is key to provide a good representation of the prediction uncertainty. Two different ensemble member generation strategies were employed during this thesis: (1) introducing a white noise (with an order of magnitude of $10^{-4}K$) only in the initial SST state. This is the procedure followed for the perfect model experiments in Chapter 3; (2) producing different initial states

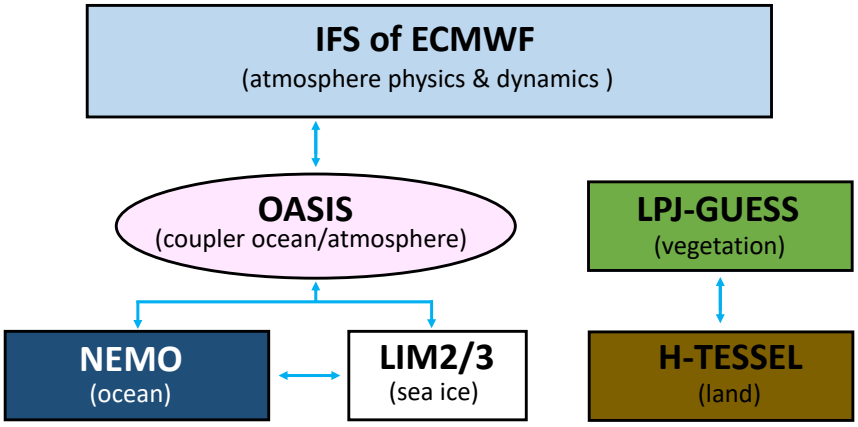


Figure 2.5: Schematic representation of the different components of EC-Earth coupled model.

for all the system components, and making combinations with them. These were obtained for each model component as follows: (a) for the ocean, the five members from the ORAS4 reanalyses are used (which differ from each other by the amount of data exploited for their generation -in order to sample the uncertainty of the observation coverage the probability of rejection is 10% for Argo observations and 5% for other platforms measuring T/S) either directly as ICs, either as a reference for a nudged ocean-sea ice reconstruction forced with the Drakkar Forcing Set (DFS); (b) singular vectors for the atmospheric component -these are perturbations designed to optimize the growth of perturbations; (c) sea ice ICs are taken from the five ocean-sea ice reconstructions nudged towards ORAS4. These also include some perturbations in the surface forcing.

2.5 Observational references

Sea Ice Concentration The two main satellite datasets used as reference to compare the modelled SIC in this thesis are the National Snow and Ice Data Center (**NSIDC**) ones: NASA-Team[1] (Cavalieri et al., 1996) and Bootstrap[2] (Comiso, 2017). The principal reason to choose these products was that they represent a continuous monthly record since the end of 1978, with daily uninterrupted measurements since 1988. Also, both together give account of the observational uncertainty related to the retrieval algorithms (which will be used in Chapter 4). Other products such as HadiSST cover a longer period of time, but

[1]https://nsidc.org/data/NSIDC-0051
[2]https://nsidc.org/data/nsidc-0079

this is not based on satellite data and it is not expected to be as homogeneous as NSIDC. Indeed, NSIDC observations are the longest *daily* SIC satellite measurements, which allowed to extend our retrospective daily predictions with respect to other satellite data sets. The data sets only differ in the algorithm employed to estimate SIC from passive microwaves measurements. The data are provided in the polar stereographic projection at a grid cell size of 25 x 25 km. Due to the inclination of the satellite orbits there is a circular sector centered over the North Pole where there are no measurements, commonly known as *pole hole*. The size of pole hole has been decreasing in time thanks to the launching of new sensors: with a radius of 611 km from 1978 to June 1987, then 311 km until December 2007, and currently 94 km. The main drawback of the NSIDC data sets is that they do not provide any uncertainty measurements to, i.e., use the products for data assimilation purposes.

ESA version 1 SIC from the Climate Change Initiative (Hollmann et al., 2013) is used in Chapter 4 as the target assimilated data. It covers 1992 to 2008 daily. The data are provided on a Lambert Azimuthal Equal Area polar projection, with a grid spacing of 25 km and no data monitoring north of 87°. Unlike the NSIDC products, ESA provides useful additional data such as an uncertainty estimate and information about the processing steps that have influenced the ice concentration value, which are needed for assimilating with the Ensemble Kalman Filter.

In this thesis, not only satellite observations have been used as reference data for model evaluation, but also climate reanalyses. The *climate reanalyses* are combinations of available observations with models and physically-consistent complete descriptions of the physical state, including variables that are scarcely observed, such as the SIT (e.g. Schweiger et al., 2011). The discipline that develops the techniques to constrain mathematical or physical models (e.g. a climate model) with a given set of observations to produce a physically plausible description of the real world is known as data assimilation (e.g. Laloyaux et al., 2016; Sugiura et al., 2008). This field is the basis for producing climate reanalyses and thus to obtain the best possible initial state for climate prediction. However, there are also relative differences in the representation of certain variables between the various reanalyses (e.g. Chevallier et al., 2017; Uotila et al., 2019), so their uncertainty has also to be taken into account.

Ocean Temperature and salinity The ocean reanalysis system 4 (**ORAS4**, Balmaseda et al., 2013) produced by the European Center for Medium-Range Weather Forecasts was used as the target reference to assimilate ocean data in Chapters 4 and 5 and produce the

initial conditions of the ocean and the sea ice. ORAS4 is a product of the NEMOVAR 3-dimensional variational data assimilation system based on NEMO version 3.0. Its ocean grid is the same as used in EC-Earth2.3, with 42 vertical levels. The reanalysis covers from 1959 to present and consists of 5 members, generated by adding wind-stress perturbations. It assimilates temperature and salinity profiles and altimeter-derived sea level anomalies. The NEMOVAR assimilation cycle is 10 days and a bias correction method is applied. Surface forcings are daily surface fluxes from ERA40 until 1989, ERA-Interim from 1989 to 2010 and the ECMWF operational archive afterwards. A strong relaxation (200 $Wm^{-2}{}^{o}C^{-1}$, equivalent to about a 2–3-day time-scale over a depth of 10 m) to gridded SST products is applied at the surface. The main limitation of this reanalysis is the absence of sea ice model: the sea ice information is added as a boundary condition derived from a mixture of satellite products. These SIC data are used to adjust the SST as follows: if the SIC is higher than a certain threshold (55%), the model SST is set to freezing point, which in a coupled model will favour sea ice formation on the whole grid cell, but not force it since a strong atmospheric forcing can always prevent sea ice formation. If SIC is below the 55% threshold but the model SST reaches the freezing point, the relaxation term to observed SST is increased to bring it below that value. These conditionals lead to a series of inconveniences for reconstructions assimilating ORAS4. More information can be found in Mogensen et al. (2012) and in Chapter 5.

EN4 (Good et al., 2013) is a global ocean analysis developed by the Met Office Hadley Centre, and it was used as the observational reference for the SST in Chapter 5. It provides a monthly coverage from 1900 to the present, with a regular 1° horizontal grid and 42 levels in the vertical. It is based on subsurface ocean temperature and salinity profile data obtained from the WOD09, GTSPP, Argo and ASBO collections. To obtain spatially complete maps the nearest neighbor algorithm was applied for unfilled locations. On top of that, a large number of quality assessment and enhancement steps were performed (see Good et al., 2013). The enhanced and cleaned profiles were interpolated to the gridded data using Objective Analysis.

Surface Air Temperature and other atmospheric variables **ERA-Interim** reanalysis (Dee et al., 2011) has been used as the atmospheric initial conditions for climate predictions and as observational reference for the SAT, as well as for the solar radiation and heat fluxes. Note that for the solar radiation no observations are assimilated, taking a constant value of 1370 Wm^{-2}). For the surface fluxes no observational data are assimilated neither, being just loosely constrained by atmospheric variables. ERA-Interim is based on the Inte-

grated Forecast System (IFS, Cy31r2), and it is performed with a 4-dimensional variational data assimilation method with a 12-hour analysis window. The variables assimilated by the atmospheric component are: surface pressure, 2 m relative humidity, 10 m wind, scatterometer wind, upper-air temperature, specific humidity and wind, ozone, brightness temperature, total precipitable water and bending angle. The horizontal resolution is approximately 80 km (T255 spectral, equivalent to 0.75º) on 60 levels in the vertical from the surface up to 0.1 hPa. It covers from January 1979 to August 2019 with 6-hourly fields.

3 Regional sea ice predictability in the Arctic ocean

This Chapter presents and expands the results of the article: Cruz-García, R., Guemas, V., Chevallier, M., and Massonnet, F. *An assessment of regional sea ice predictability in the Arctic ocean.* Climate Dynamics, pp 1–14, 2019. doi: 10.1007/s00382-018-4592-6.

3.1 Introduction

Sea ice predictability has been assessed in various frameworks, including idealized perfect-model experiments (e.g. Day et al., 2016). In such experiments, model simulations are used as a surrogate for the real climate to estimate the extent to which the model can predict itself. Ensemble perfect-model predictions are generally initialized from a control run by introducing small perturbations. Potential predictability is a measure of the amplification of those perturbations, i.e. the fraction of the signal which is inherently not predictable. Such experiments using state-of-the-art models also provide an indication of the maximum level of skill that could be achieved in real predictions if all the observations required to initialize the predictions were available, and if all processes were perfectly represented by the models.

Tietsche et al. (2014) performed the first multi-model evaluation of Arctic sea ice potential predictability on seasonal-to-interannual timescales, in a coordinated perfect-model framework defined in the Arctic Predictability and Prediction on Seasonal-to-Interannual Timescales (APPOSITE) project (Day et al., 2016; Tietsche et al., 2014). Each of the seven participating groups ran a set of 3-year long ensemble prediction experiments, initialized from a present-day control experiment near July 1. They showed that even if two models have significant predictability -based on a comparison of the prediction ensemble spread and the natural control variability- for the SIV (up to 3 years) for similar forecast times, differences in the representation of local advective processes could lead to large differences in the

regional SIT predictability. They suggested that advective sea ice processes may induce an amplification of forecast errors close to the coasts in the Arctic Ocean in winter.

A similar perfect-model approach was also followed by Day et al. (2014). In a subset of five APPOSITE models, they found similar SIE predictability reemergence mechanisms (i.e., an increase of predictability after an initial drop), consistent with the summer-to-summer and melt-to-freeze mechanisms described by Blanchard-Wrigglesworth et al. (2011a). They also found that when starting the predictions in May, the forecasts lost skill more rapidly in the first 4 months than when initialized in January or July. Another robust result was that the SIE in the MIZ of the North Atlantic region is significantly predictable 1.5–2.5 years ahead, while in the central Arctic it is less than 1 year. Note that Day et al. (2014) estimated the predictability horizon as the forecast time until which the ensemble RMSE is below the climatological RMSE, using an *F-test* for significance.

Several studies have focused on the sea ice predictability of different Arctic basins (Bushuk et al., 2019; Cheng et al., 2016; Day et al., 2014; Koenigk and Mikolajewicz, 2009). However, the mechanisms behind the regional sea ice predictability are not yet well established. To understand the Arctic sea ice cover predictability, a regional approach is needed to disentangle the different drivers of variability, which depend on the location (e.g. Bitz et al., 2005; Francis and Hunter, 2007; Schlichtholz, 2011; Tietsche et al., 2016). For instance, Bitz et al. (2005) showed that ocean heat flux convergence exerts a large influence in the marginal ice zone of the Barents Sea, but plays a relatively small role in the Labrador Sea. Likewise, Francis and Hunter (2007) suggested that the zonal wind anomalies influence the Bering Sea winter ice edge location, while the Barents Sea ice edge seems to be controlled primarily by anomalies in SST during the late winter and by anomalous meridional winds.

Regional metrics tend to have lower predictability than integrated ones (e.g. Day et al., 2014; Goessling et al., 2016). Blanchard-Wrigglesworth et al. (2016) highlighted this contrast between pan-Arctic and regional predictability with a multimodel approach, where all models initialized with identical SIT could uniformly predict September SIE anomalies, but did not show agreement regarding the spatial SIC anomaly patterns.

In the present chapter, we perform a regional sea ice predictability assessment (both for Arctic SIE and SIV) in six of the APPOSITE project models, with a special focus on EC-Earth2.3 (Hazeleger et al., 2012), since mechanisms of predictability can be investigated at greater depth in this model by, for example, projecting water mass backward trajectories for some key regions. The analyses consider sub-regions in the central and marginal Arctic ocean, exploring for which of them the same mechanisms previously attributed to pan-Arctic sea ice predictability (such as persistence and reemergence) play a role.

This chapter is structured as follows: In section 3.2, we describe our methodology, including the experimental protocol and the metrics used to quantify sea ice potential predictability. Section 3.3 presents the assessment of Arctic sea ice potential predictability at the pan–Arctic and regional scale for each of the APPOSITE models (except CanCM4). Section 3.4 discusses the mechanisms behind the regional sea ice predictability of EC-Earth2.3 and section 3.5 provides the main conclusions.

3.2 Methodology

3.2.1 Multimodel ensemble and experimental setup

An upper limit for the predictability of Arctic SIE and SIV was estimated using six of the seven coupled global climate models from the APPOSITE project (Day et al., 2016): EC-Earth2.3 (Hazeleger et al., 2012), MIROC5.2 (Watanabe et al., 2010), HadGEM1.2 (Johns et al., 2006; Shaffrey et al., 2009), GFDL-CM3 (Donner et al., 2011; Griffies et al., 2011), MPI-ESM (Jungclaus et al., 2013; Notz et al., 2013) and E6F (Sidorenko et al., 2015). The CanCM4 model was discarded because of the short length of its control simulation, which hampered the identification of start dates for all the sea ice case studies (more details below). These models have already been evaluated in the multimodel assessment of Tietsche et al. (2014), and a few characteristics of the APPOSITE simulations are shown in Day et al. (2014).

Each APPOSITE working group provided a control simulation (hereafter referred to as *ControlRun*) produced using fixed present-day radiative forcings and a set of predictions that started from the control (hereafter referred to as *IdealPred*). In this chapter we use the predictions that were started from July. Note that in this perfect-model protocol, the *ControlRun* is also the reference dataset for evaluating the performance of the idealized climate predictions.

For each model, an ensemble of between 8 and 16 members was generated, depending on the resource limitations of each modelling centre. Each prediction member has a slightly different perturbation of the initial state obtained by introducing a 10^{-4} K magnitude white noise in the SST. The number of start dates differs from model to model and ranges between 8 and 18, again depending on the resource limitations of each modelling centre. The start dates were selected to sample a range of high, low and medium sea ice states, with consideration of the Atlantic heat transport into the Arctic (AHT hereafter; calculated as the heat transport through the section formed by the sum of the Fram Strait plus the Barents Sea Opening, represented by the black thick lines in Fig. 2.1). Start dates are spaced apart in time to

make them as independent as possible. More details about the reasons behind the choice of the different start dates in EC-Earth2.3 can be found in Table A.1 in the Appendix, while further details about the APPOSITE protocol can be found in Day et al. (2016).

3.2.2 Perfect Predictability Diagnostics

Our analysis is performed using monthly data for the pan-Arctic SIE and SIV. Regional assessment is done for the basins shown in Fig. 2.1.

In this study, potential predictability is estimated both in a prognostic and diagnostic way (Boer, 2004). Simple estimates of the diagnostic potential predictability are calculated using lagged anomaly correlations in the *ControlRun* as in Blanchard-Wrigglesworth et al. (2011a). The prognostic potential predictability uses the methodology described in Pohlmann et al. (2004), and is estimated using both the control simulation and the idealized experiments.

In both cases, anomalies are calculated as follows. For each date, a 40-year-window taken from the *ControlRun* and centered around that date is used to filter out low-frequency variability and the remaining long-term drift. The mean annual cycle over that period is used as a reference to compute the anomalies in the *IdealPred* experiments. To be consistent, we also apply this protocol to determine the anomalies across the *ControlRun*, using 40-year running windows, as:

$$Z'_g = Z_g - \overline{Z}_g^{[-20y;+20y]} \tag{3.1}$$

where g is the selected month of the raw data Z, $\overline{Z_g}$ is the average of the 40 same calendar months around the selected date and Z'_g is the anomaly of month g compared to the average annual cycle of the 40-year window.

Following this, the amplitude of the natural variability is estimated as:

$$\sigma_c(g) = \sqrt{\frac{\sum_{Y_0}^{Y_l}(Z'_g(y))^2}{Y_l - Y_0}} \tag{3.2}$$

where $Z'_g(y)$ is the anomaly for month g and year y, and Y_0 and Y_l are the first and last year, respectively, considered in the summation. Note that the use of the 40-year running windows excludes 20 years at the beginning and end of the whole simulation.

The level of potential predictability is estimated using the intra-ensemble spread (i.e. the spread around the ensemble mean), as a function of the forecast time:

$$\sigma_e(t) = \sqrt{\frac{\sum_{m=1}^{M}\sum_{s=1}^{S}(Z_{m,s}(t) - \overline{Z_s(t)})^2}{M \cdot S}} \tag{3.3}$$

where M is the total number of members, S is the total number of start dates, $Z_{m,s}(t)$ is the predicted value of our variable at forecast time t for ensemble member m initialized at start date s, and $\overline{Z_s}$ is the predicted value for the same start date and forecast time averaged across the whole ensemble.

We use the prognostic potential predictability (PPP hereafter; Germe et al., 2014; Pohlmann et al., 2004). The PPP compares the ensemble spread with an estimate of the amplitude of the natural variability of the system based on the standard deviation of the control simulation (e.g. Koenigk and Mikolajewicz, 2009). It is an estimate of the predictability and is defined as:

$$PPP(t) = 1 - \frac{\sigma_e^2(t)}{\sigma_c^2(g)} \tag{3.4}$$

where σ_e^2 is the variance across the ensemble members (*IdealPred*) at forecast time t and σ_c^2 is the variance of the control integration *ControlRun* for the relevant month g. A PPP value of 1 would mean that the system is perfectly predictable (i.e. the ensemble members of the predictions did not diverge over time), whereas a PPP value of zero or less indicates that there is no predictability because the ensemble spread is equal to that expected from natural variability (Holland et al., 2011). Significance was estimated with an *F-test*. A PPP value can be found significant even though the prediction is trivial (because the sea ice variability is close to null). A criterion has been defined now to consider those as significant: for any SIE/SIV anomaly in the control simulation which would be lower than or equal to 1% of the average anomaly along the control for the same area and calendar month, the estimate of PPP is not considered as significant. Unlike the RMSE, this metric allows us to compare the dispersion of the ensemble with respect to the reference variability in a single number, giving us an idea of the proportion at every timescale, even when both of them are very small and similar.

3.3 Multimodel potential predictability of pan–Arctic and regional sea ice

Pan–Arctic sea ice The pan-Arctic SIE PPP decreases at a similar rate for the first six months after initialization for all models (Fig. 3.1a). From the first December there is a consistent predictability reemergence for HadGEM1.2, GFDL-CM3 and E6F every winter, but a lack of significant predictability during the summer. The dominance of the positive ice-albedo feedback could explain the faster intra-ensemble spread growth (the decrease of

PPP; see formula 3.4) during the melting season, whereas the increase in PPP during the freezing season could originate from the negative ice thickness-growth rate feedback, acting as amplifiers and dampeners of the initial perturbations of the sea ice conditions, respectively (Tietsche et al., 2014). This seasonality in the signal is not present in MPI-ESM, for which SIE is not significantly predictable beyond the first December. For EC-Earth2.3 the PPP decrease is sharper after the first October, when the sea ice gradually spreads across the interior of the Arctic basin and peripheral seas outside the Arctic Ocean. The significant reemergence during the second melting season seems to be characteristic of EC-Earth2.3, although MIROC5.2 also presents a significant PPP the second and third July. This might be related to an "early" summer-to-summer predictability mode for July, as can be seen in their lagged correlation matrices (Fig. A.2 in the Appendix). This mechanism does not appear for the rest of models in July, but it does in September (Fig. A.2 of the Appendix). The summer-to-summer memory reemergence has its origin in the summer SIT memory (from the central Arctic) (Blanchard-Wrigglesworth et al., 2011a). Over three continuous years, the central Arctic September SIV and the SIE are correlated in September (Fig. 3.2, red line) for all models.

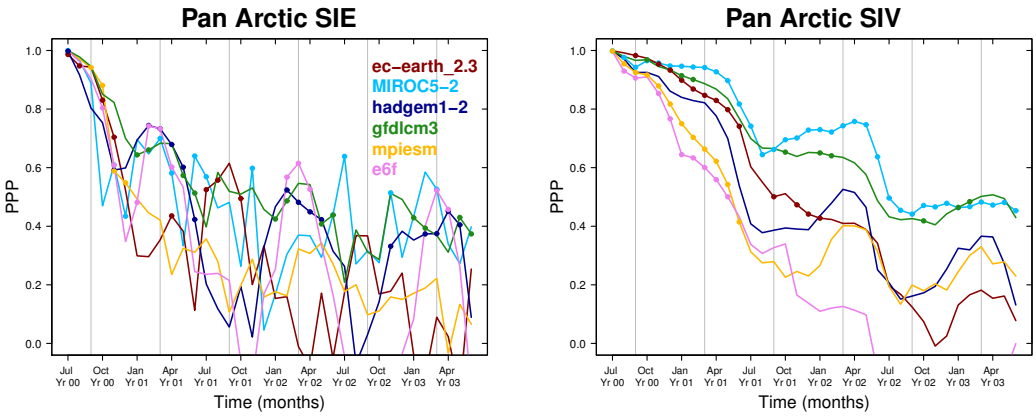


Figure 3.1: Potential predictability of the pan-Arctic **(a)** SIE and **(b)** SIV measured with the PPP of *IdealPred* using the natural variability of *ControlRun* as a reference. Dots indicate significant values at the 95% level, estimated with an *F-test*. September and March are marked by thin gray vertical lines.

The long-lasting *IdealPred* SIV potential predictability (Fig. 3.1b) is related to the persistence of the SIV, as shown by the lagged correlations calculated from the *ControlRun* (Fig. A.3 in the Appendix). The persistence of the SIV at the pan-Arctic scale arises almost entirely from the central Arctic SIV persistence (Blanchard-Wrigglesworth et al., 2011a), as

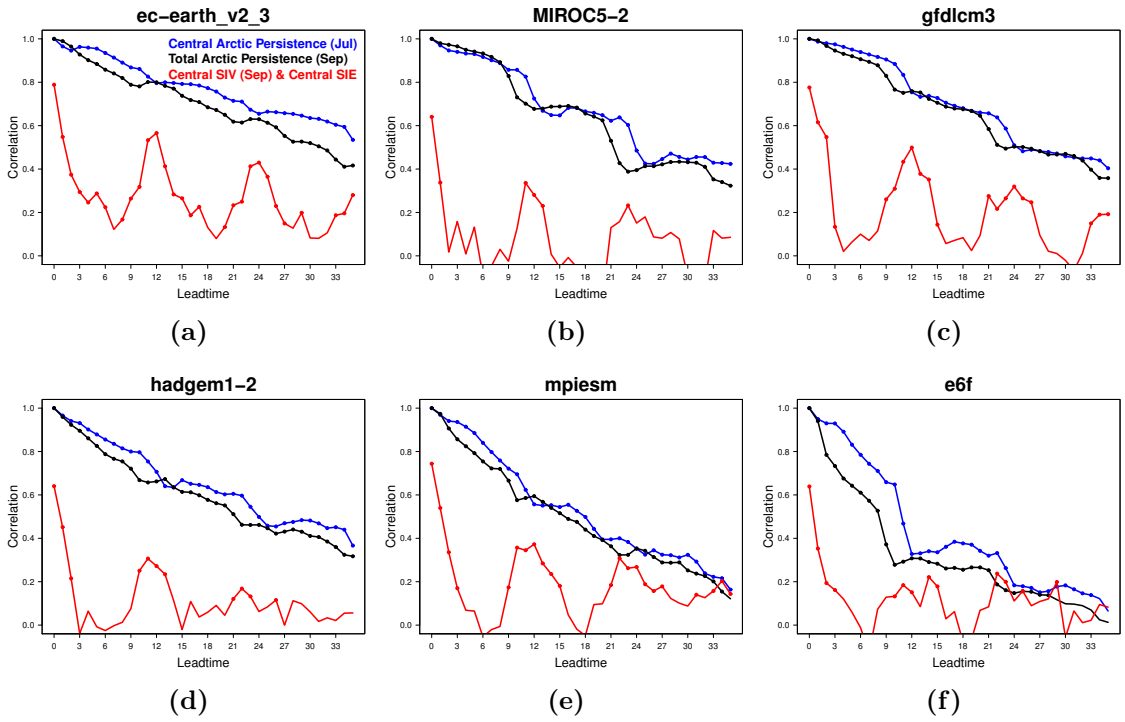


Figure 3.2: In blue, the July *ControlRun* lagged autocorrelation for the central Arctic SIV. In black, the September *ControlRun* lagged autocorrelation for the pan-Arctic SIV. In red, the lagged correlation between the September central Arctic SIV and the central Arctic SIE. Dots indicate significant values at the 95% level as estimated from a one-sided student-T distribution.

suggested when the lagged correlation of the central and pan-Arctic SIV are compared (Fig. 3.2, blue and black lines correspondingly).

In the following, we split the Arctic Ocean and surrounding basins considered in Fig. 2.1 into two groups, based on their seasonality: group one contains the peripheral basins including the summer ice-free regions (Barents Sea, Kara Sea, GIN seas, Irminger Sea, Baffin Bay, Labrador Sea, Hudson Bay, Bering Sea, Sea of Okhotsk and Chukchi Sea), and group two contains the internal Arctic seas, or the seas that are entirely ice covered during winter (the "central basins"; central Arctic, Canadian Archipelago, Beaufort Sea, East Siberian Sea, Laptev Sea).

Peripheral basins A regional analysis revealed large variations in the SIE and SIV potential predictability across the forecast systems (Fig. 3.3-3.4). Nevertheless, some common features can be extracted. For instance, in all the peripheral basins (except Hudson Bay and Chukchi; Fig. 3.3a–j) the PPP shows the same temporal pattern for the SIE as for the SIV (for each model individually), which reflects the correlation between ice concentration and

ice thickness in those regions with a thin ice cover. In the Barents, Kara and GIN seas, and in Baffin Bay, sea ice is present in July at the start of *IdealPred* for all models. The PPP initially decreases, before peaks of reemergence occur at different lead times depending on the model and basin. The Barents, Kara and Chukchi seas SIV PPP exhibit a significant predictability reemergence in summer for most models. This might be directly linked to the retreat/advance mechanism of predictability (Blanchard-Wrigglesworth et al., 2011a; Stammerjohn et al., 2012). We can cluster the GIN and Baffin Bay within the same group: there is an initial predictability drop followed by a memory reemergence in winter, which seems robust for all models. On the Pacific side (Bering and Okhotsk seas) sea ice is not present at the start of the predictions (except for HadGEM1.2). For these seas the PPP is noisier and less significant than in the Atlantic sector. We could not group the rest of peripheral seas because of the differences shown in the temporal variability of PPP between the models. This is mainly due to the differences in the mean sea ice state between the different models. To know more about these differences, please see Day et al. (2016). In the following section we consider these regions in greater detail in EC-Earth2.3.

Central basins. The SIE PPP in the interior basins other than the central Arctic (Fig. 3.3l–o) is null during the winter due to the extremely low sea ice variability. The central Arctic SIE PPP reflects how the different model sea ice conditions (and the cycle of variance) impact predictability. In most central regions, the PPP of the SIV continuously decays over time while remaining statistically significant up to 6-14 months (even until the third year for MIROC5.2 in the Canadian Archipelago). This suggests that the regional SIV is potentially predictable up to one year in advance for the seas with perennial sea ice. The significant summer reemergence of SIV PPP in the Beaufort Sea stands out for all models except for EC-Earth2.3 and HadGEM1.2. The central Arctic region exhibits the same PPP characteristics as the pan-Arctic region for the SIV, which is an indicator of the origin and sources of predictability of the pan-Arctic sea ice, as previously mentioned.

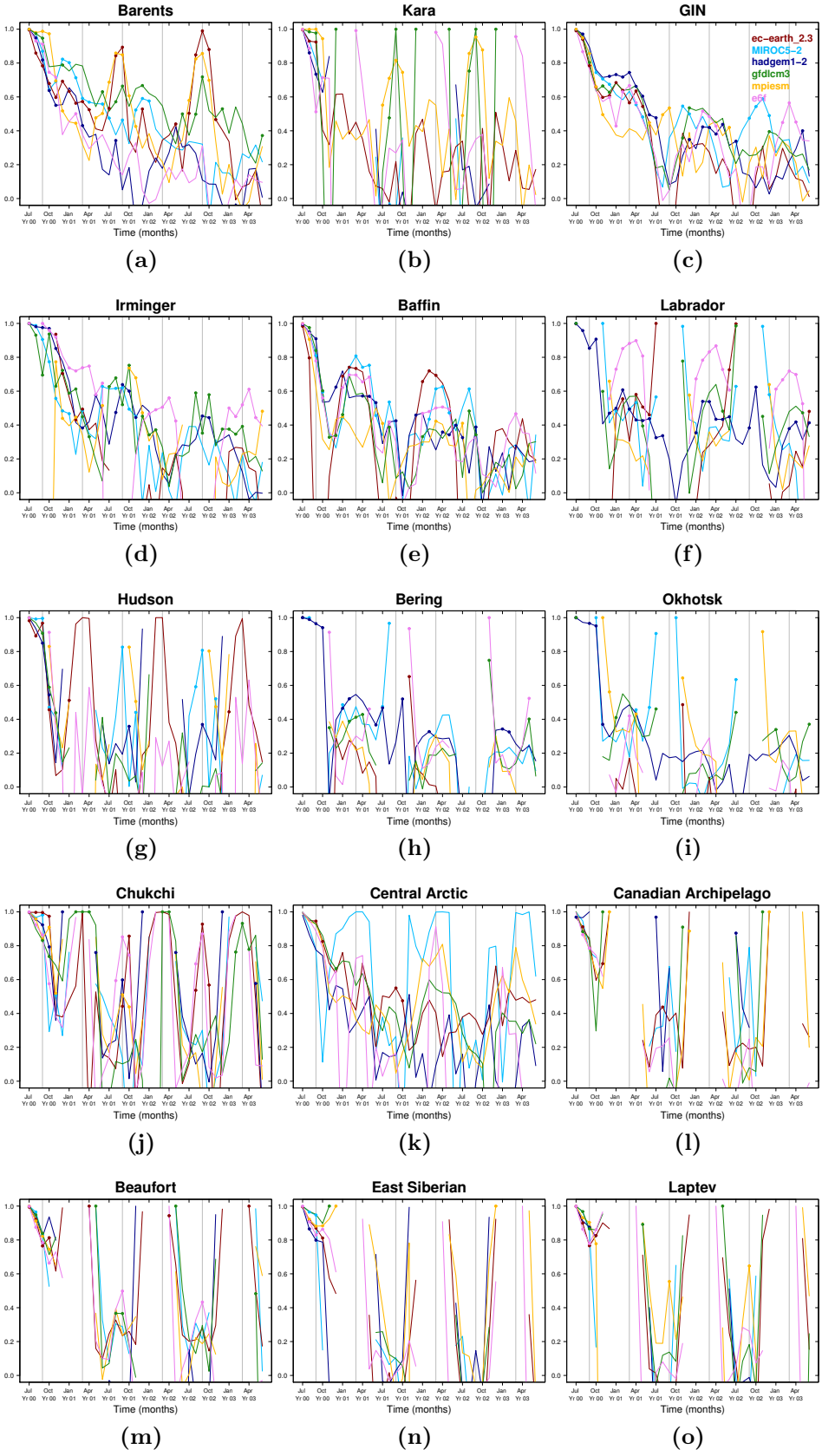


Figure 3.3: As Fig. 3.1a but for the different Arctic basins. The estimate of PPP is not considered as significant for any SIE anomaly in the control simulation which is lower than or equal to 1% of the average anomaly along the control for the same region and calendar month.

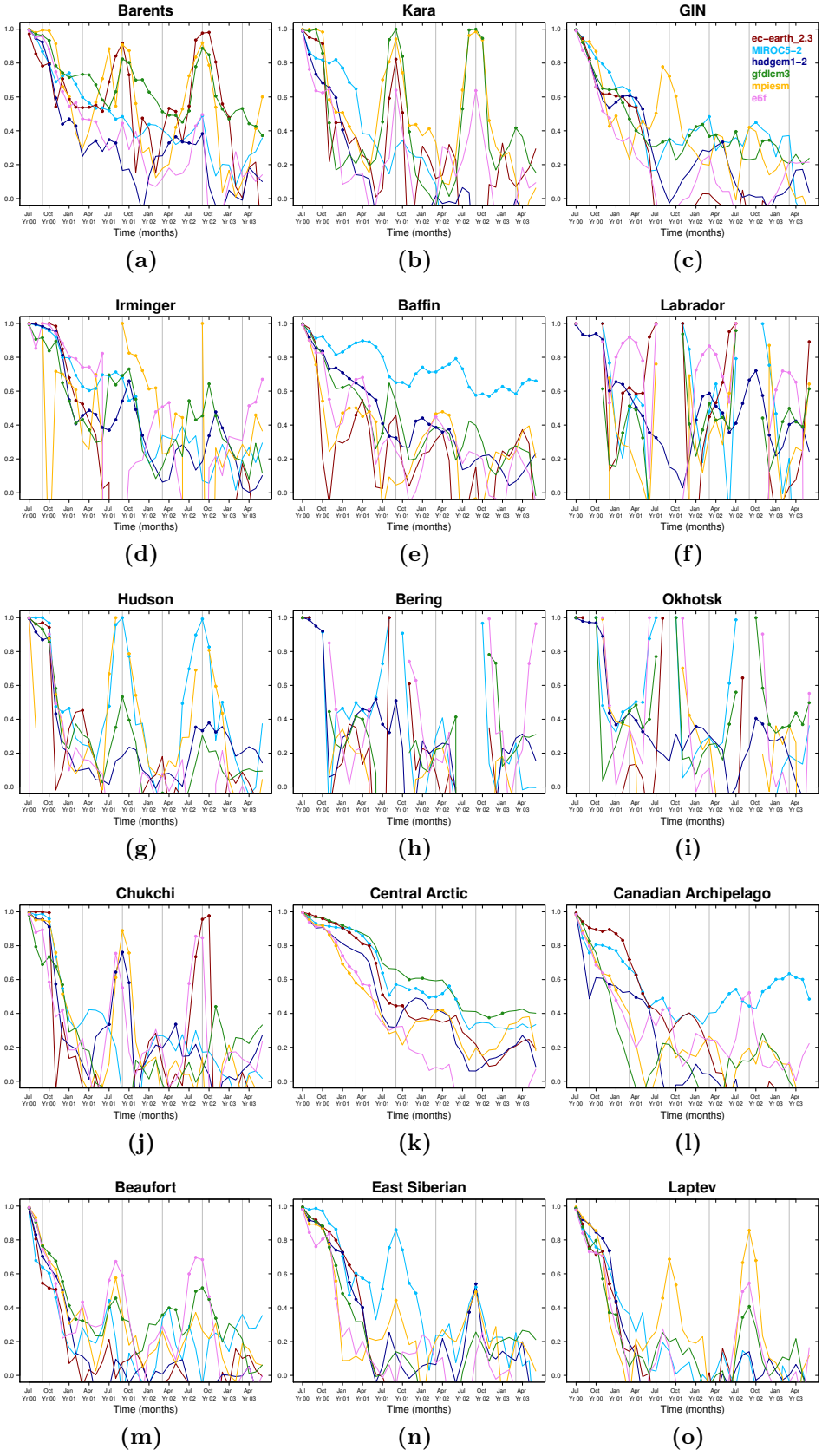


Figure 3.4: As Fig. 3.1b but for the different Arctic basins. The estimate of PPP is not considered as significant for any SIV anomaly in the control simulation which is lower than or equal to 1% of the average anomaly along the control for the same region and calendar month.

3.4 Mechanisms behind the regional sea ice predictability in EC-Earth2.3

In this section we focus on the predictability mechanisms of EC-Earth2.3 since a few regional predictability characteristics are specific -and distinct- for this model.

3.4.1 Ocean persistence in the Barents/Kara/GIN Seas/Baffin Bay

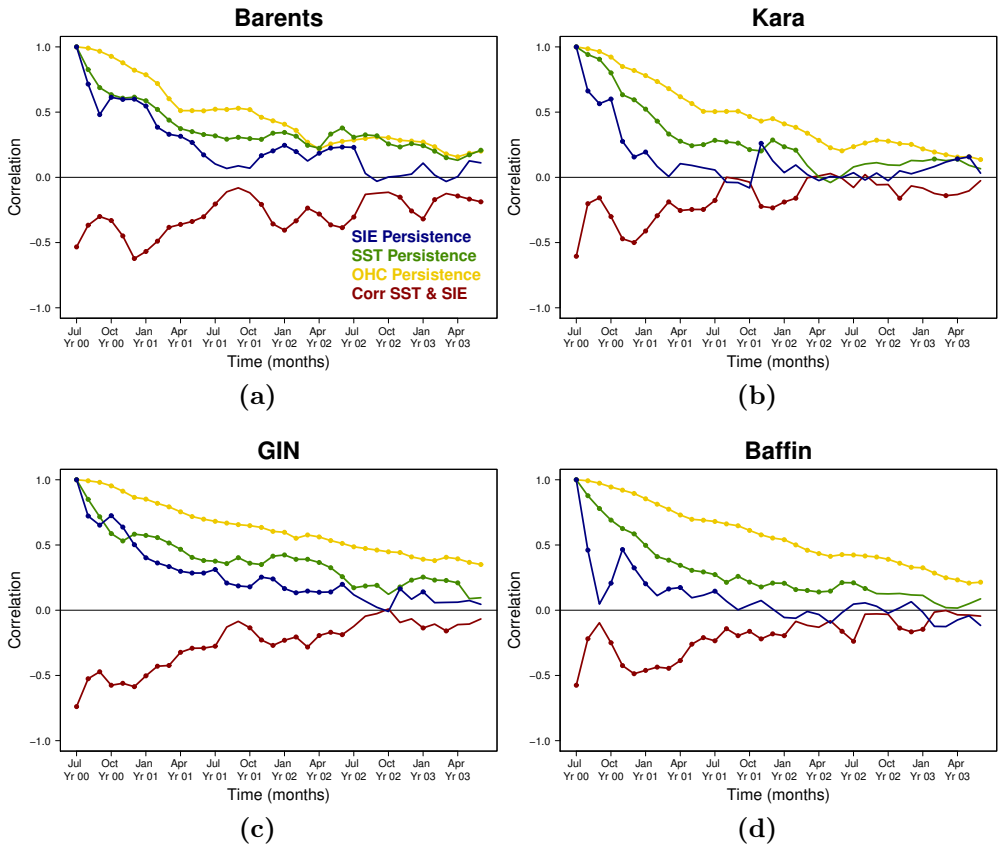


Figure 3.5: Persistence of the SIE (blue), the SST (green) and the OHC (0-300 m depth, yellow) in the **(a)** Barents, **(b)** Kara, **(c)** GIN seas and **(d)** Baffin Bay for EC-Earth2.3. In red, the lagged correlation between the July SST and the SIE for the same seas. Correlations were calculated using the *ControlRun* during the three subsequent years. Dots represent significant values at the 95% level as estimated from a one-sided student-T distribution.

In spite of the initial decay, the PPP still has significant values in the Barents, Kara and GIN seas and in Baffin Bay for the SIE and SIV up to 3 forecast years (Fig. 3.3/3.4a–c & e), despite some periods when significance is temporarily lost. The memory of the sea ice

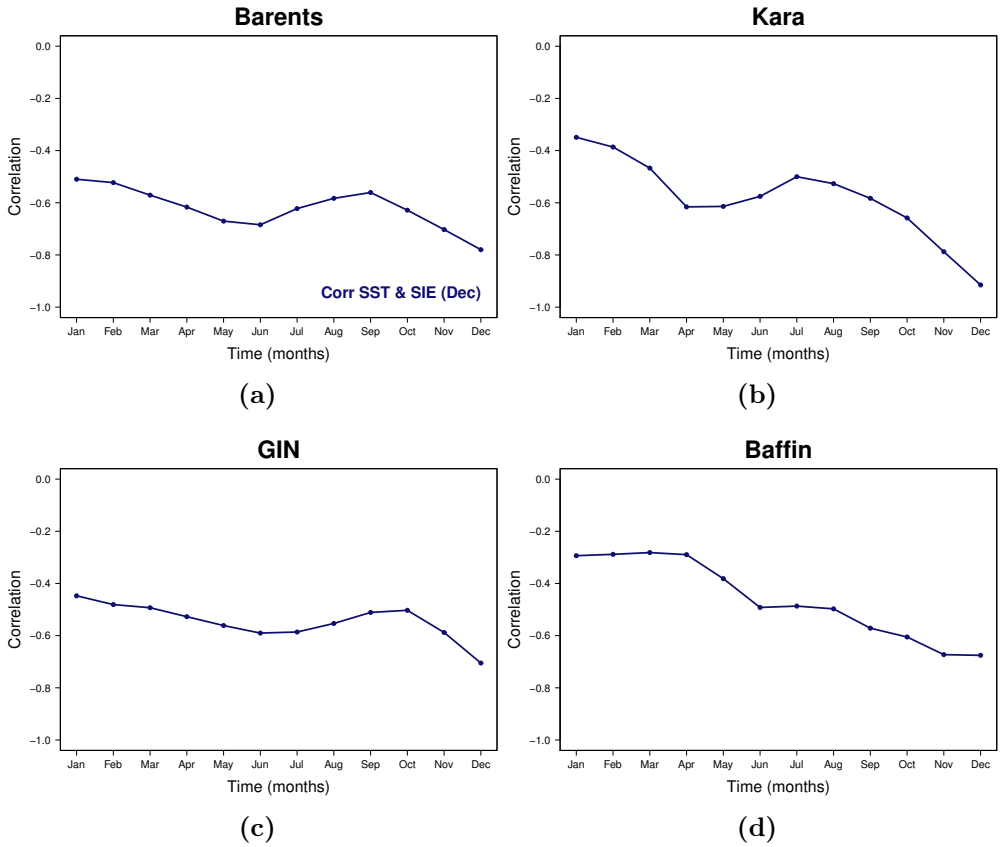


Figure 3.6: Correlation between December SIE and SST in the previous year in the **(a)** Barents, **(b)** Kara, **(c)** GIN seas and **(d)** Baffin Bay for EC-Earth2.3. Correlations were calculated using the *ControlRun*. Dots represent the significant values at the 95% level as estimated from a one-sided student-T distribution.

cover in these regions can be related to the long-term persistence of the July SST anomalies at the same location (Fig. 3.5, green lines), which varies in accordance with the ocean heat content (integrated over the first 300 meters depth; OHC hereafter). Thus, the memory of sea ice cover in the peripheral seas could be partially related to the ocean (Fig. 3.5, yellow lines).

The SIE lagged correlations (Fig. 3.5, blue lines) show a decaying but significant predictability over the first year and a melt-to-freeze reemergence the first and second winter. The SIE tends to be significantly correlated with the SST in winter (red line), which is consistent with the role of SST preventing or facilitating sea ice formation during the freezing season (the melt-to-freeze reemergence mechanism). The PPP highly depends on the start dates and the mean climate state, as we have already shown in section 3.2.2, so in this case the *ControlRun* provides a more robust idea of the regional predictability, but the mecha-

nisms are comparable. The SST during the previous spring provides predictability of the December SIE (Fig. 3.6a–c & e). The maps of the correlation between the grid point SST in December and the averaged December to February SIE in the Barents and Kara seas (Fig. 3.7a–b) agree with the time series in Fig. 3.6 .

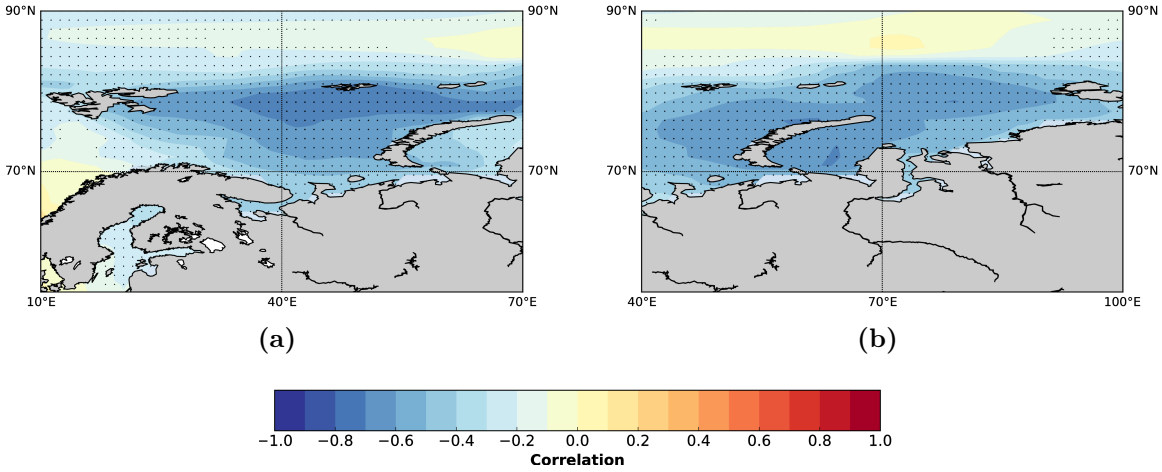


Figure 3.7: Maps of correlation between the gridpoint December SST and **(a)** Barents and **(b)** Kara SIE for December, January and February (mean correlation) for EC-Earth2.3. Black dots indicate significant correlations at the 95% level as calculated from a one-sided student-T distribution.

3.4.2 Ocean reemergence in the Labrador Sea

We investigated the peak of PPP in January–April in the Labrador Sea, which is not present in other peripheral seas and does not seem to project onto the changes in the pan-Arctic PPP in Fig. 3.1a. This peak cannot be attributed to a reemergence mechanism due to sea ice, since sea ice is not present at the start of the prediction in this area for EC-Earth2.3.

We tested the hypothesis that this memory reemergence had a remote origin. We calculated several backward water mass trajectories using the off-line mass preserving Lagrangian Ariane scheme (Blanke et al., 1999; Blanke and Raynaud, 1997). Here, 25 tracers were seeded uniformly, all at a 5-m depth (depth of the uppermost ocean level) in the Labrador Sea in February, and their origin was traced using the 40-year average monthly velocity fields from the *ControlRun*. The goal was to find the locations of those water parcels in the first July month, 8 (i.e. target February Year 1), 20 (i.e. target February Year 2) or 32 (i.e. target February Year 3) months before. The trajectories are shown in Fig. 3.8a–c.

The trajectories reveal that the water parcels present in the Labrador Sea at the time

of the first two winter SIE PPP peaks originate from the subpolar gyre area. During the first year, the local correlations, i.e. Labrador SST and OHC correlated with Labrador SIE (Fig. 3.8d, red and green lines), tend to show that the reemergence is related to the subpolar gyre persistence (Fig. 3.9a). The second year, the correlations between the Irminger Sea SST and OHC at the time of the initialization and the Labrador Sea SIE are higher than the local correlations, and they match exactly the time when the PPP reemergence in the Labrador Sea occurs (Fig. 3.8d, blue and black lines). For longer timescales, these parcels have their origins in the North Atlantic. Thus, the first winter peak of the PPP of the SIE in the Labrador Sea seems related to the subpolar gyre persistence, while the second is caused by SST anomalies that are advected from the Eastern Subpolar gyre.

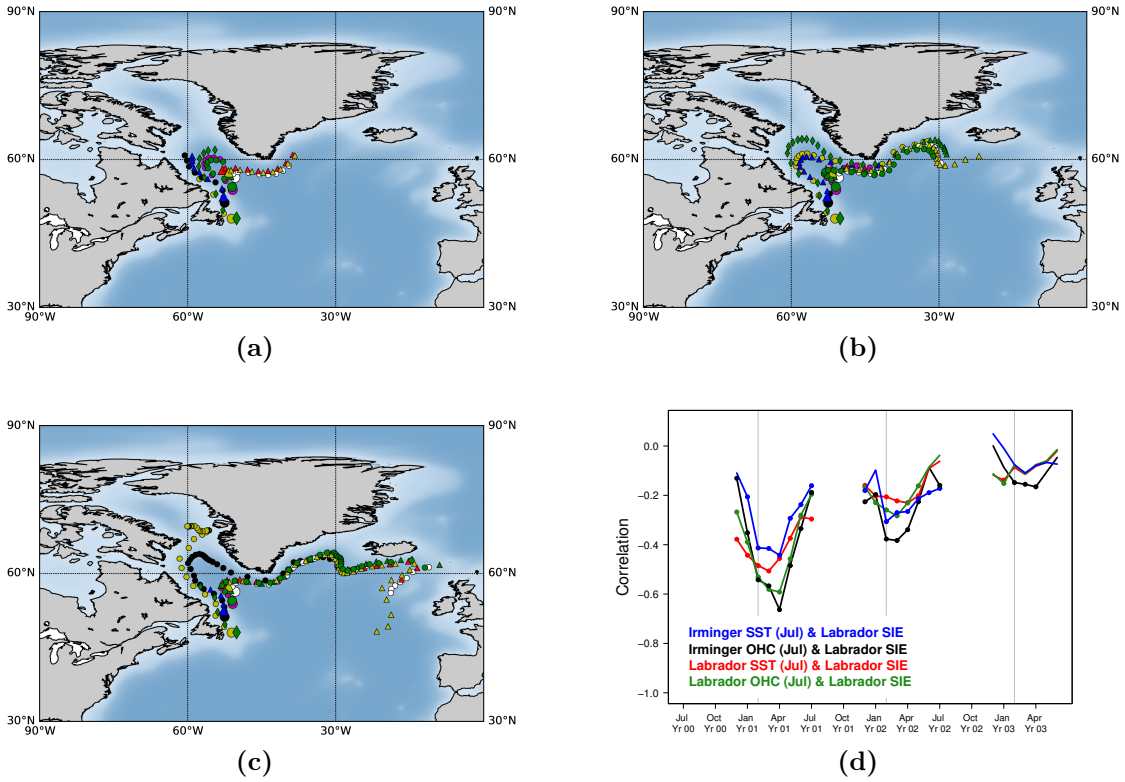


Figure 3.8: Map of the backward trajectories followed by water masses travelling from different locations in the Labrador Sea from **(a)** the first, **(b)** the second and **(c)** the third February back to the first July for EC-Earth2.3. Each lead time is marked with a dot, while the initial positions (corresponding February) are marked with bigger dots. **(d)** Correlation between the Irminger Sea SST, the Irminger Sea OHC, the Labrador Sea SST and the Labrador Sea OHC the first July and the Labrador Sea SIE the three following years for the *ControlRun*. Dots represent the significant values at the 95% level estimated from a one-sided student-T distribution. The vertical grey lines represent the months of February. The SST and OHC were integrated for the corresponding area in Fig. 2.1.

As a result of advective ocean processes, sea ice predictability in the Labrador Sea may be thus related to predictability at the other side of the subpolar gyre, propagated by the mean ocean circulation. Indeed, previous studies have highlighted the high SST predictability in the subpolar gyre area (Boer, 2004; Collins, 2002), including studies using the same climate model as ours (Wouters et al., 2013). Koenigk and Mikolajewicz (2009) confirmed that advection of SST anomalies may lead to an increase in the predictability of the Barents sea ice in winter. This result, consistent with previous studies, suggests that the initialization of the ocean is important when running real-time initialized sea ice forecasts.

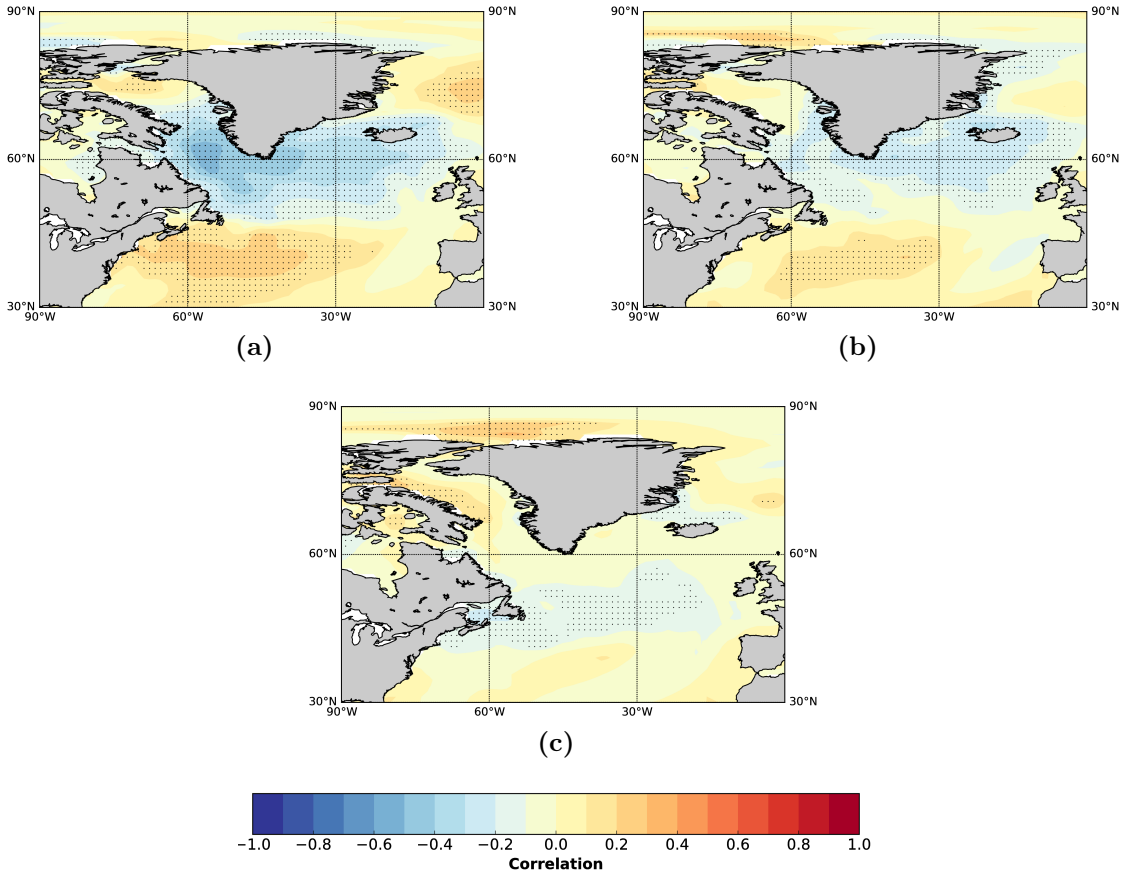


Figure 3.9: Maps of correlation between the local SST the first July and the Labrador SIE **(a)** the first, **(b)** the second and **(c)** the third February for EC-Earth2.3. Correlations were calculated using the *ControlRun*. Dots represent the significant values at the 95% level as estimated from a one-sided student-T distribution.

3.4.3 Sea ice thickness persistence in the internal Arctic basins

We mentioned above that the summer peaks of the PPP for the pan-Arctic SIE could be attributed to the persistence of the SIT in the central Arctic, as suggested by Blanchard-Wrigglesworth et al. (2011a). In September, anomalies of the pan-Arctic SIE are well correlated with anomalies of the SIT in the central Arctic, thus anomalies of the SIE re-emerge from one summer to the next due to the memory of the corresponding SIT anomaly.

In a similar way, peaks of the SIE PPP in the internal Arctic basins during the summers, can be linked to the persistence of the SIT, coherent with the long-lasting persistence of the SIV (Fig. A.1 in the Appendix). In these areas, little connection with the upper ocean should be expected, due to the insulating role played by the sea ice cover during most of the year.

3.4.4 Dependence on the initial state

Up to this point, the potential predictability of the pan-Arctic and regional SIE in EC-Earth has been assessed using the mean of all the predictions initialized at various start dates. However, it must be borne in mind that each prediction starts from a different climate state (Table A.1 in the Appendix for more details) and that we only have 9 start dates, which could impact the level of sea ice predictability. We now attempt at identifying how the sea ice and heat transport of the initial states impact the predictability.

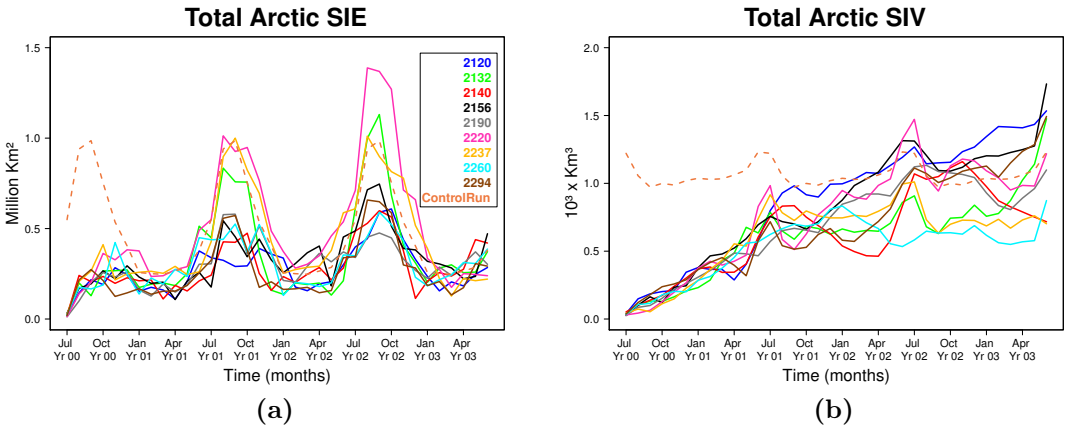


Figure 3.10: Total Arctic **(a)** SIE and **(b)** SIV ensemble spread for the nine *IdealPred* start dates. The temporal standard deviation of the *ControlRun* reference is shown in orange dashed lines.

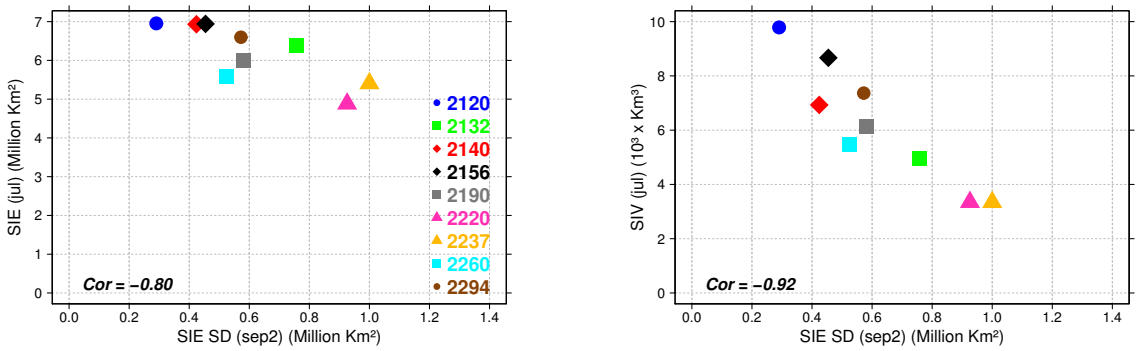
The potential predictability is estimated as the standard deviation (SD) of the long control simulation that can be explained by the intra-ensemble SD of the idealized predictions

(see section 3.2.2 for more details). The closer the ensemble SD is from the reference control simulation, the lower potential predictability will be. *ControlRun* SD and *IdealPred* SD for SIE (Fig. 3.10a) increase during the retreat of sea ice. The mechanisms behind the September SIE variance amplification were discussed by Notz (2009) and Goosse et al. (2009). The *ControlRun* SIV SD also experiences an early summer increase, followed by a slight decrease in September (Fig. 3.10b). For *IdealPred*, similar SIV SD peaks superimpose onto a slow increase in SD with the lead time. The slow SIV predictability loss (compared to SIE) results from the large SIV persistence (Fig. A.3a in the Appendix). We saw in Fig. 3.1 that SIE predictability on average is lost during the first winter and there is a reemergence the second summer. However, the reemergence in summer does not concern all of the start dates, but some have completely lost predictability such as 2237, 2220 and 2132. We are unable to estimate the significance of predictability on individual start dates (and only 8 members), so we rely partly on the average estimate of significance obtained from the mean of all the start dates. The other start dates are grouped and below *ControlRun* SD throughout most of the 36 months, without meaning that they are significant (unfortunately, no significance test was calculated for none of the SD values, but since there is no significance for the mean PPP beyond the second October, individual startdates are not expected to show significant SD neither). For the SIV, predictability seems to have been lost after the first summer for the 2120 start date. The loss of SIV predictability on average during the first spring seems to be mainly due to 2220 and 2237. The reemergence of skill in the second winter seems to concern most of the start dates except 2120, while the loss of predictability the second summer seems also mainly related to 2220, 2156 and 2120. Some of the other start dates could still have some predictability during this second spring and summer, such as 2237, 2132 and 2260, but we are unable to estimate the level of significance. The SD of the other start dates remain well below the *ControlRun* SD until the end of the forecast.

	SIE SD		
	SEP 1	MAR 1	SEP 2
SIE	-0.02	**-0.76**	**-0.80**
SIV	0.07	**-0.66**	**-0.92**
AHT	0.17	0.23	0.40

Table 3.1: Table summarizing the correlation between the SIE SD at different forecast times and the initial SIE, SIV and AHT. The significant correlations at 95% appear in bold. Significance of the correlations was calculated according to a one-sided student-T test.

Some start dates exhibit a lower SD than others for the same forecast time, even if all



(a) **(b)**

Figure 3.11: Scatter diagrams of the second September total Arctic SIE SD and the initial **(a)** SIE and **(b)** SIV states for the nine *IdealPred* start dates. The correlation coefficients are indicated at the left bottom.

of them are initialized from *ControlRun*. Correlations, across the different *IdealPred* start dates, between the initial SIE, SIV or AHT on one hand and the forecast SIE SD in the first September on the other hand (SEP 1) are not significant (Fig. 3.11a and Table 3.1). For the first forecasted March (MAR 1), there is a significant negative correlation between the initial SIE and the forecast SIE as well as between the initial SIV and the forecast SIE. These negative correlations increase in magnitude for the second September (SEP 2). No significant correlations are found with the initial AHT, suggesting that it is the initial sea ice state rather than the initial ocean that determines to which extent the ensemble members disperse along the forecast time, i.e. the level of predictability of the SIE. In conclusion, the thicker and more extensive the initial sea ice is, the larger the potential predictability of SIE will be.

3.5 Conclusions

In this chapter, we analyzed six different control experiments and their corresponding sets of 3-year ensemble predictions initialized on 1st July from the control experiment for various start dates. Using this perfect-model protocol, the main objectives of the study are to assess the regional sea ice predictability of the APPOSITE project models, to explore in EC-Earth2.3 the predictability sources and mechanisms for the SIE and SIV in different sub-basins of the Arctic Ocean, and to investigate how understanding the regional-scale mechanisms and the dependence on the initial state helps to clarify the predictability at the

pan-Arctic scale.

The potential predictability was estimated by measuring the growth of the ensemble spread in the idealized predictions, and comparing it to the natural variability derived from the control experiment. We also calculated lagged correlations in the control simulation, as a diagnostic approach to assess the persistence or lagged relationships in the control experiment and thus infer some mechanisms explaining the potential predictability in the idealized predictions. The comparison of the prognostic and diagnostic approaches indicated that lagged correlation is a good informative measure of SIE and SIV predictability.

We quantified to what extent the regional Arctic sea ice would be potentially predictable, if we had a perfect knowledge of the initial conditions of the predictions and the simulated processes matched perfectly the observed ones. More focus was put on the mechanisms behind EC-Earth2.3 predictability. For this model the regions could be clustered into three groups according to their predictability: the peripheral seas, the Labrador Sea and the interior Arctic basins.

The main conclusions from the multi-model analysis with the APPOSITE experiments are:

- Consistent SIE predictability reemergence is found in winter for HadGEM1.2, GFDL-CM3 and E6F, which could be related to the winter negative ice thickness-growth rate feedback (Tietsche et al., 2014).

- The SIV shows greater predictability, attributable to the long-lasting persistence of the SIT in the central Arctic for all models.

- The summer-to-summer reemergence of pan-Arctic SIE is consistently related in all models to the persistence of SIT anomalies in the central Arctic.

- The Baffin Bay and the GIN seas SIE exhibit a robust PPP signal among all models, characterized by a winter memory reemergence. For most models, there is SIV PPP reemergence in summer in the Barents, Kara and Chukchi seas. For the rest of the regions, we found significant inconsistencies, which we attribute to the differences in the average states of the sea ice.

- For EC-Earth2.3 and in the peripheral seas of the Atlantic Sector, significantly high PPP values over 1 year, including the 1-year reemergence, are driven by the persistence of local oceanic thermal anomalies (SST and OHC).

- In the Labrador Sea (for EC-Earth2.3), which is ice-free in July, the PPP peaks between January and April during the first year, as the result of the subpolar gyre ocean heat

persistence. However, the January to April peak of the second year seems to be related to the advection of ocean temperature anomalies from the Irminger Sea and the Eastern North Atlantic Ocean.

- In the interior Arctic seas in EC-Earth2.3, winter SIE potential predictability is trivial due to complete ice coverage. The SIV has a longer predictability in these seas than in peripheral seas as a result of the long SIT persistence.

- The larger the initial SIE and SIV in the predictions, the lower the SIE SD will be in the following March and September, i.e. the larger the potential sea ice predictability for about the first year and half.

The heterogeneity of regional mechanisms identified on EC-Earth illustrates that sea ice predictability can arise from a variety of different sources. For instance, we have seen that in EC-Earth2.3 the ocean is a potential source of predictability in the peripheral seas, while the SIT plays a dominant role in the interior seas. These results provide guidance for the design of operational forecasting systems: for lead times beyond a single season, the ocean initial state would play a key role in providing skilful forecasts in the marginal ice zone.

Moreover, there are some processes that were not investigated in this study, but that have been reported in previous studies as sources of sea ice predictability. For instance, the melt-to-freeze reemergence in the Barents and GIN seas has been related to the local SST memory (Bushuk et al., 2015; Schlichtholz, 2011). Other studies have shown that the winter Barents Sea SIA is highly correlated with the heat transport from the Atlantic waters through the Barents Sea Opening (Årthun et al., 2012; Nakanowatari et al., 2014; Onarheim et al., 2015). This inflow of warm water can also be driven by the atmospheric variability with a 1–2 year lag between the cyclonic anomalies and the ice response (Sorteberg and Kvingedal, 2006).

Previous studies have shown that different state-of-the-art coupled models exhibit similar sea ice predictability properties, like the melt-to-freeze or summer-to-summer correlation reemergence (e.g. Day et al., 2014). However, Tietsche et al. (2014) also suggested some model dependency, and that this related to specific processes: for instance, they showed that the representation of advective processes could be more model-dependent than the thermodynamic ones. The mechanisms herein documented for EC-Earth2.3 should therefore be interpreted with caution until similar diagnostics have been applied to other models. One important aspect is the possible role of model biases in shaping some mechanisms, especially on the Pacific side where there is virtually no potential predictability in EC-Earth2.3 (Guemas et al., 2014c).

By illustrating the dependence of predictability on the initial sea ice state, this work complements prior ones which addressed the dependence of predictability on the initialization month (Blanchard-Wrigglesworth et al., 2011a; Day et al., 2014) and on the mean climate. For instance, Goosse et al. (2009) suggested an initial decrease of the predictability of summer Arctic SIE due to increased variability during the twenty-first century. The present study provides further insight into possible regimes of the pan-Arctic sea ice cover variability in the future. Summer-to-summer reemergence in the pan-Arctic SIE PPP is due to the presence of perennial sea ice surviving the melt season. In a warmer climate, predictability of the Arctic sea ice cover may be closer to that of the peripheral seas, with predictability shown in this analysis suggesting that would be dominated by more ocean-related mechanisms.

4 An anatomy of Arctic sea ice forecast errors

4.1 Introduction

In the previous chapter, potential predictability has been identified up to 1.5 years in some Arctic regions. This potential predictability can be regarded as an upper bound of what can really be achieved in real time predictions, which face issues associated, among other things, with the correct initialization of the forecast system. The initialization methods typically used for seasonal climate prediction have been introduced in Chapter 1 (section 1.5.1). These can introduce different forecast errors, which are discussed in the following.

Forecast errors can arise from multiple origins, from the most commonly studied model drift to specific methodological effects due to, e.g., initial incompatibilities between the starting conditions of the different model components. The model drift manifests as a transition of the forecast from the observed climate state provided at initialization towards the model attractor (e.g. Sanchez-Gomez et al., 2016). This occurs at different rates depending on the location and the process, as well as their representation in the model and the amplitude of the inherent model biases. To obtain the highest possible forecast skill, we need as much observational information as possible in the ICs. However, the more observational information we include in these ICs, the further the initial state becomes from the model attractor, and therefore the stronger the drift will be. It becomes a trade-off between the observational initial information and the distance to the model attractor.

On the other hand, initial inconsistencies typically derive from the mismatch of ICs of two or more model components (e.g. Mulholland et al., 2015) due to their different origins. Inconsistencies could also originate from physically-inconsistent interpolations of the initialization products, like the ocean temperature and salinity giving different density profiles and leading to instabilities. These sources of errors interact with one another and disentangling them represents a challenge. This is nonetheless essential to ascertain their

consequences on the forecast quality. Ultimately, the ability of a forecast system to provide trustworthy predictions will depend jointly on the quality of the model and of its ICs.

In this chapter, we aim at decomposing the integrated Arctic sea ice forecast errors into contributions from two different sources in the latest seasonal predictions produced with EC-Earth3.2 (Acosta Navarro et al., 2019). In particular, we are interested in the development of the systematic error as a function of forecast time and the presence of initialization shocks, as well as their relative importance along the prediction. This paper is organized as follows: In Section 4.2, the EC-Earth3.2 system and the experimental and observational datasets are briefly described, together with the initialization methodology. Section 4.3 offers a detailed presentation of the different forecast errors. Their evolution and contribution to the total forecast error are discussed in Section 4.4. Finally, Section 6.1 summarizes the main conclusions of this chapter.

4.2 Methodology

4.2.1 Model description and experimental setup

The forecasts (also referred to as predictions) and their historical counterpart used in this chapter were produced with EC-Earth3.2 climate model (Doblas-Reyes et al., 2018; http://www.ec-earth.org/). The reader is referred to Chapter 2 for more details about the model. Two sets of seven month-long seasonal forecasts (*PRED* hereafter) were initialized each year from 1993 to 2008 on May 1st and November 1st, respectively, with an ensemble size of ten members. The historical experiment (*HIST*) consists of one single member.

4.2.2 Products for Forecast Initialization

The atmosphere ICs were interpolated from ERA-Interim reanalysis (Dee et al., 2011). The ocean ICs were also regridded vertically from the ORAS4 (Balmaseda et al., 2013). *ORAS4_ice* will refer to the sea ice product used to produce the ocean reanalysis ORAS4. The sea ice ICs were produced with a NEMO-LIM standalone reconstruction (hereafter referred to as *RECON*) that assimilates full-field daily SIC from ESA (Hollmann et al., 2013) the last day of every month via a 25-member EnKF (only the first 10 members were used as the ICs for the sea ice). The choice of 25 members resulted from a compromise of having a large enough ensemble to sample initial condition uncertainty, while minimizing the high computational constraints. The atmospheric fields used to force this reconstruction were taken from the Drakkar forcing set in its version 5.2 (DFS, Dussin et al., 2016). Note that it

is not possible to initialize directly LIM3 from observations, as this would require comprehensive and coherent information on many sea ice variables, which are not all conveniently observed. The advantage of the EnKF method is that through the use of model covariances, it assimilates specific observations (in this case of SIC) ensuring at the same time that the other sea ice and ocean variables evolve consistently with them. For further details on how the EnKF is implemented in the sea ice model, we redirect the reader to Mathiot et al. (2012) and Massonnet et al. (2015). Our study focuses on 1993-2008, the period covered continuously by the ESA SIC product.

4.2.3 Forecast Error Assessment

To validate the forecasts it is desirable to consider an independent dataset to the one used to produce the initial conditions. We used the satellite data from the NSIDC (Cavalieri et al., 1996). Since no uncertainty is provided for the NSIDC datasets, we estimated it as the absolute difference between the two NSIDC SIC products: NASA-Team (Cavalieri et al., 1996) and Bootstrap (Comiso, 2017), which only differ in the algorithm employed to estimate sea ice concentrations from passive microwaves measurements (the acronym NSIDC will refer to NSIDC NASA-Team hereafter). Also, to evaluate the assimilation efficiency of the EnKF technique, RECON is compared with the same ESA data used for the assimilation.

The characterization of SIC errors was complemented through the evaluation of the IIEE (Chapter 2). To track the contribution of the different sources of error to the total bias along the first month, we compute correlations between the spatial patterns that characterise the different sources of forecast errors, after applying a Gaussian interpolation to give the same weight to each grid cell, and additionally masking out from 80N to 90N due to the extremely low sea ice variability at those latitudes.

4.2.4 Limitations of the assimilation procedure

A large systematic overestimation of Arctic SIC with respect to NSIDC (PRED minus NSIDC) is already present during the first forecast day in most of the peripheral Arctic regions for the May-initialized predictions (Fig. 4.1a) and across the interior Arctic ocean for the November-initialized predictions (Fig. 4.1c). In both cases (May and November 1st, i.e. the forecasts in the first 24 hours) the largest bias ($\sim$50%) occurs at the Eastern side of the Greenland Sea. These large errors are due to the assimilation protocol itself as a consequence of large observational uncertainties over the region. Indeed, Fig. 4.1b,d illustrate the initial difference between RECON (from which sea ice ICs were extracted) and the

target observations from ESA that it assimilates, respectively for the first forecasted day of May and November. Large differences appear in the same regions where the initial forecast errors emerged. Indeed, more weight has been given to the model information than to ESA SIC data over those areas in the EnKF that produced RECON. To understand this, it is important to keep in mind how the EnKF works. The strength of assimilation relies on a subtle balance between the observational uncertainty and the simulated ensemble spread, which is assumed to represent the model uncertainty. Whenever observations are uncertain and the EnKF ensemble has little spread, the EnKF reanalysis remains close to the simulated model state. This happens, for example, in May and November along the Eastern Greenland coast, where ESA data exhibit large uncertainties (Fig. 4.2). By contrast, large model spread with well constrained observations lead to large EnKF increments and therefore closer EnKF reanalysis values to the assimilated data. This occurs in November over the Kara Sea where the model shows a large ensemble spread while there is relatively small observational uncertainty (Fig. 4.2). This large spread over the Kara Sea mostly derives from uncertainties in the DFS atmospheric forcing used to drive the different EnKF members (not shown). Larger EnKF increments are also expected in regions where the target assimilated data and RECON have large climatological differences. In regions exhibiting both important observational uncertainties and large model ensemble spread, the strength of the assimilation depends on the weights the EnKF associates to each of them. Regions with the largest observational uncertainty seem to be consistent in the NSIDC and ESA products (Fig. 4.2). These considerations suggest that in the assessment of forecast errors it is important to take into account the observational uncertainty to avoid overemphasizing model errors.

Given the large magnitude of the initial forecast errors in Fig. 4.1 with respect to NSIDC, in the remainder of the manuscript the errors will be quantified relative to RECON, which stands as the best possible estimate of the real climate state since it is thought to properly balance the observational uncertainty and the model uncertainty to produce such estimate. Our objective is to disentangle the different contributions to the bias. For each month, RECON runs in free-running mode until the last day of the month, in which satellite data is assimilated by the EnKF. This can introduce some discontinuity with respect to the past month trajectory. For this reason, we focus on the first 30 forecast days, i.e. May and November, before the next assimilation phase happens.

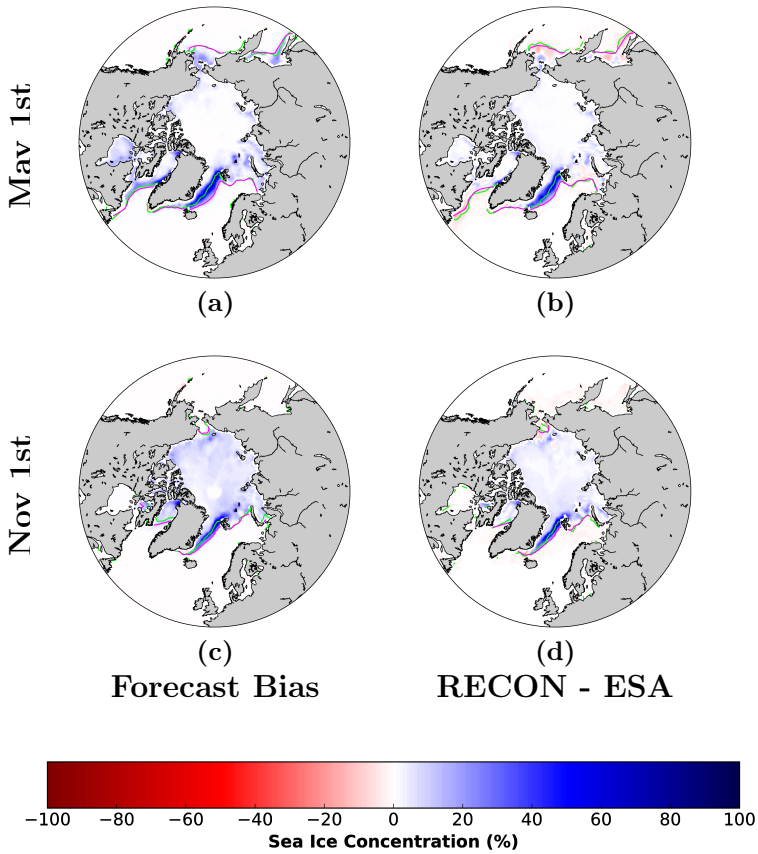


Figure 4.1: Maps of the SIC bias (PRED vs NSIDC) for **(a)** May 1st and **(c)** November 1st for the period 1993-2008. The purple line represents the model sea ice edge (15% SIC) climatology, while the green one represents the NSIDC sea ice edge. NSIDC is not the dataset assimilated for the production of the sea ice ICs. **(b)** and **(d)** show the differences in SIC between RECON and the target observations from ESA for May 1st and November 1st, correspondingly. The purple line represents the ICs sea ice edge, while the green one represents the ESA sea ice edge. In (a) and (c) blue colours represent regions with a larger sea ice in PRED than in NSIDC, while for (b) and (d) they represent regions where RECON simulates more sea ice than ESA.

4.3 Characterisation of the forecast errors

Hereinafter we will use *ICs inconsistency* to refer to the initial shock derived from the incompatibility between the sea ice and the ocean initial states (RECON for the sea ice and ORAS4 for the ocean). The adjustment towards the model attractor will be called *Systematic model error*, which will be the difference between HIST and RECON. The *Forecast error* will be defined as the difference between PRED and RECON, which will be an error deriving from the sum of the ICs inconsistency and the drift to the systematic model error.

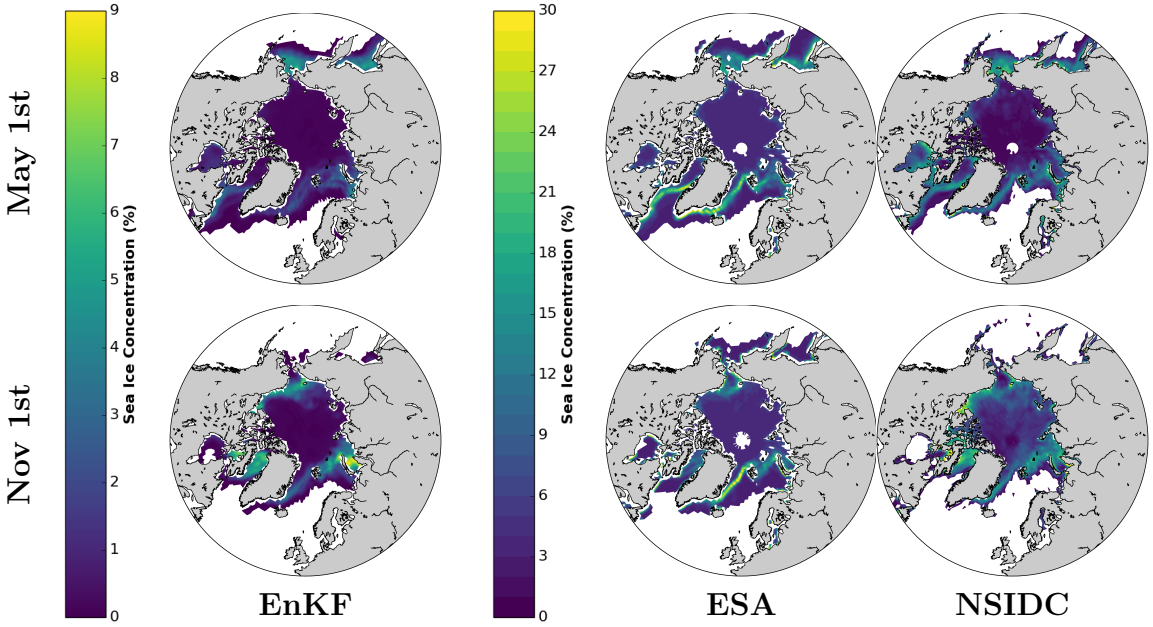


Figure 4.2: Maps of the EnKF ensemble standard deviation (10 members) and the SIC uncertainty for the ESA and NSIDC observational products for the period 1993-2008. For ESA it was calculated as the SIC standard deviation, while for NSIDC it was measured as the difference between the two NSIDC products (Bootstrap and NASA-Team). Note that the colorbars are not the same for the model and the observations.

4.3.1 Systematic model error

PRED is known to experiment a drift as it adjusts towards its own (and biased) climatology. The rate at which PRED drifts depends on how far the model is initialized from its attractor. The analysis of HIST allows to determine this attractor and therefore the systematic model biases. On the 1st of May, the climatological sea ice edge in HIST is displaced towards the north with respect to RECON in regions like the Barents and Bering Seas where the SIC are therefore largely underestimated. The opposite occurs in the Labrador Sea (Fig. 4.3b). Fig. 4.3e shows that, on the 1st of November, the HIST climatology has considerably less ice than RECON around the Arctic continental margin, in particular over the Kara Sea, and produces more sea ice in the Baffin Bay. The drift will gradually bring the predictions closer to these systematic model errors, with regional differences in speed.

4.3.2 Inconsistency between the initialization products

Using products from independent origins for initialization can introduce some shocks, as the different components adjust to each other in the first days of the forecast. We focus here on

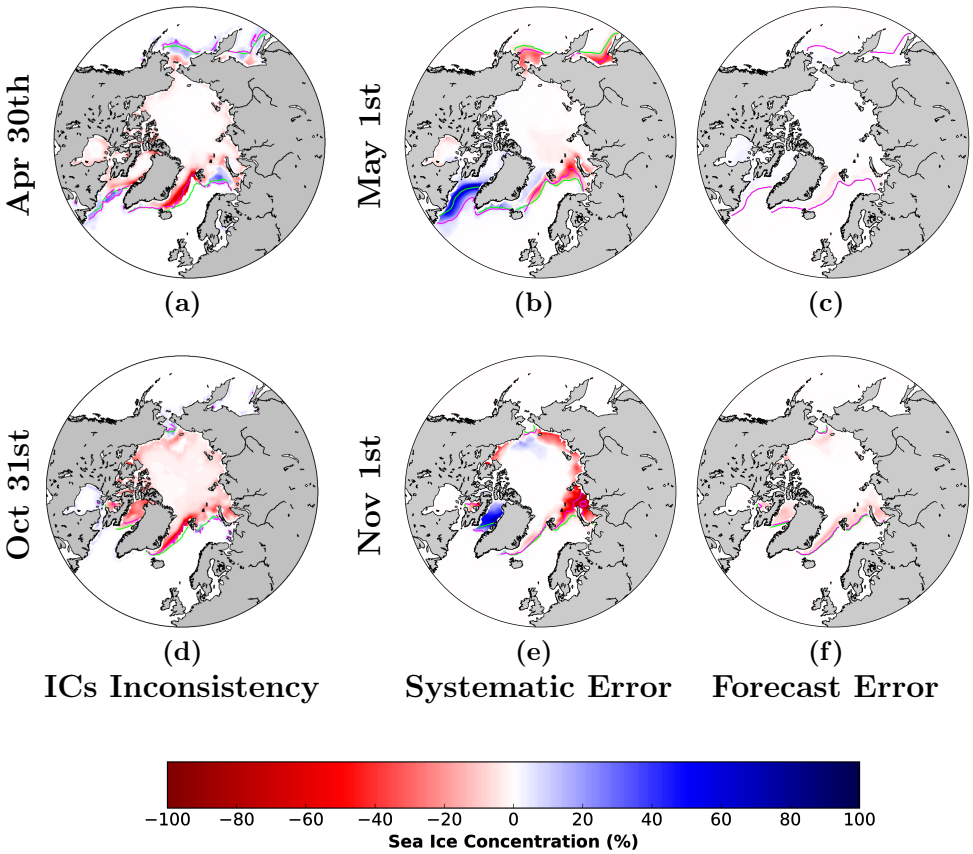


Figure 4.3: Maps of the SIC difference between the *ORAS*4_*ice* and RECON (*ORAS*4_*ice* minus RECON)the 30th of April/31st of October **(a,d)**, HIST and RECON (HIST minus RECON) **(b,e)** and PRED and RECON (PRED minus RECON) **(c,f)** the 1st of May/1st of November respectively averaged over the period 1993-2008. The green line represents RECON sea ice edge (15% SIC) climatology, while the purple one represents the *ORAS*4_*ice* (a,d), HIST (b,e) and PRED (c,f) sea ice edge. For all panels, red colours represent areas where RECON has larger SIC than (a,d) *ORAS*4_*ice*, (b,e) HIST and (c,f) PRED, while blue colours show lower SIC for RECON.

the inconsistencies of both the atmosphere and ocean ICs with those in the sea ice. There exist some notable differences between the sea ice in RECON (used as sea ice ICs) and the sea ice that was used to produce the ocean and atmospheric reanalyses (ORAS4 and ERA-Interim, respectively). If we use the sea ice from *ORAS*4_*ice* as a reference (which is really close to that from ERA-Interim), we can see that the assimilation tends to generate more sea ice in RECON both on the 30th of April and 31st of October (Fig. 4.3a,d), especially along the Eastern Greenland Coast (from Iceland to Svalbard). This reflects that the ocean state in ORAS4 is not perfectly compatible with the overlaying sea ice from RECON, and ORAS4 ocean state will tend to melt the RECON sea ice, while the RECON sea ice will

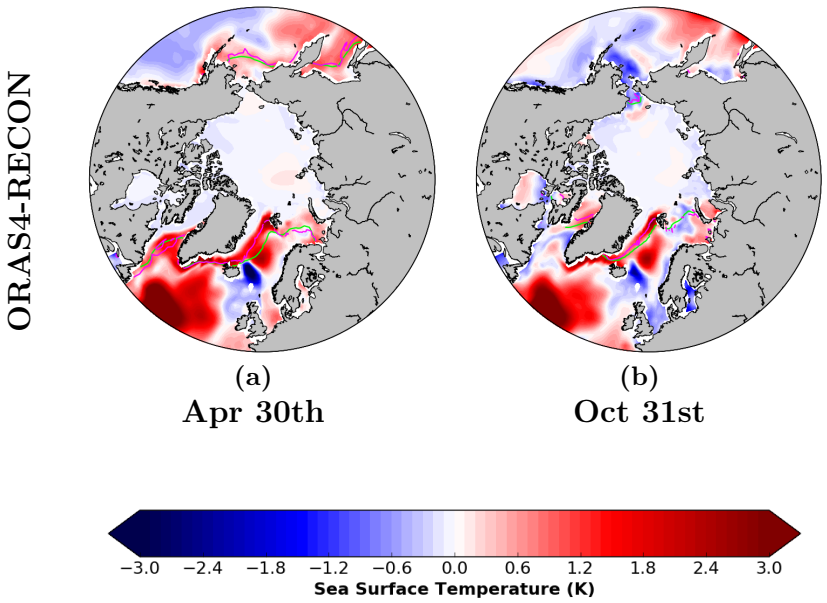


Figure 4.4: Maps of the SST difference between ORAS4 and RECON restarts for the period 1993-2008 for **(a)** April 30th and **(b)** October 31st. The green line represents RECON sea ice edge (15% SIC) climatology, while the purple one represents the *ORAS*4_*ice* sea ice edge.

tend to cool down the ORAS4 ocean state.

We remind here that as already mentioned in section 4.2.2, it is not possible to initialize the sea ice component of EC-Earth directly with a given observational dataset of SIC, like the product used to drive ORAS4, as the initialization of the sea ice model requires providing the simultaneous state of about 50 different variables which need to be physically consistent to avoid shocks. The excess of sea ice along the Eastern Greenland coast in RECON with respect to *ORAS*4_*ice* is associated with substantially colder local temperatures in the first (Fig. 4.4a,b), which implies that in the predictions the relatively warm waters ingested from ORAS4 will act to melt part of the sea ice initially imposed above, as already seen in Fig. 4.3c,f for day 1. Similarly, colder surface conditions in ORAS4 than in RECON, such as in the Bering Sea will favour sea ice formation early in the forecast. An inconsistency between *ORAS*4_*ice* SIC and SST appears in the Labrador, Barents and Bering seas and the Sea of Okhotsk, on April 30th (Fig. 4.3a & 4.4a). This issue will be explained in more details in the next chapter.

Inconsistencies between the initialization products for the sea ice and the atmosphere may also lead to some initial shocks. The atmospheric forcing of RECON is DFS, which is based on ERA-Interim but includes some bias corrections in temperature and humidity in the

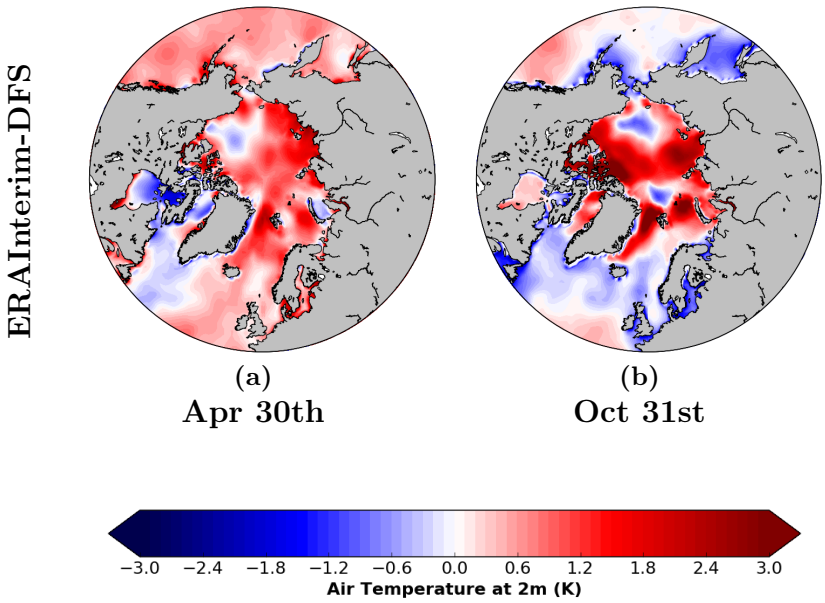


Figure 4.5: Maps of the 2-metre air temperature difference between ERA-Interim and DFS for the period 1993-2008 for **(a)** April 30th and **(b)** October 31st.

Arctic that reduce the mean air temperature by more than 0.6°C everywhere in the Arctic except in the Baffin Bay (Dussin et al., 2016). Due to these corrections, we should therefore expect during the first forecast days an overall melting effect of the atmosphere on the sea ice (that was produced with colder atmospheric conditions). The expected patterns of response can be deduced from the comparison of the 30th April/31st October air temperature fields in ERA-Interim and DFS (Fig. 4.5). These show that ERA-Interim is warmer than DFS (and therefore RECON) everywhere in the Arctic except in the Canadian Archipelago, Baffin and Hudson Bays for April 30th, and except over the Bering Sea, Sea of Okhotsk and part of the Canadian Archipelago for October 31st. The contribution of both ice-ocean and ice-atmosphere inconsistencies to the development of forecast errors will be explored in Section 4.4.2.

Additional inconsistencies could come from using atmospheric ICs produced with a different atmospheric component from the one used for the forecasts. EC-Earth3.2 includes IFS-cycle36r4 in the atmosphere, which is a more recent version than the one used for ERA-Interim (cycle31r2).

4.4 Understanding how the forecast errors develop

4.4.1 IIEE insights on the pan-Arctic sea ice biases

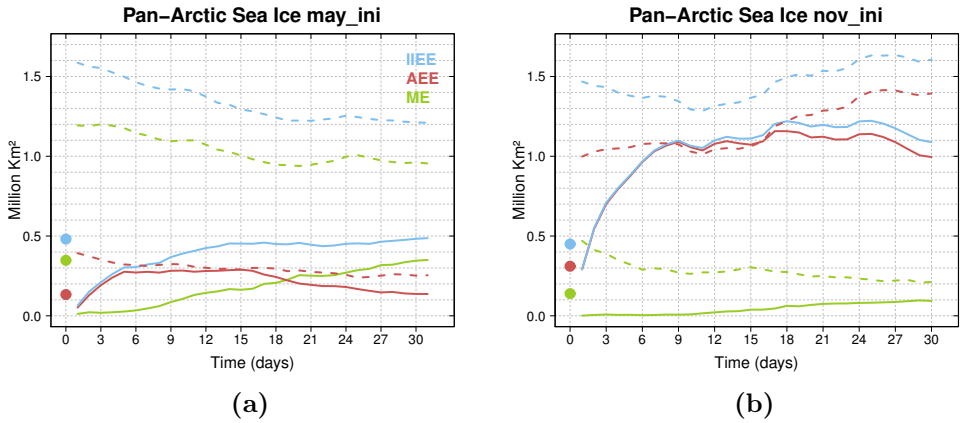


Figure 4.6: IIEE, AEE and ME as computed for the forecast error (solid line, PRED with RECON as reference), the systematic model error (dash line, HIST with RECON as reference) and the ICs inconsistency (dots, *ORAS*4_*ice* with RECON as reference). IIEE, AEE and ME are absolute errors.

The total SIE error (IIEE) can be decomposed into a contribution from the total sea ice heat budget leading to a general over/underestimation (AEE) and a contribution from incorrect processes leading to incorrect sea ice locations (ME). In May, the PRED IIEE develops slowly (Fig. 4.6a) and is still less than half of the HIST values by the end of the first forecast month, which is in stark contrast with the quick pan-Arctic SIA stabilization around the systematic model error (Fig. A.4 in the Appendix). Indeed, the IIEE in HIST is dominated in May by the ME, which is associated with an overestimated SIC in Labrador Sea/Baffin Bay and Chukchi Sea and an underestimation along the rest of the sea ice edge (Fig. 4.3b). The AEE evolution is consistent with the SIA one (Fig. A.4 in the Appendix). The development of IIEE in November is also fast (Fig. 4.6b) and mainly associated to an increase in AEE (reaching the HIST asymptotic level in about one week), which is consistent with the overall pan-Arctic SICs underestimation (Fig. A.4 in the Appendix). During this month, both the systematic model error and the ICs inconsistency are also dominated by the AEE. ME also grows throughout the month, but much more steadily and at a much lower rate, and results from overestimated SIC in Labrador and Greenland seas and Hudson Bay and underestimation along the rest of the sea ice edge (Fig. 4.3). In both May and November, the drift towards the model attractor (Fig. 4.3b,e) and ICs incompatibility (Fig. 4.3a,d) lead

to an overall sea ice underestimation (i.e. red areas dominate over the blue ones in Fig. 4.3). During the freezing season, the overall sea ice cover underestimation is reached as fast as during the melting season although the systematic model error has larger amplitude. These integrated diagnostics do not provide a comprehensive description of regional errors and their compensation which would allow to understand the ME development. These require the spatial SIC maps described in the next section.

4.4.2 Spatial evolution of the forecast errors

The spatial maps of SIC errors in May (Fig. 4.7) evidence that the ICs inconsistency could explain large forecast errors over the Greenland Sea. The SIC forecast errors are also quite close to the ones from the systematic model error by May 30th in sign and pattern (Fig. 4.7). However, for the ICs inconsistency the pattern has a larger longitudinal extension than the one from HIST. The PRED pattern of errors seems closer to the ICs inconsistency one before May 30th. In other regions, like the Barents and Kara Seas, the systematic model error dominates the forecast error in less than 10 days. By forecast day 30, the systematic model error is not yet fully developed in the forecasts, in particular in the Atlantic sector, which corresponds to regions of maximum SIC systematic model errors. For May, both the forecast and the systematic model error show a consistency between SST and SIC: warm (cold) SSTs are associated with loss (excess) of ice (Fig. 4.7 and Fig. A.5 in the Appendix). As in May, both the systematic model error and the ICs inconsistency exhibit a sea ice deficit in the Greenland Sea in November which resembles the pattern of SIC error in PRED (Fig. 4.8). The ICs inconsistency seems to explain the lack of sea ice in the forecast over the Baffin Bay, that is still present by the end of the month. The SIC errors tend to be consistent with the SST errors (Fig. 4.9): warm SST are associated with less SIC, and vice versa with the exception of the Hudson Bay and Barents and Chukchi Seas (Fig. 4.8, left column).

Estimate of SIV melt by the inconsistency between RECON and ORAS4 ICs A robust assessment of the contribution of the ICs inconsistency to the development of forecast errors require a quantification of the SIV that could be melted by the warmer ocean in ORAS4 than in RECON (orange line, Fig. 4.10a,b,c). The ICs inconsistency will lead to a cooling of the ocean and a melting of the sea ice until a new equilibrium is reached. This heat transfer from the ocean towards the sea ice could be superimposed with an additional model drift and internal variability in each forecast. To quantify the impact of the ICs inconsistency, we hypothesize that only the ocean mixed layer contributes to sea ice changes in the first month, and estimate its potential melting effect by computing the difference in heat content

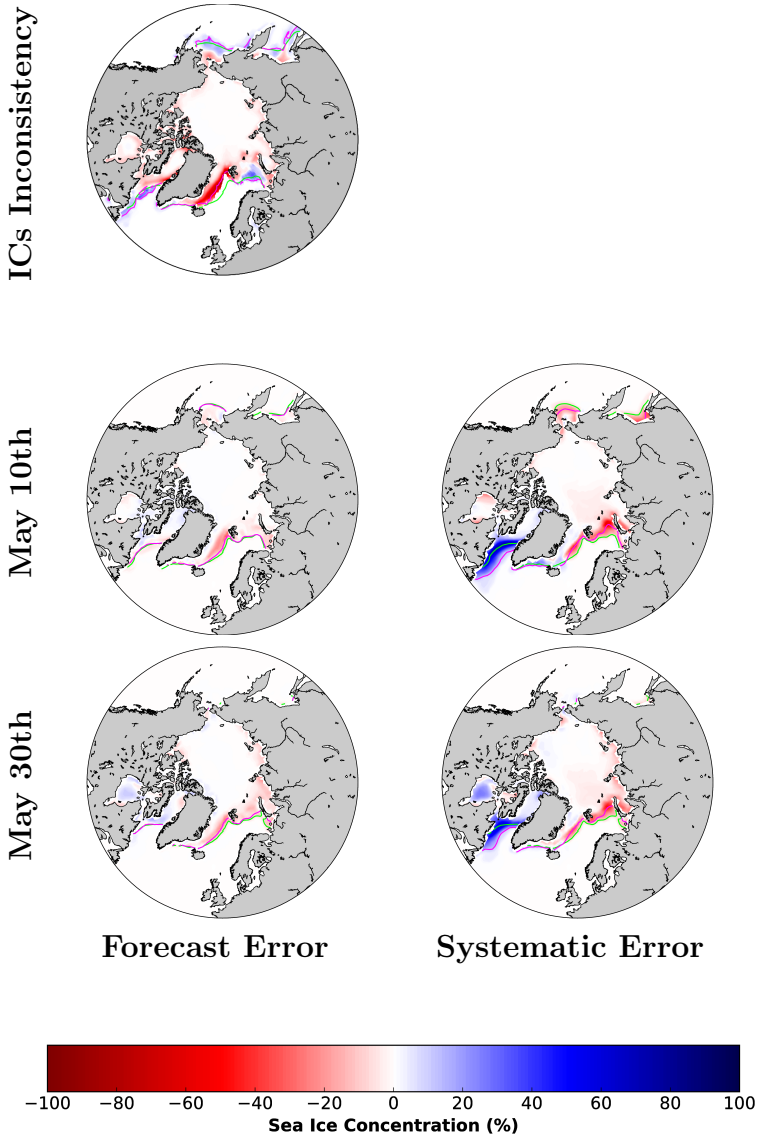


Figure 4.7: Evolution of SIC forecast errors for the predictions initialized in May. The first row corresponds to the ICs inconsistency at lag 0. The second and third rows show the SIC differences between PRED initialized in May and RECON (left column) and the systematic model error (right column) for lead times 10 and 30 for the period 1993-2008. The sea ice edge lines follow the legend of Fig. 4.3.

between ORAS4 and RECON over this layer, which in the Greenland, Baffin and Kara Seas goes down to 47, 30 and 22m respectively. This total heat is converted into a SIV. Only the ocean mixed layer is considered in this estimate because of its short timescale of interaction with the sea ice, which is supposed to be rapid enough to play a key role in the initial fast adjustment of the forecast. We focus on the Baffin Bay and Greenland and Kara seas during

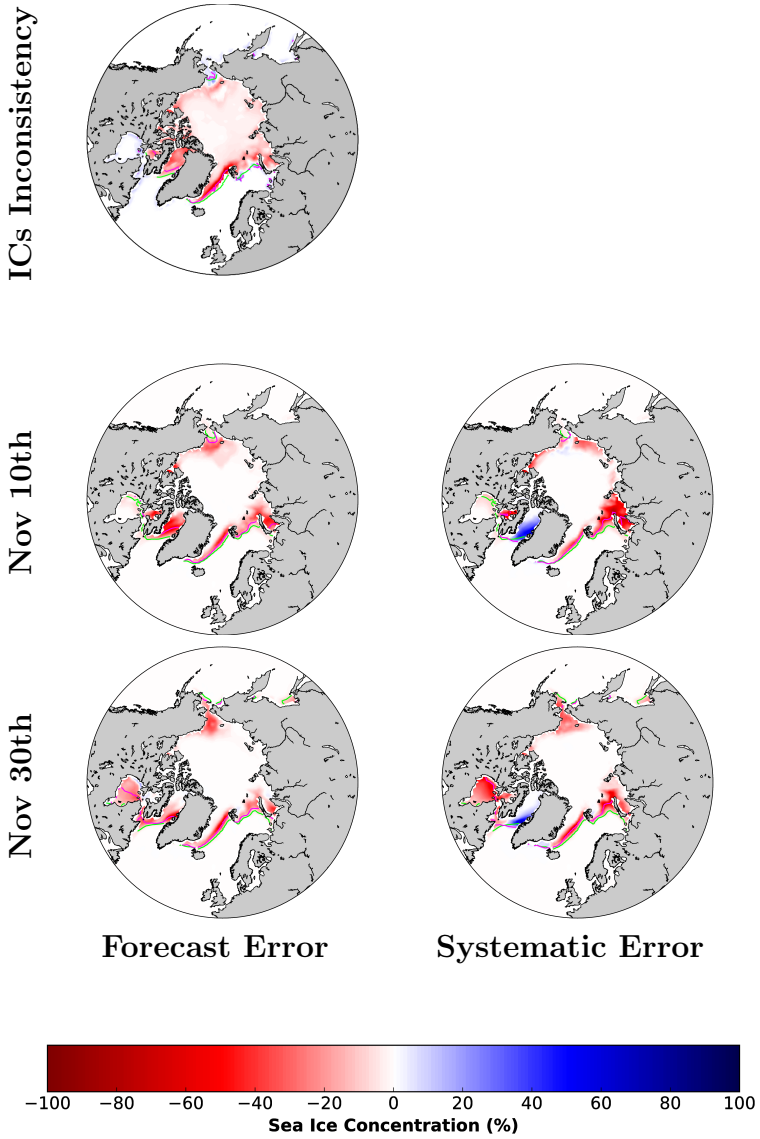


Figure 4.8: As Fig. 4.7 for PRED initialized in November.

the November forecasts because these three regions develop the largest negative SIC errors (Fig. 4.3f). In the Greenland Sea, the ORAS4 ocean holds enough excess heat to melt about three times as much sea ice than what is melt by the end of the month (Fig. 4.10a), which indicate a key contribution of the ICs inconsistency to the local SIC forecast errors. This SIV melt estimate however rely on the hypothesis of a local isolated mixed layer-ice system which did not consider potential lateral and vertical heat exchanges, nor exchanges with the atmosphere. This basic heat budget nonetheless demonstrate that enough energy is available from the warm ORAS4 ocean to explain the forecast sea ice melting. In the Baffin Bay, the

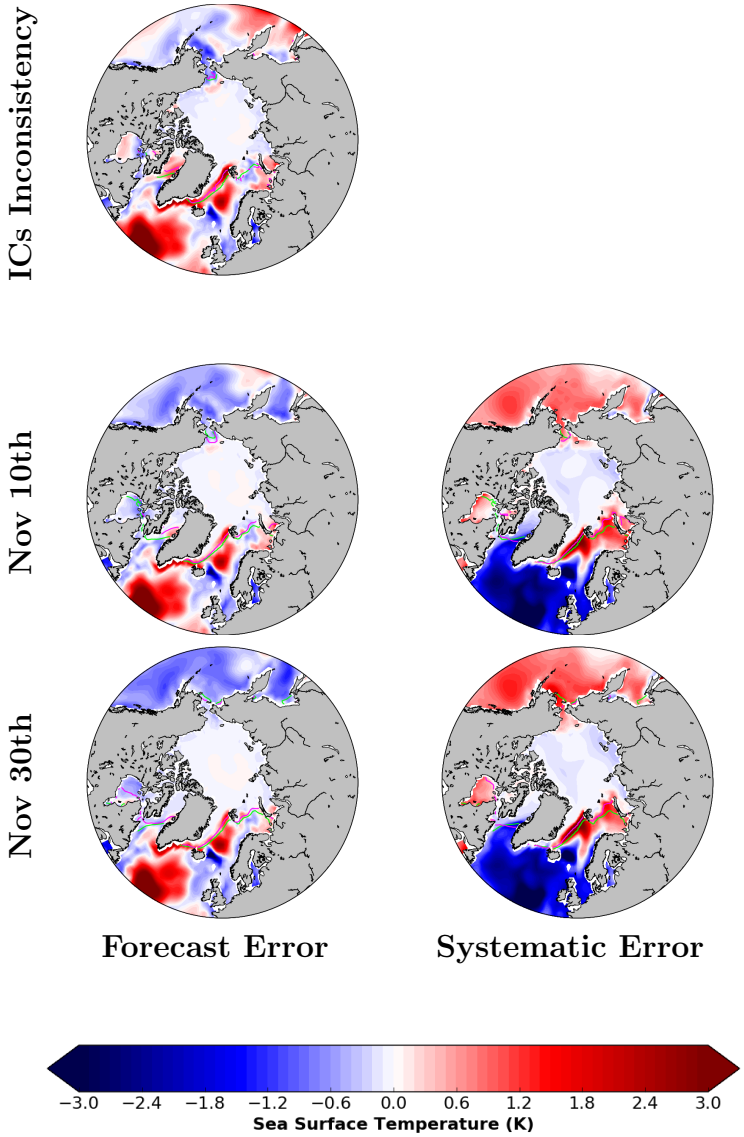


Figure 4.9: As Fig. 4.8 for the SST.

initial SIV loss in PRED amounts to about 2/3 of the estimate of SIV that can be melted by the warm ORAS4 by the end of the month (Fig. 4.10b). In the Kara Sea, ORAS4 is colder than RECON and would favour sea ice formation while the forecast sea ice melts, which discards any contribution of the ICs inconsistency (Fig. 4.10c). The development of the systematic model error would rather be the origin of the SIC forecast error in the Kara Sea (Fig. 4.8).

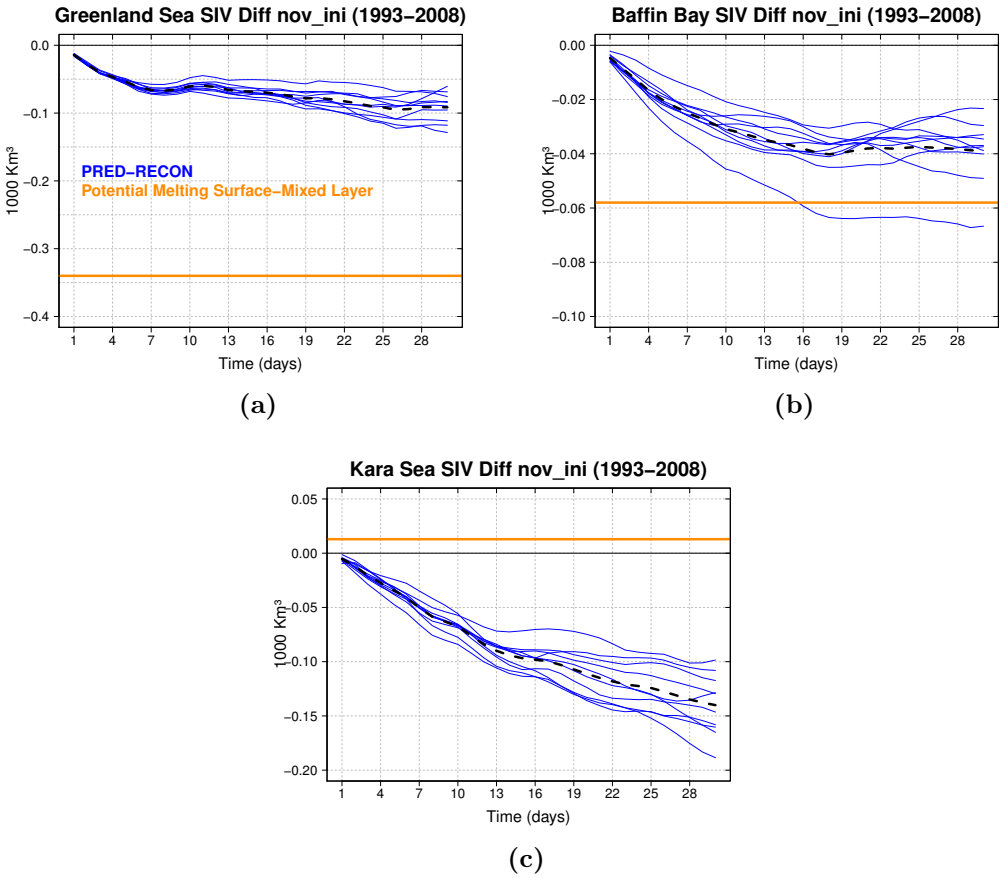


Figure 4.10: In blue, the difference in SIV between PRED initialized in November and RECON for **(a)** Greenland Sea, **(b)** Baffin Bay and **(c)** Kara Sea. The orange line represents the potential SIV melting that could be explained in PRED if all the heat difference between the sea surface and the mixed layer boundary at the ICs inconsistency was used to melt sea ice. In the case of the Kara Sea, the heat difference would involve sea ice generation.

Inconsistency between the atmosphere and sea ice ICs Inconsistencies between the atmosphere and sea ice are expected to have a lesser impact than between the ocean and sea ice since the atmospheric heat capacity is about 3-to-4 times smaller than the for the ocean and the atmosphere is also much lighter (oceans density is ∼784 times higher). To quantify the effect of the warmer atmosphere in ERA-Interim than in RECON (Fig. 4.5), we calculate the difference between the 2-metre temperature in ERA-Interim and RECON (i.e. DFS) on October 31 over the Greenland Sea, as an example, since it is the region exhibiting the largest 2-metre temperature differences in November. If all the difference in heat from the lowest 100 metres (the typical height of a stable boundary layer in polar regions) of the atmosphere was used to melt sea ice, only 0.3 km^3 would be melted, which is negligible

compared to the 100 km^3 melt by the end of the first forecast month (PRED minus RECON, Sea Ice Volume for Greenland Sea; Fig. 4.10a). It should be noted that DFS do not have temperatures above 2-metre because it is only a surface forcing, so we assuming that the temperature difference between ERA-Interim and DFS would be constant along the vertical. By contrast, the ocean is able to melt up to 350 km^3 of sea ice if all the excess of ORAS4 heat in the mixed layer (47 m) was provided to the sea ice (Fig. 4.10a). The excess of ERA-interim heat that can melt sea ice has even lower amplitude in any other Arctic region than in the Greenland sea.

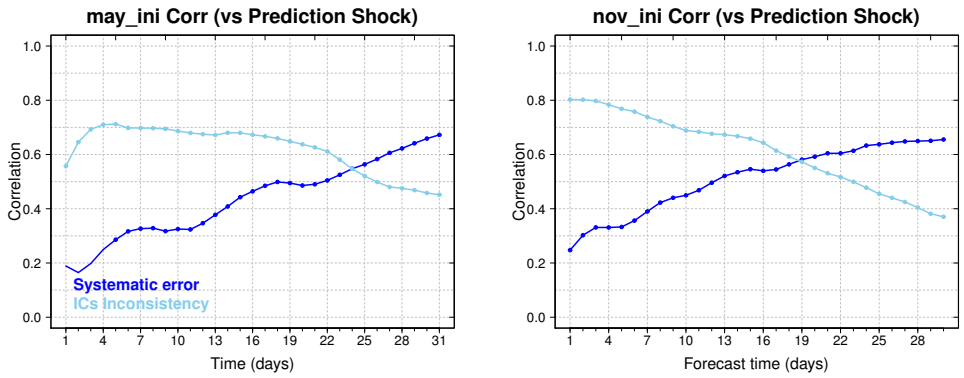


Figure 4.11: Spatial correlation between the pattern of ORAS4 minus RECON SIC at forecast day 0 and the forecast error for each day of the first month (light blue) and synchronous spatial correlation between the systematic model error and the average forecast error for each day of the first month (dark blue) for **(a)** May and **(b)** November. Dots represent the significant values at the 95% level as estimated from a one-sided student-T distribution.

Competing effect between the ICs inconsistency and the systematic model error

Our final analysis seeks to quantify how much of the forecast error can be explained in each forecast day by the ICs incompatibility and the model drift. We first compute spatial correlations between the SIC differences between $ORAS4\_ice$ and RECON at forecast time 0 and the forecast error for each of the following 31 days. We also compute the synchronous correlation between the systematic model error and the mean forecast error as a function of forecast time (Fig. 4.11). Moreover, we excluded the region from 80N to 90N to mask out areas with minimum sea ice variability. In May, the influence of ICs inconsistency on the mean forecast error increases during the first 5 days (Fig. 4.11a), which is probably the time required for the warm ocean underneath to induce a significant sea ice melting (while consuming the excess of heat from ORAS4). After 5 days, its effects seem to stabilize and

correlations slowly decrease but remain significant by the end of the month. The influence (as indicated by correlation) of systematic error grows steadily, and by 25th May it becomes predominant over the ICs inconsistency one. By the end of the month the systematic model error is largely but not yet fully developed, as evidenced by its correlation with the forecast error $\sim$0.7, as shown in Section 4.4.2. In November, the contributions from both sources of errors to the mean forecast error evolve in a similar way to May. The effect of the ICs inconsistency is, however, already maximum from the first forecast day, which could be favored by the general fast freezing in November. Furthermore, the systematic model error prevails over the ICs inconsistency after 19 days.

4.5 Conclusions

In this chapter, we have identified and quantified the contributions from the different sources of Arctic sea forecast errors that arise during the first month in a set of seasonal predictions produced with EC-Earth3.2. The main novelty of this work derives from the use of daily data, since the effect of some errors and processes is typically hidden in the monthly averages. Thanks to the use of daily data our results are equally interesting for operational and seasonal prediction. In the predictions, sea ice was initialized from a forced ocean reconstruction with Ensemble Kalman Filter (EnKF) assimilation of SIC (RECON).

The main results of the chapter are described in the following:

- The largest initial SIC errors appear in regions of large observational uncertainty and little model spread, in which the EnKF assimilation relies more strongly on the model, bringing the ICs far from the observed state.

- The comparison of a historical simulation (HIST) and the initialized predictions (PRED) has allowed us to study the model drift, i.e., how the systematic model error develops in time, and to which extent the timescales for the development of biases depend on the region. The mean forecast error in the PRED initialized in May has not reached the systematic model error (estimated as the difference between HIST and RECON) by the end of the month (Fig. 4.7), in particular in regions like the Baffin Bay and Barents and Kara Seas where the largest systematic errors can be found. For the predictions initialized in November, the model drift is accountable for the forecast errors in basins such as Barents, Kara, Chukchi and Bering Seas (Fig. 4.8), where, unlike for May, it seems to be fully developed before the end of the month.

- For both the predictions of May and November, errors in sea ice seem to be tightly linked to errors in the underlying SSTs: e.g., positive SST forecast errors collocated with negative SIC errors, and vice versa. Exceptions are found in the Hudson Bay and the Chukchi and the Barents seas in the November PRED.

- Another type of error has been identified related to an inconsistency between the initialization products used for the ocean and sea ice. In this forecast system, the ocean initial conditions, taken from ORAS4, are generally too warm for the overlying sea ice conditions from RECON, which leads to extensive sea ice melting during the first forecast days. The reason for the warm ORAS4 is explored in next chapter.

- For the month of November, we have quantified the potential SIV melting associated to these temperature differences between RECON and $ORAS4\_ice$ in the three regions with the largest SIC mean forecast errors. These estimations suggest that the ICs inconsistency could explain the sea ice melting in the Baffin Bay and the Greenland Sea by the end of the month, but not in the Kara Sea, where it is probably linked to the model drift. Incompatibilities between the sea ice and the atmosphere ICs have also been investigated, but the latter have been shown to have a negligible melting effect on the sea ice.

- By comparing the patterns of SIC forecast errors with those of the ICs inconsistency and the systematic model error, we have been able to quantify the contributions to the forecast error as a function of time. The analysis suggests that the initialization inconsistency dominates the average forecast error during the first 25 (19) days in May (November), while the systematic model error dominates afterwards. However, for particular regions like the Baffin Bay or the Greenland Sea the effect of ICs inconsistency can overshadow that of the model drift beyond one month.

Many studies have highlighted the importance of the initial conditions to improve the sea ice predictive capacity (e.g. Chevallier et al., 2013; Dirkson et al., 2017), including for record-breaking events as the Arctic sea ice minimum of 2007 (Kauker et al., 2009) and November 2016 (Acosta Navarro et al., 2019). However, initialization of sea ice can be compromised by both the initialization strategy and the observational uncertainties, as shown in this chapter. Distinct satellite products have also shown substantial differences even when measuring the same variables such as the sea ice concentration, area and extent (Ivanova et al., 2014). The retrieval algorithm to generate the sea ice products, can lead to differences in annual Arctic SIA of more than 1 million km^2 (Ivanova et al., 2014). Studies suggest that this instrumental

uncertainty can even impact substantially the seasonal forecast skill (e.g. Bunzel et al., 2016). Although the compatibility of ocean and atmosphere initialization products has been discussed in previous studies (e.g. Liu et al., 2017; Mulholland et al., 2015), it remains a topic with high potential interest for subseasonal-to-seasonal predictions that would clearly benefit from additional analyses. The role of these inconsistencies on the forecasts may have been overlooked in the past as most seasonal prediction studies focus on monthly means, and these effects seem to manifest primarily at daily timescales. Future works exploring the best products and approaches for initialization are therefore desirable.

The results from this chapter suffer from a few limitations. First, the calculation of the systematic model error has relied on the only historical simulation that had been produced with the same version of the model as used for the forecasts. Ideally, more ensemble members would be needed to filter out the internal climate variability and thus better constrain the systematic model error. This recommends some caution in the interpretation of our results, in particular at the regional level, where the true bias might not be accurately estimated. On top of that, the forecasts only expanded along 16 start dates, without covering the last decade, which could influence the robustness of the results discussed in this chapter. Another limitation is that part of the differences with respect to the sea ice reconstruction might be due to the fact that the assimilation run was not produced in coupled mode, potentially introducing an additional component to the initial shock (Mulholland et al., 2015). We keep this second potential drawback as an open question for future works. Finally, the sea ice reconstruction could have assimilated some ocean surface data to reduce sea ice uncertainty.

Despite the different errors identified and characterized in our analysis affecting the first forecast month, this same seasonal forecast system has been proven to be skilful over the Arctic and mid-latitudes, including the representation of extreme climate events (Acosta Navarro et al., 2019). There are potential ways to avoid, or at least minimize, the forecast shocks during the first month, such as initializing all system components from fully coupled assimilation runs performed with the same model version as used for the forecasts. However, reducing the initial shock does not necessarily translate into an improvement in predictive capacity. In the next chapter we will compare seasonal forecast systems initialized with several strategies based on different combinations of native and non-native products to further explore the development of the forecast errors, and whether the initial shocks have an effect on the forecast skill.

5 Comparing three sea ice-ocean initializations

5.1 Introduction

Achieving the best possible description of the initial climate state is essential to provide skill in seasonal forecasts (Balmaseda and Anderson, 2009; Guemas et al., 2014d). The impossibility to have a complete picture of the real climate evolution, in particular in remote regions such as the deep ocean, stimulated the production and use of climate reanalyses (e.g. Storto et al., 2019).

For Arctic prediction, recent years have brought encouraging progress (Guemas et al., 2014b). Sea ice initial conditions produced by assimilating ocean data, and nothing else, in ocean-sea ice models have been shown to enhance the sea ice predictive capacity with respect to a lack of assimilation. For example, using this approach Krikken et al. (2016) could predict the SIE in the Atlantic sector up to 6 months in advance; similarly Guemas et al. (2014d) showed improved skill of pan-Arctic SIA during the boreal summer when assimilating ocean data in the sea ice initial conditions in comparison to an absence of assimilation. Guemas et al. (2014d) related part of this gain in skill to the constraint exerted by the ocean nudging on the sea ice. Other studies have demonstrated that assimilating sea ice data can further increase the sea ice skill. Blockley and Peterson (2018) showed an improvement in the prediction of the September sea ice edge location when SIT data was assimilated into the GloSea seasonal prediction system, compared to a baseline with SIT assimilation excluded, the rest of the configuration being the same. Likewise, Dirkson et al. (2017) evidenced an amelioration of pan-Arctic and regional sea ice skill for all seasons, but especially during the summer, after initializing SIT from the PIOMAS reanalysis (Schweiger et al., 2011). A combination of sea ice, ocean and atmosphere data assimilation also revealed skilful predictions of September SIE 2 months in advance and up to 10 months before for winter SIE in the

seasonal forecast systems based on the CFSv2 and CanSIPS models (Sigmond et al., 2013; Wang et al., 2013, respectively).

A large variety of initialization/assimilation approaches have been considered by different prediction groups, each one with its advantages and disadvantages. Seasonal forecasts are typically initialized for the ocean component either from a non-native ocean reanalysis (i.e. built with an alternative ocean model, like ORAS4), or from a native product, i.e. a reconstruction produced with the same model as the one used for the predictions. A non-native reanalysis could lead to a fast initial adjustment because the model equilibrium might be different between the model used for the reanalysis and the one used for the prediction (e.g. Mulholland et al., 2015). However, non-native reanalysis could be chosen anyway because of their intrinsic high quality because of an extensive use of available observations and an advanced assimilation system. Furthermore, the use of compatible ICs between the various model component has been proven to benefit the forecast skill (Liu et al., 2017). ICs incompatibilities are typical of forecast systems initialized through weakly coupled assimilation, the classical approach applied in operational prediction centers (Arribas et al., 2011; Molteni et al., 2011; Saha et al., 2010), in which data assimilation is applied independently to each model component to generate the corresponding ICs (although the coupling between the various components will tend to exchange assimilated information among them). The paradigm is now changing, and some forecast centers are starting to follow the emerging technique called strongly coupled assimilation (Kimmritz et al., 2019; Penny et al., 2019; Penny and Hamill, 2017; Sugiura et al., 2008), which prevents the occurrence of ICs incompatibilities in the forecasts. As seen in Chapter 4, these incompatibilities can lead to initial shocks in the prediction. These initial shocks introduce an unpredictable behaviour in the model drift, and can be associated to different complex mechanisms, such as the presence of spurious trends in some of the initialization products (e.g. surface winds in Pohlmann et al., 2017). In some cases, the main source of error arises from the coupling itself, as the one showed by Rahmstorf (1995) between the atmosphere and the ocean, which led to an instability of the oceanic convection patterns accompanied by a transition of the model to a new equilibrium state. In other cases, initializing the ocean and the atmosphere simultaneously from their respective reanalysis is not necessarily translated into a skill improvement when compared to initializing only the ocean (as showed by Pohlmann et al. (2013) when predicting the SAT over the tropical oceans). As the only sea ice reanalysis publicly distributed is PIOMAS (unlike the oceanic or atmospheric reanalysis), the sea ice component usually requires to be initialized from in-house reconstructions (e.g. Chevallier et al., 2013; Guemas et al., 2014d). This initialization can lead to further shocks if they are not consistent with the oceanic and

atmospheric ICs. Mulholland et al. (2015) explored this issue by comparing three different initialization methods. They concluded that non-native analysis should be avoided in initialization due to differences between model attractors, both in the ocean and atmosphere, otherwise shocks are expected to occur. However, the impact on the sea ice of native and non-native initialization has not been previously investigated. To date, very few research groups have implemented a coupled data assimilation since updating the combined spatial and temporal scales of the sea ice, the ocean and the atmosphere is technically challenging and requires high computational resources. The question of how to minimize the initial shocks and the drift while maximizing the observational information provided to the forecast system remains open, especially for the Arctic sea ice. On top of that, independent initial products for the ocean, atmosphere and sea ice are provided in most cases, and the potential impact of their inconsistency has not been evaluated yet.

The previous chapter described the limitations of a single seasonal forecast system, showing in particular how the inconsistencies between the initialization products for the sea ice, ocean and atmospheric components can hinder the forecasts over certain regions, where the initialization shock causes a rapid sea ice melting. This chapter investigates two strategical choices of initialization of a seasonal forecast system built with Ec-Earth3.2. The first choice concerns the use of native versus non-native initialization products, with the latter providing an initial state that is closer to the real world but also leading to stronger shocks. The second choice concerns the production technique for native initialization products, namely the use of weak versus strong nudging coefficients when constraining the model towards the observations. In this case, stronger coefficients ensure a better fit with observations, but weaker coefficients keep the model closer to its preferred state and therefore minimize the subsequent drift in the forecasts. In our three forecast systems (i.e. non-native initialization, native with weak nudging, native with strong nudging), we assess the development of the forecast biases, the presence of initial shocks and their impact on the predictive capacity. The paper is organized as follows: In Section 5.2 the experimental setup is presented. The pan-Arctic sea ice errors are evaluated in Section 5.3.1. The daily SIC biases and their potential causes, with a particular focus on understanding the differences between the three systems, are addressed in Section 5.3.2. An evaluation of the sea ice forecast skill is presented in Section 5.3.3. Section 6.1 provides the conclusions of this chapter.

5.2 Methodology

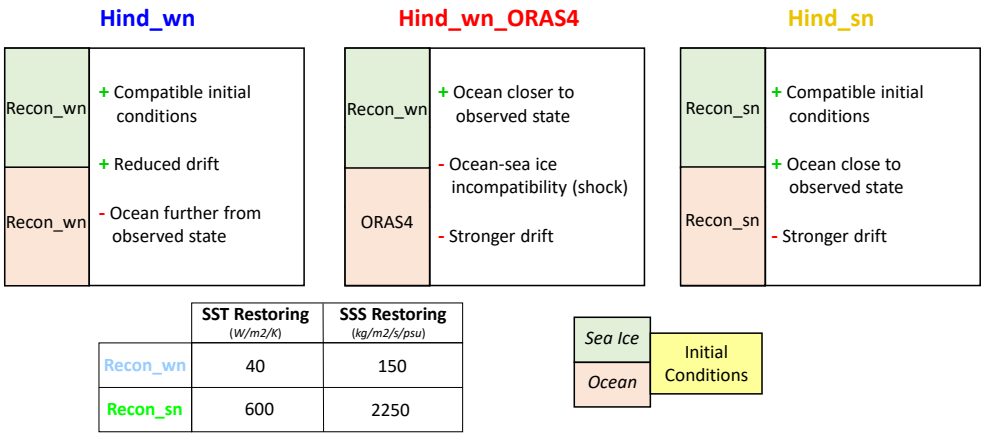


Figure 5.1: Schematic description of the main features of the three forecast systems used in this paper. Colour coding indicates the different model components (ocean and sea ice). The atmosphere initialization is from ERA-Interim for the three systems. The table at the bottom indicates the strength of the coefficients used to nudge the ocean towards ORAS4 in the ocean and sea ice reconstructions used as initial conditions for the hindcasts. The 3D nudging is the same for *Recon_wn* and *Recon_sn*, and has a depth-varying restoring timescale that starts at 3 days below an average mixed layer of 50 m, and increases monotonically to reach 329 days at 4900 m depth.

The three sets of retrospective predictions and their historical counterpart were produced with EC-Earth3.2 climate model (Doblas-Reyes et al., 2018; http://www.ec-earth.org/). The historical simulation consists of 5 members. The three hindcasts (Table 5.1; also referred to as forecasts) consist of seven month-long seasonal predictions, initialized each year from 1988 to 2012 on November 1st with an ensemble size of 16 members. All three systems have in common the use of ERA-Interim reanalysis (Dee et al., 2011) as ICs for the atmosphere. The choice of initial Gaussian perturbations in ERA-Interim air temperature intend to be representative of potential initial errors. The impossibility of initializing the sea ice model, LIM3, directly from observations (as we lack consistent information for all the prognostic sea ice variables, as e.g. sea ice thickness, in space and time), led us to produce in-house reconstructions. Our forecasts use two different sets of reconstructions for initializing the ocean and sea ice components: *Recon_wn*, produced with a weak restoring coefficient towards ORAS4 SST of 40 W/m^2K, equivalent to a time-scale of 37 days over a depth of 50 m, and towards ORAS4 SSS of 150 $kg/m^2/s/psu$; and *Recon_sn*, with a strong restoring coefficient towards ORAS4 SST of 600 W/m^2K and towards ORAS4 SSS of 2250 $kg/m^2/s/psu$, equivalent to about a 4 day time-scale over a depth of 50 m. In both reconstructions, the same 3D ocean temperature and salinity nudging is applied, based on a depth-varying restoring

time-scale that decreases monotonically from the bottom of the mixed layer (i.e. 3 days for an average mixed layer of 50m) to the bottom of the ocean (329 days at 4900 m depth). More details about the SSS restoring can be found in Table 5.1. 5 different members are generated for each reconstruction, to thus account for the uncertainty in the sea ice state, each one nudged towards a different member of the ORAS4 reanalysis.

The first hindcast, which we will refer to as *Hind_wn* hereafter, uses both the sea ice and the ocean of *Recon_wn* as ICs. The second, *Hind_wn_ORAS4*, also initializes its sea ice component from *Recon_wn*, but the ocean ICs are taken from ORAS4 (Balmaseda et al., 2013), which were previously regridded vertically to match the same resolution as in EC-Earth3.2. The third hindcast, called *Hind_sn*, is initialized from the ocean and the sea ice of *Recon_sn*. The three systems are briefly summarized in Table 5.1. The 16 ensemble members were generated combining the different members of the sea ice, ocean and atmosphere ICs. The historical simulations comprise 5 ensemble members, and their comparison with the hindcasts will allow to estimate how fast the systematic model biases develop in the forecasts and whether the hindcasts show a gain in skill derived from initialization. Our observational reference will be NSIDC (Cavalieri et al., 1996), while we will refer to *ORAS4_ice* for the sea ice product used as a boundary condition for the sea ice in ORAS4.

By comparing the first and third initialization strategies we will be able to investigate (1) the impact of the surface nudging coefficient strength on the sea ice predictions. The first and the second strategies will inform us about (2) the advantages and disadvantages of initializing the ocean from a native or non native product, in particular regarding their impact in the bias and the skill. Finally, a full comparison between the three will (3) help to identify which one is the best initialization methodology for optimizing the sea ice forecast skill.

5.3 Results

5.3.1 Development of sea ice biases

A first insight into SIE biases The model persistently overestimates the SIE from November until June as seen by comparing the historical simulations with NSIDC and *ORAS4_ice* climatologies (purple, Fig. 5.2). In the reconstructions, a weak nudging towards ORAS4 allows for drastically reducing the positive SIE bias (*Recon_wn*, light blue), while with a strong nudging, the bias becomes negative (*Recon_sn*, light green), a result that seems to be counter-intuitive. We will explain the reasons for this puzzling feature in

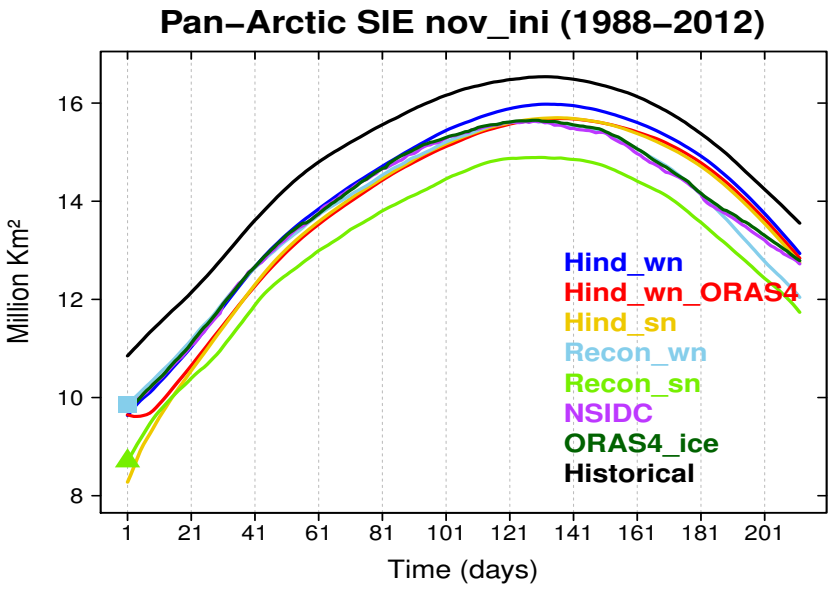


Figure 5.2: Pan–Arctic daily SIE climatology (1988-2012) for the three November-initialized hindcasts, the historical ensemble, the two reconstructions and two observational products (NSIDC and *ORAS4_ice*). Only the ensemble means of both the reconstructions and the hindcasts are shown for the sake of clarity. Their ensemble spread is comparable to the difference between the observational products.

section 5.3.2. *Hind_wn* (dark blue, Fig. 5.2) starts from *Recon_wn*, and remains as close to NSIDC as *Recon_wn* for about 3 months. This indicates a persistence of ICs information on timescales longer than monthly. Then *Hind_wn* slowly converges toward the historical experiment (black), although it is only half-way between *Recon_wn* and the historical simulations by the end of the forecast. *Hind_sn* (yellow) has, unexpectedly, a stronger initial bias than *Recon_sn*, revealing an initialization shock. In less than a month, *Hind_sn* crosses *Recon_sn* and starts approaching the observed climatologies, which it meets after four months. In the last three months, it develops a positive bias as it gradually converges towards the historical state, but without reaching it. By the end of the forecast, *Hind_wn* and *Hind_sn* nearly overlap, which indicates that their differences in ICs (mainly differences in the sea ice and ocean mixed layer) are forgotten by then. *Hind_wn_ORAS4* (red) starts from the same point as *Hind_wn* (dark blue), since they share the same sea ice ICs, but quickly develops a negative SIE bias. After about 10 days, *Hind_wn_ORAS4* joins *Hind_sn*, following afterwards a very similar trajectory, which indicates a dominant effect of ocean ICs on the sea ice since the surface ocean in ORAS4 and *Recon_sn* are very close due to the strong *Recon_sn* restoring towards ORAS4. None of the three forecasts has

completely converged towards the historical simulations after 7 months (Fig. 5.2a). Note that during the first three forecast months NSIDC, *ORAS4_ice*, *Hind_wn* and *Recon_wn* SIE overlap most of the time, making it hard to differentiate them.

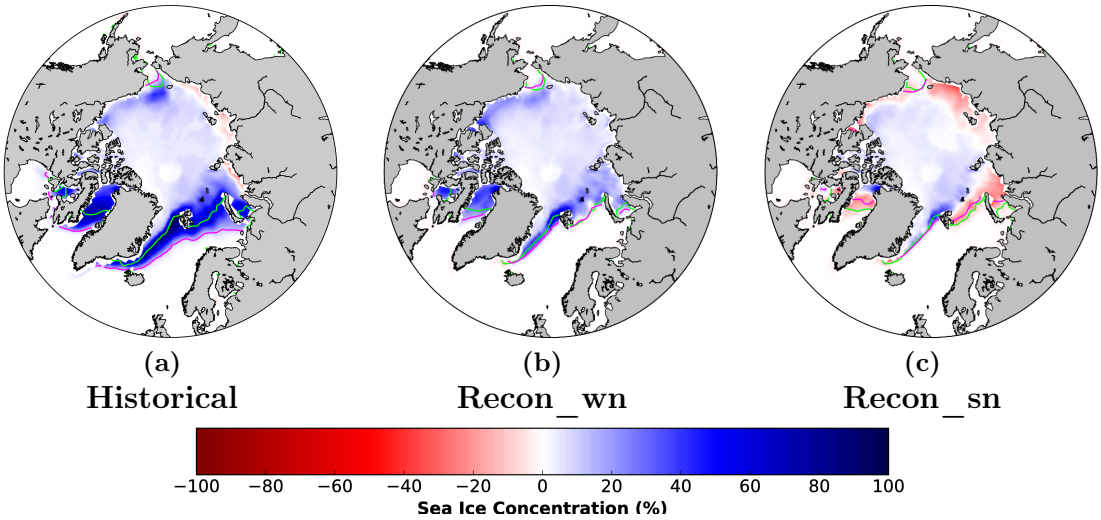


Figure 5.3: SIC differences between the **(a)** historical simulations (November 1st), **(b)** Recon_wn (October 31st), **(c)** Recon_sn (October 31st) and NSIDC. The green line represents NSIDC sea ice edge (15% SIC), while the purple one represents the sea ice edge in the (a) historical simulations, (b) Recon_wn and (c) Recon_sn.

Regional distribution of SIC biases As mentioned in Chapter 4, the SIE bias offers limited information about the local SIC biases because of the regional compensation of positive and negative SIC biases. On the other hand, spatial maps of SIC biases provide a complete description, but an extensive comparison between different forecast systems for all forecast times represent too large an amount of information. The IIEE and its components thus emerge as useful indices to synthesize and compare this information. In this section, the evolution of these indices in the three forecast systems and the historical experiment (Fig. 5.4), together with the spatial SIC biases for some selected forecast times (Figs. 5.3 and 5.5), will be described.

The initial SIC biases indicate a widespread overestimation of the SIC in the historical simulations on the 1st of November, particularly pronounced on the Atlantic side (Fig. 5.3a), with a narrow band of underestimated SICs along the Russian Coast. The overall SIC overestimation is substantially reduced in *Recon_wn* (Fig. 5.3b) thanks to the constraint of the sea ice cover via the ocean nudging. In contrast to *Recon_wn*, the stronger nudging in *Recon_sn* (Fig. 5.3c) leads to a mixture of SIC overestimation in the central Arctic and

SIC underestimation in the MIZ. Further details on the reasons behind these reconstructed biases are given in Section 5.3.2. For the first forecast day of the three prediction systems (Fig. 5.5a,b,c), most of the SIC errors originates from their counterpart in the corresponding reconstructions used for initialization (Fig. 5.3a,b).

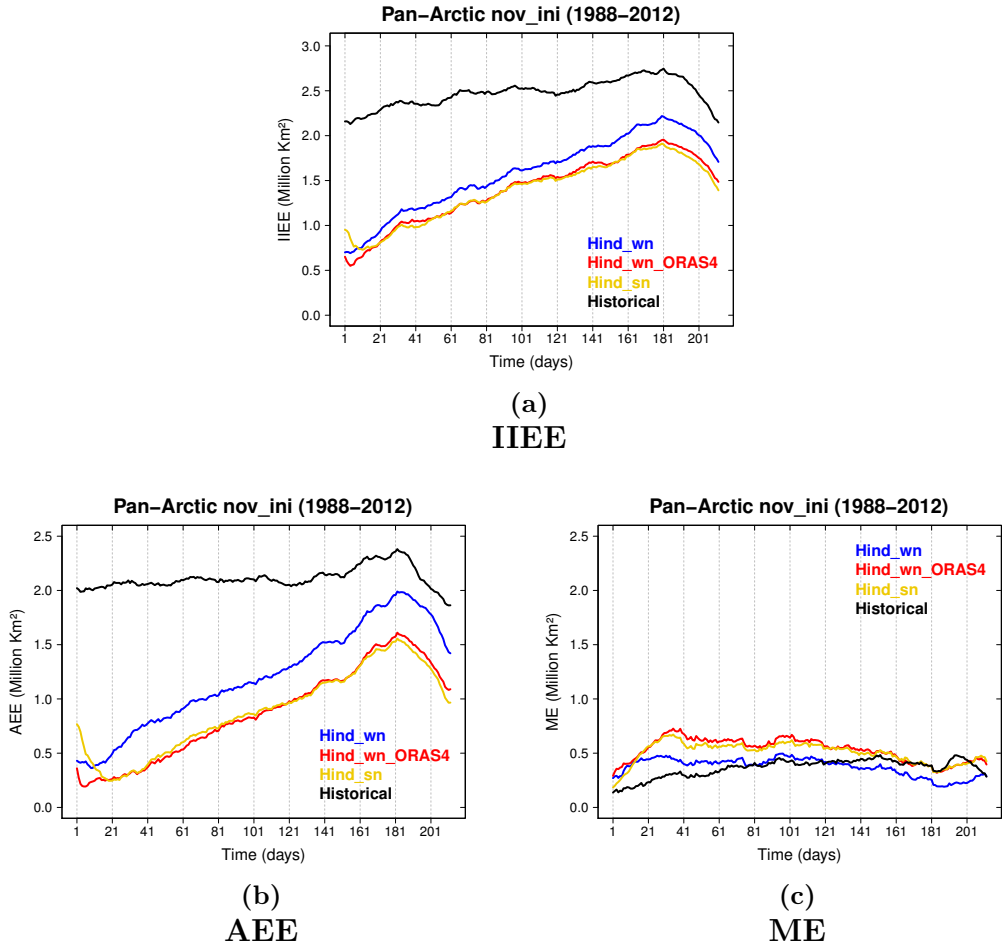


Figure 5.4: **(a)** IIEE, **(b)** AEE and **(c)** ME for the three forecast systems and the historical ensemble.

The IIEE and AEE are approximately stable in the historical simulations until approximately leadtime 180, which corresponds to a widespread SIC overestimation (Fig. 5.4a,b). The acceleration in AEE growth from lead time 150 to 181 originates from the slower model melting than in NSIDC (Fig. 5.2). Around day 181, the IIEE and AEE of the historical simulations start decreasing in association with the melt season in progress. $Hind\_wn\_ORAS4$ IIEE is initially very close to the $Hind\_wn$ one since they share the same sea ice ICs (but slightly lower for $Hind\_wn\_ORAS4$ because the ocean ICs reduce slightly the SIC overes-

timation), while *Hind_sn* IIEE is nearly twice higher (Fig. 5.4a). At forecast time 1, this is consistent with the negative *Hind_sn*/*Recon_sn* SIE bias, which is larger in absolute values than the positive *Hind_wn*/*Hind_wn_ORAS4*/*Recon_wn* bias. A sharp drop in IIEE for *Hind_sn* in the first weeks is explained by the widespread SIC underestimation from *Recon_sn* being canceled by the development of the model inherent SIC overestimation in about 2 to 3 weeks (Fig. 5.5c,g,k). The initial drop in IIEE for *Hind_wn_ORAS4* is also explained by the growing SIC underestimation induced by the ocean ICs over the Baffin Bay and in the Barents and Kara seas, as well as marginally in the Greenland and Chukchi seas (Fig. 5.5b,f,j), which partially compensates for the SIC initial overestimation (Fig. 5.4a,b). In about 3 weeks, the biases in *Hind_wn_ORAS4* have converged towards the ones of *Hind_sn* (Fig. 5.5f,g), which illustrates the key role of the ocean on the sea ice biases. *Hind_wn* IIEE evolution is more straightforward, since we only see an amplification of the initial biases to converge toward the historical biases.

The IIEE of the three forecast systems tend to approach the historical experiment one. This IIEE increase (Fig 5.4a,b) are associated to large errors emerging along the sea ice edge of the Sea of Okhotsk, Bering, Greenland, Labrador and Barents seas (Fig. A.6a,b,c in the Appendix for lead time 181). After 7 months (the end of the forecasts) none of the hindcasts has reached the IIEE values in the historical experiment (consistently with the SIE bias), which indicates that there is still some potential added value of initialization. This is because the model systematic SIC overestimation has not been completely reached in the Atlantic sector (mostly in the Labrador and Barents Seas) nor in the Bering Sea (Fig. A.6a,b,c,d in the Appendix and Fig. 5.5m,n,o,p for lead time 181). The IIEE is dominated by the AEE, which decreases during the first three weeks due to errors of opposite sign to the widespread SIC underestimation (overestimation) in *Hind_sn* (*Hind_wn_ORAS4*), but these new errors are reported in the ME.

Having a comprehensive explanation of the ME evolution represents a challenge, since the areas implied are small and the differences among experiments are marginal. *Hind_sn* AEE tends to be minimal around day 20 when the best balance between the SIC overestimation induced by the model and the underestimation coming from *Recon_sn* occurs (Fig. 5.5i,k).

Contribution of the systematic model error to the forecast SIC biases As in Chapter 4, we estimate how much of the forecast SIC bias is explained by the systematic model errors at each time step. For this, we correlate spatially the SIC bias of each forecast system at each forecast day with the historical bias for the corresponding calendar day (Fig. 5.6). The contribution of the systematic model error grows steadily for *Hind_wn*, as expected from a hindcast in which no initialization shock occurs. By contrast, *Hind_wn_ORAS4*,

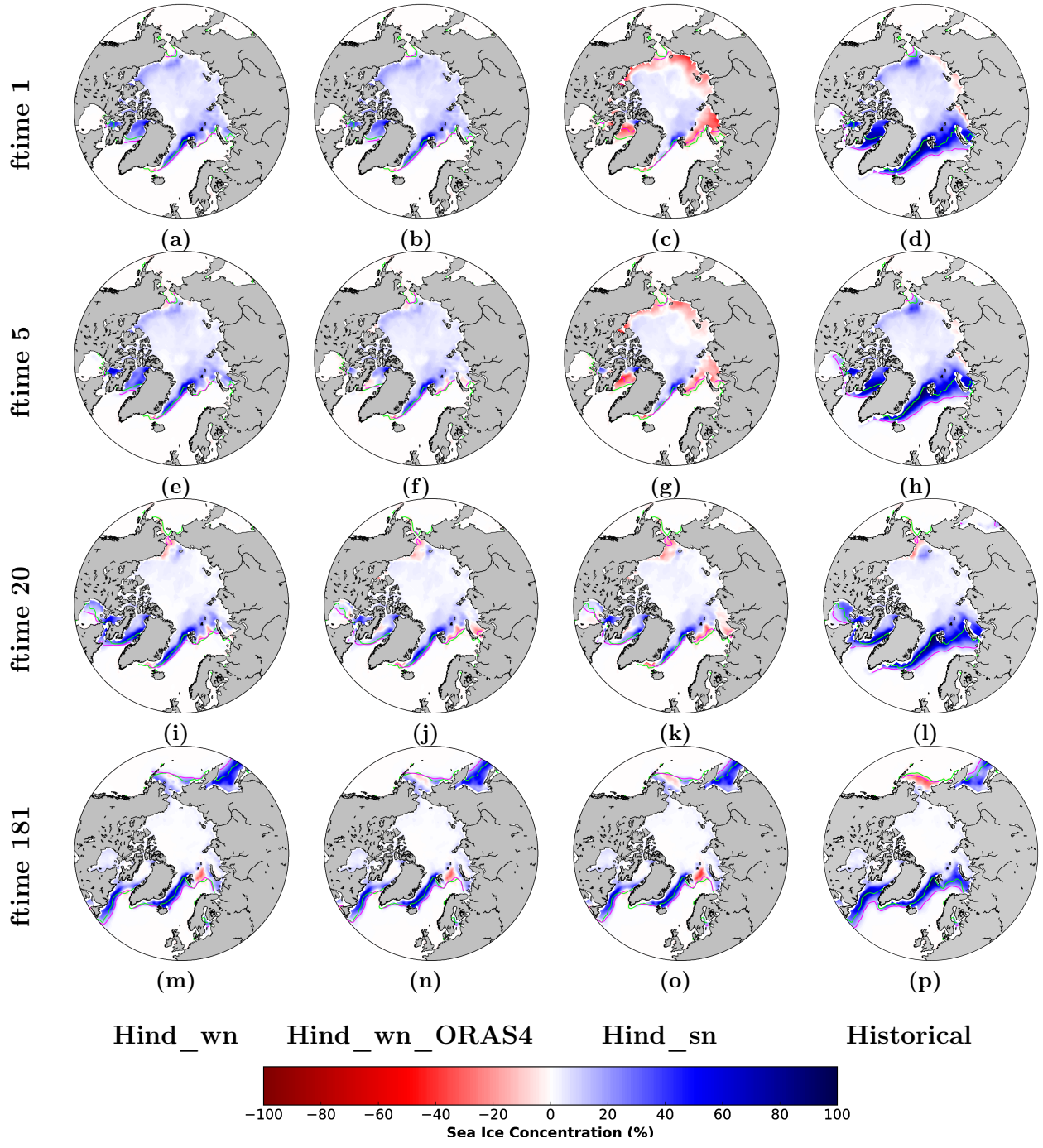


Figure 5.5: SIC biases for the three forecast systems and the historical simulations for the forecast times 1, 5, 20 and 181 of a set of November-initialized forecasts. NSIDC was used as the reference dataset. The green line represents NSIDC sea ice edge (15% SIC), while the purple one represents the sea ice edge of each forecast system.

for which the ocean and sea ice are not initially consistent, starts very close to *Hind_wn* since they share the same sea ice ICs (as described in Figs. 5.2 and 5.5), but correlations quickly drop to the *Hind_sn* correlation in approximately 10 days. This reflects a sea ice adjustment to the ocean ICs (which are very close for *Hind_wn_ORAS4* and *Hind_sn*) in about 10 days, which indicates a driving role of ocean conditions on the sea ice state. *Hind_sn* bias is initially anticorrelated with the historical one since the reconstruction providing its ICs (*Recon_sn*) generally underestimates the SIC, which opposes to the general overestimation in the historical experiment (as seen in Figs. 5.3 and 5.5). As forecast time increases, the overall *Hind_sn* SIC underestimation gradually fades away (Fig. 5.5) thus giving place to a general overestimation (as in the historical simulations) and a correlation increase in Fig. 5.6. After three months, the contribution of the systematic model error to the forecast bias of the three systems is similar, although with slightly lower values in *Hind_sn*. At the longest forecast times, the regions still showing a clear disagreement between the hindcast and the historical experiment SIC biases (Fig. 5.5) are the northern part of the Barents Sea, in which they are of opposite sign (excess of ice in the historical and deficit in the hindcasts), and the Bering Sea, in which the historical bias is much more pronounced than in the hindcasts. Given the 7-month lead time, we can expect these differences to stem from the ocean initialization, which can provide memory at seasonal and longer timescales.

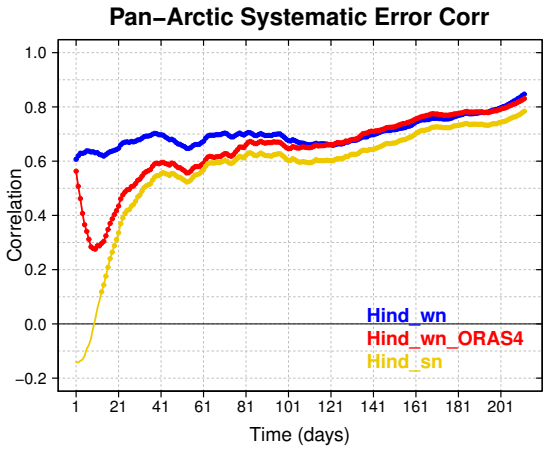


Figure 5.6: Synchronous pattern correlation between the daily model systematic error and the forecast bias for the forecasts initialized in November of the three systems. In order to weight the grid cells equally, data were interpolated previously to a Gaussian grid to ensure that they have the same area. The significance level relies on a one-sided student-T distribution and correspond to the confidence level of 95%.

5.3.2 Impact of ocean temperature on the sea ice biases

This subsection explores to which extent the initial subsurface ocean conditions can shape the SIC biases.

The weak nudging in *Recon_wn* (40 w/m^2) leads to a SST overestimation over the central Arctic compared to ORAS4, and an underestimated SST over most of the Atlantic sector (Fig. 5.7a). Despite the strong SST nudging coefficient applied in *Recon_sn* (600 w/m^2), its mean SST still differs from the one of ORAS4 (Fig. 5.7b) and its SIC with *ORAS4_ice* (Fig. A.7 in the Appendix). Indeed, the SST biases reverse over the MIZ in the Atlantic sector in comparison to *Recon_wn*, i.e. a too warm ocean, consistent with the underestimated SIC over the same region. The surprising change of sign in the SST bias with a strong nudging is related to the method used by ORAS4 to cope with the absence of sea ice model, which will be explained later in this section.

Differences in the SST at initialization between *Recon_wn* and ORAS4 (Fig. 5.7a) can also explain part of the later differences in SIC between *Hind_wn* and *Hind_wn_ORAS4* (Fig. 5.5). For example, in the Barents and Kara seas, the SSTs are warmer in ORAS4 than in *Recon_wn* (Fig. 5.7a,b), leading to less sea ice in *Hind_wn_ORAS4* than in *Hind_wn* in the first month (Fig. 5.5f,i).

Sea ice information management in ORAS4 The strong SST and SSS nudging in *Recon_sn* ensures that the reconstruction is as close as possible to the observed ocean state (ORAS4), and substantially closer than when using the weak nudging. However, it comes at the expense of producing imbalances. An imbalance could derive from the nudging of the temperature and salinity, that, even with the same restoring timescales, might create density anomalies and instabilities because of the non-linear density equation in temperature and salinity. Another example could be the creation of anomalous density current because of the density anomalies just mentioned. There is a non-linear dependence of the freezing point temperature of seawater (T_{fr}), which conditions the melting of sea ice, on the SSS. The seawater freezing point temperature is given by the following formula (Vancoppenolle et al., 2012):

$$T_{fr} = -0.0575\, SSS + 1.710523 \times 10^{-3}\, SSS^{3/2} - 2.154996 \times 10^{-4}\, SSS^2 \qquad (5.1)$$

where SSS is expressed in psu, and T_{fr} computed in Celsius degrees.

Following this formula, we find that SSTs in *Recon_sn* are above the freezing point temperature obtained from the corresponding SSS values over most of the Arctic (Fig. 5.7c),

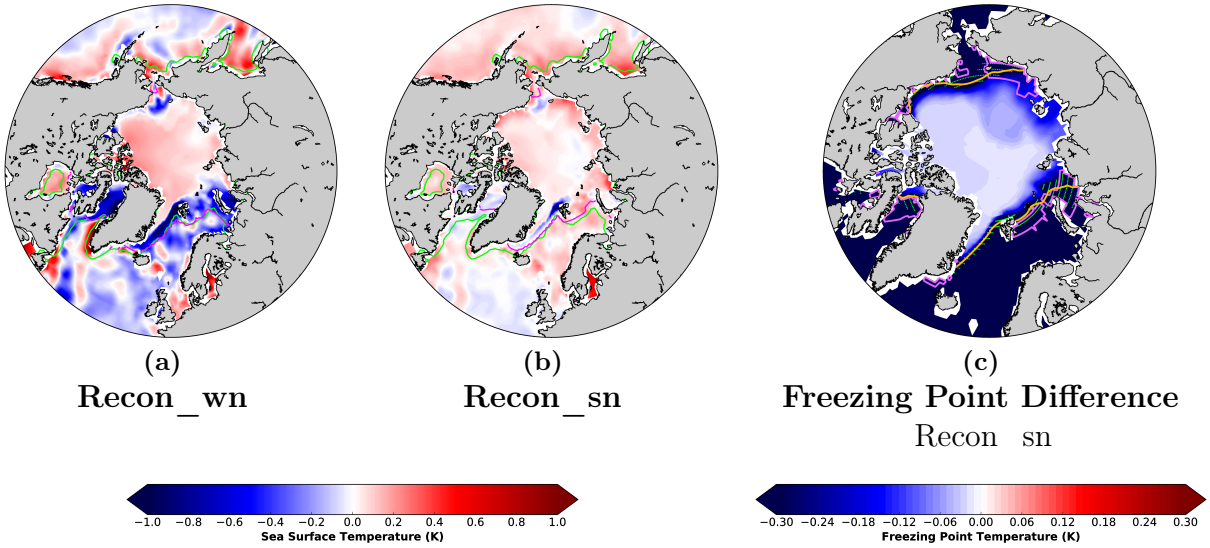


Figure 5.7: SST differences of **(a)** Recon_wn and **(b)** Recon_sn with ORAS4 for the restarts of October 31st. The green line represents *ORAS4_ice* sea ice edge (15% SIC), while the purple ones represents the sea ice edge of each reconstruction. **(c)** Difference between the freezing point temperature in Recon_sn and the SST in Recon_sn for October 31st. The freezing point temperature was calculated following formula 5.1. The golden line represents the NSIDC sea ice edge at 55% concentration. The green dots represent those grid cells which are inside the ice pack (>15% SIC) and have a SST larger than the freezing point. The pink contours represent the regions where the SIC is underestimated in Recon_sn as in Fig. 5.3b.

which thus favours the sea ice melting anywhere where SST $> T^{fr}$. Either a lower SSS or a lower SST would have been needed to avoid this sea ice melting. This incompatibility derives from the fact that ECMWF-System4 uses a different approach to LIM3 to define the freezing point temperature (Balmaseda et al., 2013; Mogensen et al., 2012). This system lacks a sea ice component, and observed SICs are imposed as a boundary condition. Because of this, the ocean model surface temperature needs to be adjusted to make it compatible with the sea ice state. For this, two criteria are applied. First, the freezing point temperature is imposed at the ocean surface whenever observed SICs overcome a threshold of 55%. In a coupled ocean-sea ice model whenever heat is extracted from the ocean surface its temperature drops until the freezing point, then heat goes on being extracted while sea ice forms, the ocean temperature being maintained at the freezing point. Since it is impossible to transform all the ocean water into sea ice before extracting more heat to decrease further the temperature, the surface ocean and the basal ice temperature can never be below the freezing point. In ORAS4, whenever heat is extracted from the ocean surface by the atmospheric forcing, the temperature can drop below the freezing point. The second criterion is that whenever SIC

is below the given threshold but the model SST is below the freezing point, the strength of the relaxation term to the observed SST is increased. When SIC $< 55\%$ and SST $>$ freezing point, the model is left with its usual ocean data assimilation with whichever biases it could incur. This means that for some SIC values below 55%, there might be a mismatch between the observed SIC and the local surface temperature in ORAS4, with the later being too warm for that particular SIC state (Fig. 5.7c and Fig. A.7 in the Appendix). This has implications for any simulation that assimilates ORAS4 temperature and salinity in polar regions, especially those in which the assimilation is strong. For example, this issue seems to affect *Recon_sn*: the 55% concentration contour in $ORAS4\_ice$ (Fig. 5.7c, golden lines) delimits fairly well the regions where the SSTs are warmer than the freezing point temperature inside the sea ice pack (i.e., $15\% < \text{SIC} < 55\%$; Fig. 5.7c, green dots). Besides, we also see that this same region coincides roughly well with the area where *Recon_sn* is underestimating the SIC (Fig. 5.7c, pink lines), with the exception of some parts of the East Siberian and Kara Seas, in which oceanic transport of warm temperatures could be playing a role. In short, Fig. 5.7c shows how using strong SST and SSS restoring coefficients towards ORAS4 derives in a sea ice melting at some locations (especially along the peripheral regions of the Arctic sea ice cover; Fig. 5.3c).

Lowest SIC biases related to SST biases We now explore which forecast system has the pan-Arctic average SIC biases, region per region during winter (December-January-March; DJF) and spring (March-April-May; MAM), the two seasons covered by the hindcasts, providing recommendations regarding the choice of an initialization strategy depending on the region of interest. Overall, $Hind\_sn$ exhibits the lowest SIC bias in DJF and MAM compared to the other two systems, in particular for most of the Atlantic sector. In the Pacific sector $Hind\_wn\_ORAS4$ prevails over the others (Fig. 5.8a,b). None of the forecast systems is found to be significantly better than the others regarding the SIC and SST biases (i.e. the differences in biases are not significant). We do highlight those grid cells in which the bias differences are the largest in magnitude, using stippling wherever the hindcast with the lowest SIC (SST) bias is at least 2% (0.1°C) smaller than the bias of the hindcast with the second lowest. Thus, the Irminger, Greenland and Barents Seas are identified both in DJF and MAM as regions where $Hind\_sn$ has the lowest SIC bias by a non-negligible amount. Interestingly, this system also leads to the lowest SST bias for DJF and MAM in the same regions (Fig. 5.8c). $Hind\_wn\_ORAS4$ exhibits the lowest SIC and SST bias in DJF over the Labrador Sea, Sea of Okhotsk, part of Hudson Bay and Kara Sea that corresponds to the lowest SST bias for the same system, and that matches the grid points with stippling. In MAM, this is reduced to isolated areas of the Sea of Okhotsk and Bering and Kara seas.

These results suggest that a strong SST nudging towards ORAS4, or using directly ORAS4 ocean ICs not only tends to constrain better the SST, but it also has the potential to decrease the SIC biases along marginal ice regions long after the initialization.

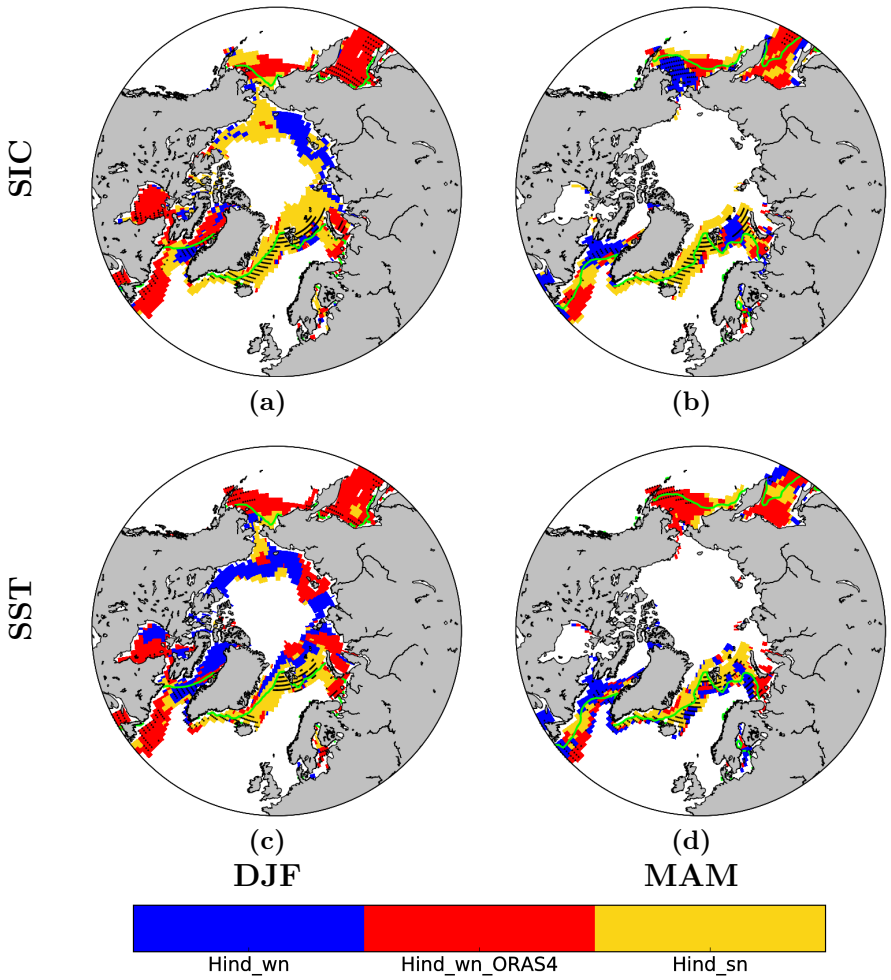


Figure 5.8: **(a-b)** Identification of the system with the lowest (absolute) SIC bias in DJF and MAM compared to NSIDC at the grid cell scale. Grid cells with climatological SIC seasonal values above 90% were masked out. Dots indicate grid cells where SIC lowest bias differs at least in 2% concentration from the second lowest bias. For DJF, the areas represented by each experiment are 1.96 million km^2 ($Hind\_wn$), 4.53 million km^2 ($Hind\_wn\_ORAS4$) and 5.19 million km^2 ($Hind\_sn$). For MAM, 1.91, 2.28 and 3.11 million km^2, correspondingly. **(c-d)** Same for SST, using EN4 data set as the reference. Dots indicate grid cells where SST lowest bias differs at least by 0.1°C from the second lowest bias. Grids cells where SIC is larger than 90% have been masked out.

5.3.3 Skill assessment

We now assess whether the same initialization approaches leading to smaller sea ice biases are also associated with higher prediction skill. For this, we compute the ACC for the pan-Arctic SIE at the monthly time scale in the three forecast systems and the historical experiment (Fig. 5.9), using NSIDC as the reference. Monthly means were used to average out the noise effects at daily scales, which introduce high differences in ACC between one day and the next, hampering the identification of skill scores that are significantly different.

Differences between the initialized forecasts and the historical simulations are small until spring, suggesting that the forced trend dominates the skill during the freezing season (Fig. 5.9a). The ACC for the three forecast systems shows very similar values, with a multi-forecast range that never exceeds 0.1 points of correlation during the seven forecast months (Fig. 5.9a). Correlations are not significantly different between the forecasts themselves nor between either of them and the historical experiments (calculated by a two-sided t-test as in Siegert et al. (2017)), in the latter case even for correlation differences larger than 0.2. This is due to our limited sample size (25 start dates). To circumvent this issue, we assess the skill of the individual members even though there is a redundancy with many forecasts (individual members) attempting at capturing a same observed trajectory. The 5 historical members are repeated until reaching 16 (the number of forecast members). This approach has the downside of leading to systematically lower skill scores, due to the presence of larger noise in the individual members, but it allows to identify significant differences thanks to the larger sample size. Indeed, the ACC calculated this way now shows significantly different correlations between the three forecast systems and the historical experiments in November and December. This implies an added value of initialization regardless of the initialization methodology (Fig. 5.9b). No significance difference in correlations has been detected between the forecast systems. When compared against the historical ACC, forecast skill values are statistically indistinguishable during January and February, but significantly different correlations appear again for March and April (except $Hind\_sn$ in April). This could be associated to a freeze-to-melt reemergence of correlations due to SST anomalies persisting over some locations across the sea ice edge (see Chapter 3). In May, no added value from initialization is seen again in the three systems.

We now repeat the ACC analysis for the individual Arctic basins, using the concatenated members to increase the sample size and thus highlight differences which are not identifiable using the correlations of the ensemble mean. Most of the central Arctic basins do not show significant skill, mainly due to the low sea ice variability, so their ACC is not shown. Regional SIE skill responds differently to initialization depending on the basin, reflecting

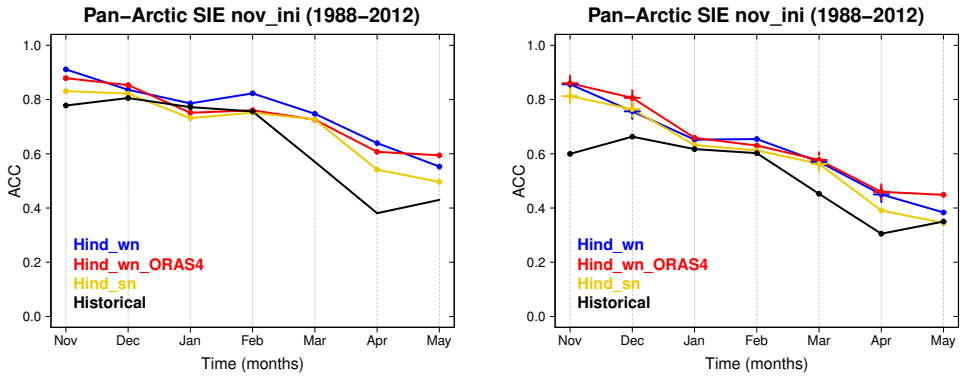


Figure 5.9: ACC of the pan-Arctic monthly SIE in the three hindcasts and the historical experiment in the period 1988-2012. NSIDC was used as the reference to evaluate the forecast skill. Panel **(a)** was calculated using the ensemble mean, while for panel **(b)** all members where concatenated to produce a longer timeseries. Dots identify the significant correlations, which rely on a one-sided student-T distribution. Crosses identify when the correlation values in the forecasts are significantly different than in the historical experiment, calculated according to a two-sided t-test.

different regimes of sea ice variability. However, a consistent result is that all the peripheral regions show added value of the initialization for at least the first forecast month (Fig. 5.10). In the Atlantic sector, the Labrador Sea stands out as the region where the three systems outperform the historical for all forecast months. The Hudson and Baffin bays and the GIN and Kara seas also show an enhanced skill with respect to the historical experiment for between 3 to 5 months, depending on the initialization method. In the Pacific sector, only the Sea of Okhotsk sees a clear improvement in skill related to initialization for the first three months. In this and the Atlantic regions previously commented the historical experiment shows weak and sometimes insignificant ACC values, which suggests that there is a moderate contribution of the forced trend in their variability. On the other hand, the trend explains most of the skill in the Barents Sea, about 1/3 in the Labrador Sea, and it also explains the skill in the second half of the forecasts in the GIN seas and the Sea of Okhotsk. The skill improvement for the basins mentioned might be related to a shift in the bias patterns derived from the ocean ICs, which could help the model variability to project onto the observed one. The SIE skill calculated using detrended anomalies agree with these results (not shown): a larger and significant skill for the three forecasts compared to the historical experiment.

To narrow down the areas where the hindcasts and the historical simulations are skilful, their spatial ACC maps for SIC are produced (Fig. 5.11), focusing on the winter and summer seasons. The ACC of the historical experiment is only strong and significant in the Kara

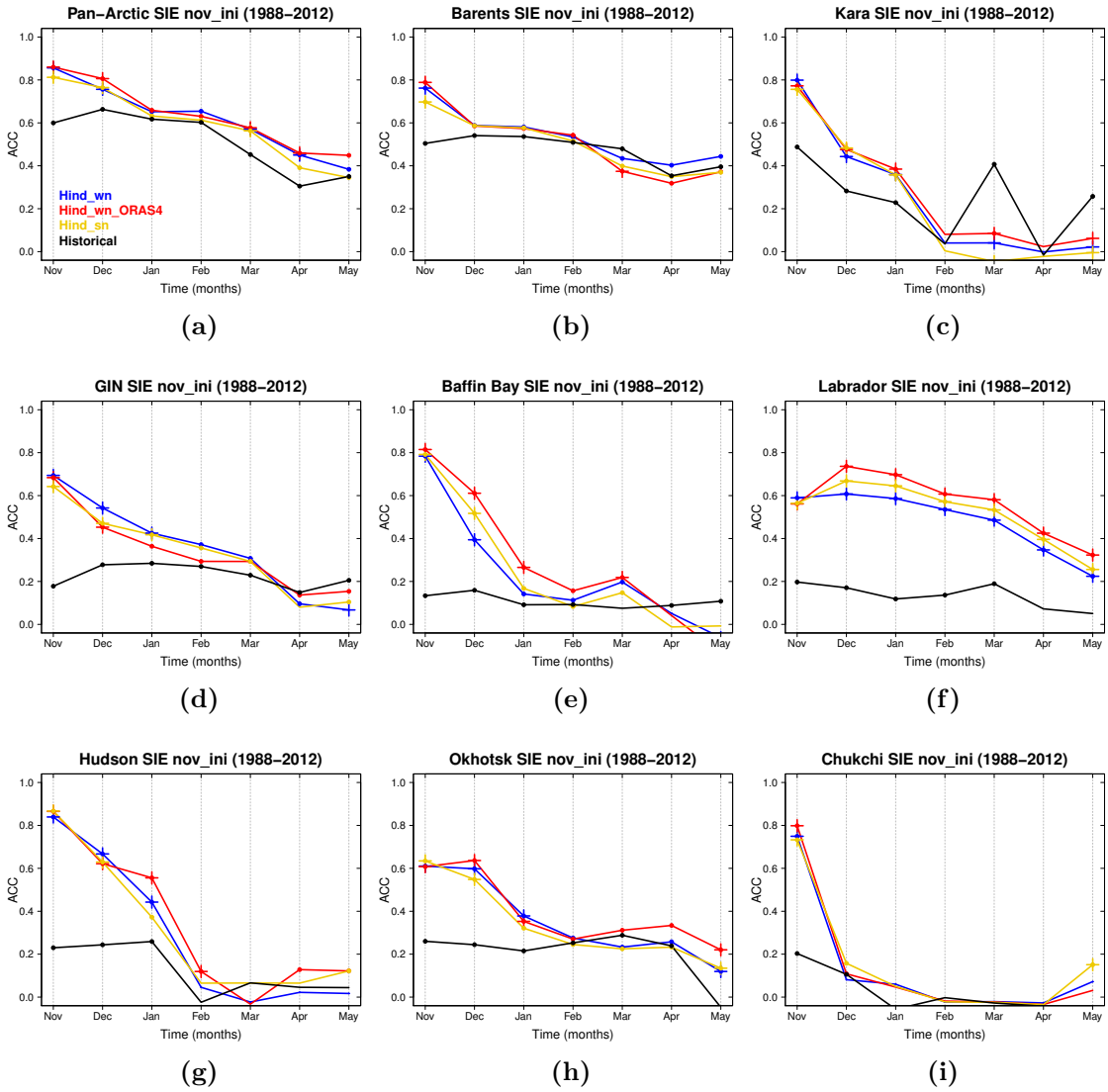


Figure 5.10: As in Figure 5.9b but for the SIE in the peripheral regions.

sea in winter, and it shifts to the Barents sea in spring, with other regions that also have significant but weaker ACC values being the Sea of Okhotsk and Baffin Bay. The three hindcasts evidence a clear gain in skill during the two seasons, particularly in the Hudson Bay, Barents, GIN and Chukchi seas in DJF in agreement with Fig. 5.10 and in the Labrador Sea for both DJF and MAM. For both seasons, the main differences in skill between the three hindcasts appear in the Sea of Okhotsk and Bering and Chukchi seas, although the ACC values are so close that it becomes hard to identify the best skill.

When compared between each other, $Hind\_wn\_ORAS4$ shows the highest skill in DJF for most of the Bering and Chukchi seas and also in the Sea of Okhotsk (Fig. 5.12a). For

the Hudson and Baffin bays as well as for the Barents and Kara seas, results are rather noisy and none of the hindcasts appear to be generally better than the other. For MAM, $Hind\_wn\_ORAS4$ again exhibits the largest ACC values in the Sea of Okhotsk (Fig. 5.12b). For the rest of regions, no hindcast stand outs clearly with respect to the others. None of the correlations shown for the hindcasts in Fig. 5.11 are significantly different from those in the other hindcast (calculated by a two-sided t-test). Also, unlike for the SIC biases in Fig. 5.5, no clear correspondence is found between the largest significant ACC for the SST and the SIC (Fig. 5.12c-d), except for some individual grid points of the Labrador Sea in DJF and the Barents Sea in MAM. A correspondence between the largest ACC and the lowest bias in SIC can only be seen for the Bering Strait in DJF and the Sea of Okhotsk in MAM.

5.4 Conclusions

In this chapter, we have compared the Arctic sea ice biases and prediction skill in three different seasonal forecast systems produced with EC-Earth3.2. The sets of predictions differ by their initialization methodology for the ocean and the sea ice, while using the same atmospheric ICs (ERA-Interim). These three initialization strategies consist of: (1) taking both the sea ice and the ocean ICs from an in-house reconstruction that assimilates ORAS4 temperature and salinity with a weak restoring coefficient (40 W/m^2K and 150 $kg/m^2/s/psu$, respectively); (2) using the sea ice ICs from the same reconstruction as in *1* and the ocean ones from ORAS4; (3) the same as *1* but from a reconstruction that uses a strong restoring coefficient (600 W/m^2K for the SST and 2250 $kg/m^2/s/psu$ for the SSS). These three experiments allowed us to investigate the effect of different surface nudging coefficients on the sea ice forecasts, the benefits and disadvantages of initializing from a native or non native ocean product and to help identifying the initialization methodology that leads to a better sea ice forecast skill.

Concerning the sea ice biases, the main findings of this study are:

- The model systematically overestimates the sea ice across the whole Arctic, as seen in the historical experiment. This bias is drastically reduced with initialization of the sea ice from a weakly nudged reconstruction (by more than the half, as indicated by the integrated edge errors). After 7 months, biases have not been developed completely in any of the 3 hindcasts, an indication that the benefits from initialization can persist throughout several seasons in the Arctic.

- The initial shock for strategy *2* is caused by an inconsistency between the ICs of the

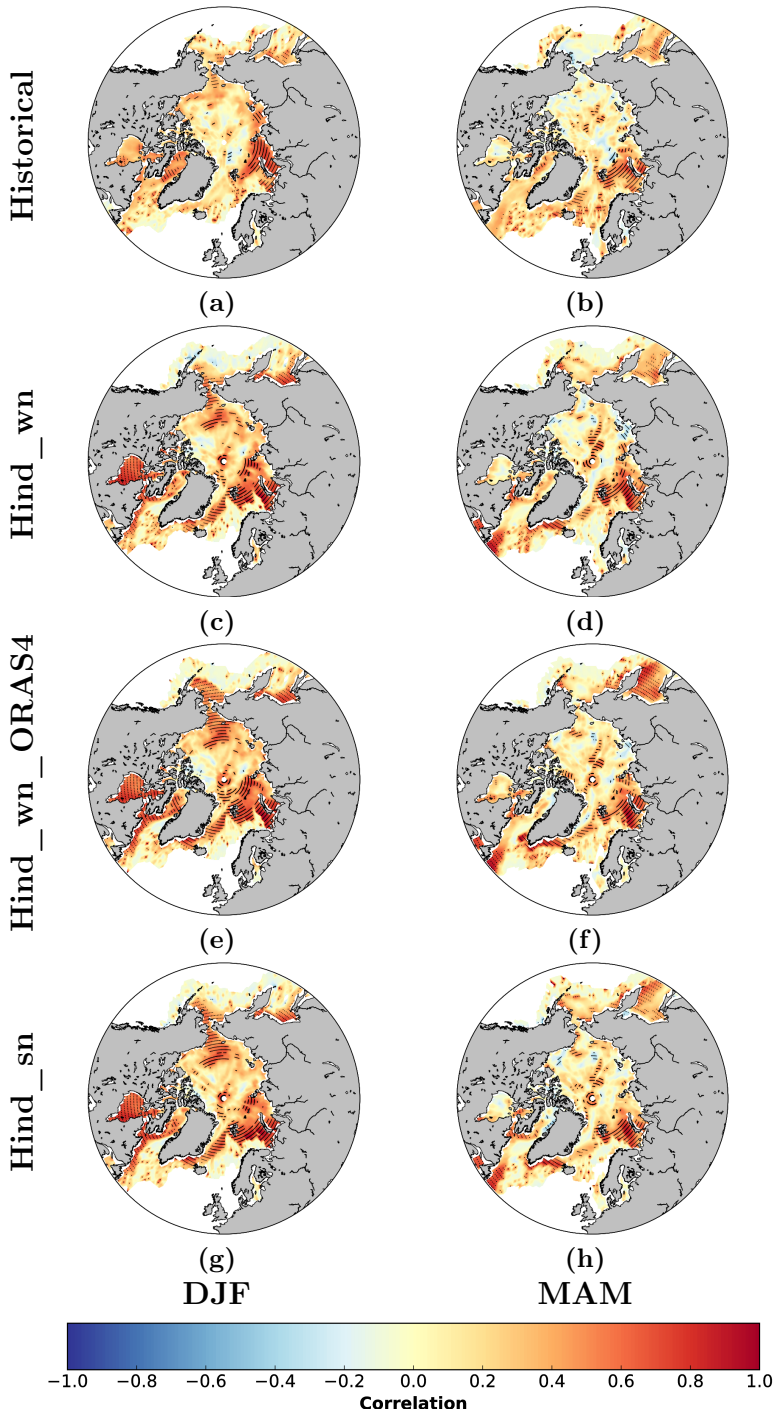


Figure 5.11: Spatial ACC maps of SIC in DJF (left) and MMA (right) in the three hindcasts and the historical experiment. NSIDC was used as the reference to evaluate the forecast skill. These plots are based on ensemble means. Dots represent the significant values at the 95% level as estimated from a one-sided student-T distribution.

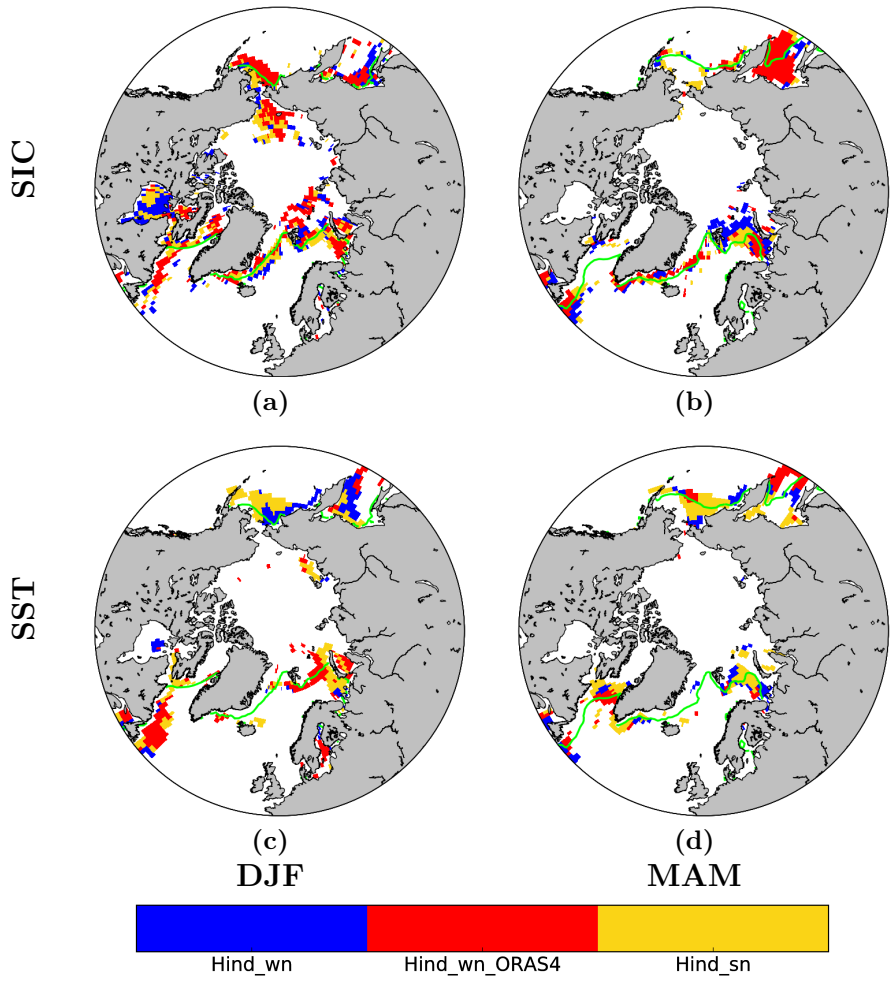


Figure 5.12: **(a-b)** Hindcast with the largest and significant ACC of SIC in DJF and MAM. NSIDC is used as reference. **(c-d)** Same but for SST, using EN4 as reference. Grids cells where SIC is larger than 90% have been masked out to exclude the central Arctic, where the systems are unskilful.

ocean and the sea ice. The ocean ICs from ORAS4 are affected by a warm bias in the high latitudes. Although the reconstruction with weak nudging uses ORAS4, its ocean is not as warm as ORAS4 so that when its sea ice is combined with ORAS4 to produce a set of ICs, an initial shock is triggered.

- For strategy *3*, the initial shock derives from the fact that both the ocean and sea ice are affected by the warm ORAS4 bias in this reconstruction, so there is also a strong sea ice deficit which is not consistent with observations. This explains a very strong drift, which is in fact derived from the lack of sea ice model in ORAS4, which is handled

with an ad-hoc algorithm providing erroneous temperature.

- In the first two weeks of forecast, sea ice forecast biases are dependent on the sea ice reconstruction from which they are initialized.

- After 20 days into the forecast, differences between hindcasts are substantially reduced. The two hindcasts in which the ocean is initially closer to observations (strategies *2* and *3*) nearly overlap with consistently lower sea ice biases until the end of the forecast than the hindcast *1*, started from the reconstruction with weak ocean nudging. This demonstrates the crucial role of the ocean in driving the sea ice biases.

- In winter (DJF) and spring (MAM) seasons, and at the regional level, hindcasts with the lowest SIC biases tend to coincide with those with the lowest SST biases. For example, the *2nd* initialization strategy in the Sea of Okhotsk and the Bering Sea, or in the *3rd* along the Atlantic sector, from Southern Greenland to Svalbard. Methods initializing the ocean closer to the observed state are beneficial to ultimately reduce the sea ice biases.

Regarding the impact of initialization on the prediction skill, the following conclusions have been drawn:

- No significant differences between the ACC of the three hindcasts and the historical experiment have been found using the ensemble means due to the limited sample size. An alternative way, using not the ensemble means as predictions but the individual members each correlated with the observations, was considered to increase the sample size. This method does not provide a fair estimate of prediction skill but allows for a fair comparison between systems. It allowed to see significant differences in pan-Arctic SIE skill between the three hindcasts and the historical experiment during the first two forecast months, but none between either of them.

- The Labrador Sea stands out as the only region in which the ACC for the three hindcasts is significantly different from the skill in the historical experiment during the whole prediction. For the rest of peripheral basins, the added value of initialization in the three forecasts typically lasts 3 months.

- Unlike for the bias, there is no initialization strategy that clearly outperforms or underperforms the rest, either for the pan-Arctic or the individual regions. Moreover, no correspondence has been found between the locations and hindcasts with the largest SST skill and the largest SIC skill.

- The analysis has also been unable to detect any clear regional association, for any initialization strategy, between sea ice bias and skill improvements.

Due to the lack of a sea ice model during the production of ORAS4, excessively warm temperature provided by an ad-hoc surface algorithm have shown an impact for the reconstruction strongly assimilating it. This problem will hopefully be solved with the new ORAS5 (Zuo et al., 2018), which includes LIM2 as sea ice model.

Preliminary results of an on-going study show that SIC biases can be considerably reduced when using sea ice nudging. However, the consistency between reconstructions using sea ice restoring and independent ocean reanalyses (e.g. ORAS4) is not guaranteed at initialization and shocks are expected to occur (as in Chapter 4). To reduce the initialization shocks, weakly or strongly coupled assimilation, in which the ocean/atmosphere/sea ice observations are simultaneously assimilated, are an optimal way forward, and EC-Earth community is making progress in that direction. This analysis has only focused on the impact of these initialization strategies on Arctic sea ice. Larger (and possibly significant) differences across strategies can be expected when focusing on other variables, regions and timescales (e.g. SST in the North Atlantic, a prominent region of multidecadal variability).

6 Conclusions and perspectives

This thesis has focused on estimating the level of predictability and the mechanisms behind it at the pan-Arctic and regional scale, exploiting both idealized experiments and real predictions. We have also identified the contributions to the development of forecast biases coming from the initial conditions and the model inherent biases. This work finally aimed at providing recommendations regarding initialization strategies to minimize the forecast biases and maximize the predictive skill. In Chapter 1, we raised three main objectives, that now will be revisited in the form of various questions with their respective answers (section 6.1). Section 6.2 exposes the remaining questions that have not been answered yet or that need to be further investigated to obtain more conclusive answers.

6.1 Conclusions

❶ Investigating the sources of sea ice predictability at pan-Arctic and regional scales using a perfect model approach

▶ Is the level of potential predictability of the pan-Arctic and regional sea ice model dependent?

This question has been answered in Chapter 3 by comparing an ensemble of idealized predictions with 6 different GCMs, all coordinated by the APPOSITE project and initialized in July. A reemergence in pan-Arctic SIE predictability was consistently found the first and the second winter for HadGEM1.2, GFDL-CM3 and E6F models. Long control simulations allowed us to identify a summer-to-summer reemergence of pan-Arctic SIE with similar amplitude in all models, which was related to the summer SIT memory, as suggested by Blanchard-Wrigglesworth et al. (2011a). For the pan-Arctic SIV, the 6 APPOSITE models

showed a persistence that ranged from 1 to 3 years. This SIV persistence was related to the long-lasting persistence of SIT anomalies in the central Arctic.

Despite the different climatological sea ice states between the models, robust commonalities have been found for the predictability signal of some regions: a winter SIE reemergence in the Baffin Bay and GIN seas, and a summer SIV reemergence in the Barents, Kara and Chukchi seas.

▶ Is there a dependence of potential predictability on the initial sea ice state?

The start dates of the idealized predictions from the APPOSITE project were chosen to cover a wide range of different sea ice and AHT states and to be sufficiently separated in time along the present-day control simulation to be considered as independent (Day et al., 2016). Correlating the initial SIE and SIV states with the SIE ensemble spread at different lead times has shown that the initial state can significantly condition the level of skill: the larger the SIE and SIV at initialization, the larger the potential predictability. This illustrates how crucial it is to include as many start dates as possible to smooth out the dependence of predictability on the initial sea ice state and thus produce robust estimates of predictability or level of skill.

▶ What are the main mechanisms of sea ice predictability at regional scale for EC-Earth?

The analysis of the regional sea ice predictability has demonstrated that Arctic regions can be classified according to three distinct regimes: (1) in the majority of the peripheral basins of the Atlantic sector, the melt-to-freeze memory reemergence dominates, which could be associated with the persistence of local oceanic thermal anomalies; (2) in the central Arctic basins, a reemergence of predictability occurs during the summer and is associated with the large SIT persistence in the central Arctic; (3) in the Labrador Sea, the predictability reemergence is related to the persistence of ocean heat content anomalies in the subpolar gyre, which are mostly local in the first winter and advected from the Irminger Sea and Eastern North Atlantic in the second and third winters, respectively.

❷ Identifying and estimating different sources of Arctic sea ice forecast biases

▶ How does an inconsistency between the initial conditions of the different model components affect the development of the Arctic sea ice biases during the first forecast month?

A large mismatch was found in Chapter 4 between the sea ice and ocean initial conditions in two sets of seasonal predictions run with EC-Earth3.2, initialized on May 1st and on November 1st, respectively. For both, the incompatibility occurs because the ocean initial state (from ORAS4) is too warm across the whole Arctic for the overlaying sea ice conditions (from an in-house reconstruction). The initial ocean impacted the forecast by producing a fast sea ice melting, which was of larger amplitude in November due to the opposing effect of the sudden melting and the seasonal sea ice expansion. To understand how much SIV loss can be accounted for by the excessive heat in the ocean, we computed the differences in ocean heat content in the mixed layer between ORAS4 and the in-house reconstruction from which the sea ice initial conditions are extracted, and estimated the maximum volume of sea ice that it could melt. This budget could explain the SIV melting in the Baffin Bay and Greenland Sea, but not over other regions, such as the Barents, Kara, Chukchi and Bering Seas, where the underestimated SIC anomalies developing during the first forecast month seem to be associated with systematic model errors. A mismatch between the initial sea ice and atmosphere was also identified, but it was found to play a small contribution to the initial forecasted sea ice melting, mainly because of the low air heat capacity.

▶ How does the model drift compete with the initial conditions inconsistency to generate the first month sea ice biases?

The initial conditions inconsistency impacts the sea ice bias beyond the first forecast day. A comparison between the patterns of SIC forecast bias and the patterns of initial conditions inconsistency on one hand, or the patterns of model systematic error on the other hand showed that the initial conditions inconsistency dominates the bias during the first 25 (19) days in May (November), while the model systematic error dominates the bias afterwards. In the Baffin Bay and the Greenland Sea, the forecast error is still mostly explained by the initial shock by the end of the month, which suggests that such initialization errors can affect the quality of the seasonal predictions. The forecast bias has not reached the model attractor by the end of the month.

❸ Quantifying and comparing the Arctic sea biases and prediction skill of three seasonal forecast systems initialized through different strategies

▶ Should we systematically expect a reduction of the model inherent biases with a strong nudging?

In chapter 5, two different consistent ocean-sea ice initialization strategies were designed, only differing in the strength of their ocean surface restoring towards ORAS4 (strong vs weak). Sea ice biases are reduced for the reconstruction with the weak restoring coefficient compared to a historical experiment, i.e. still too much ice but to a lesser extent, while the biases have opposite sign with the strong nudging, i.e. deficit in sea ice. The biases of opposite sign for the strong nudging reconstruction are derived from an inconsistency between the ocean and sea ice fields in ORAS4, which are due to the lack of a sea ice model in ORAS4. To constraint ORAS4 towards the SIC observations, an ad-hoc algorithm is used instead, which works as follows: SST is systematically set at the freezing point whenever the observed SIC $> 55\%$, while SST is nudged towards observed SSTs whenever observed SIC $< 55\%$. This method does not prevent the SST from being above the freezing point when there is sea ice with observed SIC $< 55\%$. This algorithm leads to a warm bias in ORAS4 over some locations, and a strong constraint of the ocean towards ORAS4 produces an ocean and sea ice state with a deficit of sea ice and too warm ocean in the North Pole. These initial biases which are not aligned with the inherent model bias lead to a strong forecast drift.

▶ Which initialization strategy leads to the lowest forecast sea ice bias? Does a strong restoring coefficient to produce the initial conditions reduce the subsequent forecast sea ice biases in comparison to a weak nudging?

During the first week after initialization, the forecast sea ice bias depends mostly on the sea ice reconstruction used as initial conditions, with the prediction using ORAS4 as the initial ocean showing the lowest error. This is due to a compensation of errors between its sea ice and ocean initial conditions. After 20 days, this prediction and the one initialized from the strong nudging coefficient reconstruction converge. This convergence illustrates the key role of the initial ocean, which is close for both predictions, on the sea ice state. The prediction initialized from the reconstruction with a weak restoring coefficient shows a larger error than the other two, since the weak ocean nudging kept the reconstruction closer to

the model attractor (i.e. the error in the historical experiment). In the Atlantic sector, the prediction initialized from the strong nudging reconstruction is the one exhibiting the lowest SIC and SST biases in both DJF and MAM. On the Pacific coast, it is the one initialized from ORAS4 ocean and the sea ice from the weak nudging reconstruction which provides the best SIC and SST climatological estimates. This indicates, once again, that constraining the initial ocean closer to observations helps reducing the sea ice biases, and therefore, using a reconstruction with strong ocean nudging or an interpolated reanalysis at initialization is more efficient for minimizing forecast biases.

▶ Is there an initialization method that outperforms the others in terms of sea ice forecast skill?

For the whole Arctic, no significant differences were found between the sea ice skill of the three forecast systems. However, during the first two forecast months, all three systems showed significantly improved SIE skill when compared to the historical experiments. Regionally, the Labrador Sea outstands as the only basin with a significant improvement in skill for the three systems when compared to the historical during the whole prediction (7 months). For the rest of peripheral seas, the significant benefits of initialization last between 2-3 months. The system directly initialized from ORAS4 has significantly higher skill than the other two in the Bering Sea during DJF and the Sea of Okhotsk during MAM. For the other regions no significant differences in skill among the forecast systems are found for either of the two seasons. Regionally, the best performing systems in terms of SIC skill do not correspond with the best ones for SST. Our analysis suggests that a larger sample (more ensemble members and/or start dates) is needed to identify more robust significant differences between the three forecast systems. It also implies that for prediction systems covering relatively short training periods and with a relatively small ensemble size, the methodological choices herein tested for initialization do not influence substantially the forecast skill.

6.2 Perspectives

The mechanisms proposed in Chapter 3 (reemergence of predictability in the MIZ due to ocean thermal anomalies, reemergence of predictability in the Central Arctic thanks to SIT persistence and memory reemergence in the Labrador Sea related to advection of SST anomalies) have been identified in EC-Earth. It remains to be investigated whether these mechanisms are present in the other APPOSITE models, as well as whether there is a dependency

of these mechanisms on the model background state. Additionally, some of these predictability mechanisms might be misrepresented, or represented differently in the various EC-Earth forecast systems analyzed in Chapter 4 and 5, due to the effect of their systematic model biases. Understanding whether (and in this case why) this happens would be of great interest to understand the limitations that current prediction systems face.

The level of potential predictability that has been estimated in Chapter 3 seems higher than the level of skill estimated in Chapters 4 and 5, which indicates that there is potential room for improvement. In order to make a fair comparison between the predictive capacity and the potential predictability obtained in a perfect model approach, two different experiments with the same forecast system could be conducted, as in Bushuk et al. (2019). These "twin" experiments would open the possibility to investigate robustly whether the mechanisms of potential predictability identified for EC-Earth are also present under non-perfect initialization. They would also allow to quantify how far the forecast skill in real hindcasts is from the one in idealized ones, through sensitivity experiments to the initial conditions. A complementary experiment would consist of degrading the initial conditions used in the perfect model initialization setup to match the levels of observational uncertainty present in real hindcasts. The impact of the initial observational uncertainty on the prediction skill could then be mapped.

To further understand the physical processes behind the model biases, the experiments from Chapters 4 and 5 can be further exploited. Additional focus could be put on the chronology of the development of the forecast biases, namely whether the sea ice biases are established slightly before the ocean ones or viceversa in the adjustment towards the systematic model error. This would allow to determine causes and consequences in the development of the model inherent biases. As there was a limitation in the number and frequency of the output variables in the forecast analyzed in Chapter 4, a deeper analysis on the mechanisms leading to its sea ice forecast biases would require to produce equivalent simulations with a complete set of ocean and sea ice variables at daily (or even subdaily) output.

In Chapter 5, the impacts of native and non-native sea ice-ocean initialization on the sea ice bias and skill have been compared. However, the potential added-value of using initial conditions from a native atmosphere has not been estimated, which could be particularly important during the first days of the forecast. A robust assessment of the impact of initialization inconsistencies between the atmosphere and other model components would require a comparison of Chapter 5 experiments with another forecast system produced with consistent

(and native) atmosphere, sea ice and ocean initial conditions.

Non-native and inconsistent initial conditions between model components have been shown to be detrimental to forecast quality. The latest developments for EC-Earth are heading towards a strongly coupled data assimilation system, which aims at reducing initial shocks by building consistent information at initialization. This will definitely allow to investigate if a strong assimilation towards observational data has a beneficial effect, without the presence of initial shocks.

Analysis in Chapter 5 has shown how some inconsistencies between the sea ice boundary conditions and the reconstructed SSTs in ORAS4 can be inherited when producing in-house reconstructions, in particular when strong nudging coefficients are considered. Unlike ORAS4, the production of the new ORAS5 reanalysis includes the LIM2 sea ice model. This augurs the end of inconsistencies between the reanalyzed surface temperatures and sea ice cover in the polar regions. Hence, it is envisaged to repeat the set of reconstructions and hindcasts using the strong nudging to ORAS5, to see if in this case more clear improvements with respect to the other initialization strategies are identified.

Appendix

This appendix includes the additional figures mentioned along this thesis, which provide supportive information to the different results discussed in chapters 3 to 5.

Extra figures for Chapter 3

	SIE	SIV	AHT	SIE/SIV trend	AHT Trend
2120	7(H)	10(H)	1.5(H)	D	D
2132	6.3(M)	5(M)	1.6(H)	D	D
2140	7(H)	7(H)	0.8(L)	D	I
2156	7(H)	8.6(H)	0.6(L)	D	I
2190	6(M)	6(M)	1.1(M)	I	I
2220	4.9(L)	3.5(L)	1.3(M)	D-I	I
2237	5.2(L)	3.5(L)	1.5(H)	I	D
2260	5.5(L)	5.5(M)	1.1(M)	D-I	D
2294	6.5(H)	7.5(H)	1.4(H)	D	D

Table A.1: Table summarizing the reasons for the choice of the different start dates in EC-Earth2.3. These focus on the Arctic SIE (in millions of square kilometers), SIV (in thousands of cubic kilometers) and Atlantic heat transport (AHT) into the Arctic (in PW), at the start of the prediction in July, and their correspondent trends. Here, H is high, M is medium, L is low, I is increasing, and D is decreasing.

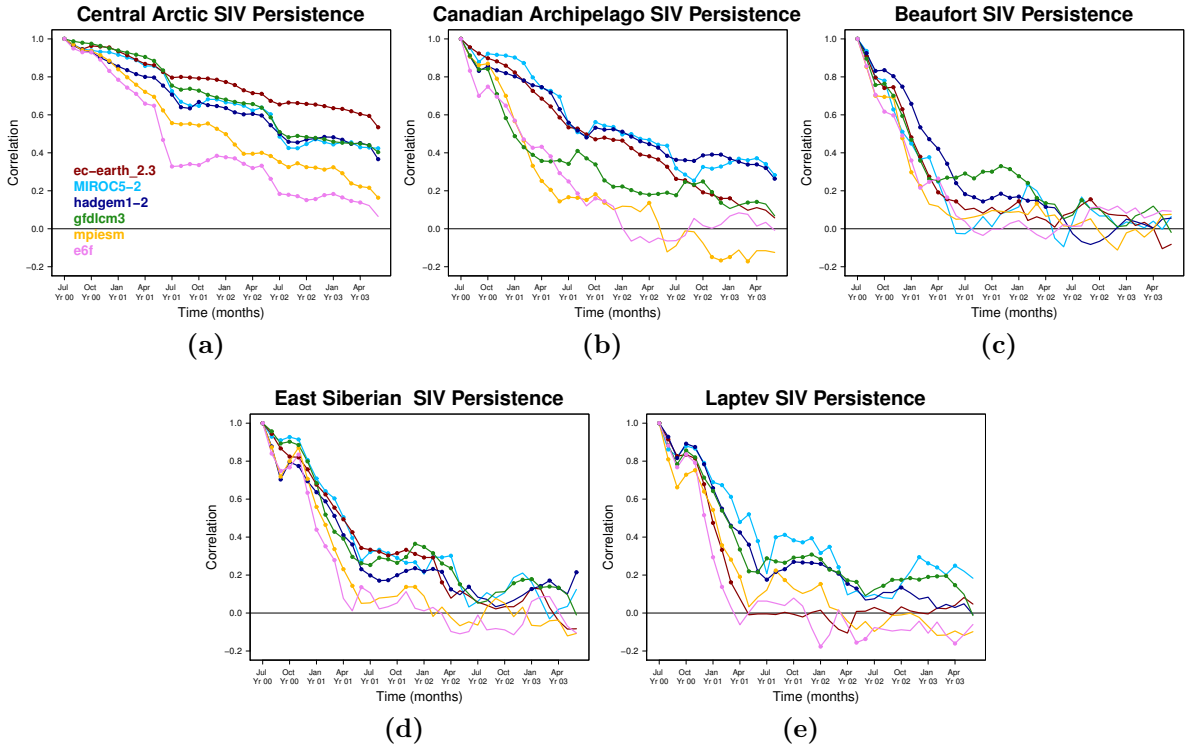


Figure A.1: SIV persistence for the **(a)** Central Arctic, **(b)** Canadian Archipelago, **(c)** Beaufort Sea, **(d)** East Siberian Sea and **(e)** Laptev Sea. Correlations were calculated using the *ControlRun* during the three subsequent years. The dots represent significant values at the 95% level as estimated from a one-sided student-T distribution.

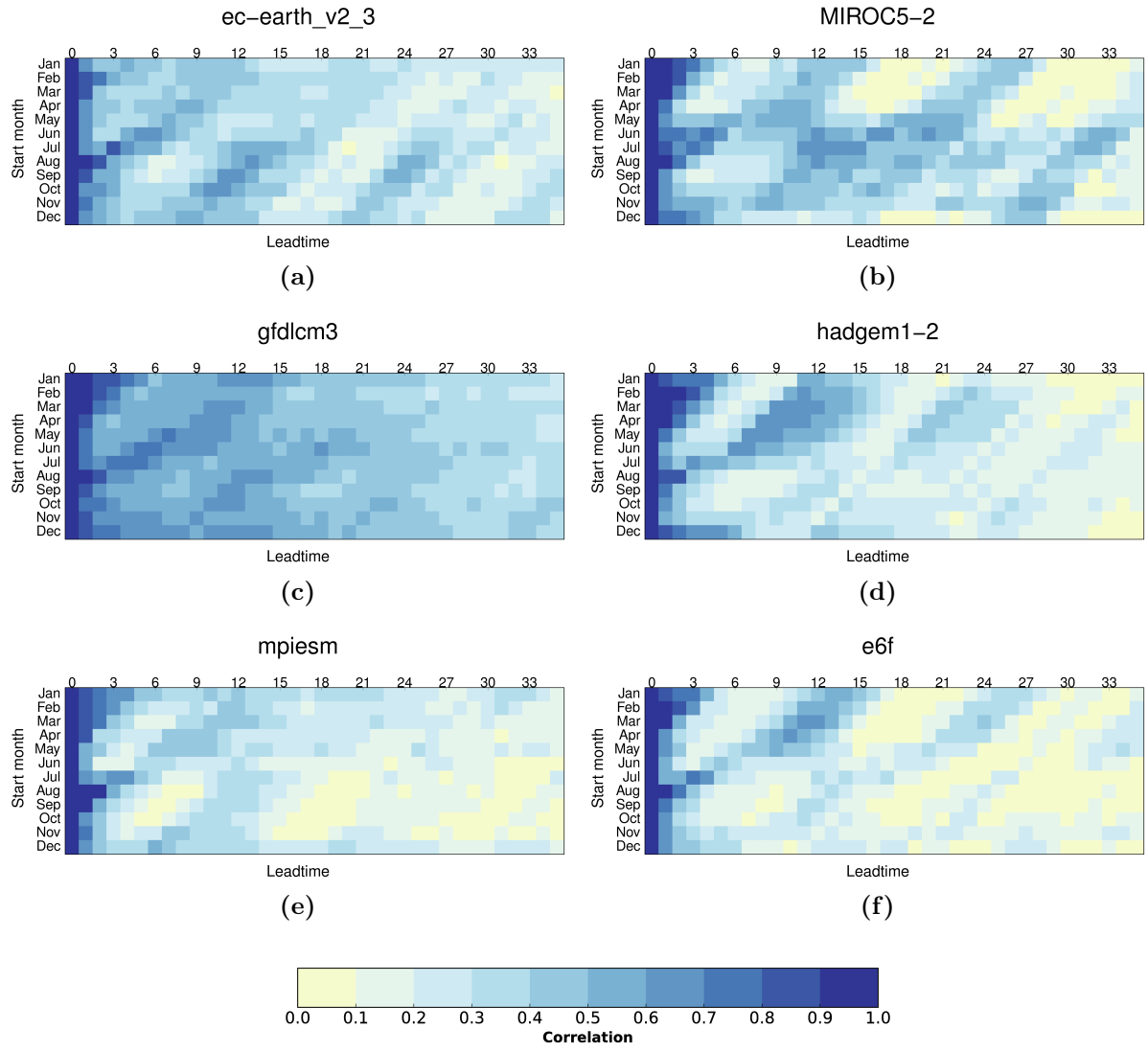


Figure A.2: SIE *ControlRun* autocorrelation of the anomalies for each month against increasing lead times. Numbers along the horizontal axis indicate the lead time (in months) from the start date indicated on the vertical axis.

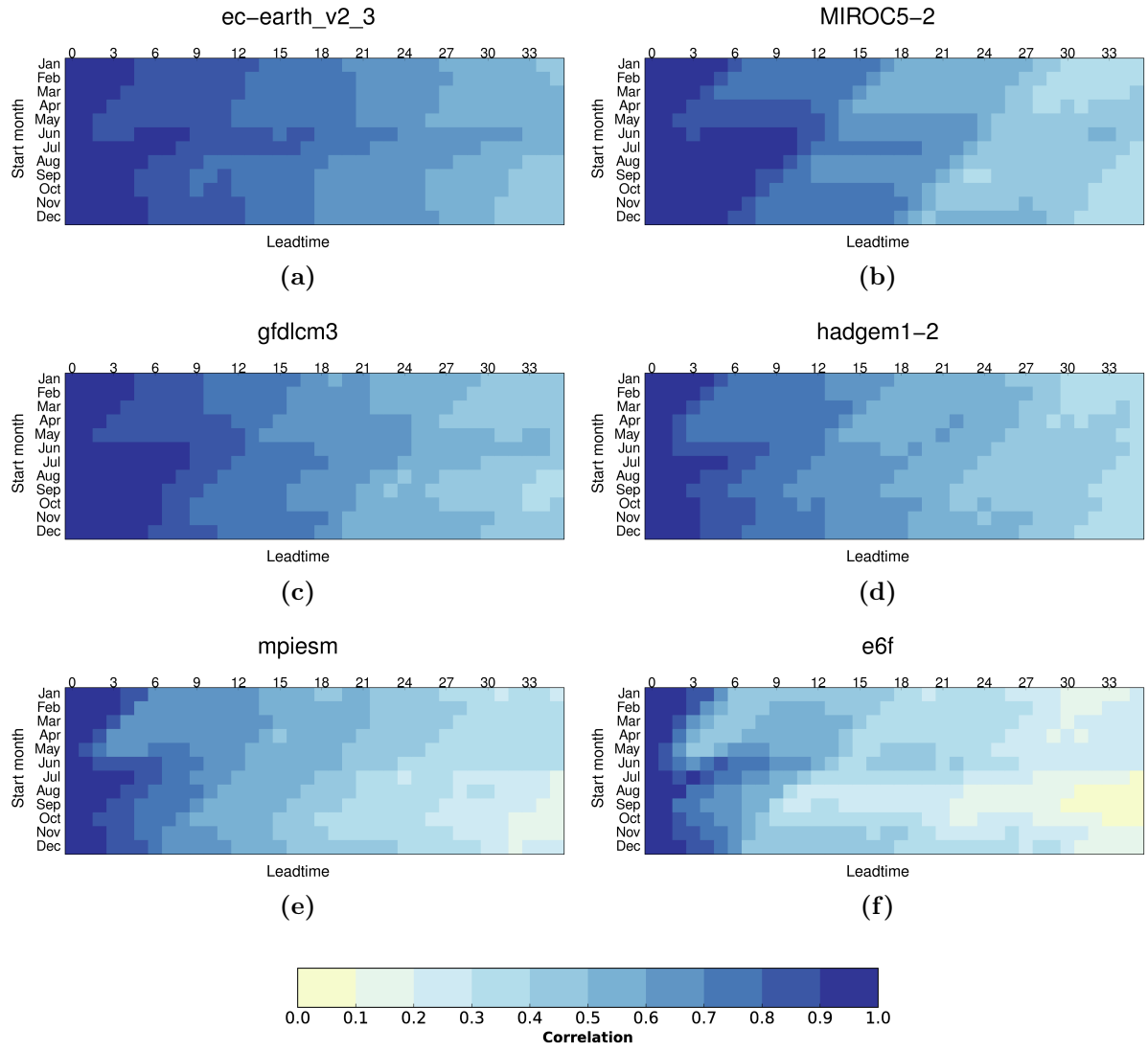
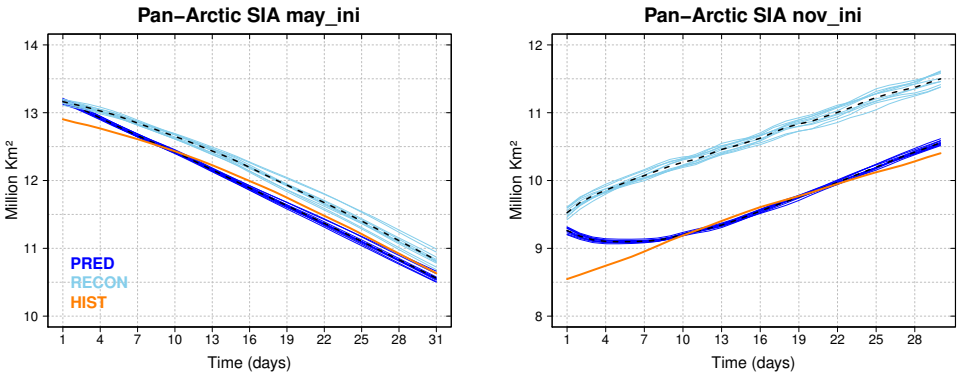


Figure A.3: SIV *ControlRun* autocorrelation of the anomalies for each month against increasing lead times. Numbers along the horizontal axis indicate the lead time (in months) from the start date indicated on the vertical axis.

Extra figures for Chapter 4



(a) **(b)**

Figure A.4: Pan-Arctic SIA daily climatology (1993-2008) for **(a)** May and **(b)** November for PRED (dark blue), RECON (light blue) and HIST (orange).

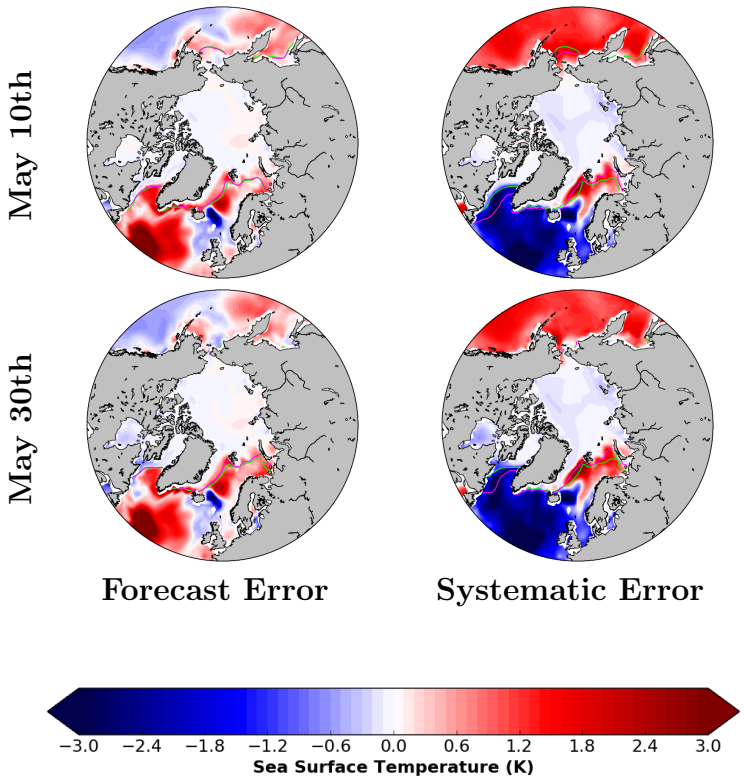


Figure A.5: Evolution of SST forecast errors for the predictions initialized in May. SST difference between PRED initialized in May and RECON (PRED minus RECON, left column) and the model systematic error (HIST minus RECON, right column) for lead times 10 and 30 for the period 1993-2008. The green line represents RECON sea ice edge (15% SIC) climatology in all panels. For the first row, the purple line represents $ORAS4\_ice$ sea ice edge. For the second to fourth rows, the purple line represents the sea ice edge of PRED (left column) and HIST (right column). For all panels blue colours represent areas where RECON has colder SST than the corresponding experiment, while red colours show warmer SST for RECON.

Extra figures for Chapter 5

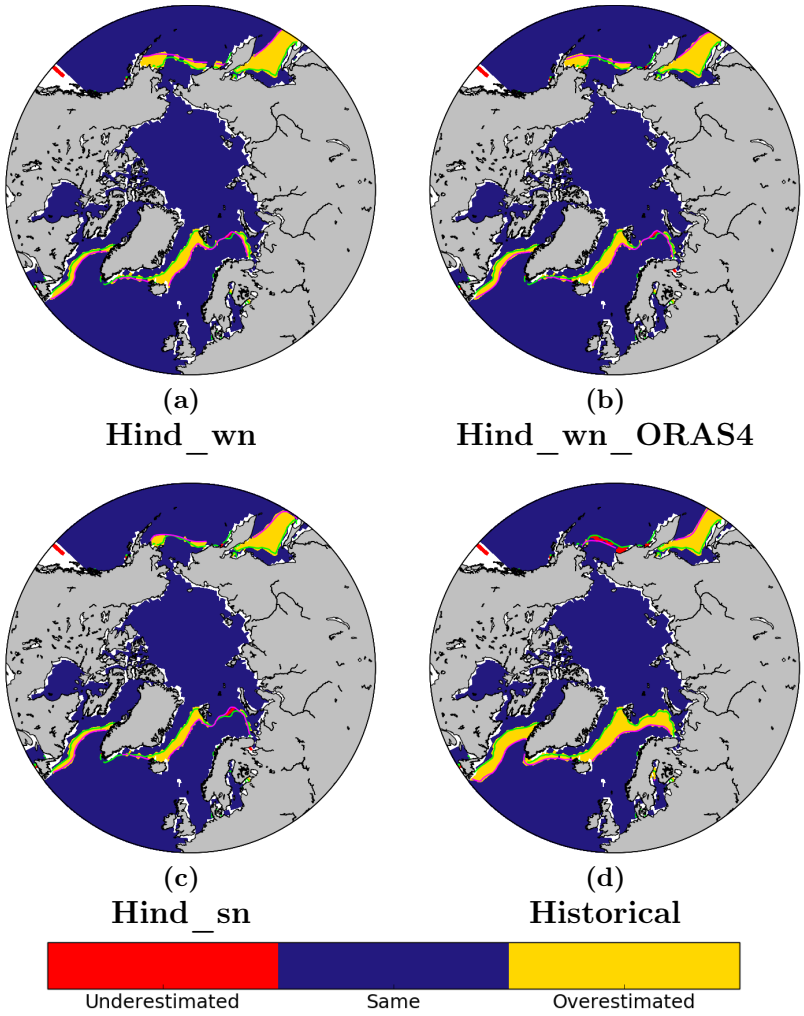


Figure A.6: IIEE maps for lead time 181.

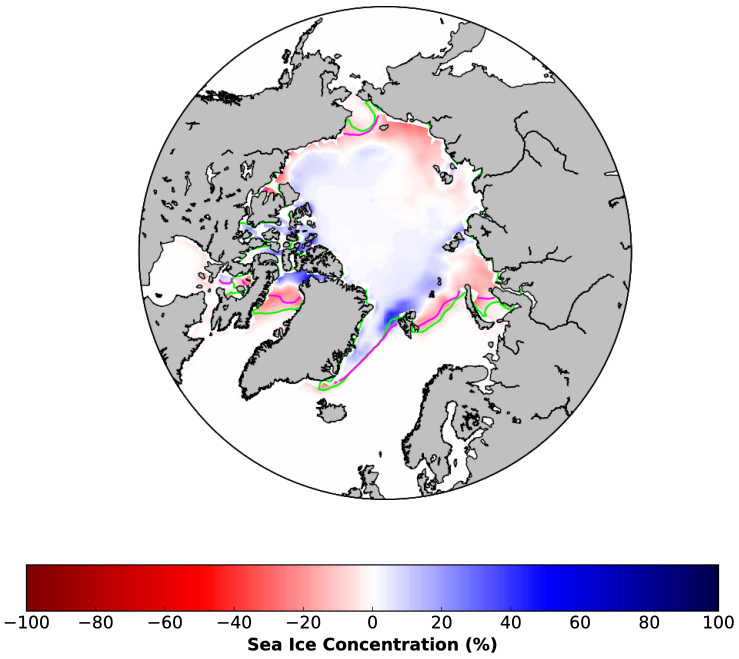


Figure A.7: SIC difference between Recon_sn and ORAS4 for November 1st averaged over the period 1988-2012.

List of acronyms

AA	Arctic amplification
ACC	Anomaly Correlation Coefficient
AEE	Absolute Extent Error
AHT	Atlantic Heat Transport into the Arctic
APPOSITE	Arctic Predictability and Prediction on Seasonal-to-Interannual Timescales
DFS	Drakkar Forcing Set
DJF	December-January-February
ECMWF	European Center for Medium-Range Weather Forecasts
EnKF	Ensemble Kalman Filter
ENSO	El Niño–Southern Oscillation
ESA	European Space Agency
GHG	Greenhouse Gas
ICs	Initial Conditions
IFS	Integrated Forecasting System
IIEE	Integrated Ice Edge Error
ITD	Ice Thickness Distribution
LIM	Louvain-la-Neuve Sea Ice Model
MAM	March-April-May
ME	Misplacement Error
MIZ	Marginal Ice Zone
NAO	North Atlantic Oscillation

NASA	National Aeronautics and Space Administration
NEMO	Nucleus for European Modelling of the Ocean
NSIDC	National Snow and Ice Data Center
ORAS4	Ocean Reanalysis System 4
PPP	Prognostic Potential Predictability
RMSE	Root Mean-Square Error
SAT	Surface Air Temperature
SD	Standard Deviation
SIA	Sea Ice Area
SIC	Sea Ice Concentration
SIE	Sea Ice Extent
SIT	Sea Ice Thickness
SIV	Sea Ice Volume
SST	Sea Surface Temperature

References

Acosta Navarro, J. C., Ortega, P., García-Serrano, J., Guemas, V., Tourigny, E., Cruz-García, R., Massonnet, F., and Doblas-Reyes, F. J. (2019). December 2016: Linking the lowest arctic sea-ice extent on record with the lowest european precipitation event on record. *Bulletin of the American Meteorological Society*, 100(1):S43–S48.

Arribas, A., Glover, M., Maidens, A., Peterson, K., Gordon, M., MacLachlan, C., Graham, R., Fereday, D., Camp, J., Scaife, A., et al. (2011). The glosea4 ensemble prediction system for seasonal forecasting. *Monthly Weather Review*, 139(6):1891–1910.

Årthun, M., Eldevik, T., Smedsrud, L. H., Skagseth, Ø., and Ingvaldsen, R. (2012). Quantifying the influence of atlantic heat on barents sea ice variability and retreat. *Journal of Climate*, 25(13):4736–4743.

Atger, F. (1999). The skill of ensemble prediction systems. *Monthly Weather Review*, 127(9):1941–1953.

Balmaseda, M. and Anderson, D. (2009). Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, 36(1).

Balmaseda, M. A., Alves, O. J., Arribas, A., Awaji, T., Behringer, D. W., Ferry, N., Fujii, Y., Lee, T., Rienecker, M., Rosati, T., et al. (2009). Ocean initialization for seasonal forecasts. *Oceanography*, 22(3):154–159.

Balmaseda, M. A., Ferranti, L., Molteni, F., and Palmer, T. N. (2010). Impact of 2007 and 2008 arctic ice anomalies on the atmospheric circulation: Implications for long-range predictions. *Quarterly Journal of the Royal Meteorological Society*, 136(652):1655–1664.

Balmaseda, M. A., Mogensen, K., and Weaver, A. T. (2013). Evaluation of the ecmwf ocean reanalysis system oras4. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1132–1161.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K. (2009). A revised hydrology for the ecmwf model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of hydrometeorology*, 10(3):623–643.

Barry, R. G., Serreze, M. C., Maslanik, J. A., and Preller, R. H. (1993). The arctic sea ice-climate system: Observations and modeling. *Reviews of Geophysics*, 31(4):397–422.

Bekryaev, R. V., Polyakov, I. V., and Alexeev, V. A. (2010). Role of polar amplification in long-term surface air temperature variations and modern arctic warming. *Journal of Climate*, 23(14):3888–3906.

Bellprat, O., García-Serrano, J., Fučkar, N. S., Massonnet, F., Guemas, V., and Doblas-Reyes, F. J. (2016). The role of arctic sea ice and sea surface temperatures on the cold 2015 february over north america. *Bulletin of the American Meteorological Society*, 97(12):S36–S41.

Bitz, C., Holland, M., Hunke, E., and Moritz, R. (2005). Maintenance of the sea-ice edge. *Journal of climate*, 18(15):2903–2921.

Bitz, C., Holland, M., Weaver, A., and Eby, M. (2001). Simulating the ice-thickness distribution in a coupled climate model. *Journal of Geophysical Research: Oceans*, 106(C2):2441–2463.

Bitz, C. and Roe, G. (2004). A mechanism for the high rate of sea ice thinning in the arctic ocean. *Journal of Climate*, 17(18):3623–3632.

Blanchard-Wrigglesworth, E., Armour, K. C., Bitz, C. M., and DeWeaver, E. (2011a). Persistence and inherent predictability of arctic sea ice in a gcm ensemble and observations. *Journal of Climate*, 24(1):231–250.

Blanchard-Wrigglesworth, E., Barthélemy, A., Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., Posey, P., Wang, W., Zhang, J., Ardilouze, C., et al. (2016). Multi-model seasonal forecast of arctic sea-ice: forecast uncertainty at pan-arctic and regional scales. *Climate Dynamics*, pages 1–12.

Blanchard-Wrigglesworth, E., Bitz, C., and Holland, M. (2011b). Influence of initial conditions and climate forcing on predicting arctic sea ice. *Geophysical Research Letters*, 38(18).

Blanke, B., Arhan, M., Madec, G., and Roche, S. (1999). Warm water paths in the equatorial atlantic as diagnosed with a general circulation model. *Journal of Physical Oceanography*, 29(11):2753–2768.

Blanke, B. and Raynaud, S. (1997). Kinematics of the pacific equatorial undercurrent: An eulerian and lagrangian approach from gcm results. *Journal of Physical Oceanography*, 27(6):1038–1053.

Blockley, E. W. and Peterson, K. A. (2018). Improving met office seasonal predictions of arctic sea ice using assimilation of cryosat-2 thickness. *The Cryosphere*, 12(11):3419–3438.

Blyverket, J., Hamer, P. D., Bertino, L., Albergel, C., Fairbairn, D., and Lahoz, W. A. (2019). An evaluation of the enkf vs. enoi and the assimilation of smap, smos and esa cci soil moisture data over the contiguous us. *Remote Sensing*, 11(5):478.

Boer, G. J. (2004). Long time-scale potential predictability in an ensemble of coupled climate models. *Climate dynamics*, 23(1):29–44.

Bouillon, S., Maqueda, M. A. M., Legat, V., and Fichefet, T. (2009). An elastic–viscous–plastic sea ice model formulated on arakawa b and c grids. *Ocean Modelling*, 27(3):174–184.

Bunzel, F., Notz, D., Baehr, J., Müller, W. A., and Fröhlich, K. (2016). Seasonal climate forecasts significantly affected by observational uncertainty of arctic sea ice concentration. *Geophysical Research Letters*, 43(2):852–859.

Bushuk, M., Giannakis, D., and Majda, A. J. (2015). Arctic sea ice reemergence: The role of large-scale oceanic and atmospheric variability. *Journal of Climate*, 28(14):5477–5509.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Yang, X., Rosati, A., and Gudgel, R. (2019). Regional arctic sea–ice prediction: potential versus operational seasonal forecast skill. *Climate Dynamics*, 52(5-6):2721–2743.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., and Yang, X. (2017). Skillful regional prediction of arctic sea ice on seasonal timescales. *Geophysical Research Letters*, 44(10):4953–4964.

Carmack, E., Polyakov, I., Padman, L., Fer, I., Hunke, E., Hutchings, J., Jackson, J., Kelley, D., Kwok, R., Layton, C., et al. (2015). Toward quantifying the increasing role of oceanic heat in sea ice loss in the new arctic. *Bulletin of the American Meteorological Society*, 96(12):2079–2105.

Cavalieri, D., Parkinson, C., Gloersen, P., and Zwally, H. (1996). Sea ice concentrations from nimbus-7 smmr and dmsp ssm/i passive microwave data. *Digital Media, National Snow and Ice Data Center*.

Cavalieri, D., Parkinson, C., and Vinnikov, K. Y. (2003). 30-year satellite record reveals contrasting arctic and antarctic decadal sea ice variability. *Geophysical Research Letters*, 30(18).

Chen, M., Wang, W., and Kumar, A. (2013). Lagged ensembles, forecast configuration, and seasonal predictions. *Monthly Weather Review*, 141(10):3477–3497.

Cheng, W., Blanchard-Wrigglesworth, E., Bitz, C. M., Ladd, C., and Stabeno, P. J. (2016). Diagnostic sea ice predictability in the pan-arctic and us arctic regional seas. *Geophysical Research Letters*, 43(22).

Chevallier, M. and Salas y Mélia, D. (2012). The role of sea ice thickness distribution in the arctic sea ice potential predictability: A diagnostic approach with a coupled gcm. *Journal of Climate*, 25(8):3025–3038.

Chevallier, M., Salas y Mélia, D., Voldoire, A., Déqué, M., and Garric, G. (2013). Seasonal forecasts of the pan-arctic sea ice extent using a gcm-based seasonal prediction system. *Journal of Climate*, 26(16):6092–6104.

Chevallier, M., Smith, G. C., Dupont, F., Lemieux, J.-F., Forget, G., Fujii, Y., Hernandez, F., Msadek, R., Peterson, K. A., Storto, A., et al. (2017). Intercomparison of the arctic sea ice cover in global ocean–sea ice reanalyses from the ora-ip project. *Climate Dynamics*, 49(3):1107–1136.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., et al. (2014). Recent arctic amplification and extreme mid-latitude weather. *Nature geoscience*, 7(9):627.

Cohen, J. L., Furtado, J. C., Barlow, M. A., Alexeev, V. A., and Cherry, J. E. (2012). Arctic warming, increasing snow cover and widespread boreal winter cooling. *Environmental Research Letters*, 7(1):014007.

Collins, M. (2002). Climate predictability on interannual to decadal time scales: the initial value problem. *Climate Dynamics*, 19(8):671–692.

Collow, T. W., Wang, W., Kumar, A., and Zhang, J. (2015). Improving arctic sea ice prediction using piomas initial sea ice thickness in a coupled ocean–atmosphere model. *Monthly Weather Review*, 143(11):4618–4630.

Comiso, J. (2017). Bootstrap sea ice concentrations from nimbus-7 smmr and dmsp ssm/i-ssmis, version 3. *NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, CO*.

Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L. (2008). Accelerated decline in the arctic sea ice cover. *Geophysical research letters*, 35(1).

Coumou, D., Di Capua, G., Vavrus, S., Wang, L., and Wang, S. (2018). The influence of arctic amplification on mid-latitude summer circulation. *Nature communications*, 9(1):2959.

Cvijanovic, I., Santer, B. D., Bonfils, C., Lucas, D. D., Chiang, J. C., and Zimmerman, S. (2017). Future loss of arctic sea-ice cover could drive a substantial decrease in california's rainfall. *Nature communications*, 8(1):1947.

Day, J., Hargreaves, J., Annan, J., and Abe-Ouchi, A. (2012). Sources of multi-decadal variability in arctic sea ice extent. *Environmental Research Letters*, 7(3):034011.

Day, J., Tietsche, S., Collins, M., Goessling, H., Guemas, V., Guillory, A., Hurlin, W., Ishii, M., Keeley, S., Matei, D., et al. (2016). The arctic predictability and prediction on seasonal-to-interannual timescales (apposite) data set. *Geoscientific Model Development*, 8(10).

Day, J., Tietsche, S., and Hawkins, E. (2014). Pan-arctic and regional sea ice predictability: Initialization month dependence. *Journal of Climate*, 27(12):4371–4390.

Decremer, D., Chung, C. E., Ekman, A. M., and Brandefelt, J. (2014). Which significance test performs the best in climate simulations? *Tellus A: Dynamic Meteorology and Oceanography*, 66(1):23139.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Deser, C., Tomas, R., Alexander, M., and Lawrence, D. (2010). The seasonal atmospheric response to projected arctic sea ice loss in the late twenty-first century. *Journal of Climate*, 23(2):333–351.

Ding, Q., Schweiger, A., L'Heureux, M., Battisti, D. S., Po-Chedley, S., Johnson, N. C., Blanchard-Wrigglesworth, E., Harnos, K., Zhang, Q., Eastman, R., et al. (2017). Influence of high-latitude atmospheric circulation changes on summertime arctic sea ice. *Nature Climate Change*, 7(4):289.

Dirkson, A., Denis, B., and Merryfield, W. (2019). A multimodel approach for improving seasonal probabilistic forecasts of regional arctic sea ice. *Geophysical Research Letters*, 46(19):10844–10853.

Dirkson, A., Merryfield, W. J., and Monahan, A. (2017). Impacts of sea ice thickness initialization on seasonal arctic sea ice predictions. *Journal of Climate*, 30(3):1001–1017.

Divine, D. and Dick, C. (2007). March through august ice edge positions in the nordic seas, 1750-2002, version 1. *NSIDC National Snow and Ice Data Center*.

Doblas-Reyes, F., Acosta Navarro, J., Acosta, M., Bellprat, O., Bilbao, R., Castrillo, M., Fuckar, N., Guemas, V., Lledó, L., Ménégoz, M., et al. (2018). Using EC-Earth for climate prediction research. Technical Report 154, ECMWF.

Doblas-Reyes, F., Pavan, V., and Stephenson, D. (2003). The skill of multi-model seasonal forecasts of the wintertime north atlantic oscillation. *Climate dynamics*, 21(5-6):501–514.

Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., and Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4):245–268.

Docquier, D., Grist, J. P., Roberts, M. J., Roberts, C. D., Semmler, T., Ponsoni, L., Massonnet, F., Sidorenko, D., Sein, D. V., Iovino, D., et al. (2019). Impact of model resolution on arctic sea ice and north atlantic ocean heat transport. *Climate Dynamics*, pages 1–29.

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., Golaz, J.-C., Ginoux, P., Lin, S.-J., Schwarzkopf, M. D., et al. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component am3 of the gfdl global coupled model cm3. *Journal of Climate*, 24(13):3484–3519.

Dulière, V. and Fichefet, T. (2007). On the assimilation of ice velocity and concentration data into large-scale sea ice models. *Ocean Science Discussions*, 4(2):265–301.

Dussin, R., Barnier, B., Brodeau, L., and Molines, J. (2016). The making of the drakkar forcing set dfs5. *DRAKKAR/MyOcean Rep*, pages 01–04.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., and Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, 41(15):5620–5628.

Ebert, E. E. and Curry, J. A. (1993). An intermediate one-dimensional thermodynamic sea ice model for investigating ice-atmosphere interactions. *Journal of Geophysical Research: Oceans (1978–2012)*, 98(C6):10085–10109.

Eiríksson, J., Knudsen, K. L., Haflidason, H., and Henriksen, P. (2000). Late-glacial and holocene palaeoceanography of the north icelandic shelf. *Journal of Quaternary Science: Published for the Quaternary Research Association*, 15(1):23–42.

Ethé, C., Aumont, O., Foujols, M., and Lévy, M. (2006). Nemo reference manual, tracer component: Nemo-top. *Preliminary version, Note du Pole de Modélisation, Institut Pierre-Simon Laplace*.

Evensen, G. (2003). The ensemble kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, 53(4):343–367.

Fauria, M. M., Grinsted, A., Helama, S., Moore, J., Timonen, M., Martma, T., Isaksson, E., and Eronen, M. (2010). Unprecedented low twentieth century winter sea ice extent in the western nordic seas since ad 1200. *Climate Dynamics*, 34(6):781–795.

Fichefet, T. and Maqueda, M. (1997a). Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research: Oceans (1978–2012)*, 102(C6):12609–12646.

Fichefet, T. and Maqueda, M. (1997b). Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *Journal of Geophysical Research: Oceans*, 102(C6):12609–12646.

Fleming, G. H. and Semtner Jr, A. J. (1991). A numerical study of interannual ocean forcing on arctic ice. *Journal of Geophysical Research: Oceans*, 96(C3):4589–4603.

Francis, J. A. and Hunter, E. (2007). Drivers of declining sea ice in the arctic winter: A tale of two seas. *Geophysical Research Letters*, 34(17).

Francis, J. A. and Vavrus, S. J. (2012). Evidence linking arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, 39(6).

Fu, Q. and Lin, P. (2011). Poleward shift of subtropical jets inferred from satellite-observed lower-stratospheric temperatures. *Journal of climate*, 24(21):5597–5603.

Furevik, T. (2001). Annual and interannual variability of atlantic water temperatures in the norwegian and barents seas: 1980–1996. *Deep Sea Research Part I: Oceanographic Research Papers*, 48(2):383–404.

García-Serrano, J. and Doblas-Reyes, F. (2012). On the assessment of near-surface global temperature and north atlantic multi-decadal variability in the ensembles decadal hindcast. *Climate dynamics*, 39(7-8):2025–2040.

García-Serrano, J., Guemas, V., and Doblas-Reyes, F. (2015). Added-value from initialization in predictions of atlantic multi-decadal variability. *Climate Dynamics*, 44(9-10):2539–2555.

Germe, A., Chevallier, M., y Mélia, D. S., Sanchez-Gomez, E., and Cassou, C. (2014). Interannual predictability of arctic sea ice in a global climate model: regional contrasts and temporal evolution. *Climate Dynamics*, 43(9-10):2519–2538.

Ghent, D., Kaduk, J., Remedios, J., and Balzter, H. (2011). Data assimilation into land surface models: the implications for climate feedbacks. *International journal of remote sensing*, 32(3):617–632.

Gloersen, P. and Campbell, W. J. (1988). Variations in the arctic, antarctic, and global sea ice covers during 1978–1987 as observed with the nimbus 7 scanning multichannel microwave radiometer. *Journal of Geophysical Research: Oceans*, 93(C9):10666–10674.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T. (2016). Predictability of the arctic sea ice edge. *Geophysical Research Letters*, 43(4):1642–1650.

Good, S. A., Martin, M. J., and Rayner, N. A. (2013). En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716.

Goosse, H., Arzel, O., Bitz, C. M., de Montety, A., and Vancoppenolle, M. (2009). Increased variability of the arctic summer ice extent in a warmer climate. *Geophysical Research Letters*, 36(23).

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., Jonko, A., Kushner, P. J., Lecomte, O., Massonnet, F., et al. (2018). Quantifying climate feedbacks in polar regions. *Nature communications*, 9(1):1919.

Griffies, S. M., Winton, M., Donner, L. J., Horowitz, L. W., Downes, S. M., Farneti, R., Gnanadesikan, A., Hurlin, W. J., Lee, H.-C., Liang, Z., et al. (2011). The gfdl cm3 coupled climate model: characteristics of the ocean and sea ice simulations. *Journal of Climate*, 24(13):3520–3544.

Grumet, N. S., Wake, C. P., Mayewski, P. A., Zielinski, G. A., Whitlow, S. I., Koerner, R. M., Fisher, D. A., and Woollett, J. M. (2001). Variability of sea-ice extent in baffin bay over the last millennium. *Climatic Change*, 49(1-2):129–145.

Gudkovich, Z. (1961a). On the nature of pacific currents in the bering strait and the causes of seasonal variations in their intensity. *Okeanologiya (Oceanology)*, 1(4):608.

Gudkovich, Z. (1961b). Relation of the ice drift in the arctic basin to ice conditions in the soviet arctic seas. *Tr. Okeanogr. Kom. Akad. Nauk SSSR*, 11:14–21.

Guemas, V., Auger, L., and Doblas-Reyes, F. J. (2014a). Hypothesis testing for auto-correlated short climate time series. *Journal of Applied Meteorology and Climatology*, 53(3):637–651.

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., et al. (2014b). A review on arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*.

Guemas, V., Doblas-Reyes, F. J., Mogensen, K., Keeley, S., and Tang, Y. (2014c). Ensemble of sea ice initial conditions for interannual climate predictions. *Climate dynamics*, 43(9-10):2813–2829.

Guemas, V., Doblas-Reyes, F. J., Mogensen, K., Keeley, S., and Tang, Y. (2014d). Ensemble of sea ice initial conditions for interannual climate predictions. *Climate dynamics*, 43(9-10):2813–2829.

Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—i. basic concept. *Tellus A: Dynamic Meteorology and Oceanography*, 57(3):219–233.

Hakkinen, S., Mellor, G., and Kantha, L. (1992). Modeling deep convection in the greenland sea. *Journal of Geophysical Research: Oceans*, 97(C4):5389–5408.

Hall, A. (2004). The role of surface albedo feedback in climate. *Journal of Climate*, 17(7):1550–1568.

Hall, C. M. and Saarinen, J. (2010). Polar tourism: Definitions and dimensions. *Scandinavian Journal of Hospitality and Tourism*, 10(4):448–467.

Hansen, J. E., Takahashi, T., et al. (1984). Climate processes and climate sensitivity. *Washington DC American Geophysical Union Geophysical Monograph Series, 29*.

Hawkins, E., Tietsche, S., Day, J. J., Melia, N., Haines, K., and Keeley, S. (2016). Aspects of designing and evaluating seasonal-to-interannual arctic sea-ice prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 142(695):672–683.

Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., et al. (2012). Ec-earth v2. 2: description and validation of a new seamless earth system prediction model. *Climate Dynamics*, 39(11):2611–2629.

He, Y., Wang, B., Liu, M., Liu, L., Yu, Y., Liu, J., Li, R., Zhang, C., Xu, S., Huang, W., et al. (2017). Reduction of initial shock in decadal predictions using a new initialization strategy. *Geophysical Research Letters*, 44(16):8538–8547.

Hibler III, W. (1979). A dynamic thermodynamic sea ice model. *Journal of physical oceanography*, 9(4):815–846.

Hibler III, W. (1980). Modeling a variable thickness sea ice cover. *Monthly weather review*, 108(12):1943–1973.

Hibler III, W. and Bryan, K. (1987). A diagnostic ice–ocean model. *Journal of Physical Oceanography*, 17(7):987–1015.

Hinzman, L. D., Deal, C. J., McGuire, A. D., Mernild, S. H., Polyakov, I. V., and Walsh, J. E. (2013). Trajectory of the arctic as an integrated system. *Ecological Applications*, 23(8):1837–1868.

Holland, M. M., Bailey, D. A., and Vavrus, S. (2011). Inherent sea ice predictability in the rapidly changing arctic environment of the community climate system model, version 3. *Climate dynamics*, 36(7-8):1239–1253.

Holland, M. M. and Stroeve, J. (2011). Changing seasonal sea ice predictor relationships in a changing arctic climate. *Geophysical Research Letters*, 38(18).

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R., et al. (2013). The esa climate change initiative: Satellite data records for essential climate variables. *Bulletin of the American Meteorological Society*, 94(10):1541–1552.

Honda, M., Inoue, J., and Yamane, S. (2009). Influence of low arctic sea-ice minima on anomalously cold eurasian winters. *Geophysical Research Letters*, 36(8).

Huang, J., Zhang, X., Zhang, Q., Lin, Y., Hao, M., Luo, Y., Zhao, Z., Yao, Y., Chen, X., Wang, L., et al. (2017). Recently amplified arctic warming has contributed to a continual global warming trend. *Nature Climate Change*, 7(12):875.

Hunke, E. C., Hebert, D. A., and Lecomte, O. (2013). Level-ice melt ponds in the los alamos sea ice model, cice. *Ocean Modelling*, 71:26–42.

Hunke, E. C., Lipscomb, W. H., and Turner, A. K. (2010). Sea-ice models for climate study: retrospective and new directions. *Journal of Glaciology*, 56(200):1162–1172.

Huntington, H. P., Boyle, M., Flowers, G. E., Weatherly, J. W., Hamilton, L. C., Hinzman, L., Gerlach, C., Zulueta, R., Nicolson, C., and Overpeck, J. (2007). The influence of human activity in the arctic on climate and climate impacts. *Climatic Change*, 82(1-2):77.

ICPO (2011). Data and bias correction for decadal climate predictions. *International CLIVAR Project Office Publication Series*, 150:5.

Ineson, S. and Scaife, A. (2009). The role of the stratosphere in the european climate response to el niño. *Nature Geoscience*, 2(1):32.

Infanti, J. M. and Kirtman, B. P. (2016). Prediction and predictability of land and atmosphere initialized ccsm4 climate forecasts over north america. *Journal of Geophysical Research: Atmospheres*, 121(21):12–690.

Ivanov, V., Alexeev, V., Koldunov, N. V., Repina, I., Sandø, A. B., Smedsrud, L. H., and Smirnov, A. (2016). Arctic ocean heat impact on regional ice decay: A suggested positive feedback. *Journal of Physical Oceanography*, 46(5):1437–1456.

Ivanova, N., Johannessen, O. M., Pedersen, L. T., and Tonboe, R. T. (2014). Retrieval of arctic sea ice parameters by satellite passive microwave sensors: A comparison of eleven

sea ice concentration algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):7233–7246.

Jackson, L., Kahana, R., Graham, T., Ringer, M., Woollings, T., Mecking, J., and Wood, R. (2015). Global and european climate impacts of a slowdown of the amoc in a high resolution gcm. *Climate dynamics*, 45(11-12):3299–3316.

Jahn, A. (2018). Reduced probability of ice-free summers for 1.5 c compared to 2 c warming. *Nature Climate Change*, 8(5):409.

Jin, E. K., Kinter, J. L., Wang, B., Park, C.-K., Kang, I.-S., Kirtman, B., Kug, J.-S., Kumar, A., Luo, J.-J., Schemm, J., et al. (2008). Current status of enso prediction skill in coupled ocean–atmosphere models. *Climate Dynamics*, 31(6):647–664.

Johns, T. C., Durman, C. F., Banks, H. T., Roberts, M. J., McLaren, A. J., Ridley, J. K., Senior, C. A., Williams, K., Jones, A., Rickard, G., et al. (2006). The new hadley centre climate model (hadgem1): Evaluation of coupled simulations. *Journal of Climate*, 19(7):1327–1353.

Johnson, C., Lemke, P., and Barnett, T. (1985). Linear prediction of sea ice anomalies. *Journal of Geophysical Research: Atmospheres*, 90(D3):5665–5675.

Jun, S.-Y., Ho, C.-H., Jeong, J.-H., Choi, Y.-S., and Kim, B.-M. (2016). Recent changes in winter arctic clouds and their relationships with sea ice and atmospheric conditions. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1):29130.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., et al. (2016). Advancing polar prediction capabilities on daily to seasonal time scales. *Bulletin of the American Meteorological Society*, 97(9):1631–1647.

Jung, T., Miller, M., Palmer, T., Towers, P., Wedi, N., Achuthavarier, D., Adams, J., Altshuler, E., Cash, B., Kinter Iii, J., et al. (2012). High-resolution global climate simulations with the ecmwf model in project athena: Experimental design, model climate, and seasonal forecast skill. *Journal of Climate*, 25(9):3155–3172.

Jungclaus, J., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and Storch, J. (2013). Characteristics of the ocean simulations in the max planck institute ocean model (mpiom) the ocean component of the mpi-earth system model. *Journal of Advances in Modeling Earth Systems*, 5(2):422–446.

Kam, J., Sheffield, J., Yuan, X., and Wood, E. F. (2014). Did a skillful prediction of sea surface temperatures help or hinder forecasting of the 2012 midwestern us drought? *Environmental Research Letters*, 9(3):034005.

Kapsch, M.-L., Graversen, R. G., Tjernström, M., and Bintanja, R. (2016). The effect of downwelling longwave and shortwave radiation on arctic summer sea ice. *Journal of Climate*, 29(3):1143–1159.

Karklin, V. (1977). State and future of using of geliogeophysical factors for ice forecasts. *Probl. Arkt. Antarkt*, 50:45–48.

Kashiwase, H., Ohshima, K. I., Nihashi, S., and Eicken, H. (2017). Evidence for ice-ocean albedo feedback in the arctic ocean shifting to a seasonal ice zone. *Scientific reports*, 7(1):8170.

Kauker, F., Kaminski, T., Karcher, M., Giering, R., Gerdes, R., and Voßbeck, M. (2009). Adjoint analysis of the 2007 all time arctic sea-ice minimum. *Geophysical Research Letters*, 36(3).

Kim, B.-M., Son, S.-W., Min, S.-K., Jeong, J.-H., Kim, S.-J., Zhang, X., Shim, T., and Yoon, J.-H. (2014). Weakening of the stratospheric polar vortex by arctic sea-ice loss. *Nature communications*, 5:4646.

Kim, Y., Kim, T.-H., and Ergün, T. (2015). The instability of the pearson correlation coefficient in the presence of coincidental outliers. *Finance Research Letters*, 13:243–257.

Kimmritz, M., Counillon, F., Smedsrud, L., Bethke, I., Keenlyside, N., Ogawa, F., and Wang, Y. (2019). Impact of ocean and sea ice initialisation on seasonal prediction skill in the arctic. *Journal of Advances in Modeling Earth Systems*.

Kinnard, C., Zdanowicz, C. M., Fisher, D. A., and Wake, C. P. (2006). Calibration of an ice-core glaciochemical (sea-salt) record with sea-ice variability in the canadian arctic. *Annals of Glaciology*, 44:383–390.

Knippertz, P., Ulbrich, U., Marques, F., and Corte-Real, J. (2003). Decadal changes in the link between el niño and springtime north atlantic oscillation and european–north african rainfall. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 23(11):1293–1311.

Koenigk, T. and Mikolajewicz, U. (2009). Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Climate dynamics*, 32(6):783.

Koster, R. D., Mahanama, S., Yamada, T., Balsamo, G., Berg, A., Boisserie, M., Dirmeyer, P., Doblas-Reyes, F., Drewitt, G., Gordon, C., et al. (2010). Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophysical Research Letters*, 37(2).

Krikken, F., Schmeits, M., Vlot, W., Guemas, V., and Hazeleger, W. (2016). Skill improvement of dynamical seasonal arctic sea ice forecasts. *Geophysical Research Letters*, 43(10):5124–5132.

Kröger, J., Pohlmann, H., Sienz, F., Marotzke, J., Baehr, J., Köhl, A., Modali, K., Polkova, I., Stammer, D., Vamborg, F. S., et al. (2017). Full-field initialized decadal predictions with the mpi earth system model: An initial shock in the north atlantic. *Climate Dynamics*, pages 1–16.

Kumar, A., Peng, P., and Chen, M. (2014). Is there a relationship between potential and actual skill? *Monthly Weather Review*, 142(6):2220–2227.

Labe, Z., Magnusdottir, G., and Stern, H. (2018). Variability of arctic sea ice thickness using piomas and the cesm large ensemble. *Journal of Climate*, 31(8):3233–3247.

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., and Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(694):65–78.

Lam, V. W., Cheung, W. W., Reygondeau, G., and Sumaila, U. R. (2016). Projected change in global fisheries revenues under climate change. *Scientific Reports*, 6:32607.

Lemke, P., Trinkl, E., and Hasselmann, K. (1980). Stochastic dynamic analysis of polar sea ice variability. *Journal of Physical Oceanography*, 10(12):2100–2120.

Leutbecher, M. (2018). Ensemble size: How suboptimal is less than infinity? *Quarterly Journal of the Royal Meteorological Society*.

Lindsay, R. and Zhang, J. (2006). Assimilation of ice concentration in an ice–ocean model. *Journal of Atmospheric and Oceanic Technology*, 23(5):742–749.

Liu, J., Curry, J. A., Wang, H., Song, M., and Horton, R. M. (2012). Impact of declining arctic sea ice on winter snowfall. *Proceedings of the National Academy of Sciences*, 109(11):4074–4079.

Liu, X., Köhl, A., Stammer, D., Masuda, S., Ishikawa, Y., and Mochizuki, T. (2017). Impact of in-consistency between the climate model and its initial conditions on climate prediction. *Climate Dynamics*, 49(3):1061–1075.

Madec, G. et al. (2015). Nemo ocean engine. *Note du Pole de modélisation de l'Institut Pierre-Simon Laplace No 27, ISSN No 1288-1619*.

Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., and Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate dynamics*, 41(9-10):2393–2409.

Maher, P. T. (2016). Tourism futures in the arctic. In *The interconnected Arctic UArctic Congress*, pages 213–220.

Manabe, S. and Stouffer, R. J. (1980). Sensitivity of a global climate model to an increase of co2 concentration in the atmosphere. *Journal of Geophysical Research: Oceans*, 85(C10):5529–5554.

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

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

Mathiot, P., König Beatty, C., Fichefet, T., Goosse, H., Massonnet, F., and Vancoppenolle, M. (2012). Better constraints on the sea-ice state using global sea-ice data assimilation. *Geoscientific Model Development*, 5(6):1501–1515.

McKinnon, K. A., Rhines, A., Tingley, M., and Huybers, P. (2016). Long-lead predictions of eastern united states hot days from pacific sea surface temperatures. *Nature Geoscience*, 9(5):389.

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., et al. (2014). Decadal climate prediction: an update from the trenches. *Bulletin of the American Meteorological Society*, 95(2):243–267.

Meehl, G. A., Teng, H., and Branstator, G. (2006). Future changes of el niño in two global coupled climate models. *Climate Dynamics*, 26(6):549–566.

Meier, W. N., Hovelsrud, G. K., van Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., et al. (2014). Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Reviews of Geophysics*, 52(3):185–217.

Melia, N., Haines, K., and Hawkins, E. (2016). Sea ice decline and 21st century trans-arctic shipping routes. *Geophysical Research Letters*, 43(18):9720–9728.

Ménégoz, M., Cassou, C., Swingedouw, D., Ruprich-Robert, Y., Bretonnière, P.-A., and Doblas-Reyes, F. (2018). Role of the atlantic multidecadal variability in modulating the climate response to a pinatubo-like volcanic eruption. *Climate dynamics*, 51(5-6):1863–1883.

Mikolajewicz, U., Sein, D. V., Jacob, D., Königk, T., Podzun, R., and Semmler, T. (2005). Simulating arctic sea ice variability with a coupled regional atmosphere-ocean-sea ice model. *Meteorologische Zeitschrift*, 14(6):793–800.

Mitchell, J. F., Johns, T., Gregory, J. M., and Tett, S. (1995). Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature*, 376(6540):501.

Mogensen, K., Balmaseda, M. A., Weaver, A., et al. (2012). The nemovar ocean data assimilation system as implemented in the ecmwf ocean analysis for system 4. Technical report, Technical Memorandum 668.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F. (2011). *The new ECMWF seasonal forecast system (System 4)*, volume 49. European Centre for Medium-Range Weather Forecasts Reading.

Monitoring, A. et al. (2011). *Snow, Water, Ice and Permafrost in the Arctic (SWIPA): Climate Change and the Cryosphere.* Arctic Monitoring and Assessment Programme (AMAP).

Mulholland, D. P., Laloyaux, P., Haines, K., and Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Monthly Weather Review*, 143(11):4631–4644.

Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768.

Nakamura, T., Yamazaki, K., Iwamoto, K., Honda, M., Miyoshi, Y., Ogawa, Y., Tomikawa, Y., and Ukita, J. (2016). The stratospheric pathway for arctic impacts on midlatitude climate. *Geophysical Research Letters*, 43(7):3494–3501.

Nakanowatari, T., Sato, K., and Inoue, J. (2014). Predictability of the barents sea ice in early winter: Remote effects of oceanic and atmospheric thermal conditions from the north atlantic. *Journal of Climate*, 27(23):8884–8901.

Notz, D. (2009). The future of ice sheets and sea ice: Between reversible retreat and unstoppable loss. *Proceedings of the National Academy of Sciences*, 106(49):20590–20595.

Notz, D. (2012). Challenges in simulating sea ice in earth system models. *Wiley Interdisciplinary Reviews: Climate Change*, 3(6):509–526.

Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H., and Marotzke, J. (2013). Arctic sea-ice evolution as modeled by max planck institute for meteorology's earth system model. *Journal of Advances in Modeling Earth Systems*, 5(2):173–194.

Notz, D. and Marotzke, J. (2012). Observations reveal external driver for arctic sea-ice retreat. *Geophysical Research Letters*, 39(8).

Notz, D. and Stroeve, J. (2016). Observed arctic sea-ice loss directly follows anthropogenic co2 emission. *Science*, 354(6313):747–750.

Notz, D. and Stroeve, J. (2018). The trajectory towards a seasonally ice-free arctic ocean. *Current Climate Change Reports*, 4(4):407–416.

Ogilvie, A. (1996). Sea-ice conditions off the coasts of iceland ad 1601–1850 with special reference to part of the maunder minimum period (1675–1715). *North European climate data in the latter part of the Maunder Minimum period AD*, pages 1675–1715.

Onarheim, I. H., Eldevik, T., Årthun, M., Ingvaldsen, R. B., and Smedsrud, L. H. (2015). Skillful prediction of barents sea ice cover. *Geophysical Research Letters*, 42(13):5364–5371.

Outten, S. and Esau, I. (2012). A link between arctic sea ice and recent cooling trends over eurasia. *Climatic Change*, 110(3-4):1069–1075.

Overland, J. E., Dethloff, K., Francis, J. A., Hall, R. J., Hanna, E., Kim, S.-J., Screen, J. A., Shepherd, T. G., and Vihma, T. (2016). Nonlinear response of mid-latitude weather to the changing arctic. *Nature Climate Change*, 6(11):992.

Overland, J. E., Spillane, M. C., and Soreide, N. N. (2004). Integrated analysis of physical and biological pan-arctic change. *Climatic Change*, 63(3):291–322.

Overpeck, J. T., Meehl, G. A., Bony, S., and Easterling, D. R. (2011). Climate data challenges in the 21st century. *science*, 331(6018):700–702.

Palmer, T. (2014). Climate forecasting: Build high-resolution global climate models. *Nature News*, 515(7527):338.

Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., and Leutbecher, M. (2005). Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, 33:163–193.

Parkinson, C. L. and Cavalieri, D. J. (2008). Arctic sea ice variability and trends, 1979–2006. *Journal of Geophysical Research: Oceans (1978–2012)*, 113(C7).

Peings, Y., Cattiaux, J., Vavrus, S. J., and Magnusdottir, G. (2018). Projected squeezing of the wintertime north-atlantic jet. *Environmental Research Letters*, 13(7):074016.

Penny, S., Bach, E., Bhargava, K., Chang, C.-C., Da, C., Sun, L., and Yoshida, T. (2019). Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, 11(6):1803–1829.

Penny, S. G. and Hamill, T. M. (2017). Coupled data assimilation for integrated earth system analysis and prediction. *Bulletin of the American Meteorological Society*, 97(7):ES169–ES172.

Perovich, D. K., Light, B., Eicken, H., Jones, K. F., Runciman, K., and Nghiem, S. V. (2007). Increasing solar heating of the arctic ocean and adjacent seas, 1979–2005: Attribution and role in the ice-albedo feedback. *Geophysical Research Letters*, 34(19).

Petoukhov, V. and Semenov, V. A. (2010). A link between reduced barents-kara sea ice and cold winter extremes over northern continents. *Journal of Geophysical Research: Atmospheres*, 115(D21).

Petty, A. A., Holland, M. M., Bailey, D. A., and Kurtz, N. T. (2018). Warm arctic, increased winter sea ice growth? *Geophysical Research Letters*, 45(23):12–922.

Pithan, F. and Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, 7(3):181.

Pohlmann, H., Botzet, M., Latif, M., Roesch, A., Wild, M., and Tschuck, P. (2004). Estimating the decadal predictability of a coupled aogcm. *Journal of Climate*, 17(22):4463–4472.

Pohlmann, H., Kröger, J., Greatbatch, R. J., and Müller, W. A. (2017). Initialization shock in decadal hindcasts due to errors in wind stress over the tropical pacific. *Climate Dynamics*, 49(7-8):2685–2693.

Pohlmann, H., Mueller, W. A., Kulkarni, K., Kameswarrao, M., Matei, D., Vamborg, F., Kadow, C., Illing, S., and Marotzke, J. (2013). Improved forecast skill in the tropics in the new miklip decadal climate predictions. *Geophysical Research Letters*, 40(21):5798–5802.

Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S., Fitzpatrick, J. J., Funder, S., Holland, M., et al. (2010). History of sea ice in the arctic. *Quaternary Science Reviews*, 29(15-16):1757–1778.

Polyakov, I. V., Proshutinsky, A. Y., and Johnson, M. A. (1999). Seasonal cycles in two regimes of arctic climate. *Journal of Geophysical Research: Oceans*, 104(C11):25761–25788.

Preller, R. H. and Posey, P. G. (1989). The polar ice prediction system-a sea ice forecasting system. Technical report, NAVAL OCEAN RESEARCH AND DEVELOPMENT ACTIVITY STENNIS SPACE CENTER MS.

Prodhomme, C., Batté, L., Massonnet, F., Davini, P., Bellprat, O., Guemas, V., and Doblas-Reyes, F. J. (2016). Benefits of increasing the model resolution for the seasonal forecast quality in ec-earth. *Journal of Climate*, 29(24):9141–9162.

Quinn, W. (1974). Outlook for el niño-like conditions in 1975. *NORPAX Highlights*, 2(6):2–3.

Rahmstorf, S. (1995). Climate drift in an ocean model coupled to a simple, perfectly matched atmosphere. *Climate Dynamics*, 11(8):447–458.

Rayner, N., Parker, D. E., Horton, E., Folland, C., Alexander, L., Rowell, D., Kent, E., and Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 108(D14).

Rigor, I. G., Wallace, J. M., and Colony, R. L. (2002). Response of sea ice to the arctic oscillation. *Journal of Climate*, 15(18):2648–2663.

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M., Gilbert, D., Xu, J., Pouliquen, S., Thresher, A., et al. (2016). Fifteen years of ocean observations with the global argo array. *Nature Climate Change*, 6(2):145.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., et al. (2010). The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015.

Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., and Terray, L. (2016). Drift dynamics in a coupled model initialized for decadal forecasts. *Climate Dynamics*, 46(5-6):1819–1840.

Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., Eade, R., Fereday, D., Folland, C., Gordon, M., et al. (2014). Skillful long-range prediction of european and north american winters. *Geophysical Research Letters*, 41(7):2514–2519.

Scaife, A. A. and Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, 1(1):1–8.

Schlichtholz, P. (2011). Influence of oceanic heat variability on sea ice anomalies in the nordic seas. *Geophysical Research Letters*, 38(5).

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R. (2011). Uncertainty in modeled arctic sea ice volume. *Journal of Geophysical Research: Oceans*, 116(C8).

Screen, J. and Deser, C. (2019). Pacific ocean variability influences the time of emergence of a seasonally ice-free arctic ocean. *Geophysical Research Letters*, 46(4):2222–2231.

Screen, J. A. (2013). Influence of arctic sea ice on european summer precipitation. *Environmental Research Letters*, 8(4):044015.

Screen, J. A., Deser, C., Smith, D. M., Zhang, X., Blackport, R., Kushner, P. J., Oudar, T., McCusker, K. E., and Sun, L. (2018). Consistency and discrepancy in the atmospheric response to arctic sea-ice loss across climate models. *Nature Geoscience*, 11(3):155.

Serreze, M., Barrett, A., Stroeve, J., Kindig, D., and Holland, M. (2009). The emergence of surface-based arctic amplification. *The Cryosphere*, 3(1):11–19.

Serreze, M. C. and Barry, R. G. (2011). Processes and impacts of arctic amplification: A research synthesis. *Global and Planetary Change*, 77(1):85–96.

Serreze, M. C. and Francis, J. A. (2006). The arctic amplification debate. *Climatic change*, 76(3-4):241–264.

Sévellec, F., Fedorov, A. V., and Liu, W. (2017). Arctic sea-ice decline weakens the atlantic meridional overturning circulation. *Nature Climate Change*, 7(8):604.

Shaffrey, L. C., Stevens, I., Norton, W., Roberts, M., Vidale, P. L., Harle, J., Jrrar, A., Stevens, D., Woodage, M. J., Demory, M.-E., et al. (2009). Uk higem: The new uk high-resolution global environment model—model description and basic evaluation. *Journal of Climate*, 22(8):1861–1896.

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Gößling, H. F., et al. (2015). Towards multi-resolution global climate modeling with echam6–fesom. part i: model formulation and mean climate. *Climate Dynamics*, 44(3-4):757–780.

Siegert, S., Bellprat, O., Ménégoz, M., Stephenson, D. B., and Doblas-Reyes, F. J. (2017). Detecting improvements in forecast correlation skill: Statistical testing and power analysis. *Monthly Weather Review*, 145(2):437–450.

Sigmond, M., Fyfe, J., Flato, G., Kharin, V., and Merryfield, W. (2013). Seasonal forecast skill of arctic sea ice area in a dynamical forecast system. *Geophysical Research Letters*, 40(3):529–534.

Sigmond, M., Fyfe, J. C., and Swart, N. C. (2018). Ice-free arctic projections under the paris agreement. *Nature Climate Change*, 8(5):404.

Slingo, J. and Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4751–4767.

Smith, B. (2001). Lpj-guess-an ecosystem modelling framework. *Department of Physical Geography and Ecosystems Analysis. INES, Sölvegatan*, 12:22362.

Smith, D. M., Eade, R., and Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate dynamics*, 41(11-12):3325–3338.

Smith, L. C. and Stephenson, S. R. (2013). New trans-arctic shipping routes navigable by midcentury. *Proceedings of the National Academy of Sciences*, 110(13):E1191–E1195.

Sokolov, A. (1962). Drift of ice in the arctic basin and changes in ice conditions over the northern sea route (english transl.). *Probl. Arct. Antarct*, 11:j1–j20.

Sorteberg, A. and Kvingedal, B. (2006). Atmospheric forcing on the barents sea winter ice extent. *Journal of Climate*, 19(19):4772–4784.

Stammerjohn, S., Massom, R., Rind, D., and Martinson, D. (2012). Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters*, 39(6).

Stark, J. D., Ridley, J., Martin, M., and Hines, A. (2008). Sea ice concentration and motion assimilation in a sea ice- ocean model. *Journal of Geophysical Research: Oceans*, 113(C5).

Stephenson, S. R., Smith, L. C., and Agnew, J. A. (2011). Divergent long-term trajectories of human access to the arctic. *Nature Climate Change*, 1(3):156.

Storto, A., Alvera-Azcárate, A., Balmaseda, M. A., Barth, A., Chevallier, M., Counillon, F., Domingues, C. M., Drevillon, M., Drillet, Y., Forget, G., et al. (2019). Ocean reanalyses: Recent advances and unsolved challenges. *Frontiers in Marine Science*, 6:418.

Stroeve, J., Hamilton, L. C., Bitz, C. M., and Blanchard-Wrigglesworth, E. (2014). Predicting september sea ice: Ensemble skill of the search sea ice outlook 2008–2013. *Geophysical Research Letters*, 41(7):2411–2418.

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M. (2007). Arctic sea ice decline: Faster than forecast. *Geophysical research letters*, 34(9).

Stroeve, J. and Notz, D. (2018). Changing state of arctic sea ice across all seasons. *Environmental Research Letters*, 13(10):103001.

Stroeve, J. C., Schroder, D., Tsamados, M., and Feltham, D. (2018). Warm winter, thin ice? *The Cryosphere*, 12(5):1791–1809.

Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Malanik, J., and Barrett, A. P. (2012). The arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, 110(3-4):1005–1027.

Stuecker, M. F., Bitz, C. M., Armour, K. C., Proistosescu, C., Kang, S. M., Xie, S.-P., Kim, D., McGregor, S., Zhang, W., Zhao, S., et al. (2018). Polar amplification dominated by local forcing and feedbacks. *Nature Climate Change*, 8(12):1076.

Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama, T., Igarashi, H., and Ishikawa, Y. (2008). Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. *Journal of Geophysical Research: Oceans*, 113(C10).

Tietsche, S., Day, J., Guemas, V., Hurlin, W., Keeley, S., Matei, D., Msadek, R., Collins, M., and Hawkins, E. (2014). Seasonal to interannual arctic sea ice predictability in current global climate models. *Geophysical Research Letters*, 41(3):1035–1043.

Tietsche, S., Hawkins, E., and Day, J. J. (2016). Atmospheric and oceanic contributions to irreducible forecast uncertainty of arctic surface climate. *Journal of Climate*, 29(1):331–346.

Tietsche, S., Notz, D., Jungclaus, J., and Marotzke, J. (2013). Assimilation of sea-ice concentration in a global climate model–physical and statistical aspects. *Ocean science*, 9:19–36.

Tintó-Prims, O., Castrillo, M., Acosta, M. C., Mula-Valls, O., Lorente, A. S., Serradell, K., Cortés, A., and Doblas-Reyes, F. J. (2018). Finding, analysing and solving mpi communication bottlenecks in earth system models. *Journal of Computational Science*.

Tsamados, M., Feltham, D., Petty, A., Schroeder, D., and Flocco, D. (2015). Processes controlling surface, bottom and lateral melt of arctic sea ice in a state of the art sea ice model. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2052):20140167.

Underhill, V., Fetterer, F., and Petersen, C. (2014). Arctic sea ice concentration and extent from Danish Meteorological Institute Sea Ice Charts, 1901-1956, version 1. *NSIDC National Snow and Ice Data Center*.

Uotila, P., Goosse, H., Haines, K., Chevallier, M., Barthélemy, A., Bricaud, C., Carton, J., Fučkar, N., Garric, G., Iovino, D., et al. (2019). An assessment of ten ocean reanalyses in the polar regions. *Climate Dynamics*, 52(3-4):1613–1650.

Valcke, S. (2006). Oasis3 user guide (prism_2-5). *PRISM support initiative report*, 3:64.

Vancoppenolle, M., Bouillon, S., Fichefet, T., Goosse, H., Lecomte, O., Morales Maqueda, M., and Madec, G. (2012). The louvain-la-neuve sea ice model. *Notes du pole de modélisation, Institut Pierre-Simon Laplace (IPSL), Paris, France*, ISSN No 1288-1619(31).

Vancoppenolle, M., Fichefet, T., and Bitz, C. M. (2005). On the sensitivity of undeformed arctic sea ice to its vertical salinity profile. *Geophysical Research Letters*, 32(16).

Vancoppenolle, M., Fichefet, T., and Bitz, C. M. (2006). Modeling the salinity profile of undeformed arctic sea ice. *Geophysical Research Letters*, 33(21).

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., and Maqueda, M. A. M. (2009). Simulating the mass balance and salinity of arctic and antarctic sea ice. 1. model description and validation. *Ocean Modelling*, 27(1):33–53.

Vancoppenolle, M., Goosse, H., De Montety, A., Fichefet, T., Tremblay, B., and Tison, J.-L. (2010). Modeling brine and nutrient dynamics in antarctic sea ice: The case of dissolved silica. *Journal of Geophysical Research: Oceans*, 115(C2).

Vaughan, D. G., Comiso, J. C., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., et al. (2013). Observations: cryosphere. *Climate change*, 2103:317–382.

Vihma, T. (2014). Effects of arctic sea ice decline on weather and climate: a review. *Surveys in Geophysics*, 35(5):1175–1214.

Vinje, T. (2001). Fram strait ice fluxes and atmospheric circulation: 1950-2000. *Journal of Climate*, 14(16):3508–3517.

Volpi, D., Guemas, V., and Doblas-Reyes, F. J. (2016). Comparison of full field and anomaly initialisation for decadal climate prediction: towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. *Climate Dynamics*, 49(4):1181–1195.

Von Storch, H. and Zwiers, F. W. (2001). *Statistical analysis in climate research*. Cambridge university press.

Walsh, J. (1978). A data set on northern hemisphere sea ice extent. *Glaciological Data, Report GD*, 2:49–51.

Walsh, J. E. (1980). Empirical orthogonal functions and the statistical predictability of sea ice extent. *Sea ice processes and models. University of Washington Press, Seattle*, pages 373–384.

Walsh, J. E., Fetterer, F., Scott Stewart, J., and Chapman, W. L. (2017). A database for depicting arctic sea ice variations back to 1850. *Geographical Review*, 107(1):89–107.

Walsh, J. E., Hibler III, W. D., and Ross, B. (1985). Numerical simulation of northern hemisphere sea ice variability, 1951–1980. *Journal of Geophysical Research: Oceans*, 90(C3):4847–4865.

Wang, W., Chen, M., and Kumar, A. (2013). Seasonal prediction of arctic sea ice extent from a coupled dynamical forecast system. *Monthly Weather Review*, 141(4):1375–1394.

Washington, W. M. and Meehl, G. A. (1996). High-latitude climate change in a global coupled ocean-atmosphere-sea ice model with increased atmospheric co2. *Journal of Geophysical Research: Atmospheres*, 101(D8):12795–12801.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., et al. (2010). Improved climate simulation by miroc5: Mean states, variability, and climate sensitivity. *Journal of Climate*, 23(23):6312–6335.

Weigel, A. P., Liniger, M., and Appenzeller, C. (2008). Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 134(630):241–260.

Woodgate, R. A., Weingartner, T., and Lindsay, R. (2010). The 2007 bering strait oceanic heat flux and anomalous arctic sea-ice retreat. *Geophysical Research Letters*, 37(1).

Wouters, B., Hazeleger, W., Drijfhout, S., Oldenborgh, G., and Guemas, V. (2013). Multiyear predictability of the north atlantic subpolar gyre. *Geophysical Research Letters*, 40(12):3080–3084.

Wu, R., Kirtman, B. P., and van den Dool, H. (2009). An analysis of enso prediction skill in the cfs retrospective forecasts. *Journal of Climate*, 22(7):1801–1818.

Wyrtki, K., Stroup, E., Patzert, W., Williams, R., and Quinn, W. (1976). Predicting and observing el niño. *Science*, 191(4225):343–346.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature climate change*, 3(10):875.

Yang, S. and Christensen, J. H. (2012). Arctic sea ice reduction and european cold winters in cmip5 climate change experiments. *Geophysical Research Letters*, 39(20).

Yeager, S., Danabasoglu, G., Rosenbloom, N., Strand, W., Bates, S., Meehl, G., Karspeck, A., Lindsay, K., Long, M., Teng, H., et al. (2018). Predicting near-term changes in the earth system: A large ensemble of initialized decadal prediction simulations using

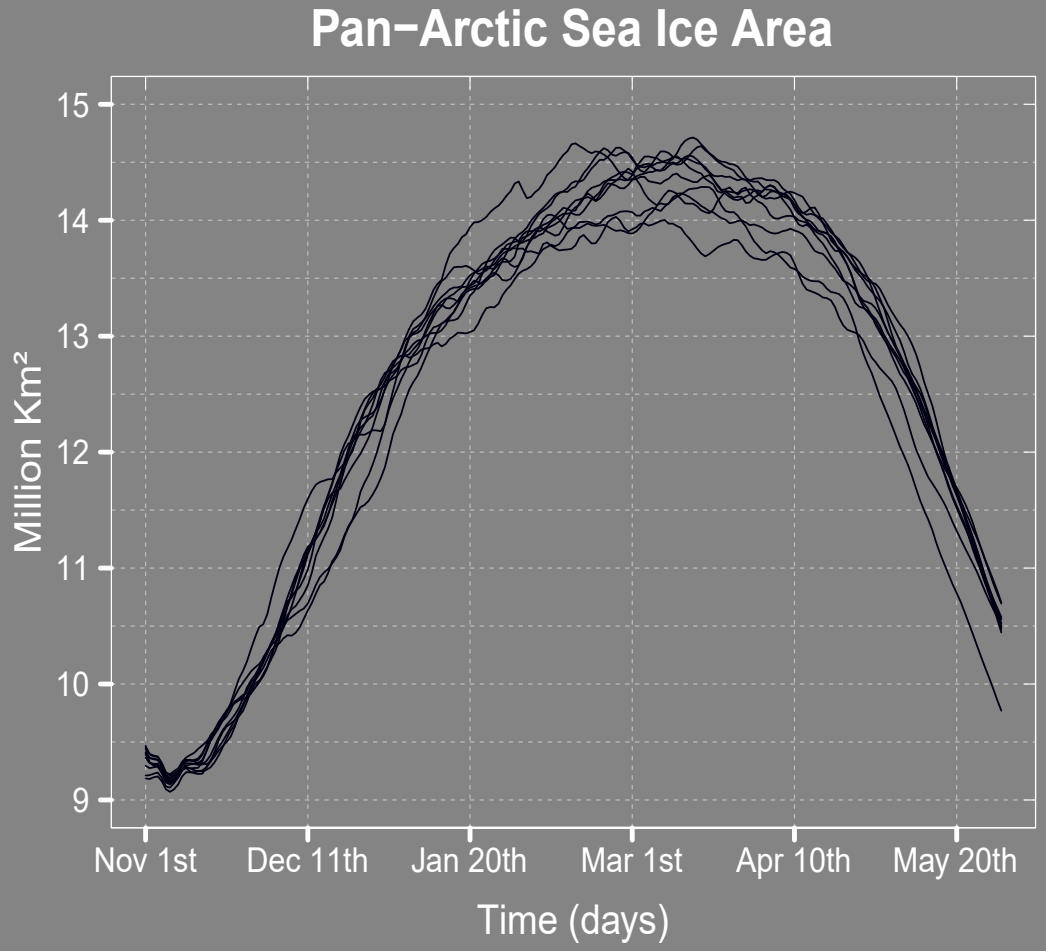
the community earth system model. *Bulletin of the American Meteorological Society*, 99(9):1867–1886.

Zappa, G. and Shepherd, T. G. (2017). Storylines of atmospheric circulation change for european regional climate impact assessment. *Journal of Climate*, 30(16):6561–6577.

Zhang, J., Steele, M., Lindsay, R., Schweiger, A., and Morison, J. (2008). Ensemble 1-year predictions of arctic sea ice for the spring and summer of 2008. *Geophysical Research Letters*, 35(8).

Zhang, R. (2015). Mechanisms for low-frequency variability of summer arctic sea ice extent. *Proceedings of the National Academy of Sciences*, 112(15):4570–4575.

Zuo, H., Balmaseda, M., Mogensen, K., and Tietsche, S. (2018). *OCEAN5: the ECMWF Ocean Reanalysis System and its Real-Time analysis component*. European Centre for Medium-Range Weather Forecasts.



BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación