

Arctic weather  
and climate:  
from  
mechanisms  
to forecasts

Yang Liu

1. Independent observations are indispensable for a better assessment of the scientific outcome obtained from reanalysis products.  
(this thesis)
2. Deep learning is an imperfect solution for nonlinearity representation in models of physical systems.  
(this thesis)
3. Making software compliant with research software development standards is more important for scientific progress than many innovative ideas.
4. Instead of being charged by publishers, open science deserves to be rewarded.
5. Artificial intelligence is revolutionary because humans cannot understand non-linear relations well.
6. Emphasizing the intricacy of concepts in science using elegant practical examples in daily life enhances public support for science.

Propositions belonging to the thesis, entitled  
Arctic weather and climate: from mechanisms to forecasts  
Yang Liu  
Wageningen, 12 October 2021



# Arctic weather and climate: from mechanisms to forecasts

Yang Liu

## **Thesis committee**

### **Promotor:**

Prof. Dr W. Hazeleger  
Special Professor, Climate Dynamics  
Wageningen University & Research  
Professor of Climate System Science  
Utrecht University

### **Co-promotor:**

Dr J. J. Attema  
Programme Manager  
Netherlands eScience Center, Amsterdam

### **Other members:**

Dr A. J. Teuling, Wageningen University & Research  
Prof. Dr R. Bintanja, University of Groningen  
Prof. Dr P. Grünwald, University of Leiden  
Prof. Dr J. Stroeve, University College London, United Kingdom

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE)

# Arctic weather and climate: from mechanisms to forecasts

Yang Liu

## **Thesis**

submitted in fulfilment of the requirements for the degree of doctor  
at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 12 October 2021

at 1:30 p.m. in the Aula.

Yang Liu

Arctic weather and climate: from mechanisms to forecasts

182 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)

With references, with summary in English

ISBN: 978-94-6395-775-5

DOI: <https://doi.org/10.18174/545045>



# Summary

Since the mid-20th Century, the Arctic is warming about two times faster than the rest of the world. This is widely known as Arctic Amplification (AA) and has drawn lots of attention recently. Apart from a dominant impact on the local weather and climate, its influence has been extended beyond the polar region through significant modification on the circulation in both the atmosphere and ocean. In order to identify the drivers of Arctic climate change and the connections between the Arctic and global climate system, it is important to understand the Arctic weather and climate.

The Arctic weather and climate system is manifested by the cryosphere. Given the essential role of sea ice and the key factors that regulate the variability of sea ice, it is of vital importance to shed new light on the energy budget, as well as the atmosphere-ocean interaction in the perspective of the energy budget in the Northern Hemisphere. A better quantification of the Arctic sea ice variability and the Arctic energy budget will give us a better understanding of the mechanisms behind Arctic climate warming. This in turn would help improve the weather forecasts for the Northern Hemisphere.

This thesis, I mainly focus on the variations and forecasting of the Arctic sea ice. Given the close relation between the Arctic energy budget and sea ice variability, an emphasis is placed on the meridional energy transport (MET). In chapter 2, I quantified meridional energy transport in the atmosphere (AMET) and ocean (OMET) at subpolar latitudes using the latest methods and six reanalysis data sets. An intercomparison of the results from the chosen reanalysis products indicates that although the mean transport in all data sets agrees well, the spatial distributions and temporal variations of AMET and OMET differ substantially among the reanalysis data sets. I further investigated the coherence between MET and the Arctic climate variability at interannual timescales. The regressions of climatological fields in the Arctic on both AMET and OMET suggest that the Arctic climate is sensitive to changes of meridional energy transport at subpolar latitudes in winter. Our study confirms that the analyzed reanalysis products are useful for the diagnostics of energy transport. However, beyond interannual timescales, the results must be interpreted with caution, especially when studying variability and interactions between the Arctic and midlatitudes.

The analysis of energy transport is extended to the whole energy budget and the interactions between atmosphere and ocean in the Northern Hemisphere in chapter 3. Based on our findings in chapter 2, I examined the compensation between heat transport variations in the atmosphere and ocean. Different from studies in the past, which were mostly based on numerical climate

models, we provided new insights into the so-called Bjerknes compensation using multiple re-analysis products. Our study shows that Bjerknes compensation is present at almost all latitudes from  $40^{\circ}$  to  $70^{\circ}\text{N}$  in the Northern Hemisphere from interannual to decadal time scales. We found that the response of the mean flow to OMET variability leads to the Bjerknes compensation. It is the shift of the Ferrel cell at midlatitudes at decadal time scales in winter, which is driven by the eddy momentum flux that causes the compensation. Our findings are different from some experiments with numerical climate models, which attribute the compensation to the variation of transient eddy transports in response to the changes of OMET at multidecadal time scales. Limited by the short historical records, the cause–effect relation between the atmosphere and ocean cannot be resolved.

On the basis of physical insights into the energy transport and sea ice variability shown in chapter 2 and 3, the study continues in the direction of sea ice forecasts and a series of experiments of extended range sea ice forecasts with novel deep neural networks are presented in chapter 4. Inspired by the rapid developments in deep learning techniques, in this chapter I proposed Convolutional Long Short Term Memory Networks (ConvLSTM) to forecast sea ice in the Barents Sea at weather to sub-seasonal time scales. The architecture of this neural network is designed to exploit the covariances between different variables, including spatial and temporal relations, which is suitable for the prediction of spatial-temporal sequential data. Using reanalysis products, we demonstrated that ConvLSTM is able to learn the variability of the Arctic sea ice and can forecast regional sea ice concentration skillfully at weekly to monthly time scales. In general, forecasts from ConvLSTM outperform those with climatology, persistence, and a statistical model, and they are comparable to the forecasts from operational sub-seasonal to seasonal weather forecast systems. Furthermore, we demonstrated that the ConvLSTMs are able to preserve the physical consistency between predictors and predictands in their forecasts. A sensitivity test aiming at an evaluation of the impact of different predictors on the quality of forecasts shows that the surface energy budget components have a significant impact on the predictability of sea ice at weather time scales. Our findings indicate that such data-driven methods with deep neural networks are promising tools for enhancing operational Arctic sea ice forecasting.

Our exploration is further extended to probabilistic deep learning in chapter 5. Motivated by the need for uncertainty quantification in weather forecasts and the limit on uncertainty estimation reflected by previous studies using deterministic neural networks, we explored probabilistic deep neural networks for weather forecasting. By replacing fixed weights with distributions following the methodology in Bayesian deep learning (BDL), we turned normal Long-Short Term Memory neural networks (LSTM) into Bayesian Long-Short Term Memory neural networks (BayesLSTMs). With an aim to understand the characteristics of BDL within a simplified dynamical system that represents the essence of midlatitude atmospheric dynamics, we used BayesLSTMs to forecast output from the Lorenz 84 system with seasonal forcing. We showed that forecasts with the BayesLSTM can stay close to the attractor of the Lorenz model and concluded that they represent the nonlinear relations between each components in this simplified atmospheric circulation system. We further demonstrated that the BayesLSTMs are able to produce reliable probabilistic forecasts and address uncertainties relevant to weather forecasting. Our study indicates that BDL is an easy and fast solution for probabilistic weather

forecast and is promising to enhance weather forecasting capabilities at short to medium-range timescales.

Finally, a synthesis of this thesis and an outlook for suggested future research are shown in chapter 6.





# Contents

	Page
<b>Summary</b>	<b>vii</b>
<b>Contents</b>	<b>xi</b>
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 Understanding the Arctic climate system . . . . .	2
1.1.1 A picture of the Arctic climate system . . . . .	3
1.1.2 Meridional energy transport and the Arctic climate variability . . . . .	6
1.2 Improving weather forecast in the Arctic with deep learning . . . . .	11
1.2.1 Predictability of Arctic weather and climate . . . . .	11
1.2.2 Weather forecasting with numerical weather prediction system . . . . .	13
1.2.3 Deep neural networks for weather forecasting . . . . .	14
1.3 Thesis Overview . . . . .	19
<b>Chapter 2 Synthesis and evaluation of historical meridional heat transport from midlatitudes towards the Arctic</b>	<b>23</b>
2.1 Introduction . . . . .	25
2.2 Data and Methodology . . . . .	27
2.2.1 Reanalyses . . . . .	27
2.2.2 Oceanic Observations and OGCM Hindcast . . . . .	29
2.2.3 Computation of Meridional Energy Transport . . . . .	30
2.2.4 Statistical Analysis . . . . .	33
2.3 Results . . . . .	34
2.3.1 Overview of AMET and OMET . . . . .	34
2.3.2 Sources of Disparity . . . . .	38
2.3.3 MET and the Arctic . . . . .	43
2.4 Discussion . . . . .	45
2.5 Conclusions . . . . .	47
<b>Chapter 3 Atmosphere-ocean interactions and their footprint on heat transport variability in the Northern Hemisphere</b>	<b>49</b>
3.1 Introduction . . . . .	51
3.2 Data and Methodology . . . . .	52

3.2.1	Reanalysis Datasets . . . . .	53
3.2.2	Methodology . . . . .	53
3.3	Results . . . . .	55
3.3.1	Meridional Energy Transport . . . . .	55
3.3.2	Bjerknes Compensation and Atmospheric Response to Ocean Variability .	59
3.3.3	Drivers for the OMET variations . . . . .	68
3.4	Conclusion and Discussion . . . . .	70
<b>Chapter 4 Extended Range Arctic Sea Ice Forecast with Convolutional Long-Short Term Memory Networks</b>		<b>75</b>
4.1	Introduction . . . . .	77
4.2	Data and Methodology . . . . .	79
4.2.1	Convolutional Long-Short Term Memory Networks . . . . .	79
4.2.2	Reanalysis Data Sets . . . . .	81
4.2.3	Evaluation Metrics . . . . .	82
4.2.4	Training and Hyperparameter Tuning . . . . .	85
4.2.5	Baseline Statistical Model Analysis . . . . .	86
4.3	Results . . . . .	87
4.3.1	Constrained Predictability . . . . .	87
4.3.2	Sensitivity Analysis of Predictors . . . . .	91
4.3.3	Lead-Time Dependent Operational Forecasts . . . . .	94
4.3.4	Physical Consistency of ConvLSTM Forecasts . . . . .	100
4.4	Discussion . . . . .	103
4.5	Conclusion . . . . .	104
<b>Chapter 5 Exploring Bayesian deep learning for weather forecasting with the Lorenz 84 system</b>		<b>107</b>
5.1	Introduction . . . . .	109
5.2	Methodology . . . . .	110
5.2.1	Lorenz 84 Model with Seasonal Forcing . . . . .	110
5.2.2	BayesLSTM and Bayes by Backprop . . . . .	111
5.2.3	Ensemble Forecasting with BDL and Numerical Configurations . . . . .	113
5.2.4	Vector Autoregressive Model . . . . .	114
5.3	Results . . . . .	114
5.3.1	Representing the Evolution of Lorenz 84 Model . . . . .	115
5.3.2	Evaluate the BayesLSTM ensemble forecasts . . . . .	117
5.3.3	Probabilistic Sea Ice Forecasting with BDL . . . . .	118
5.4	Discussion . . . . .	121
5.5	Conclusion . . . . .	122
<b>Chapter 6 Synthesis and Outlook</b>		<b>129</b>
6.1	Physical insights into Arctic climate . . . . .	130
6.2	Improving weather forecast in the Arctic . . . . .	133
6.3	Outlook . . . . .	135

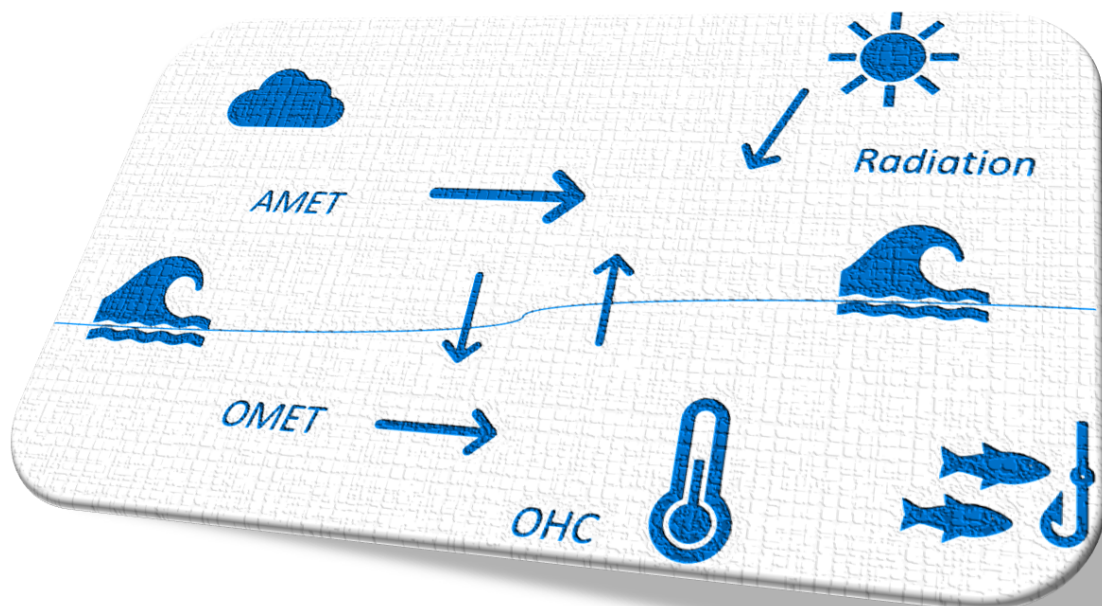
<b>CONTENTS</b>	<b>xiii</b>
<b>References</b>	<b>139</b>
<b>Acknowledgements</b>	<b>159</b>
<b>About the author</b>	<b>163</b>





## Chapter 1

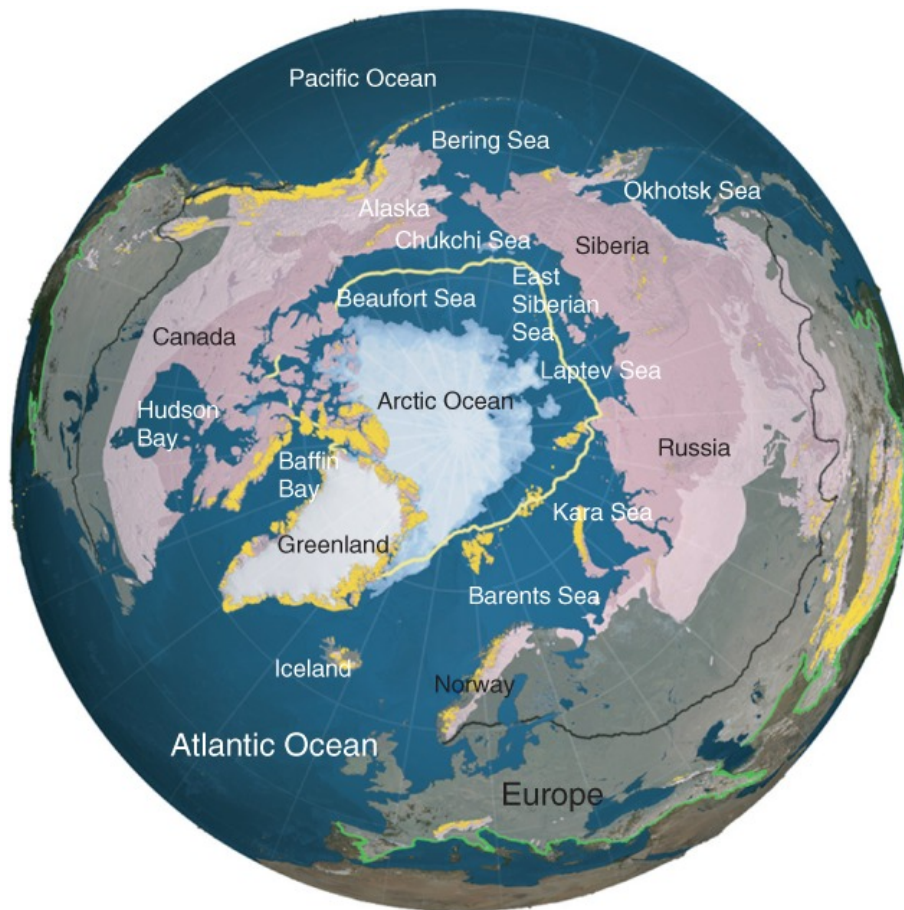
# Introduction



## 1.1 Understanding the Arctic climate system

Although there is hardly any footprint of humans in the Arctic area, there are footprints of human civilization everywhere in the Arctic weather and climate system. The urgent necessity of identifying the drivers for climate change and understanding the Arctic climate system motivates this study, which aims to provide more physical insights into the Arctic climate and improve weather forecasts in the Arctic.

Covered by snow, sea ice, glaciers and permafrost, the Arctic is the “frozen heart” of the planet. Geographically, it is located at the north end of the world and stays relatively far away from the centres of human civilization. Characterized by the midnight sun and polar night, the Arctic area is conceptually defined as north of the polar circle (see Figure 1.1). It is surrounded by the Eurasian and American continents and it covers the Arctic Ocean. The special location and complex geophysical composition of the Arctic, including land, sea and ice, lead to a unique climate system in this area.



**Figure 1.1:** Map of the Arctic region including average sea ice extent (yellow line), sea ice cover during record minimum in summer of 2012 (shades of white), continuous and discontinuous permafrost (shades of pink), glacier locations (gold dots), and snow-cover (average location of 50% snow line in black and maximum snow line in green as inferred from moderate-resolution imaging spectroradiometer data). Excerpt from [Comiso and Hall, 2014].

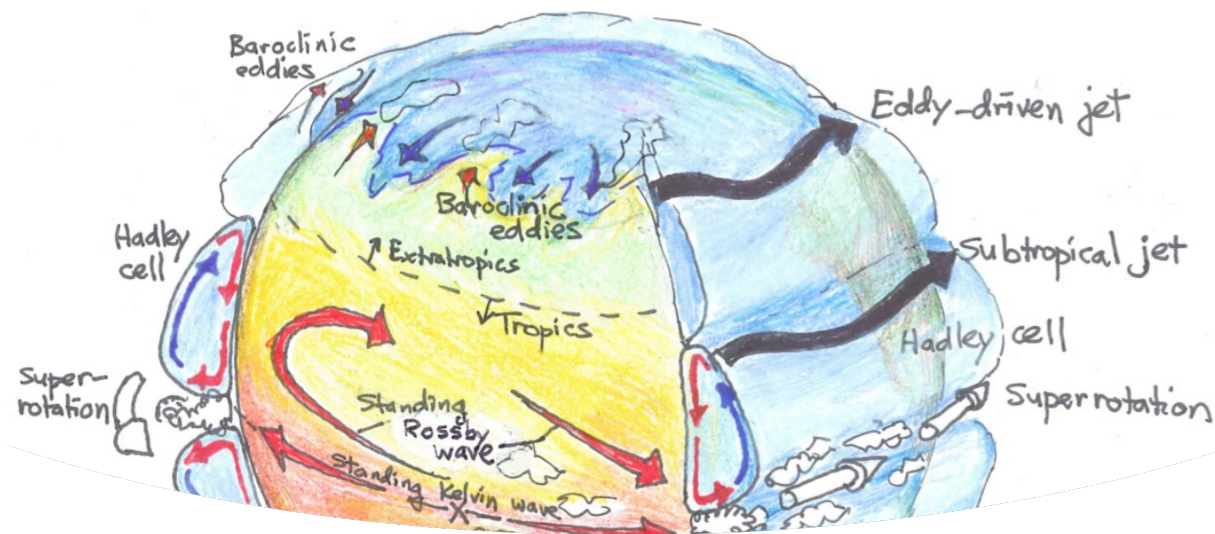
In the Arctic, relatively long cold winters and short warm summers are the most noticeable features of the climate [Serreze and Barry, 2014]. While there is substantial variability of temperature with latitude and location, for most of the regions the mean temperature is below freezing point in winter and above freezing point in summer. Given the low temperature in this area, the precipitation always falls as snow in winter and partially as rain in summer. Overall, the Arctic has an extremely dry atmosphere since the descending cold air does not carry much moisture [Serreze and Barry, 2014].

### 1.1.1 A picture of the Arctic climate system

Weather and climate on our planet is driven by the uneven distribution of solar radiation. The Arctic acts as a heat sink in the global climate system due to the net radiation loss to space from the top of the atmosphere (TOA). The polar area “imports” heat from lower latitudes and this process is manifested by the general circulation of the atmosphere and ocean. Although the atmospheric circulation and the ocean circulation are varying all the time, their large-scale structures remain fairly constant. Latitudinally, the atmospheric circulation in the polar cap is featured by the Polar cell. It is recognized as a weak meridional circulation characterized by ascending motion of air in the subpolar latitudes and descending motion of air over the North Pole, in the troposphere [Holton, 1973]. It is weaker than the tropical/subtropical Hadley cell and is barely detectable due to the presence of strong midlatitude cyclones and anticyclones in this region. At lower latitudes, the energy redistribution is mainly carried out by the zonal mean flow, thus the Hadley cell. The transport of thermal energy peaks at midlatitudes (around 40°N), which is associated with the occurrence of the largest temperature gradient [Liu et al., 2020b]. This strong temperature gradient creates environment for baroclinic instability and this allows the growth of mid-latitude weather systems. As a result, from midlatitudes towards the Arctic, transient eddies play a crucial role in transporting heat, momentum and moisture. A schematic illustration of this dynamical process is shown in Figure 1.2. The contribution from the mean flow is relatively small in terms of heat transport in midlatitudes.

Horizontally, the unique characteristics of atmospheric circulation in the Arctic can be represented by typical polar weather and climate patterns. One of the most persistent patterns is the polar vortex [Jaiser et al., 2013]. It is a low pressure area that occurs at the upper-level of the atmosphere in the Arctic. Its strength is determined by the movement of mass and the transfer of heat in the polar region. Through modifying other elements and phenomena in the polar area, like sudden stratospheric warmings (SSW) and the jet stream [Jaiser et al., 2013], the variability of the Arctic polar vortex has a strong impact on the dynamics of the atmospheric circulation. Changes in the atmospheric circulation can cause anomalous heat transport towards the North Pole and therefore influence the energy budget in the polar area.

Another large-scale climate pattern that dominates the whole polar cap is the Arctic Oscillation (AO). The AO was first proposed by Thompson and Wallace [1998] to describe the alternated opposing patterns of pressure fields in the Arctic and midlatitudes. When the atmospheric pressure is low in the Arctic and high at midlatitudes, the AO is in the positive phase and this brings more storms, carrying warm and humid air, to the north. As a result, it will cause an increase in the atmospheric energy transport from midlatitudes towards the Arctic. In the negative phase,



**Figure 1.2:** Schematic illustration of atmospheric circulation in the Northern Hemisphere. Excerpt from [Showman et al., 2013].

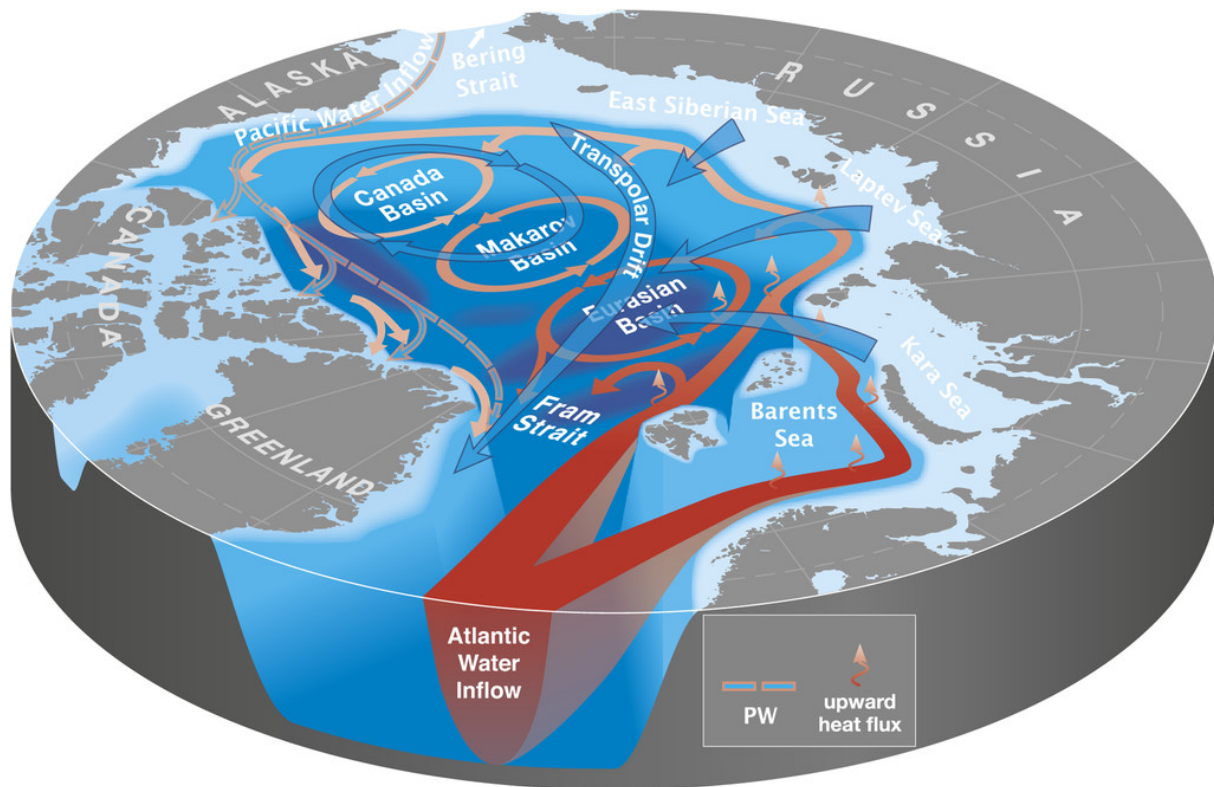
the pattern is reversed. Its phase is linked to the location and intensity of the storm tracks and the midlatitude jet stream. Given the close relation between the AO and the atmospheric circulation in the Arctic, this pattern has strong impact on the variations of heat transport towards the polar cap in the atmosphere. Moreover, by modifying the sea level pressure (SLP) field, the AO pattern also affects the surface wind field and its influence is therefore extended to the thermohaline circulation and heat transport in the ocean [Jungclauss and Koenigk, 2010]. Previous studies have shown that the variability of AO is corresponding to the strength of the Atlantic Meridional Overturning Circulation (AMOC) [Gastineau et al., 2013]. Obviously, this pattern reflects the climate variability in the Northern Hemisphere [Delworth and Zeng, 2016].

Apart from these large-scale patterns, there are also locally constrained patterns, like polar lows, which are intense mesoscale cyclones developed at high latitudes [Sergeev et al., 2018]. They are formed when cold Arctic air flows over relatively warm open water during the cold season. The polar lows are “fuelled” by sensible and latent heat fluxes and therefore they are also coupled to the surface energy budget in the Arctic.

In addition to the atmospheric circulation, the ocean also plays a crucial role in shaping the Arctic climate system. The ocean mainly exerts its impact on the variability of the Arctic climate through the heat transport and its circulation. These processes are associated with specific “characteristics” of the system, like surface buoyancy fluxes, stratification of the ocean, meridional energy transport, ocean heat storage, and air-sea-ice interactions [Aagaard and Carmack, 1994]. A schematic illustration of the ocean circulation and heat transport in the Arctic is shown in Figure 1.3. The Arctic Ocean mainly receives inflows from the Atlantic and Pacific oceans. River runoff from North America and Siberia also contributes to it. The stratification of the Arctic Ocean is predominantly set by salinity, which is mainly controlled by freezing and melting of sea ice and wind stress at the surface [Timmermans and Marshall, 2020].

Regarding the energy transport in the ocean, most of the energy that reaches the Arctic is





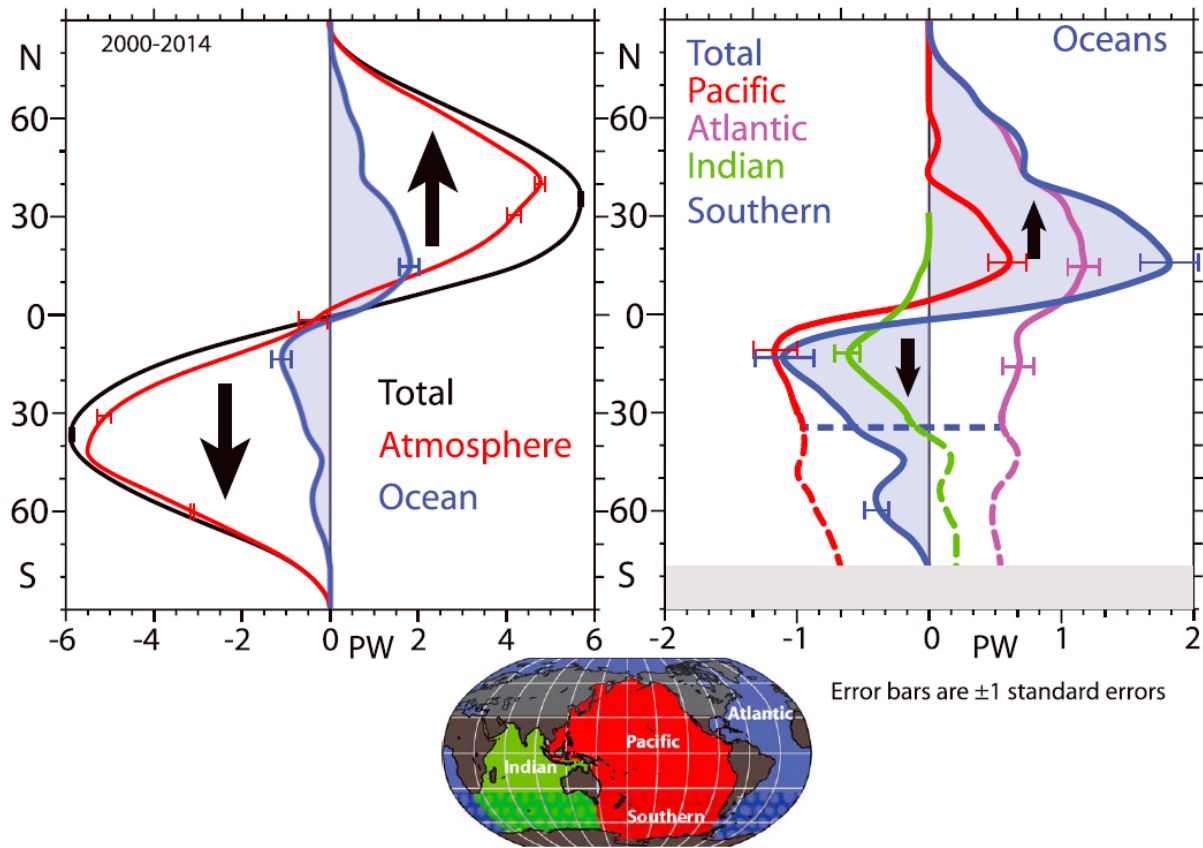
**Figure 1.3:** Schematic illustration of ocean circulation and heat transport in the North Pole. Excerpt from [Carmack et al., 2015].

transported through the North Atlantic current via the Fram Strait and the Barents Sea opening. It originates from the heat transported by the AMOC, which is the zonally-integrated component of surface and deep currents in the Atlantic Ocean that is responsible for most of the meridional energy transport by the ocean at midlatitudes in the Northern Hemisphere [Trenberth and Fasullo, 2008]. Consequently, the energy transport in the ocean towards the Arctic is strongly coupled to the strength of AMOC. Observational and modelling evidence suggest that poleward meridional energy transport in the ocean considerably contributes to the warming of the Arctic. Given the essential role of the ocean in the Arctic climate system, the Arctic ocean has become a focal point of climate change, with ocean warming, freshening, and sea ice decline [Timmermans and Marshall, 2020].

The phenomena mentioned above are all crucial in shaping the energy balance of the Arctic. And the energy balance is fundamental to understand and predict the Arctic climate. Featured by the interplay of heat budget and general circulation, the variability of the Arctic weather and climate is closely related to the variations of meridional energy transport. This underlines the importance of quantifying energy transport in the atmosphere and ocean towards the Arctic and identifying their impact on the general circulation and the total energy budget.

### 1.1.2 Meridional energy transport and the Arctic climate variability

The sources and sinks of net radiation in the tropics and the Arctic, respectively, lead to the poleward energy transport, which is also called the meridional energy transport (MET). The balance of the global energy budget is maintained by the meridional energy transport in the atmosphere (AMET) and ocean (OMET), and the energy flow in-between. Figure 1.4 provides a typical view of annual and zonal mean of AMET and OMET as a function of latitude. Anomalous energy transport can cause large variations in the weather and climate system, like the expansion or reduction of the overturning circulation, changes of eddy behaviours, and the shift of large-scale climate patterns [e.g. D'Andrea et al., 2005, Trenberth and Stepaniak, 2003].



**Figure 1.4:** The annual and zonal means of the meridional energy transports for 2000–2014 in PW for (left) the total Earth system (black), the atmosphere (red) and the ocean (blue). (right) The ocean component broken down into the contributions from the Atlantic (violet), Pacific (red), and Indian (green) Oceans which combine south of 35°S to give the southern ocean value, as given in the small map below. Excerpt from [Trenberth and Fasullo, 2017].

Given the unique features of the cryosphere and the polar weather and climate system, the Arctic is extremely susceptible to the variability of MET [Serreze et al., 2007]. Consequently, to understand the variability of Arctic climate, we can examine the energy transport towards the Arctic and identify the crucial role of the interactions between AMET and OMET. This leads to a need for reliable quantification of meridional energy transport in the Northern Hemisphere.

*Quantifying meridional energy transport*

The energy transport in the climate system is obtained through either integral of all the energy components, or by calculating the residual between sources and sinks of energy in the system, which are often referred to as direct and indirect methods, respectively. The variations of the energy budget ( $\frac{\partial E}{\partial t}$ ) can be approximated as  $\frac{\partial E}{\partial t} = \nabla \mathbf{v} \cdot \mathbf{E} + F_s$ . Using direct method, we include all the relevant components (e.g. dry static energy, latent energy) and compute the divergence of energy in the system ( $\nabla \mathbf{v} \cdot \mathbf{E}$ ). While using indirect method, we calculate the energy transport as the difference between the energy sources and sinks ( $F_s$ , e.g. solar radiation at the top of the atmosphere), and the variations of energy ( $\frac{\partial E}{\partial t}$ ) in the system.

However, these are not easy tasks due to a lack of reliable data, which is either related to the sparseness of observations or caused by the defects in the numerical models used to construct these data sets. The latter includes issues like coarse model resolution, imperfect physical modelling, mass imbalance, unrealistic moisture budget [Trenberth, 1991]. Some of these issues are rooted in our incomplete knowledge of the energy budget, for instance, the uncertainty related to the observed radiation at top of the atmosphere (TOA) and the heat storage in the ocean [Loeb et al., 2012, Trenberth et al., 2014].

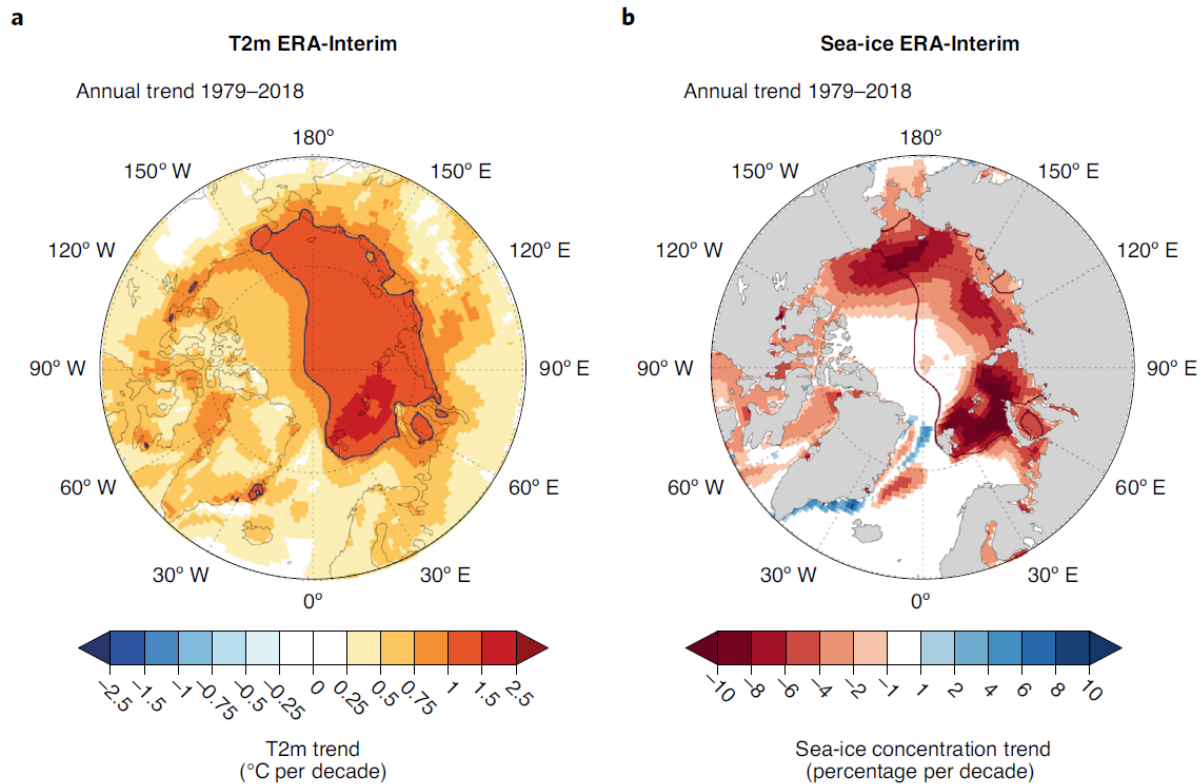
Since the early 20th century, many efforts have been made to reproduce the AMET and OMET, even with very limited availability of observational data [e.g., Vonder Haar and Oort, 1973]. After entering the satellite era, the increased data richness boosted the studies of AMET and OMET, and the quality and reliability of the estimations of MET have been improved. However, these estimations still have limitations. For example, most of these studies were conducted using outdated data sets and old numerical models [e.g., Trenberth and Solomon, 1994, Wunsch, 2005, Mayer and Haimberger, 2012]. These computations did not benefit from the latest numerical weather prediction and ocean models, the increased number and coverage of observations in the recent decade, the improved knowledge of physical modelling, and the increase in resolution and length of the covered time span. Although there were some attempts based on the improved models and methodologies, these studies were subject to a single reanalysis product or model simulations, which are prone to the uncertainty associated with data sources and provide little room for the inspection of the robustness and reliability of their conclusions [e.g., Fasullo and Trenberth, 2008, Mayer and Haimberger, 2012].

This motivates our research on the quantification of AMET and OMET using multiple state-of-the-art reanalysis data sets, and on the examination of the coherence between MET and the Arctic climate variability. By incorporating the latest reanalysis products, it paves the way for a better estimation of AMET and OMET, as well as the atmosphere and ocean interactions considering the energy compensation.

*Meridional energy transport and the Arctic warming*

Driven by the combination of natural variability and anthropogenic forcing, global warming became the main trend of climate system since mid-20th century and it is enhanced by the increase of greenhouse gas emissions. Given the fact that Arctic is extremely susceptible to global warming and climate change, the Arctic has experienced big changes over time [Overpeck

et al., 1997] and it is warming much faster than the rest of the world, which is an effect known as Arctic Amplification (AA) [Comiso and Hall, 2014, Francis et al., 2017]. This unprecedented trend of warming in the Arctic climate has been captured by both the satellite and in-situ observations, and the increase in surface temperature has left their footprints on the cryosphere [Comiso and Hall, 2014, Wang et al., 2016b, Stroeve and Notz, 2018]. Figure 1.5 shows the trend of surface air temperature and sea ice concentration (SIC) in the Arctic from 1979 to 2018.



**Figure 1.5:** (a) Near-surface air temperature (2 meter temperature) trend (degree per decade) and (b) Sea-ice concentration trend (percentage per decade) over 1979–2018 from ERA-Interim. Excerpt from [Jansen et al., 2020].

The Arctic sea ice extent is in sharp decline. Since the beginning of the satellite era, there is significant sea ice loss in the Arctic across all seasons. The most dramatic decrease of sea ice up to approximately 50% has been observed in summer and autumn [Vihma, 2014]. Large anomalies in spring and winter sea ice coverage have been found in both the historical records and numerical simulations, and they are as significant as the observed drop in summer sea ice extent [Stroeve and Notz, 2018]. Although data available on sea ice thickness is not as abundant as sea ice extent, recent studies have agreed on the fact that sea ice thickness in the Arctic has experienced a remarkable decrease, especially in winter the reduction of sea ice thickness is approximately 50% compared to the early records around 1980 [Kwok and Rothrock, 2009, Lindsay and Schweiger, 2015].

Previous studies have shown that the AA and sea ice decline are coupled with poleward energy

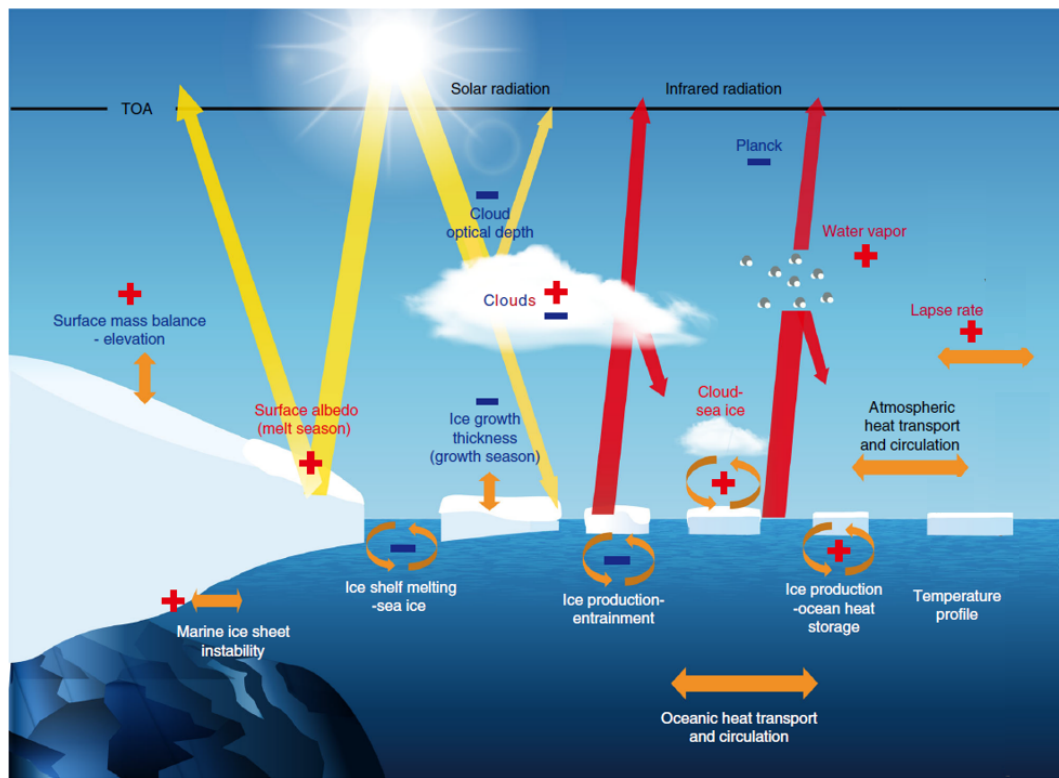
transport [Hwang et al., 2011]. In the atmosphere, AMET plays an important role in the Arctic temperature evolution [Yang et al., 2010] and therefore leverages the advance or retreat of the Arctic sea ice in different seasons [Kapsch et al., 2013]. A link between the vertical structure of Arctic warming at the surface and in the troposphere, and the advection of AMET towards the Arctic was found in the historical records [Graversen et al., 2008].

In the ocean, specifically in the Atlantic sector, increasing energy transports can contribute to sea ice melting and Arctic warming. OMET through the Barents Sea opening and the Fram Strait has a strong impact on the sea ice variability in the Barents Sea and the Central Arctic Ocean [Mahlstein and Knutti, 2011, Årthun et al., 2012, Sandø et al., 2014]. An increased OMET into the Nordic Seas and Arctic Ocean in recent decades was reported using more than 20 years of volume transport measurements [Tsubouchi et al., 2021]. It was confirmed by climate model projections that this trend of increasing OMET in the Atlantic sector is determined by warmer water as the current itself is found to weaken due to a slowdown of the AMOC and a future long-term increase is likely to continue in a warming background [Årthun et al., 2019]. The increase is sufficient to account for the recent accumulation of heat in the northern seas in recent decades [Tsubouchi et al., 2021]. Most of this extra heat warmed the ocean and a fraction of it caused sea ice melt [Mayer et al., 2016].

#### *Exploring compensation between energy transport variations in the historical records*

Changes in AMET and OMET have been proposed to explain the climate variability in the Arctic. This is relevant for the balance of surface energy budget in the Arctic, and it is inextricably linked with complex feedback mechanisms and ultimately determines the growth and decay of sea ice in the polar region. The decline in sea ice cover due to the increased OMET exposes the ocean to the atmosphere. It causes a decrease in surface albedo. This leads to an increase in surface solar radiation absorption and oceanic heat uptake, which amplifies warming and drives further melt [Curry et al., 1995, Mayer et al., 2016]. This is a positive feedback mechanism called the sea-ice albedo feedback, which serves as one of the key drivers for sea ice variations [Curry et al., 1995]. It also interacts with other feedback processes in the Arctic, like the temperature feedback [Pithan and Mauritsen, 2014], and lapse-rate feedback [Feldl et al., 2020]. These feedback processes are corresponding to the non-linear relation between sea ice change and Arctic warming [Goosse et al., 2018], as shown in Figure 1.6. The impact of OMET is extended to the atmosphere through open water via surface flux and this allows intensified flow of heat between the ocean and atmosphere. As a result, it greatly contributes to the complexity of the ocean-atmosphere coupling in the Arctic [Deser et al., 2015]. The resulting decrease of temperature gradient at high latitudes due to the Arctic warming weakens the AMOC and its poleward energy transport [Sévellec et al., 2017, Caesar et al., 2018, Liu et al., 2019]. This affects the thermohaline circulation and passes its influence to the atmosphere through air-sea interactions.

At relatively large time scales (e.g. interannual to decadal time scales), the compensation between the AMET and OMET via air-sea interactions leads to a closed energy budget at the top of the atmosphere and hence a stable climate system. The variability of the total heat transport is much smaller than the variability in either AMET or OMET over interannual time scales. This



**Figure 1.6:** A schematic of some important radiative and non-radiative feedbacks in polar regions involving the atmosphere, the ocean, sea ice and ice sheets. Excerpt from Goosse et al. [2018].

is a feature known as Bjerknes compensation [Bjerknes, 1964]. This phenomenon is characterized by an interplay of AMET, OMET, and air-sea interaction in the climate system. Furthermore, it is associated with the response of atmosphere and ocean to the changes of energy transport over decadal time scales. Intrinsically, these adjustments are driven by physical processes involved in the atmospheric and oceanic circulation, such as wind-driven Ekman transport and Sverdrup transport, anomalous buoyancy flux and relevant impact on the thermohaline processes, changes in baroclinicity and variability of heat and momentum transport by eddies [Shaffrey and Sutton, 2006, Van der Swaluw et al., 2007, Liu et al., 2020a]. As a consequence, a deep look at the mechanism of Bjerknes compensation can provide insight into the general circulation and energy budget in the Arctic. However, due to the involvement of complex climate feedback mechanisms [Liu et al., 2016], intricate air-sea interaction processes [Czaja and Blunt, 2011], and influence from the global warming [He et al., 2019], it is a challenging task.

This topic has been addressed by many studies and most of them demonstrated that compensation between AMET and OMET occurs from low latitudes to high latitudes across many different time scales [e.g. Shaffrey and Sutton, 2006, Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, Farneti and Vallis, 2013, Liu et al., 2016, Outten and Esau, 2017, Outten et al., 2018]. However, these studies mostly relied on numerical climate models with relatively long records. An intercomparison of Bjerknes compensation found in different numerical models shows that these results are model dependent and they can hardly reach a consensus on the mechanism of the Bjerknes compensation [Outten et al., 2018].

Given the differences in representing Bjerknes compensation in the numerical models, this motivates our interests in the interaction between atmosphere and ocean, and Bjerknes feedback in the historical climate records. The use of multiple reanalysis data sets will allow us to investigate the Bjerknes feedback in each reanalysis product and intercompare the results. This can help us examine the robustness of the mechanism of Bjerknes compensation found in each dataset, and explore the large scale variability of the Arctic climate associated with the interactions between the atmosphere and ocean.

Almost all the mechanisms listed above, which describe the evolution of the Arctic climate system and explain Arctic warming, are coupled to each other via the energy budget. This points to a way for research on the Arctic climate, which is to quantify the energy transport and identify their roles in the polar area. Furthermore, insight into the MET and its influence on the Arctic climate can assist the forecasting of the evolution of climatological fields in the polar area, for instance, the variability of Arctic sea ice. It is particularly informative for data-driven weather forecast approaches, whose forecast skill is strongly dependent on the selection of input variables. The lessons learnt from the quantification and evaluation of AMET and OMET towards the Arctic, as well as the compensation between AMET and OMET variations, form the basis for feature selection, which serves as the starting point of weather forecasting based on deep learning approaches. This initiates the second part of this thesis, in which I elaborate upon our attempts to push the boundary of weather forecasts in the polar region using deep neural networks.

## 1.2 Improving weather forecast in the Arctic with deep learning

Driven by the demand from societal, scientific and commercial sectors, improving weather and climate forecasts in the Arctic is one of the most challenging tasks. It is of paramount importance to the science community and the society. In this section, I will describe the predictability of the Arctic weather and climate in general and discuss the current techniques for weather forecasting, which are mostly based on the numerical weather prediction systems (NWP). Finally, I will explain the reason for introducing data-driven approaches to climate science and elaborate on the potential of deep neural networks for weather forecasting.

### 1.2.1 Predictability of Arctic weather and climate

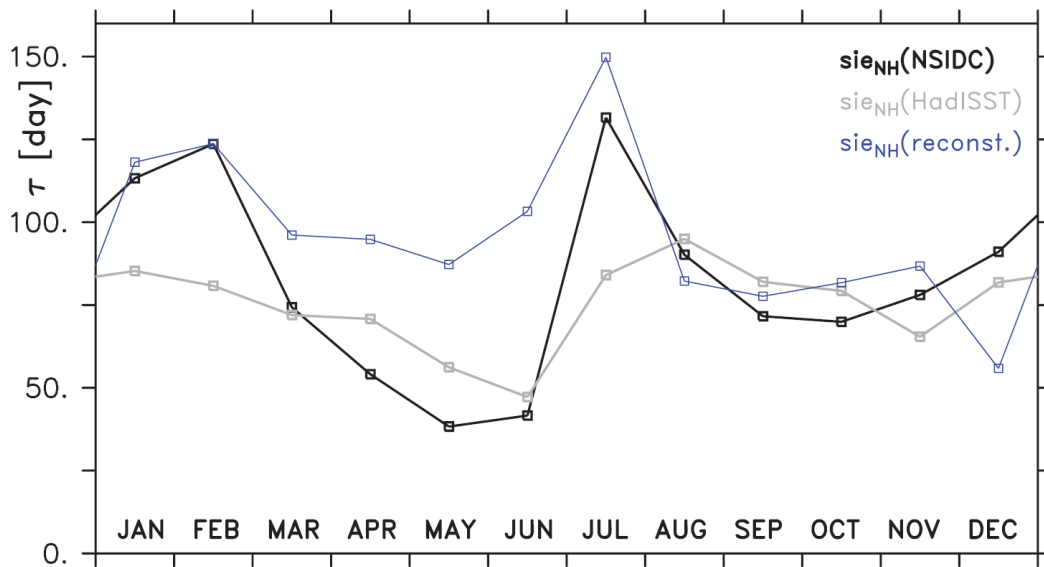
The basis for Arctic weather and climate forecasting at different time scales and locations is the predictability coming from the physical processes and interactions between the elements in the integrated climate system, including atmosphere, ocean, cryosphere and anthropogenic influence [Guemas et al., 2016]. Despite the chaotic nature of the atmosphere, large-scale atmospheric flow still acts as the major source of predictability at hourly to weekly time scales in the Arctic. Also, large-scale patterns in the atmospheric system, like the AO and NAO, can provide predictability to the forecasts from seasonal to decadal time scales [Bader et al., 2011, Gastineau and Frankignoul, 2015, Gong et al., 2018]. Over decadal time scales, changes in the ocean are expected to have an effect on predictability because of the relatively slow changes of ocean heat



content (OHC) due to the large heat capacity and the effect of ocean heat transport convergence on OHC. In the Arctic, the persistence and propagation of anomalies in SIC and sea surface temperature (SST) can form the basis for the forecasts from sub-seasonal to seasonal time scales [Blanchard-Wrigglesworth et al., 2011b,a, Day et al., 2014b]. Moreover, some large scale patterns and processes in the ocean, like Pacific Decadal Oscillation (PDO), Atlantic Multidecadal Oscillation (AMO) and AMOC, are useful predictors for the Arctic climate forecasts at longer time scales [Newman et al., 2016, Delworth et al., 2017]. In this thesis, I mainly focus on the Arctic sea ice predictability and forecasting.

### *Arctic sea ice predictability*

Given the close relations between sea ice and other components in the Arctic climate system, there are multiple sources that provide predictability for the Arctic sea ice. The first source of predictability of the Arctic sea ice lies in the variability of sea ice itself. The Arctic sea ice shows strong persistence at sub-monthly to monthly scales in winter and summer [Blanchard-Wrigglesworth et al., 2011a]. In general, this statement holds for the sea ice extent (SIE), sea ice thickness (SIT) and sea ice volumes (SIV) for most of the regions in the pan-Arctic [Guemas et al., 2016]. A relatively large autocorrelation of SIC was identified in spring and autumn with time scales of 3-7 weeks in the Arctic area. Using both the observations and model output, previous studies have shown that a characteristic e-folding time of the total Arctic SIE ranges from 2-5 months across different seasons [Blanchard-Wrigglesworth et al., 2011a], as shown in Figure 1.7. Such persistence patterns of SIC are consistent in both the satellite observations and numerical model simulations [Blanchard-Wrigglesworth et al., 2011a, Day et al., 2014b].



**Figure 1.7:** E-folding time-scale of the correlation between the monthly linearly detrended Arctic sea-ice-extent anomalies and the monthly linearly detrended sea-ice-extent anomalies of the following month from the NSIDC observational dataset in black, the HadISST dataset in grey and the HistDfsNudg sea-ice reconstruction from Guemas et al. [2014] in blue. The figure is excerpted from Guemas et al. [2016].



As an extension to the persistence, re-emergence of sea ice anomalies can also provide predictability to the Arctic sea ice forecast. Depending on the mechanism, there are a few types of re-emergence in models. On the basis of a slow decorrelation of thickness anomalies, a summer to summer re-emergence exists in the model simulations [Blanchard-Wrigglesworth et al., 2011a]. There is also re-emergence associated with the link between the anomalies in the growth season and its origin in the melt season. The latter stems from the persistence of SST and it is confirmed in both the observations and model studies [Day et al., 2014b].

Similar to the re-emergence, there is another source of predictability related to the variations of sea ice, which is the advection of sea ice anomalies. Characterized by the anticyclonic and cyclonic regimes of the mean Arctic ocean circulation, the sea ice drift occurs periodically, following the shifts of the regimes. Some studies suggest a periodicity of 5-7 years [e.g., Proshutinsky and Johnson, 1997], but this is still debatable. This phenomenon can provide predictability in the sea-ice conditions at interannual time-scales. Such sea ice variations are associated with the sea ice transport via Transpolar Drift Stream and sea ice export through Fram Strait, which are primarily wind-driven. This reveals the essential role of the atmosphere in driving Arctic sea ice variability, especially for these large-scale atmospheric patterns characterizing the Arctic surface circulation, like the NAO and AO [e.g. Hilmer and Jung, 2000]. The predictability of the atmospheric circulation and the large-scale patterns can therefore contribute to the predictability of the sea-ice conditions, as well.

Given the slow changes of OHC due to the large heat capacity, the ocean possesses a relatively “long memory” and it can act as a major source of sea ice predictability. For instance, the persistence associated with vast heat storage and slow propagation of heat anomalies in the ocean allows for a winter-to-winter memory of the presence sea ice edge [Guemas et al., 2016]. The source of predictability from the ocean is rooted in either the dynamics process (e.g. drag effect due to the ocean circulation) or the thermodynamics process (e.g. heat exchange and heat transport). Given the susceptibility of sea ice advance and retreat to the variations of heat budget, the latter can be used to forecast sea ice for large time scales. For instance, with observed OMET in the Atlantic sectors, recent studies have shown considerable prediction skill of SIE and SIV in the Barents Sea at interannual time scales [Årthun et al., 2012, Onarheim et al., 2015]. Furthermore, it was found by some model-based experiments that OMET in the Barents Sea Opening and through Fram Strait could even be used to analyse and predict the sea ice variability concerning complex morphological changes, like ponding, frazil, congelation growth and bottom melting in the Central Arctic Ocean [Sandø et al., 2014].

### 1.2.2 Weather forecasting with numerical weather prediction system

Modern weather forecasts are usually based on numerical weather predictions. It is a computation-intensive process of solving mathematical equations representing the climate system governed by the laws of physics. This process always relies on NWP systems, which are integrated numerical systems combining the art of mathematical representation of physical processes, parameterization schemes, the representation of spatial and temporal variability by the discretizations, data assimilation frameworks, the treatment of boundary conditions and the initialization approaches [Bauer et al., 2015]. Together with ensemble approaches [Gneiting and

Raftery, 2005], the deterministic NWP-based weather forecasts can be further extended to the probabilistic domain and therefore provide knowledge about uncertainty associated with the evolution of a chaotic system. However, this system also has some disadvantages. For instance, these numerical climate models require near real-time input of environmental conditions and accurate models to assure the quality of the forecasts, and the simulations are computationally expensive, especially considering the implementation of a large ensemble [Milinski et al., 2020] and high resolution. Moreover, incomplete physical knowledge and imperfect approximation of the physical processes on the basis of parameterization schemes can also influence the forecasts quality [Stensrud, 2009]. Despite of all these drawbacks, NWP-based ensemble forecasting, at this moment, is still the most used solution for operational weather forecasts.

In practice, apart from the limits placed by the potential predictability attributed to physical processes and the defects in models, the configuration of numerical experiments can also significantly influence the forecasting performance. This includes the initialization strategy [Day et al., 2014a,b], richness of observations for the data assimilation, model spin-up and initial shocks and drifts, ensemble generation approaches [Guemas et al., 2016]. In order to generate reliable sea ice forecast, it is necessary to take these factors into consideration.

The sparseness of observations available in the Arctic and the resulting large uncertainty in the initial state significantly limit the skill of sea ice forecasts in the Arctic based on NWPs [Sato et al., 2017]. Also, modelling challenges of some complex physical processes, for instance, wave-ice interaction, landfast ice, melt ponding, and floe size distribution rheology, ice thickness distribution, are not well addressed in the numerical climate models [Leppäranta, 2011, Leppäranta et al., 2020]. Similarly, the imperfect approximations of many physical processes via parameterization, either related to sea ice variations, or associated with changes in the atmosphere and ocean, also hamper the forecast quality in this region. As a result, for NWP-based forecasting, the forecast quality and even the potential sea ice predictability show strong dependency on the numerical models [Blanchard-Wrigglesworth and Bushuk, 2019]. Note that forecasts with fully coupled systems need to undergo bias correction due to the drift and biases growth [Gneiting et al., 2007]. The forecast quality is thus prone to the shortcomings of the bias correction methods, as well.

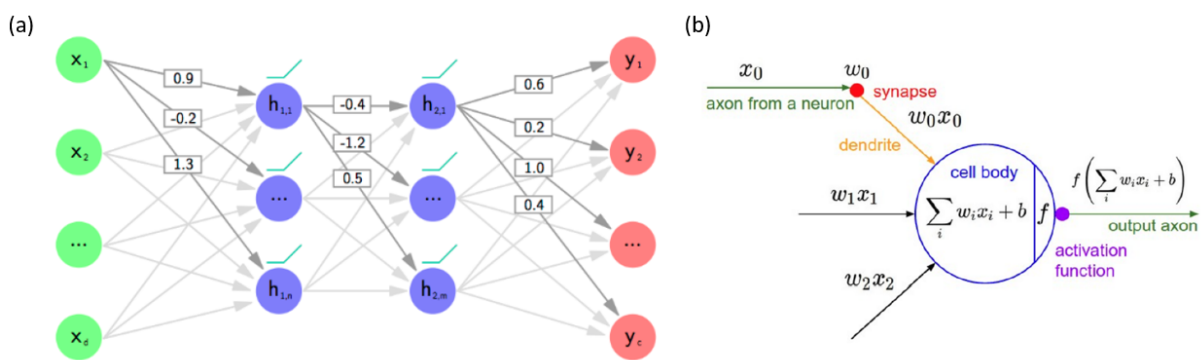
Although the advances in numerical weather prediction have pushed the NWP-based weather forecast to a new level in the recent decade, given the defects of the NWP-based approaches, the community continues to search for new techniques and methodologies to assist, improve and even replace this system. Considering that the essence of nonlinear feedback mechanisms in the integrated Arctic climate system are reflected in the historical records and model outputs, we might need nonlinear approaches to replicate the coupling between atmosphere, ocean and sea ice and forecast the evolution of the system. This motivates us to search for a data-driven method that can account for complex nonlinear relations and finally brings the deep learning into our scope.

### 1.2.3 Deep neural networks for weather forecasting

Unlike NWP systems, deep learning approaches do not require prior physical knowledge to build a model. They try to learn the non-linear relations between given fields and determine

the parameters of the predefined neural networks via an optimization procedure. The distilled non-linear relationships are stored into the chosen neural network structure. Hence, the neural networks try to approximate the physical consistency of the given system and can be used to produce forecasts. Also, these deep neural networks are trained with a relatively large data base and therefore do not rely on a single initial state. A bias correction method is not needed for deep learning. Moreover, deep neural networks are flexible and adaptive tools that can deal with heterogeneous data. As a consequence, they can make use of abundant historical records of weather and climate states. The rapid development of computer hardware and machine learning methods leads to the increasing trend of adapting latest techniques in deep learning to the conventional weather and forecasting applications [e.g Salman et al., 2015, Kim et al., 2017, Rasp et al., 2018, Kim et al., 2019, 2020, Ham et al., 2019, Reichstein et al., 2019]. Consequently, deep neural networks become promising tools to deal with complex weather and climate forecasting tasks, like Arctic sea ice forecasts.

Known for their “brute-force” problem-solving abilities, deep learning benefits from big leaps in the development of computing infrastructures and have become a powerful toolkit for multi-disciplinary applications [LeCun et al., 2015]. The core of deep learning is neural networks (NN). Neural networks are constructed based on many simple neurons. Mathematically, these neurons are weight matrices and each can produce a sequence of activations depending on its structure and the connection with other neurons [Schmidhuber, 2015]. During the so-called training process, these weight matrices will be updated according to the chosen optimization strategy and they are trained to represent the features of training data. In terms of the training methods, machine learning approaches can be categorized as supervised learning, unsupervised learning and reinforcement learning [Schmidhuber, 2015]. For example, Figure 1.8 provides a simple multilayer neural network and a close look at the mathematical process in one neuron, which illustrates the concept of supervised learning. Deep learning is part of the family of machine learning and its types are often defined in terms of the architecture of deep neural networks.

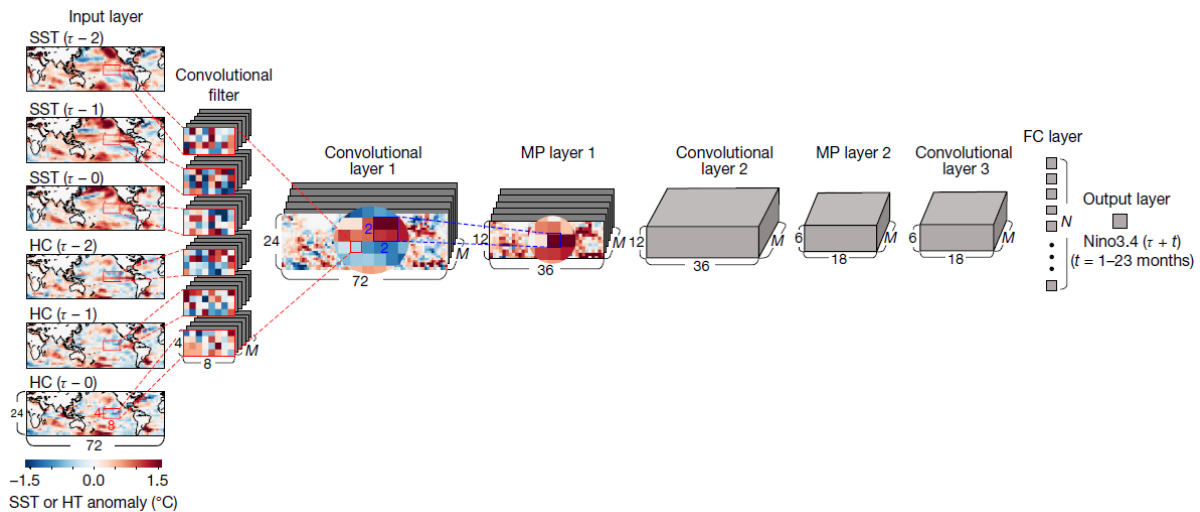


**Figure 1.8:** (a) An illustration of multilayer neural networks and (b) a close look at the mathematical process inside a neuron.  $x$  and  $y$  are input and output values, respectively.  $h$  represents the hidden unit.  $w$  and  $b$  are weight matrices and bias. The figure is excerpted from Shridhar et al. [2019].

In this thesis, I mainly focus on supervised learning. Supervised learning is the most common form of deep learning, or machine learning in general. During training, the training data will be

fed to the neural network and the neural network will produce an output in a desired form. A loss function is used to measure the error between the output and the ground truth. This procedure is called the forward process. Then we need to modify the internal adjustable parameters in the neural network to reduce the error. Several mathematical approaches exist for this purpose and most of them are derived from stochastic gradient descent (SGD) [LeCun et al., 2015]. This consists of including a few examples as the input vector, computing the outputs and the errors, computing the average gradient for those examples, and adjusting the weights accordingly. In combination with a procedure called the back-propagation, which computes the gradient of an objective function with respect to the weights of a neural network based on the chain rule of derivatives, we are able to adjust the weights to reduce the error. By repeating the whole procedure for a number of times, the trained network will learn to account for the non-linear relations between given variables. Note that in this procedure, the ground truth is shown to the neural network. This is the essence of the concept of supervised learning.

The main function of a neural network is to extract relevant information from the training set and represent the useful information with multiple levels of abstraction in a latent space [LeCun et al., 2015]. Depending on the implementation of different types of feature extractors and the corresponding architectures, these neural networks can be divided to different categories. Among all of them, the Convolutional Neural Networks (CNN) and the Recurrent Neural Networks (RNN) are most prevalent in weather forecasting, due to their capacities in handling spatial and temporal sequential data [Reichstein et al., 2019]. We will use these two types of deep neural networks to forecast sea ice in the Arctic.

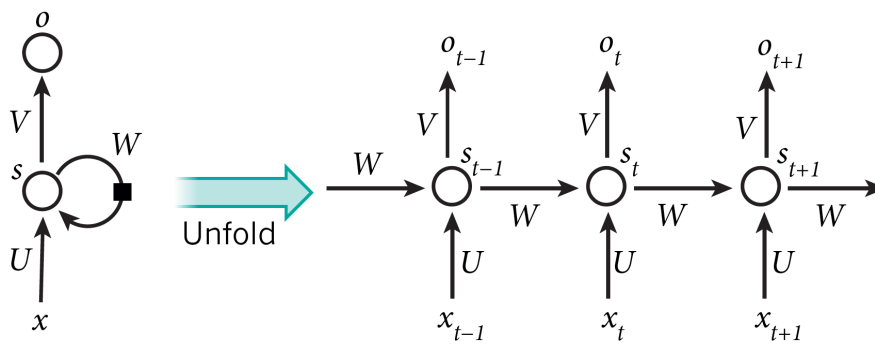


**Figure 1.9:** Architecture of the CNN model used for the ENSO forecasts. The CNN model consists of one input layer (the predictor), three convolutional layers, two max-pooling (MP) layers, one fully connected (FC) layer, and one output layer (the predictand). The figure is excerpted from Ham et al. [2019].

CNN is a typical type of feed-forward network that is designed to process data in the form of stacked layers [Rawat and Wang, 2017]. In general, CNN architectures consist of multiple layers with different functions, which are convolutional layers, pooling layers and dense layers.

Convolutional layers act as feature extractors. They are used to detect local conjunctions of features from the previous layer. The role of the pooling layer is to achieve spatial invariance from input distortions and translations. They can merge semantically similar features into one and therefore reduce the spatial resolution of the feature maps [LeCun et al., 2015]. The dense layers, also known as fully-connected layers, are often placed as the last layer before the output. They serve to interpret the feature representations extracted by the convolutional layers and pooling layers, and perform the function of high-level reasoning [Rawat and Wang, 2017]. By stacking these three type of layers, we can build a CNN that is able to exploit the hierarchic properties of the data and perform regression or classification with the trained network. For example, an architecture of a CNN model used for the ENSO forecasts is shown in Figure 1.9. Based on the classical architecture of CNN, many variants have been proposed to improve the classification accuracy or to reduce the computational cost [Rawat and Wang, 2017].

Serving as an “expert” in sequential data processing, RNNs are popular feed-forward neural networks which are good at capturing the dynamics of sequences via cycles in the network nodes [Lipton et al., 2015]. Their ability comes from the basic structure of a RNN, which are the recurrent layers or hidden layers. These layers consist of recurrent cells whose states are affected by both past states and current input with feedback connections [Yu et al., 2019]. A typical RNN architecture and the unfolding in time of the computation involved in its forward computation is shown in Figure 1.10. In each forward process, RNNs process an input sequence one element at a time. A summary of the history of all the past elements of the sequence is maintained in their hidden units implicitly [LeCun et al., 2015]. By organizing the recurrent layers into various architectures, it is possible to obtain RNNs in different forms. Long-Short Term Memory Network (LSTM) is one of those variants [Hochreiter and Schmidhuber, 1997], which preserves the powerful dynamic properties of RNNs while it solves the problem of gradient exploding or vanishing during back-propagation. In this thesis, we will apply LSTM and more details about the implementation of this neural network to weather forecasting is given in the Chapter 4 and 5.



**Figure 1.10:** A recurrent neural network and the unfolding in time of the computation involved in its forward computation.  $x_t$  and  $o_t$  are inputs and outputs at step  $t$ .  $U$ ,  $V$  and  $W$  are weight matrices.  $s_t$  is the state at step  $t$ . The figure is excerpted from LeCun et al. [2015].

So far, these neural networks have shown promising skill in many climate science related studies, like short term weather forecasting [Dalto et al., 2015], extreme events forecasting [Gope et al., 2016], forecasting of large-scale climate patterns [Salman et al., 2015, Feng et al., 2016, Ham

et al., 2019], representation of physical processes in climate models [Rasp et al., 2018], missing observation data reconstruction [Kadow et al., 2020], post-processing of ensemble forecasts [Rasp and Lerch, 2018]. Deep learning is data-driven and it can make use of the large amount of historical records in a way different from NWP-based approaches. It is very promising to extend the application of NN to Arctic sea ice forecasts, as mentioned in the beginning of this section.

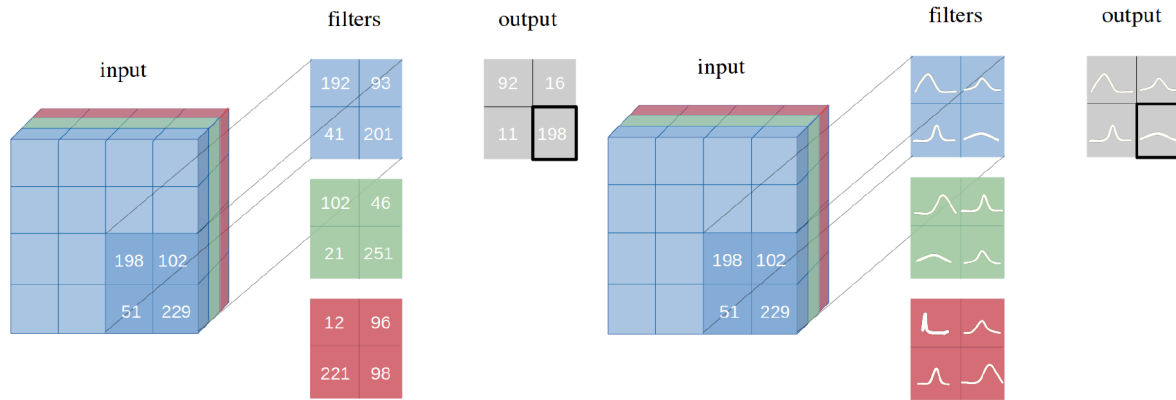
### *Probabilistic deep learning*

Trustworthy weather forecasting involves uncertainty estimation related to uncertainties in the observations, models formulations and stochastic nature of flows. Usually, for a NWP-based forecast system, this is achieved by using ensemble approaches. It is also possible to extend deterministic forecasts with NN to probabilistic forecasts through ensemble approaches. However, for a deterministic NN, it is very expensive to generate an ensemble. In practice, a NN-based ensemble is obtained through perturbing either the training data or the structure of NN [e.g. Zaier et al., 2010, Wang et al., 2017]. Depending on the types of perturbation, these approaches can be further categorized as randomization, bagging and boosting Zaier et al. [2010]. However, these methods require multiple times of training to obtain a certain number of ensemble members and therefore they can be very costly.

Fortunately, a probabilistic deep learning approach, namely the Bayesian deep learning (BDL), provides a feasible solution for NN-based probabilistic weather forecast [Blundell et al., 2015]. Unlike deterministic neural networks, where the weight matrices are fixed, within BDL the weight matrices are distributions. An ensemble forecast can be made by sampling the weight distribution [Shridhar et al., 2019]. The difference between a deterministic NN and a probabilistic NN is illustrated in Figure 1.11. There are several variational inference schemes available for the training of Bayesian deep neural networks (BNN), like Monte Carlo Markov Chain (MCMC) [Salimans et al., 2015], Monte Carlo dropout (MCD) [Gal and Ghahramani, 2015, 2016] and Bayes by Backprop (BBB) [Blundell et al., 2015, Shridhar et al., 2019].

The ability of uncertainty representation is fundamental for any probabilistic forecast and therefore makes BDL a candidate for weather forecasting. Typically, three types of uncertainties are considered: (1) Uncertainties in the initial conditions (2) necessary approximations and corresponding uncertainties in the construction of a numerical model of the real atmosphere (3) uncertainties posed by external forcing, boundaries and the chaotic nature of flows. The last one is often ignored in an operational weather forecast system, but relevant for climate studies and local forecasts. In NWP systems, uncertainties in the initial conditions and model parameters are projected by ensemble forecasts with perturbations in the initial conditions and model formulations Palmer [2002], Milinski et al. [2020]. While with BDL, model uncertainty is addressed via placing a prior distribution over a model's weight, and initial condition uncertainty is represented by generating a distribution over the output of the model [Kendall and Gal, 2017]. Consequently, uncertainties in initial conditions and model parameters can both be addressed by BDL, albeit differently than in operational NWP approaches.

Applied to different types of neural networks, for instance, the CNN and the LSTM, BDL can gain sufficient capacity in representing the variance of intricate spatial-temporal sequential data



**Figure 1.11:** Difference of forward procedures between a deterministic NN and a probabilistic NN. The figure is excerpted from Shridhar et al. [2019].

and provide adequate knowledge of uncertainty. These all make BDL a good candidate for the Arctic sea ice forecast.

### 1.3 Thesis Overview

In this thesis, I present our attempts to gain more physical knowledge about the Arctic weather and climate, and our endeavours to improve the weather forecast with novel statistical approaches in this region. At the beginning of this chapter, I have explained the essential role of the meridional energy transport in the Northern Hemisphere. It has strong impact on the general circulation of the atmosphere and ocean in the Arctic. The AMET and OMET are key components in the energy budget. They regulate the variability of sea ice in the polar region and are related to the warming in this area. In order to improve our understanding about the weather and climate system in the Arctic, it is necessary to have reliable quantification of heat transport. Thanks to the developments in modelling techniques and an increase of data richness, it is promising to have better quantification of AMET and OMET using modern reanalysis products. Our motivation and the improved conditions naturally lead to the research question:

- *Can we obtain a reliable and consistent estimation of meridional energy transport in the atmosphere and ocean using multiple modern reanalysis data sets, and are the results coherent with the Arctic climate variability at different time scales?*

To provide an answer, we will compute energy transport using reanalysis products and a synthesis and evaluation of the results is given in Chapter 2. Specifically, we will calculate AMET and OMET in the Northern Hemisphere towards the Arctic using six reanalysis products. The assessment will be based on the inter-comparison among the chosen data sets and the OMET will be further evaluated against independent observations in the Atlantic ocean. An extended analysis of the linkages between the evolution of the Arctic climatological fields and the variability of AMET and OMET is also conducted. Better quantification of AMET and OMET

allows us to continue our study on the relationship between the variability of the Arctic climate and changes in the energy budget. In the first section of this chapter, I have described the air-sea interactions in the perspective of heat transport and introduced the Bjerknes compensation. Corresponding to complex air-sea interaction processes and climate feedback mechanisms, Bjerknes compensation is responsible for maintaining the energy balance between the atmosphere and ocean, and therefore helps to stabilize the climate system. More knowledge about the occurrence of Bjerknes compensation and the mechanisms related to it in the Northern Hemisphere can potentially shed lights on the climate variability in the Arctic. Studies with numerical climate models fail to reach an agreement on the presence and mechanism of the Bjerknes compensation. Their results are model-dependent and are not fully convincing. Different from these studies, we will revisit this topic with multiple modern reanalysis products and try to answer the following question:

- *Does the compensation between energy transport variations in the atmosphere and ocean in the Northern Hemisphere occur in the historical records and is there a robust mechanism found in multiple modern reanalysis products?*

This question is addressed in Chapter 3, by investigating the energy compensation between the atmosphere and ocean from annual to decadal time scales using the same reanalysis products as in Chapter 2. For the first time, the Bjerknes compensation is studied using reanalysis data. I will focus on the response of the atmosphere and ocean to the variations of heat fluxes and study the consistency of the physical pictures of the integrated climate system within each process.

By quantifying the energy transport and inspecting the Bjerknes compensation in Chapter 2 and 3, we will have a good overview of the coupling between the Arctic climate variability and energy budget variations in the polar region. This link is useful for weather forecasting in the Arctic and it can provide practical information for feature selection, which is vital for data driven approaches. In the last section, we have elaborated on our motivation for introducing the deep neural networks to weather forecasting. Physical knowledge of the climate system serves as the key to improve weather forecasts. The influence of energy budget components on the evolution of climatological fields in the Arctic, like sea ice advance and retreat in accordance with the varying heat transport, is valuable for determining the predictors as input fields for deep learning. It can further assist the analysis of physical consistency in predictands. Based on the physical pictures, I will examine the weather forecasting skill of deep learning approaches through a case study. It focuses on sea ice forecasting in the Barents Sea and the research question reads:

- *Can we produce skillful Arctic sea ice forecasts at weekly time scales using deep neural networks, and is the physical consistency preserved in these forecasts?*

I will present a special NN to work with complex spatial temporal sequential data and explore its capacity in sea ice forecasting in Chapter 4. It will focus on the sea ice in the Barents sea considering its important role in AA and Arctic heat budget since the heat exchange between ocean and atmosphere in this region is controlled by coupled feedback processes [Smedsrud et al., 2013]. The forecast with the NN will be assessed against the operational weather forecast with



numerical models and the forecast with an intricate non-linear statistical model.

A sophisticated weather forecast includes a good estimation of uncertainty. Such demand motivated us to search for a probabilistic deep learning technique that can satisfy this requirement in a feasible way with a relatively lower cost compared to an ensemble based on a deterministic NN, and therefore leads to the research question:

- *Can probabilistic deep neural networks learn to represent the evolution of weather and climate system with an appropriate estimation of uncertainty, and can we use them to generate reliable weather forecasts?*

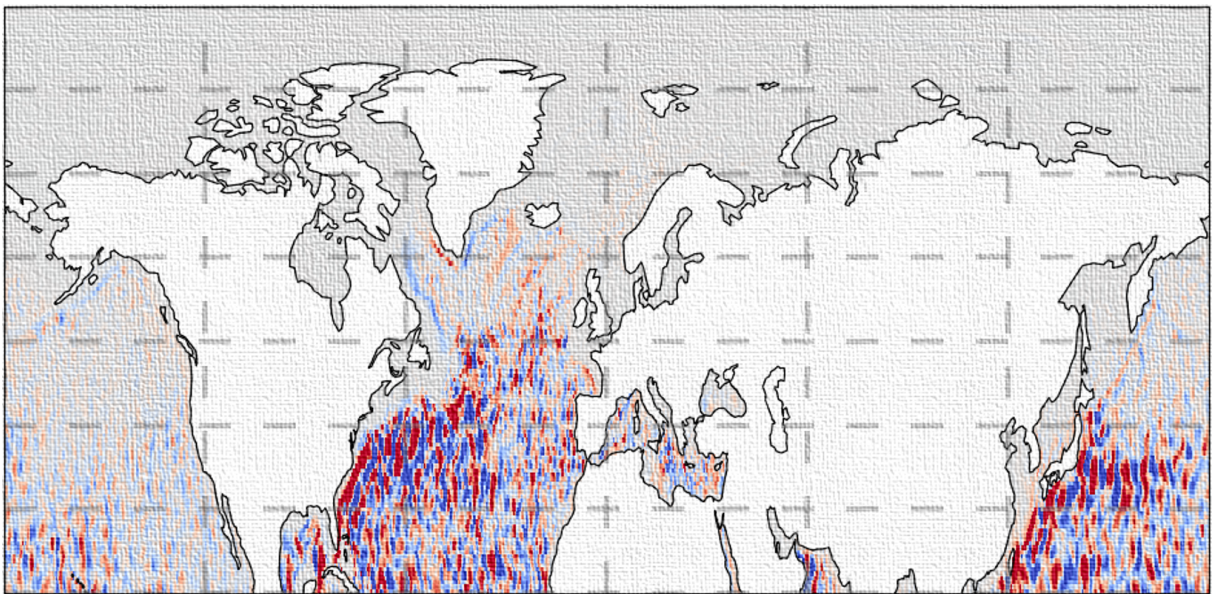
In Chapter 5, we will try to answer this follow-up question by exploring BDL with a simple Rossby wave system, namely the Lorenz 84 model. We will examine if the Bayesian deep neural networks are able to represent the evolution of the large scale eddies on a westerly jet and preserve the physical consistency between each field.

Finally, I will conclude our findings in Chapter 6. It serves as a synthesis of the whole thesis and provides an outlook with respect to the theme “Arctic weather and climate from mechanisms to forecasts”.



## Chapter 2

# Synthesis and evaluation of historical meridional heat transport from midlatitudes towards the Arctic



This chapter is based on:

Yang Liu, Jisk Attema, Ben Moat, and Wilco Hazeleger. Synthesis and evaluation of historical meridional heat transport from midlatitudes towards the arctic. *Earth System Dynamics*, 11(1):77–96, 2020b

## Abstract

Meridional Energy Transport (MET), both in the atmosphere (AMET) and ocean (OMET), has significant impact on the climate in the Arctic. In this study, we quantify AMET and OMET at subpolar latitudes from six reanalysis data sets. We investigate the differences between the data sets and we check the coherence between MET and the Arctic climate variability at interannual time scales. The results indicate that, although the mean transports in all data sets agree well, the spatial distributions and temporal variations of AMET and OMET differ substantially among the reanalysis data sets. For the ocean, only after 2007 the low frequency signals in all reanalysis products agree well. A further comparison with observed heat transports at 26.5°N and the subpolar Atlantic, and a high resolution ocean model hindcast confirms that the OMET estimated from the reanalysis data sets are consistent with the observations. For the atmosphere, the differences between ERA-Interim and JRA55 are small, while MERRA2 differs from them. An extended analysis of linkages between Arctic climate variability and AMET shows that atmospheric reanalyses differ substantially from each other. Among the chosen atmospheric products, ERA-Interim and JRA55 results are most consistent with those from coupled climate models. For the ocean, ORAS4 and SODA3 agree well on the relation between OMET and sea ice concentration (SIC), while GLORYS2V3 deviates from those data sets. The regressions of multiple fields in the Arctic on both AMET and OMET suggest that the Arctic climate is sensitive to changes of meridional energy transports at subpolar latitudes in winter. Given the good agreements on the diagnostics among assessed reanalysis products, our study suggests that the reanalysis products are useful for the evaluation of energy transports. However, assessments of products with the AMET and OMET estimated from reanalysis data sets beyond interannual time scales should be conducted with great care and the robustness of results should be evaluated through intercomparison, especially when studying variability and interactions between the Arctic and midlatitudes.

## 2.1 Introduction

Poleward meridional energy transport, both in the atmosphere (AMET) and ocean (OMET), is one of the most fundamental aspects of the climate system. It is closely linked to the changes of weather and climate at different latitudes. The quantifications of AMET and OMET have been studied extensively. In the 1980s, many efforts were made to reproduce the AMET and OMET with very limited observational data available [Vonder Haar and Oort, 1973, Oort and Vonder Haar, 1976]. After entering the satellite era, much progress has been made in particular during the recent two data-rich decades. Using the radiation at the top of the atmosphere from satellite data and the reanalysis data, a complete picture of AMET and OMET is given by Trenberth and Caron [2001]. Following their work, rapid progress was made using similar methodologies and new data sets of observations [Ganachaud and Wunsch, 2000, 2003, Wunsch, 2005, Fasullo and Trenberth, 2008, Zheng and Giese, 2009, Mayer and Haimberger, 2012]. Nevertheless, these estimations still suffered from problems like mass imbalance, unrealistic moisture budget, coarse resolution, and sparseness of observations [Trenberth, 1991, Trenberth and Solomon, 1994]. Fortunately, recent improvements in numerical weather prediction and ocean models, and increased data coverage of observations provide a basis to improve the estimation of AMET and OMET. As a result of an increase of available reanalysis products, an increase in resolution and length of the covered time span, and an increase of components of the Earth system that are included in the products [Dee et al., 2011, Gelaro et al., 2017, Harada et al., 2016, Balmaseda et al., 2013, Ferry et al., 2012b, Carton et al., 2018], it is very promising to have better quantification of AMET and OMET using the latest reanalysis data sets. In this study, we will provide further insights into MET from midlatitudes towards the Arctic, with the state-of-the-art reanalysis products.

To support the examination of MET from midlatitudes towards the Arctic, it is worth investigating the AMET and OMET in relation to climate variability at different time scales in the Arctic region. In recent decades, the Arctic is warming twice as fast as the global average [Comiso and Hall, 2014, Francis et al., 2017]. This phenomenon is known as Arctic Amplification (AA) and it has an impact far beyond the Arctic [Miller et al., 2010, Serreze and Barry, 2011]. In order to understand the warming, the processes behind the AA, its wider consequences and to make reliable predictions of the Arctic climate, it is crucial to understand Arctic climate variability. Among all factors responsible for the variability in the processes described above, meridional energy transport, from midlatitudes toward the Arctic, plays a significant role [Graversen et al., 2008, Kapsch et al., 2013, Zhang, 2015]. There is a large volume of published studies describing the impacts of AMET and OMET on the variation of sea ice and the warming in the Arctic. Using reanalysis data, Yang et al. [2010] showed that poleward AMET is linked with the evolution of temperature in the free troposphere at decadal time scales. By separating the planetary and synoptic-scale waves, Graversen and Burtu [2016] showed that latent heat transport, as a component of AMET, influences the Arctic warming with reanalysis data. Gimeno-Sotelo et al. [2019] studied moisture transport with reanalysis data and observations, and showed that the moisture sources in the Arctic region are linked with interannual fluctuations of Arctic sea ice. Nummelin et al. [2017] analyzed the linkages between OMET, Ocean Heat Content (OHC) and AA through climate model simulations within the Coupled Model Intercomparison Project

Phase 5 (CMIP5). They reported an enhancement of OMET as a result of heat loss in the subpolar ocean and the contribution of OMET to the AA through increasing OHC in the Arctic ocean. Also by analyzing CMIP5 simulations, Sandø et al. [2014] showed a large impact of heat transport in the Barents Sea on sea ice loss. However, ocean reanalyses do not show a clear sign of AA in the Arctic OHC increases [Mayer et al., 2016, von Schuckmann et al., 2018, Mayer et al., 2019]. Consequently, knowledge on poleward AMET and OMET at subpolar and polar latitudes will aid in the understanding of AA.

Global climate models show compensations between variations in atmospheric and oceanic heat transports at subpolar and midlatitudes [Outten et al., 2018]. This is indicative of positive feedbacks between the ocean and atmosphere, and it has been associated with variations in sea ice by some studies [Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, van der Linden et al., 2016]. These studies all point to connection between energy transports and variations of the Arctic climate. However, these results are mostly based on numerical model simulations and they tend to differ among these models. In contrast to numerical modeling studies, here we intend to examine AMET and OMET variability and their relation with the Arctic using reanalysis data sets, which are regarded as the best estimates of the historical variability.

In this paper, we quantify AMET and OMET using multiple state-of-the-art reanalysis products. These are representations of the historical state of the atmosphere and ocean optimally combining available observations and numerical simulations using data assimilation techniques. Emphasis is placed on the variation of AMET and OMET from midlatitudes to the Arctic at interannual time scales ( $\sim 5$  yr). Different from earlier studies, we include multiple reanalysis data sets for intercomparison. Independent observations in the Atlantic from the Rapid Climate Change-Meridional Overturning Circulation and Heatflux Array (RAPID array) and the Overturning in the Subpolar North Atlantic Program (OSNAP) are included in the comparison. The RAPID array is a trans-basin observing array along  $26.5^\circ$  N in the Atlantic [Johns et al., 2011, McCarthy et al., 2015]. It operates since 2004 and provides the volume and heat transport in the Atlantic basin. OSNAP is an ocean observation program designed to provide a continuous record of the trans-basin fluxes of heat, mass and freshwater in the subpolar North Atlantic [Susan Lozier et al., 2017, Lozier et al., 2019]. Moreover, a state-of-the-art NEMO-LIM2 1/12° ocean circulation/sea ice model simulation forced by Drakkar Surface Forcing data set version 5.2 [Moat et al., 2016] is also employed in the comparison. Based on the intercomparison of reanalysis data, especially with the independent observation data, we will be able to identify the sources of uncertainty. To support our comparison of AMET and OMET, we also investigate the interactions between oceanic and atmospheric variations and remote responses. The correlations between the variability of AMET and OMET, and the changes in the Arctic climate are compared to literature. This is motivated by previous studies that explain those connections with only numerical models or a single reanalysis dataset [Graversen, 2006, Van der Swaluw et al., 2007, Graversen et al., 2008, Jungclaus and Koenigk, 2010, Kapsch et al., 2013].

The paper is organized as follows: Section 2 presents the data and our methodology. Results and analysis are given in Section 3. It includes AMET and OMET calculated from reanalysis data and an intercomparison of them. The correlation between the variability of AMET and OMET, and the Arctic climate is elaborated upon in detail. Finally, remarks are given in Section 4 and

conclusions are provided in Section 5.

## 2.2 Data and Methodology

The reanalysis data sets used in this study are introduced in this section. Moreover, the methodology for the quantification of AMET and OMET is also included in this section. The statistical tests performed in this study are elucidated in detail.

### 2.2.1 Reanalyses

In order to make use of observations and advanced numerical models, six state-of-the-art reanalysis data sets are used in this study. The chosen reanalysis products have a high temporal and spatial resolution, thus are suitable for the computation of energy transport (see section 2.3). We chose three atmosphere reanalysis data sets: ERA-Interim, MERRA2, and JRA55 (references below) and three ocean reanalysis data sets: ORAS4, GLORYS2V3, and SODA3 (references below). To avoid interpolation errors and imbalances in the mass budget introduced by regridding, the calculations are based on data from the original model grid. Note that the latest atmospheric reanalysis ERA5 from ECMWF is not included here since the model level data has not been opened to the public yet [ECMWF, 2017]. In addition, the computation is too expensive to achieve a longer time series for the study of the interannual variability of AMET using ERA5. As a synthesis, Table 2.1 shows the basic specifications of the reanalysis products contained in this study.

**Table 2.1:** Basic specification of reanalyses products included in this study

Type	Product Name	Producer	Period	Temporal Resolution	Spatial Resolution / Grid
Atmosphere	ERA-Interim	ECMWF	1979 - 2017	6-hourly	TL255, L60 up to 0.1 hPa
	MERRA2	NASA	1980 - 2017	3-hourly	0.5° x 0.625°, L72 up to 0.01 hPa
	JRA55	JMA	1979 - 2016	6-hourly	TL319, L60 up to 0.1hPa
Ocean	ORAS4	ECMWF	1979 - 2016	Monthly	ORCA1
	GLORYS2V3	Mercator-Ocean	1993 - 2014	Monthly	ORCA025
	SODA3	Univ. of Maryland	1980 - 2014	5-daily	MOM5

#### *ERA-Interim*

ERA-Interim is a global reanalysis dataset produced by the European Center for Medium Range Weather Forecasts (ECMWF) [Dee et al., 2011], which covers the data-rich period since 1979. It employs the cycle 31r2 of ECMWF's Integrated Forecast System (IFS) and generates atmospheric state estimates using 4D-Var data assimilation with a T255 ( $\sim 79\text{km}$ ) horizontal resolution on 60 vertical levels [Berrisford et al., 2009]. Compared with its predecessor, ERA-40 [Uppala et al., 2005], ERA-Interim is superior in quality in terms of the atmospheric properties like mass, moisture and energy [Berrisford et al., 2011]. The improvement in observations and the ability of 4D-Var contributes a lot to the quality of the divergent wind [Berrisford et al., 2011], which is significant for the mass budget and hence the energy budget. We use the data that is provided on a  $256 \times 512$  Gaussian grid, with a  $0.75^\circ \times 0.75^\circ$  horizontal resolution and 60 vertical hybrid model levels. We take 6-hourly data with a range from 1979 to 2016.

*MERRA2*

The Modern-Era Retrospective Analysis for Research and Applications version 2 [Gelaro et al., 2017], in short MERRA2, is the successor of MERRA from the Global Modeling and Assimilation Office (GMAO) of the National Aeronautics and Space Administration (NASA). It assimilates observational data with the Goddard Earth Observing System (GEOS) model and analysis scheme [Molod et al., 2015, Gelaro et al., 2017]. The atmospheric state estimates are produced by a 3D-Var Incremental Analysis Update (IAU) assimilation scheme and have coverage from 1980 until present. Unlike most of the reanalysis products, the GEOS atmospheric model includes a finite-volume dynamical core that uses a cube-sphere horizontal-discretization [Gelaro et al., 2017]. The model grid has a resolution of  $0.5^\circ \times 0.625^\circ$  with 72 hybrid levels. For this study, we use the 3-hourly assimilation data on the native model grid from 1980 to 2016.

*JRA55*

Extending back to 1958, Japanese 55-year reanalyses (JRA55) is the second reanalysis product made by the Japan Meteorological Agency (JMA) [Kobayashi et al., 2015, Harada et al., 2016]. JRA55 applies 4D-Var assimilation and it is generated on TL319 horizontal resolution with 60 hybrid levels. Before entering the satellite era in 1979, the assimilated upper air observations mainly come from radiosonde data. In this project we take 6-hourly data from 1979 to 2015 on the original model grid, which has a horizontal resolution of  $0.5625^\circ \times 0.5625^\circ$  with 60 hybrid model levels.

*ORAS4*

Serving as the historical reconstruction of the ocean's climate, the Ocean reanalyses System 4, in short ORAS4, is the replacement of its predecessor used by the ECMWF, the reanalyses system ORAS3 [Balmaseda et al., 2013]. It implements Nucleus for European Modelling of the Ocean (NEMO) as ocean model [Madec, 2008, Ferry et al., 2012a] and uses NEMOVAR as the data assimilation system [Mogensen et al., 2012]. The model is forced by atmosphere-derived daily surface fluxes, from ERA-40 from 1957 to 1989 and ERA-Interim from 1989 to 2010. Since 2010, the forcing is changed to operational forcing [Balmaseda et al., 2013]. ORAS4 produces analyses with a 3D-Var FGAT assimilation scheme and spans from 1958 to the present. ORAS4 runs on the ORCA1 grid, which is associated with a horizontal resolution of  $1^\circ$  in the extratropics and a refined meridional resolution up to  $0.3^\circ$  in the tropics. It has 42 vertical levels, 18 of which are located in the upper 200m. Here we skip the first two decades and use the monthly data from 1979 to 2014 to avoid the uncertainties reported by Balmaseda et al. [2013]. We use the monthly mean fields on the native model grid.

*GLORYS2V3*

GLORYS2V3, which is short for GLobal Ocean reanalyses and Simulations version 3, is a global ocean and sea-ice eddy permitting reanalysis system that yielded from the collaboration between the Mercator Ocean, the Drakkar consortium and Coriolis Data center [Ferry et al., 2010, 2012b]. It spans the altimeter and Argo eras, from 1993 until present. The NEMO ocean model is implemented on the ORCA025 grid (approximately  $0.25^\circ \times 0.25^\circ$  with 75 vertical levels). The



model is forced by a combination of ERA-Interim fluxes (e.g., shortwave radiation) and turbulent fluxes obtained with bulk formulae using ERA-Interim near-surface parameters. The data is generated by a 3D-Var assimilation scheme with temperature and salinity profiles assimilated from the CORA3.3 database [Ferry et al., 2012b]. In this study, monthly data from 1993 to 2014 on the original ORCA025 grid is used.

### *SODA3*

SODA3 is the latest version of Simple Ocean Data Assimilation (SODA) ocean reanalyses conducted mainly at the University of Maryland [Carton et al., 2018]. SODA3 is built on the Modular Ocean Model v5 (MOM5) ocean component of the Geophysical Fluid Dynamics Laboratory CM2.5 coupled model [Delworth et al., 2012] with a grid configuration of approximately  $0.25^\circ$  (latitude)  $\times$   $0.25^\circ$  (longitude)  $\times$  50 levels resolution [Carton et al., 2018]. To be consistent with the other two reanalysis data sets assessed in this study, the SODA 3.4.1 is chosen since it applies surface forcing from ERA-Interim. For this specific version, the 5-daily data is available from 1980 to 2015. Reanalysis data from this period on the original MOM5 grid is used in this case.

#### **2.2.2 Oceanic Observations and OGCM Hindcast**

For independent examination of the OMET calculated from reanalysis data sets, observations of the meridional transport of mass and heat throughout the Atlantic basin are used here. We use data from the RAPID-MOCHA-WBTS program [Johns et al., 2011, McCarthy et al., 2015] and the OSNAP program [Susan Lozier et al., 2017, Lozier et al., 2019]. The RAPID-MOCHA-WBTS program, which is known as the RAPID array, employs a transbasin observing array along  $26.5^\circ\text{N}$  and it is in operation since 2004. The OMET from the RAPID array available to this study is from April 2004 to March 2016. The OSNAP program has an observing system that comprises an integrated coast-to-coast array extending from the southeastern Labrador shelf to the southwestern tip of Greenland, and from the southeastern tip of Greenland to the Scottish shelf. So far, it provides OMET data from the full installation of the array in 2014 until the first complete data recovery in 2016, 21 months in total. Although it is too short to provide a good estimate of the interannual variability of OMET, we still include it as it is a unique observation system for OMET in the subpolar Atlantic.

Apart from the RAPID array and OSNAP observational data, a NEMO ORCA hindcast is also included here to provide more insights since two of the chosen reanalysis products are also built on the NEMO ocean circulation model [Moat et al., 2016, Marzocchi et al., 2015]. This forced model simulation implements the NEMO ORCA global ocean circulation model version 3.6 [Madec, 2008]. It is configured with the ORCA0083 grid, which has a nominal resolution of  $1/12^\circ$ , on 75 vertical levels. Climatological initial conditions for temperature and salinity were taken in January from PHC2.1 at high latitudes [Steele et al., 2001], MEDATLAS in the Mediterranean [Jourdan et al., 1998], and the rest from Levitus et al. [1998]. It is forced by the surface fields produced by the Drakkar project, which supplies surface air temperature, winds, humidity, surface radiative heat fluxes and precipitation, and a formulation that parameterizes the turbulent surface heat fluxes and is provided for the period 1958 to 2012 (dataset version

5.2) [Brodeau et al., 2010, Dussin et al., 2016]. More information about this hindcast is given by Moat et al. [2016]. We take monthly mean data from the hindcast, which spans from 1979 to 2012. For clarity, this hindcast will be referred to as the Oceanic General Circulation Model (OGCM) simulation in this paper.

### 2.2.3 Computation of Meridional Energy Transport

The methods for quantification of AMET and OMET with atmospheric and oceanic reanalyses are included in this section, respectively.

#### *Energy Budget in the Atmosphere*

The total energy per unit mass of air has four major components: internal energy ( $I$ ), latent heat ( $H$ ), geopotential energy ( $\phi$ ) and kinetic energy ( $k$ ). They are defined as:

$$\begin{aligned} I &= c_v T \\ H &= L_v q \\ \Phi &= gz \\ k &= \frac{1}{2} \mathbf{v} \cdot \mathbf{v} \end{aligned} \tag{2.1}$$

with  $c_v$  the specific heat capacity of dry air for constant volume ( $J/(kgK)$ ),  $T$  the absolute temperature ( $K$ ),  $L_v$  the specific heat of condensation ( $J/kg$ ),  $q$  the specific humidity  $kg/kg$ ,  $g$  the gravitational acceleration ( $kg/(ms^2)$ ),  $z$  the altitude ( $m$ ) and  $v$  the zonal/meridional wind velocity ( $m/s$ ). The northward propagation is positive. In addition, these four quantities can be divided into three groups: the dry static energy  $I + \phi$ , the moist static energy  $I + \phi + H$  and the kinetic energy  $k$ . A constant value of  $L_v = 2500 KJ/kg$  was used to compute the AMET with the atmosphere reanalysis data sets. In addition, recently improved formulations of energy budget equations proposed by Mayer et al. [2017] and Trenberth and Fasullo [2018] are addressed here. We use an updated formulation of AMET as a combination of the divergence of dry-air enthalpy, latent heat, geopotential and kinetic energy transports, which is suggested by Mayer et al. [2017]. Note that in this case the enthalpy transports associated with vapor fluxes are neglected.

In pressure coordinates, the total energy transport at a given latitude  $\Phi_i$  can be expressed as [Mayer et al., 2017]:

$$E = \oint_{\Phi=\Phi_i} \int_{p_s}^{p_t} [(1-q)c_p T + L_v q + gz + \frac{1}{2} \mathbf{v} \cdot \mathbf{v}] v \frac{dp}{g} dx \tag{2.2}$$

with  $c_p$  the specific heat capacity of dry air at constant pressure,  $p_t$  the pressure level at the top of the atmosphere ( $Pa$ ) and  $p_s$  the pressure at the surface ( $Pa$ ). A constant value of  $c_p = 1004.64 J/(kgK)$  was used. Since we work on the native hybrid model coordinate with

each atmosphere reanalysis product, the equation can be adjusted as follows (see Graversen [2006]):

$$E = \oint_{\Phi=\Phi_i} \frac{1}{g} \int_0^1 [(1-q)c_p T + L_v q + gz + \frac{1}{2} \mathbf{v} \cdot \mathbf{v}] v \frac{\partial p}{\partial \eta} d\eta dx \quad (2.3)$$

where  $\eta$  indicates the number of the hybrid level.

Unfortunately, a direct estimation of AMET based on the equations above cannot provide meaningful energy transports obtained from reanalysis data. It has been widely reported that reanalysis products suffer from mass inconsistency [Trenberth, 1991, Trenberth et al., 2002, Graversen, 2006, Graversen et al., 2007, Chiodo and Haimberger, 2010, Berrisford et al., 2011]. Spurious sinks and sources mainly come from low spatial and temporal resolution, interpolation and re-gridding, and data assimilation. The interpolation from the original model level to pressure level can introduce considerable errors to the mass budget [Trenberth et al., 2002]. Therefore we prevent interpolations onto the pressure levels and use data on the native model levels with a high temporal resolution. Trenberth [1991] provided a method to correct the mass budget through the use of the continuity equation. The method assumes that the mass imbalance mainly comes from the divergent wind fields and corrects the overall mass budget by adjusting the barotropic wind. The conservation of mass for a unit column of air can be represented as:

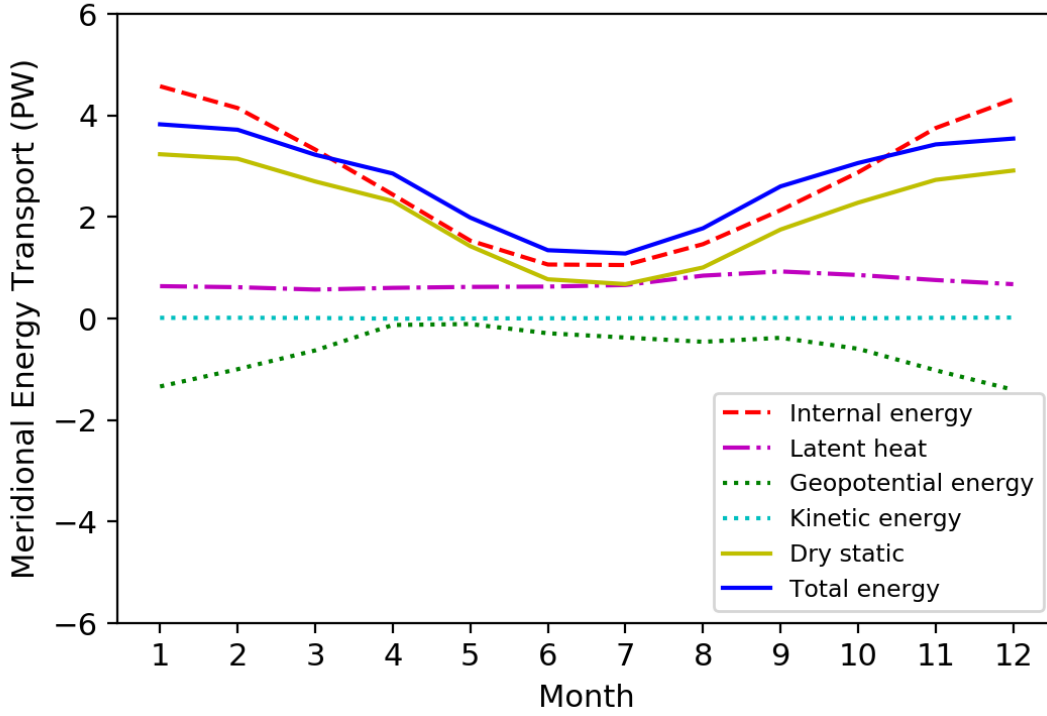
$$\frac{\partial p_s}{\partial t} + \nabla \cdot \int_{p_s}^{p_t} \mathbf{v} dp = g(E - P) \quad (2.4)$$

Where  $E$  stands for evaporation and  $P$  denotes precipitation. It has been noticed that big uncertainties reside in the evaporation and precipitation of global reanalyses [Graversen, 2006]. Hence we use the moisture budget to derive the net moisture change in the air column, according to:

$$E - P = \frac{\partial}{\partial t} \left( \int_{p_s}^{p_t} q \frac{dp}{g} \right) + \nabla \cdot \int_{p_s}^{p_t} (\mathbf{v} \cdot \mathbf{q}) \frac{dp}{g} \quad (2.5)$$

The related fields for the mass budget correction are surface pressure ( $p_s$ ), meridional and zonal winds ( $u, v$ ), and specific humidity ( $q$ ). After determining the mass budget imbalance, we correct the barotropic wind fields ( $u_c, v_c$ ), with  $u_c$  and  $v_c$  indicating the correction terms for zonal and meridional wind components as a result of the barotropic mass budget correction, and then calculate AMET [Trenberth, 1991]. Note that all the computations regarding barotropic mass budget correction were performed in the spectral domain via spherical harmonics. Figure 2.1 shows the mean AMET and each component in each month at 60°N estimated from ERA-Interim.

It is worth mentioning that MERRA2 is very different from ERA-Interim and JRA55, in terms of the discretization method and grid incorporated by the dynamical core. The dynamical core for MERRA2 is the GEOS-5 model and it computes all fields on a cubed-sphere grid with a resolution of  $50 \times 50 km$  [Gelaro et al., 2017], while in ERA-Interim and JRA55 the



**Figure 2.1:** Estimation of mean AMET and each component in each month at 60°N with ERA-Interim from 1979 to 2017.

computations were performed in the spectral domain. However, the data collections are saved only on the latitude-longitude grid after interpolation. Thus the data cannot be transferred back to the cubed-sphere grid without loss of information. Moreover, the vector field computations on the cubed-sphere grid are not divergence-free due to the implementation of finite volume discretization methods [Putman and Lin, 2007]. Consequently, we transferred MERRA2 fields to the spectral domain and performed vector field computations via spherical harmonics to minimize the numerical errors, the same treatment as ERA-Interim and JRA55.

#### *Energy Budget in the Ocean*

Unlike the atmosphere, energy transport in the ocean can be well represented by the internal energy itself. Consequently, the total energy transport in the ocean at a given latitude  $\phi_i$  can be expressed in terms of the temperature transport [Hall and Bryden, 1982]:

$$E = \oint_{\Phi=\Phi_i} \int_{z_b}^{z_0} \rho_0 c_{p_0} \theta \cdot v dz d\phi \quad (2.6)$$

where  $\rho_0$  is the seawater density ( $kg/m^3$ ),  $c_{p_0}$  is the specific heat capacity of seawater ( $J/(kg^\circ C)$ ),  $\theta$  is the potential temperature ( $^\circ C$ ),  $v$  is the meridional current velocity ( $m/s$ ),  $z_0$  and  $z_b$  are sea surface and the depth till the bottom (m), respectively. A constant value of  $c_{p_0} = 3987 J/(kg^\circ C)$

was used in all the calculations of OMET. Ocean heat content (OHC, with unit  $J$ ) is another variable that plays a role in the ocean heat budget. The total OHC between certain latitudes can be calculated by:

$$OHC = \int_{\Phi_i}^{\Phi_0} \int_{z_b}^{z_0} \rho_0 c_{p0} \theta dz d\phi \quad (2.7)$$

Our computation of OMET suffers from a small mass imbalance (e.g., mass imbalance coming from the difference between precipitation and evaporation [Mayer et al., 2017]). In the ocean, with its strong boundary circulations even the smallest imbalance can lead to large errors in the heat flux. However, the barotropic correction method adopted by the atmosphere is not feasible here due to the mass imbalance coming from the residual between precipitation and evaporation, and some budget terms that are hard to diagnose. In oceanographic literature it is common to use a reference temperature when calculating OMET in both observations and model diagnostics [Bryan, 1962, Hall and Bryden, 1982, Zheng and Giese, 2009]. Here, we also take a reference temperature  $\theta_r$  ( $^{\circ}C$ ). Note that the influence of taking a reference temperature on a zonally integrated transport is smaller than that on a single strait [Schauer and Beszczynska-Möller, 2009]. Then the quantification of OMET becomes:

$$E = \oint_{\Phi=\Phi_i} \int_{z_b}^{z_0} \rho_0 c_{p0} (\theta - \theta_r) \cdot v dz d\phi \quad (2.8)$$

Here, we take  $\theta_r$  equal to  $0^{\circ}C$ . Finally, operations in the zonal direction are different from their conventional meaning. As the three ocean reanalysis products used here are all built on a curvilinear grid, the zonal direction on the native model grid is curvilinear as well. Similar to the considerations made in Section 2.1, regridding from the native curvilinear grid to a uniform geographical grid will introduce large errors. So, we worked on the original multi-pole grid and followed a zig-zag setup when taking zonal integrals. The method is illustrated by Outten et al. [2018] in their Figure 2. After applying this method the resulting OMET values are comparable to those in earlier publications [Trenberth and Caron, 2001, Wunsch, 2005, Trenberth and Fasullo, 2008]. Note that we only have access to sub-monthly data for SODA3. The computation of OMET using monthly data in GLORYS2V3 could miss part of heat transport by eddies, while ORAS4 does not include the heat transport from the eddy parameterization scheme [Gent and McWilliams, 1990] as the related eddy-induced velocity field was not archived.

### 2.2.4 Statistical Analysis

In order to understand the connection between MET and changes in the Arctic, and to compare to the results from numerical climate models or single reanalysis dataset [Graversen, 2006, Van der Swaluw et al., 2007, Graversen et al., 2008, Jungclaus and Koenigk, 2010, Kapsch et al., 2013], in the following section we performed linear regressions on multiple fields with AMET and OMET. To test the significance of the regressions, we use the student's t-test. The

autocorrelations are taken into account. Note that all the reanalysis data sets included in this study have relatively short time series (no more than 456 months, see Table 2.1).

## 2.3 Results

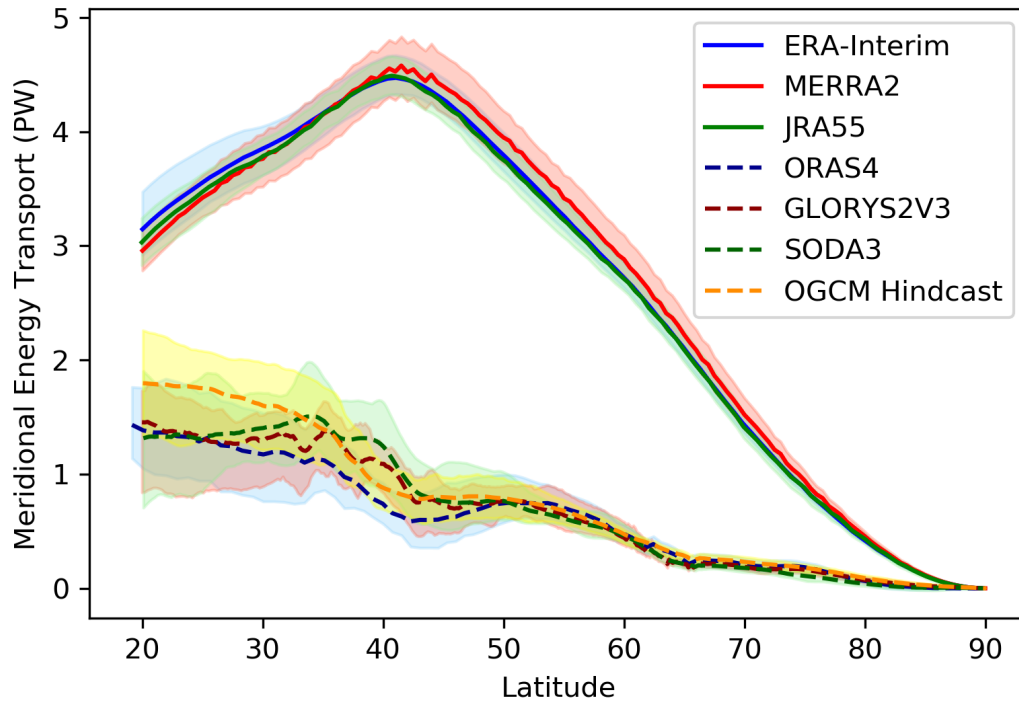
Unless specifically noted, the results shown in this section are all based on monthly mean fields with low pass filter of 5 years, which will be referred to as interannual time scales for the rest of the paper.

### 2.3.1 Overview of AMET and OMET

Globally, MET is driven by the unequal distribution of net solar radiation and thermal radiation. There are transports from regions with positive net TOA radiation to regions with negative net TOA radiation. Figure 2.2 shows the mean AMET and OMET over the entire time series of every product at each latitude in the Northern Hemisphere. For the atmosphere, all three data sets agree very well. The results differ a bit in amplitude but capture similar variations at each latitude. The peak of AMET is around  $41^{\circ}\text{N}$ , after which it starts to decrease towards the north pole. In ERA-Interim and JRA55, AMET peaks at 4.45 PW at  $41^{\circ}\text{N}$ , while in MERRA2 AMET peaks at 4.5 PW at  $41.5^{\circ}\text{N}$ . These findings are consistent with previous work [e.g. Trenberth and Caron, 2001, Fasullo and Trenberth, 2008, Mayer and Haimberger, 2012, and many others].

Apart from the climatology of MET, we are particularly interested in the variations across different time scales from midlatitudes towards the Arctic. The time series of AMET, integrated zonally over  $60^{\circ}\text{N}$ , are shown in Figure 2.3a. The seasonal cycle is dominant in each component, as expected, and the phase is very similar, but differences in the amplitudes are noted. The mean AMET provided by the chosen three atmospheric reanalysis data sets agrees well. However, their variations differ from each other. In ERA-Interim, the standard deviation (std) of AMET is 0.92 PW, while MERRA2 has a relatively large std of 0.97 PW, and in JRA55 the std is 0.91 PW. Hence, it can be concluded that the seasonal cycles of AMET presented by the chosen atmospheric reanalysis data sets are similar. After removing the seasonal cycles and applying a 5-year low pass filter, we obtain the low frequency signals of AMET anomalies at interannual time scales (see Figure 2.3b). ERA-Interim and JRA55 agree well, and the correlation coefficient between them is 0.82. MERRA2 provides a different result, and the correlation coefficient between ERA-Interim and MERRA2 is -0.53. The std of AMET anomaly in ERA-Interim is 0.02 PW, while in MERRA2 the std is 0.04 PW and in JRA55 the std is 0.03 PW. This implies that the variations of AMET anomalies at large time scales are similar in ERA-Interim and JRA55, but not in MERRA2. We further assess the sources of the difference in the next section.

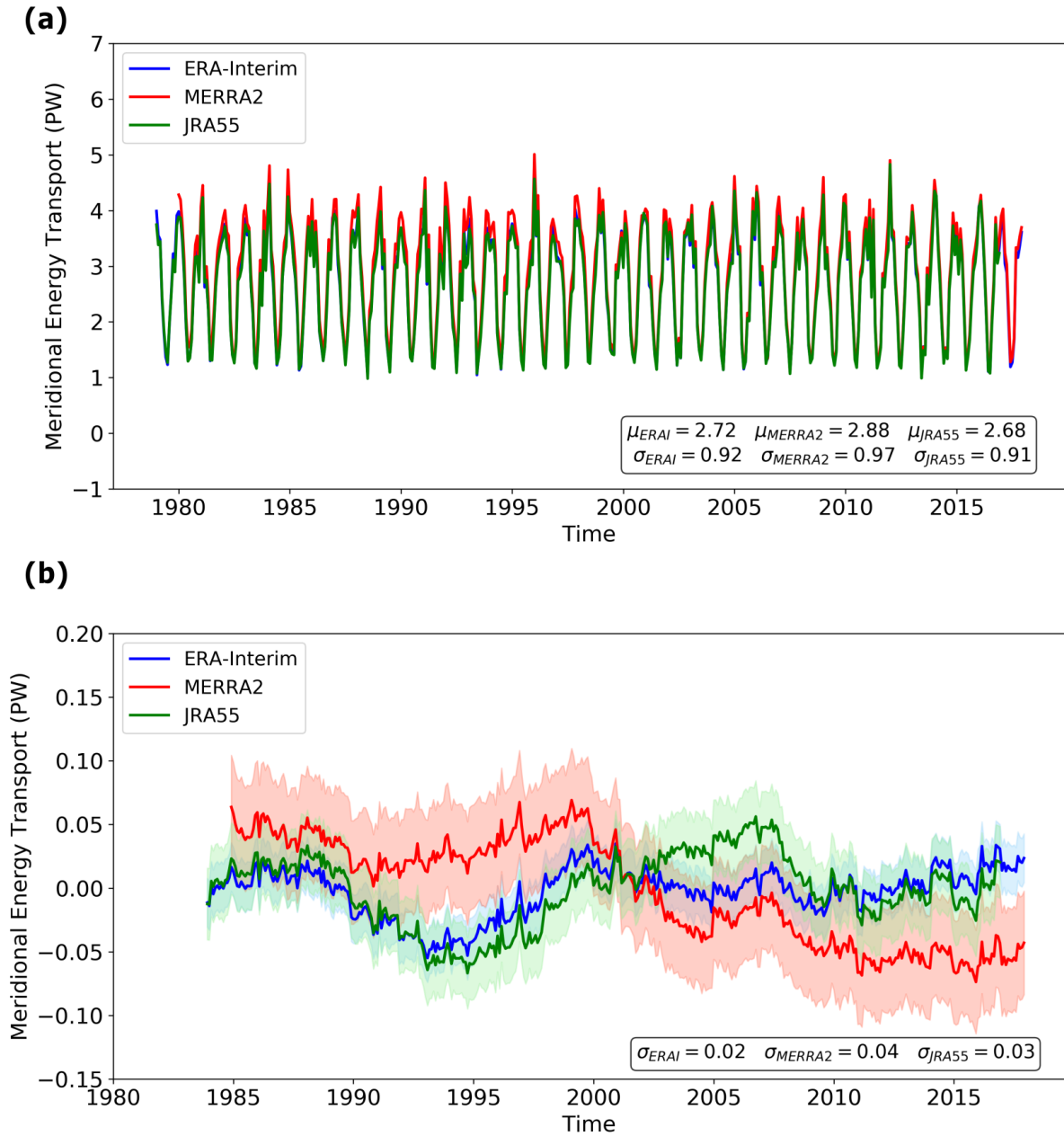
For the ocean, all the reanalysis data sets agree well at almost all the latitudes, except for the OMET between  $30^{\circ}\text{N}$  and  $40^{\circ}\text{N}$ , where the Gulf Stream resides (Figure 2.2). One possible explanation is that GLORYS2V3 and SODA3 both have been generated with eddy-permitting models while ORAS4 has not. In ORAS4, an eddy parameterization scheme from Gent and McWilliams [1990] is implemented. The implementation of this eddy parameterization scheme can lead to a big difference in heat transport, compared to eddy-permitted models [Stepanov and



**Figure 2.2:** Mean AMET and OMET over the entire time span of each product as function of latitude in the Northern Hemisphere. AMET are illustrated with solid lines while OMET with dash lines. The shades represent the full range of MET across the entire time series at each latitude. The time span of each product used in this study is given in Table 2.1.

Haines, 2014]. However, in this case the computation of OMET with ORAS4 does not include the contribution from eddy-induced velocity as the fields related to the use of eddy advection schemes were not archived. The eddy-permitting reanalysis data sets with high resolution, like GLORYS2V3 and SODA3, are capable of addressing the large scale geostrophic turbulence. It has been shown that their eddy-permitting capacity can account for the large scale eddy variability and represent the eddy energy associated with both the Gulf Stream and the Kuroshio pathways well [Masina et al., 2017]. Consequently, at the latitude of the Gulf Stream (between 30°N and 40°N), a strong spatial variability, which might represent more realistic patterns of the large scale eddy variability, is apparent in all data sets but ORAS4.

Similarly, we show the zonal integral of the OMET at 60°N in Figure 2.4. Differences in amplitudes and trends can be observed in the unfiltered time series. The mean and std of all the OMET time series are similar (see Figure 2.4a). The mean of OMET in ORAS4 is 0.47 PW, in GLORYS2V3 it is 0.44 PW and in SODA3 it is 0.46 PW. The OGCM hindcast gives a similar result, which is also 0.47 PW. The std of OMET in ORAS4 and the OGCM hindcast is 0.06 PW, while in GLORYS2V3 and SODA3 the std is 0.07 PW. The OMET anomalies with a 5-year low pass filter are shown in Figure 2.4b. OMET anomalies in ORAS4 resemble that in SODA3, especially after 1998. While OMET anomalies in GLORYS2V3 are very different from

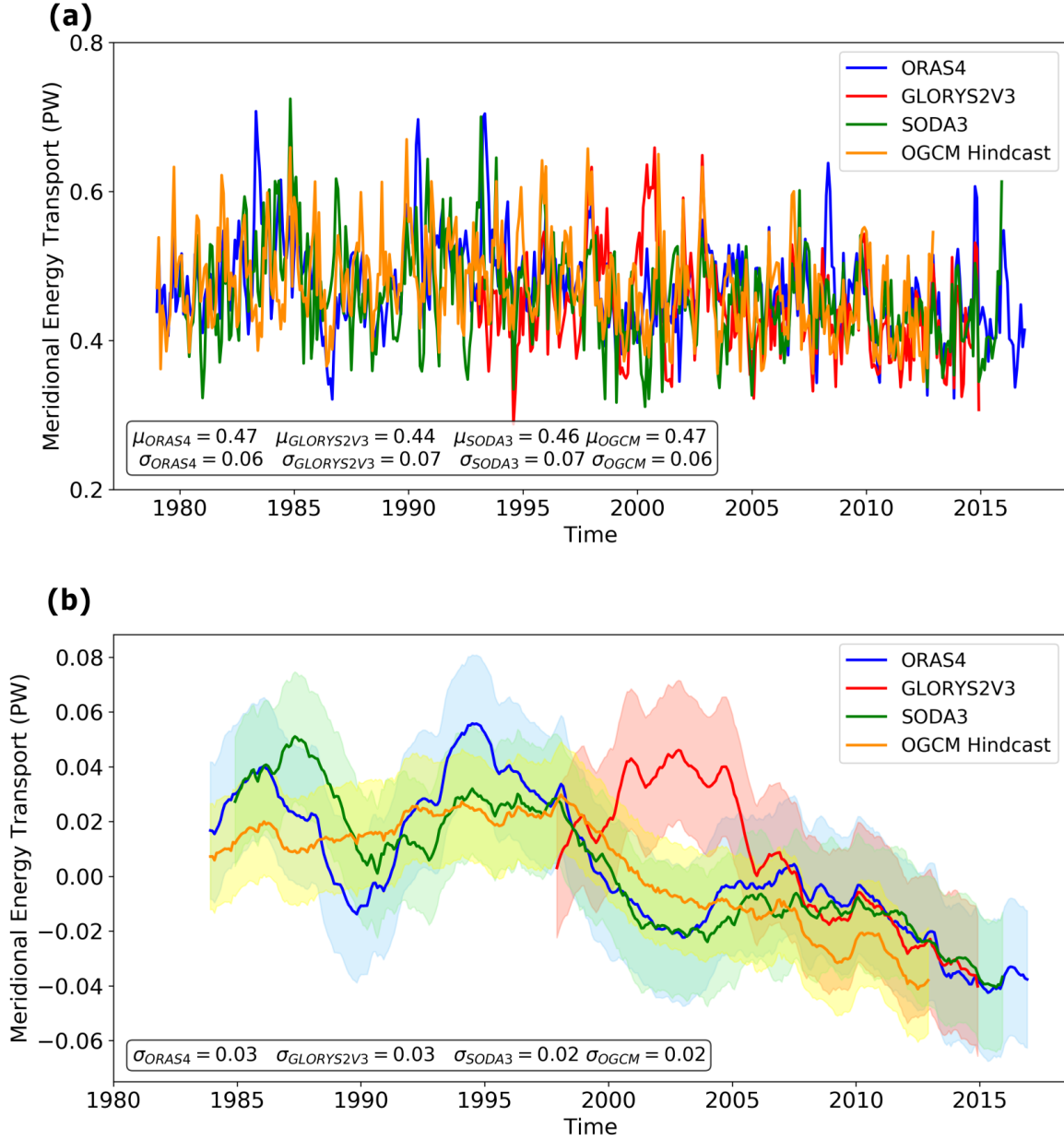


**Figure 2.3:** Time series of zonal integral of AMET at 60°N without/with low pass filter. (a) The original time series and (b) the ones with low pass filter include signals from ERA-Interim (blue), MERRA2 (red) and JRA55 (green). For the low pass filtered ones, we take a running mean of 5 years. The shades represent the confidence intervals with one standard deviation.  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

that in ORAS4 and SODA3 from 1998 to 2006. The differences reveal that the first 10 years in GLORYS2V3 are quite suspicious because of its large deviation from the other products. Such large differences should be noticeable in the heat content changes or surface fluxes. We find that OHC anomalies in GLORYS2V3 is indeed very different from ORAS4 and SODA3 during this period (see Figure 2.8). Nevertheless, after 2007 all the oceanic reanalyses agree well and the



OGCM hindcast deviates from the reanalyses. It is noteworthy that the observations improve considerably around that period due to an increasing number of Argo floats in use [Riser et al., 2016]. The reanalysis products used here are greatly influenced by the number of available in-situ observations. We further assess the sources of differences in the next section.



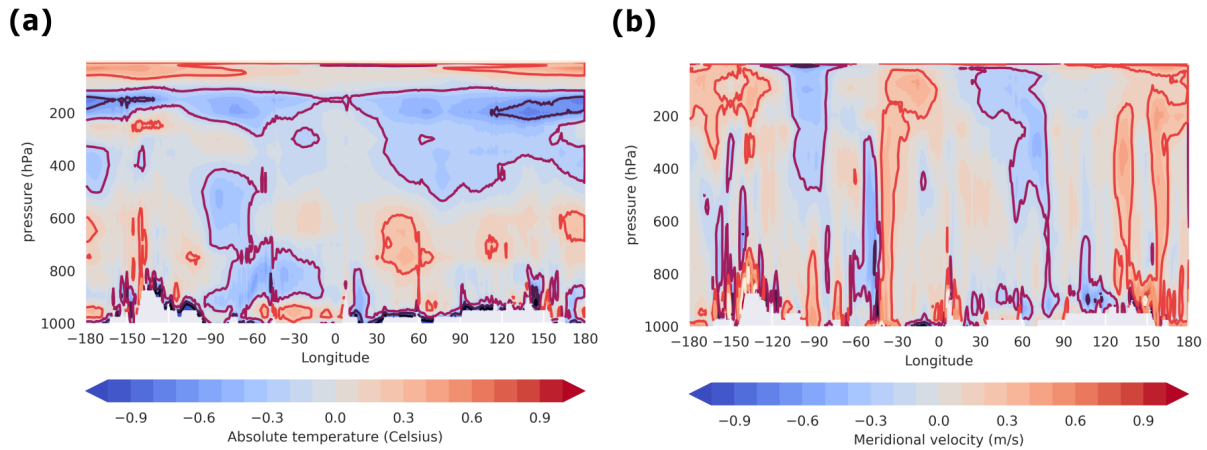
**Figure 2.4:** Time series of zonal integral of OMET at 60°N without/with low pass filter. (a) The original time series and (b) the ones with low pass filter include signals from ORAS4 (blue), GLORYS2v3 (red), SODA3 (green) and the OGCM hindcast (yellow). For the low pass filtered ones, we take a running mean of 5 years. The shades represent the confidence intervals with one standard deviation.  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

### 2.3.2 Sources of Disparity

In order to further understand the difference between the AMET estimated from each atmosphere reanalysis product, we compare each component of AMET separately. We investigate the difference between each component of AMET at 60°N estimated from ERA-Interim against those from MERRA2 and JRA55. It is noticed that the differences mainly originate from meridional temperature transport ( $vc_pT$ ) and geopotential energy transport ( $vgz$ ). We find that the correlation between the difference in total energy transport and the difference in meridional temperature transport between ERA-interim and MERRA2 is 0.55, while between ERA-Interim and JRA55 that is 0.21. In addition, the correlation between the difference in total energy transport and the difference in geopotential energy transport ( $vgz$ ) between ERA-Interim and MERRA2 is 0.56, while between ERA-Interim and JRA55 that is 0.60. For the other components, the correlations between them and the total difference are small. The results are all obtained with a confidence interval of 95%. Large differences in temperature transport among reanalysis products are found at almost all latitudes (not shown). Such differences are consistent with the fact that the temperature transport and geopotential energy transport have a large contribution to the total AMET (see Figure 2.1). Note that the differences in each AMET component are of the same order of magnitude as AMET. Besides, the mean and anomalous latent heat transport agree well between the chosen atmospheric products (not shown). A similar result was found by Dufour et al. [2016] in their study using more reanalysis data sets.

In order to know the relative contribution of each field to the difference of the mean total AMET among the chosen reanalyses, a direct comparison of the vertical profile of temperature and meridional velocity fields between ERA-Interim and MERRA2 is presented in Figure 2.5. We compare the monthly mean temperature and velocity fields of ERA-Interim and MERRA2 from 1994 to 1998, in which the biggest difference was observed (Figure 2.3, taking into account the running mean of 5 years). To accommodate a point-wise comparison, the fields from MERRA2 are interpolated onto the vertical grid of ERA-Interim. It shows that these two reanalysis products differ substantially regarding each variable field (Figure 2.5a and b). Big differences in temperature reside mostly at the tropopause. Large differences in meridional wind components are distributed over the entire vertical column of the tropopause. Such differences in both fields are expected to be responsible for the difference in mean temperature transports ( $vc_pT$ ). Large differences are found in geopotential height fields, too (not shown). It should be noted that this comparison is carried out on pressure levels and mass conservation is not ensured. Therefore it can only provide insight qualitatively, and a quantitative contribution of the difference in every single field to the mean temperature transports can not be identified here.

Differences between every two chosen atmospheric products are found at nearly each pressure level. This analysis is not sufficient to explain conclusively where the uncertainty mainly comes from in terms of the dynamics and physics in the atmosphere model and data assimilation system. We do find that uncertainties, as indicated by the spread between the data sets, in both the temperature and meridional velocity fields, are too large to constrain the AMET. Note that the difference in horizontal advection schemes can also influence the results. The chosen atmospheric reanalyses systems use Semi-Lagrangian advection schemes, but this is not the case for MERRA2. Hence studies on low frequency variability of energy transports and associated

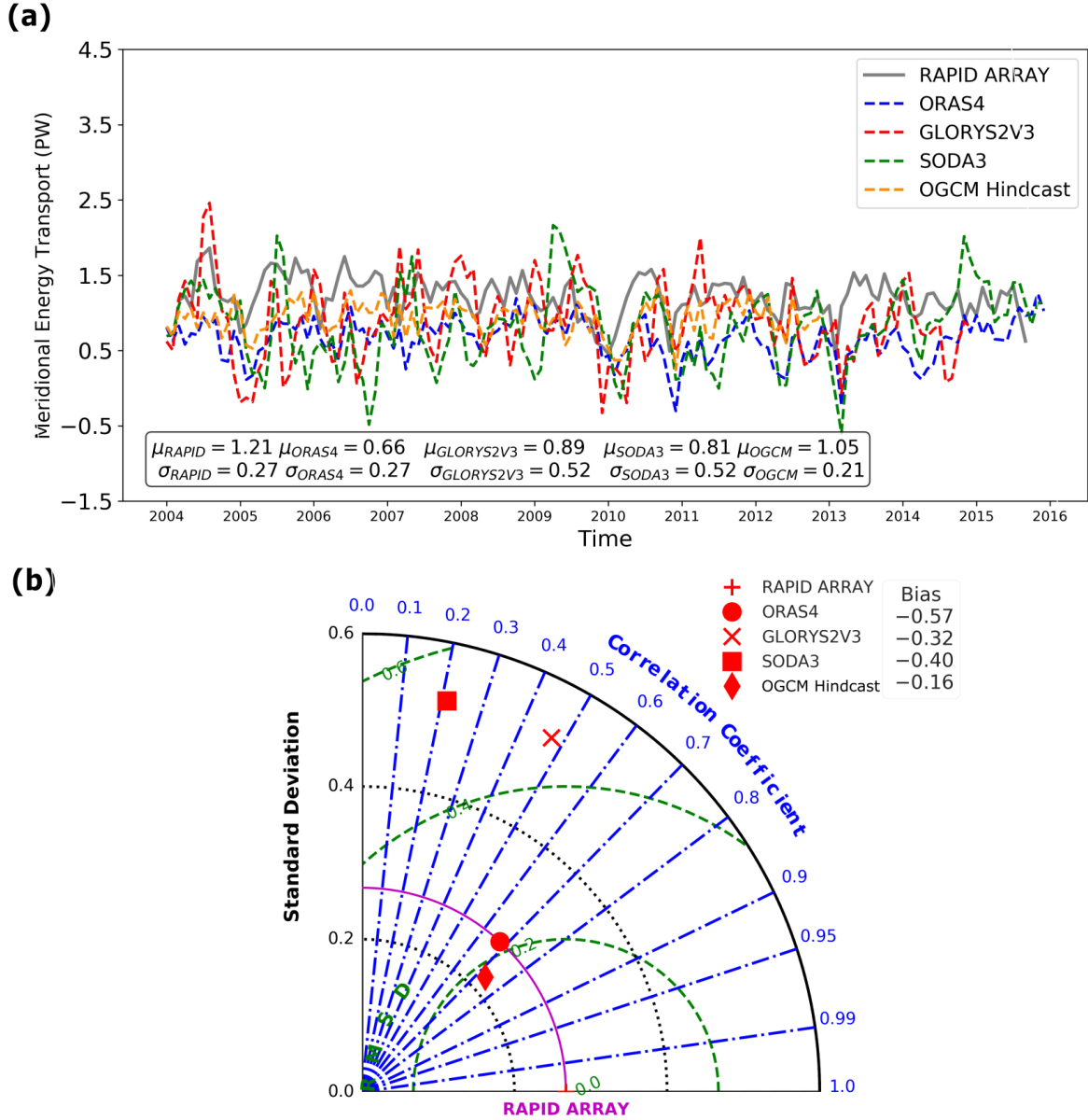


**Figure 2.5:** Difference in temperature, meridional wind velocity and temperature transport between MERRA2 and ERA-Interim at 60°N. The vertical profile of (a) temperature difference and (b) meridional wind velocity difference are calculated from the climatology of each fields from 1994 to 1998, respectively.

variables should be interpreted with care as the reanalysis products differ substantially, and we cannot judge a priori how close they are to actual energy transports since independent direct observations are not available.

For the ocean, fortunately observations of OMET in the Atlantic Ocean are available. First, OMET estimated from ORAS4, GLORYS2V3, SODA3 and the OGCM hindcast is evaluated against OMET measured at 26.5°N. The intercomparison shows that the reanalysis products capture roughly the mean amplitude of the OMET (Figure 2.6). Some large events are captured as well, such as the strong weakening in 2009. Statistically, the mean OMET provided by the RAPID array is  $1.21 \pm 0.27 PW$ . It is higher than the chosen products here. The mean OMET in ORAS4 is  $0.66 \pm 0.27 PW$ , in GLORYS2V3 it is  $0.89 \pm 0.52 PW$ , in SODA3 it is  $0.81 \pm 0.52 PW$  and in OGCM hindcast it is  $1.05 \pm 0.21 PW$ . This means that all chosen products underestimate the mean OMET at 26.5°N in the Atlantic basin. Of all products, ORAS4 has the largest bias. The std of OMET given by ORAS4 is the same as that from the RAPID array, while in GLORYS2V3 and SODA3 we find a higher std of OMET. The OGCM hindcast has a relatively small OMET std, which is 0.21 PW. In terms of the correlation and standard deviation, ORAS4 and the OGCM hindcast agree well with observations. It is noteworthy that the OGCM does not assimilate ocean data. The simulation is only constrained by the surface fluxes and this suggests that the surface forcing is a very important driver of OMET variability. To conclude, the heat transport at 26.5°N is too low in these products, but ORAS4 and OGCM hindcast appear to have reasonable variability.

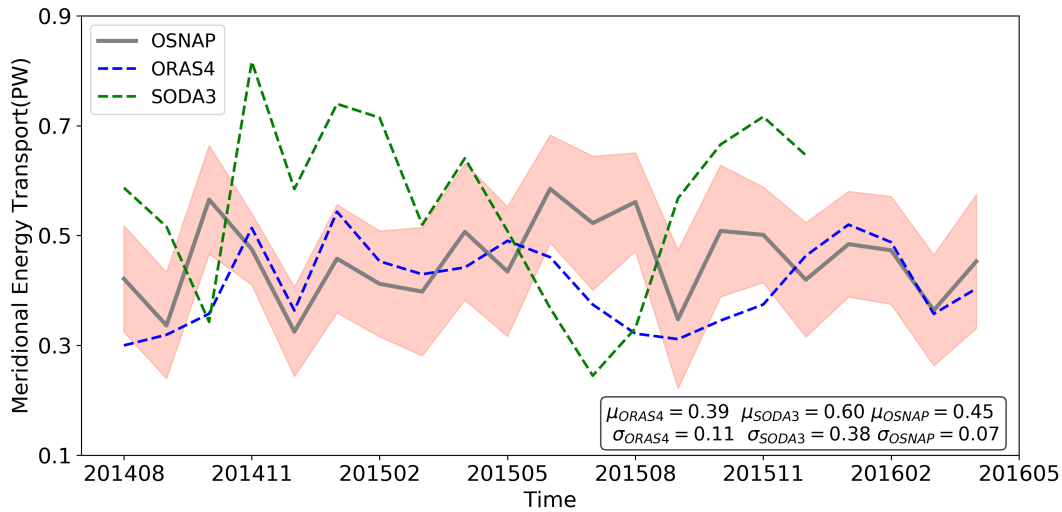
Moreover, the comparison of time series in the chosen reanalyses and OSNAP observations is given in Figure 2.7. Due to the limited length of OMET time series, only ORAS4 and SODA3 are included in the comparison. It can be noticed that the OMET given by ORAS4 is comparable to that in OSNAP in terms of the amplitude and variability. For most of the time within the observation period, OMET in ORAS4 falls into the range of the OSNAP observation including the uncertainty margins. The mean of OMET in ORAS4 is  $0.39 \pm 0.11 PW$ , which is quite similar



**Figure 2.6:** OMET estimated from ORAS4 (blue), GLORYS2V3 (red), SODA3 (green) and the OGCM hindcast (orange) compared to the RAPID array observation (gray) at 26.5°N across the Atlantic basin. The time series of OMET is presented in (a). The statistical properties are shown in (b) Taylor Diagram, including bias, correlation (blue), standard deviation (black) and root mean square deviation (green).  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

to the mean OMET  $0.45 \pm 0.07 \text{PW}$  of OSNAP. However, OMET in SODA3 has a larger mean and standard deviation than that in OSNAP and thus deviates from the observations.

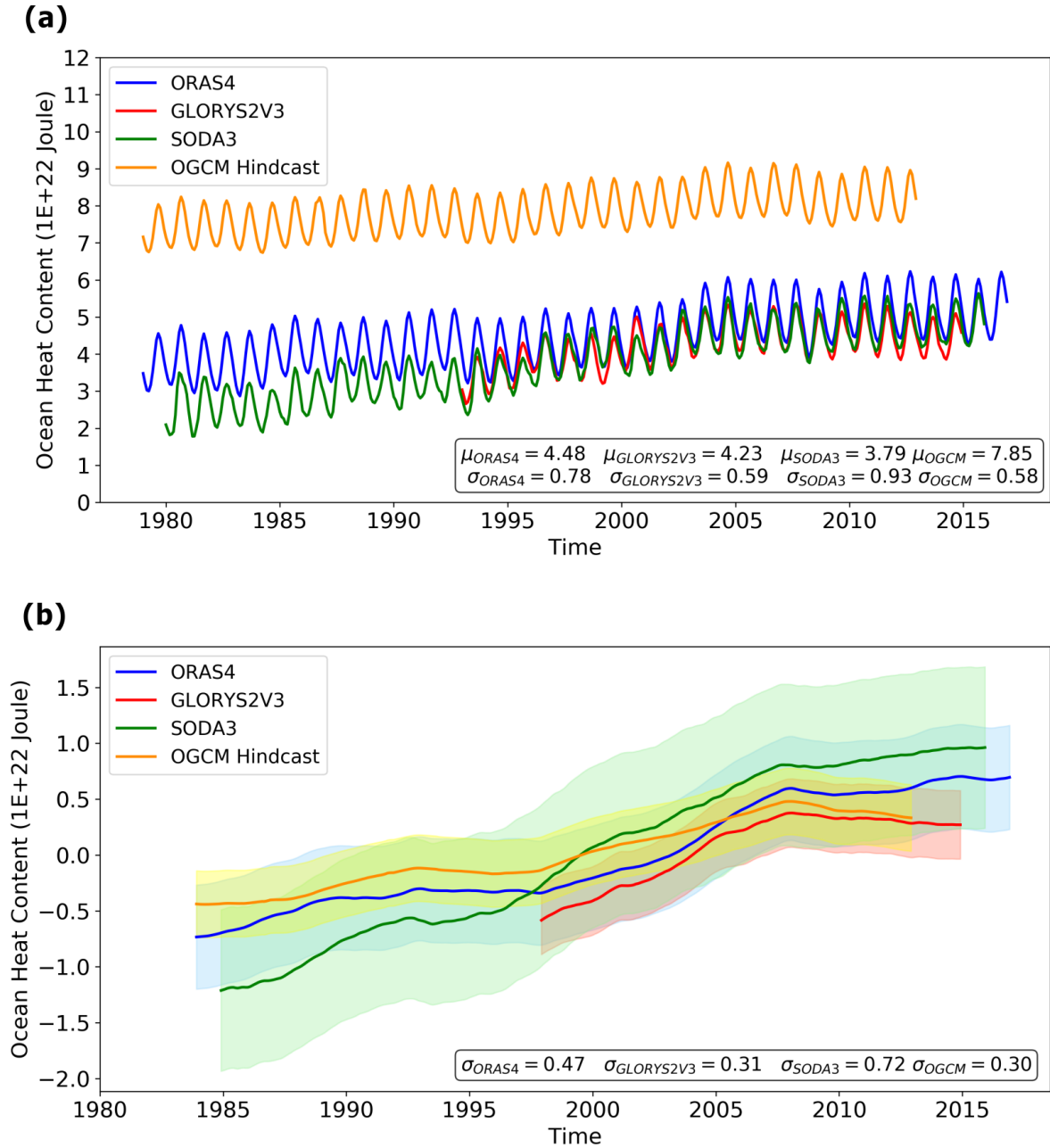
Just as in the atmosphere, we would like to study the temperature and meridional current velocity contributions to the ocean heat transport to identify the sources of the difference between products. However, due to the nature of the curvilinear grid, the comparison of local fields after interpolation is not trustworthy. To get further insight, we calculate the ocean heat content



**Figure 2.7:** OMET estimated from ORAS4 (blue), SODA3 (green) and compared to the OSNAP observation (gray) at subpolar Atlantic basin. The range of uncertainty from OSNAP observation is marked by the red shade.  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

(OHC), since the convergence of the heat transports is likely related to OHC change. A full budget analysis was not feasible as most data sets did not include the surface fluxes. Figure 2.8 illustrates the OHC (Figure 2.8a) and OHC anomalies (Figure 2.8b) quantified from ORAS4, GLORYS2V3, SODA3 and the OGCM hindcast. It depicts the OHC integrated in the polar cap (from  $60^{\circ}\text{N}$  to  $90^{\circ}\text{N}$ ) over all depths. The mean OHC in ORAS4 is  $4.48 \pm 0.78 * 10^{22} J$ , in GLORYS2V3 it is  $4.23 \pm 0.59 * 10^{22} J$  and in SODA3 it is  $3.79 \pm 0.93 * 10^{22} J$ , while the OGCM hindcast shows a much larger mean OHC of  $7.85 \pm 0.58 * 10^{22} J$ . Actually, we found that the OHC between  $60^{\circ}\text{N}$  to  $70^{\circ}\text{N}$  in the OGCM hindcast agrees well with the reanalyses (not shown). Thus the difference seems to be associated with changes to the sea ice distribution. Given the limited observations in the Arctic to constrain the reanalyses and different assumptions about the Arctic sea ice due to the differences in spatial resolutions, it is more complex than just concluding that the reanalyses are better. The variations of OHC are similar between chosen products. Regarding the OHC anomalies in Figure 2.8b, a positive trend of OHC anomalies in the polar cap is captured by each product. However, the variability is different and these are reflected in the standard deviation of OHC anomalies time series. Increases in surface temperature and OHC are often taken as a sign of AA in many papers [e.g. Serreze and Barry, 2011]. Qualitatively, the trends of OHC in the chosen reanalyses at the polar cap could be taken as a sign of the AA, but it might be just Arctic warming and not necessarily a higher warming rate than the global mean temperature change. A quantitative evaluation of the AA is not possible due to large differences between products. To conclude, there are large differences in OHC between chosen products, while their variations agree relatively well. Since OHC is a function of temperature fields only, this can imply that temperature profiles are different among the chosen ocean reanalysis data sets. The differences of OHC between chosen products are partially consistent with the differences that we found for OMET. However, the OHC anomalies

agree better among reanalysis products than the absolute OHC, which indicates that the trend of OHC is captured in a similar way among all the ocean reanalysis products.



**Figure 2.8:** Time series of (a) ocean heat content (OHC) and (b) OHC anomalies with a low pass filter at the polar cap. The OHC is integrated from surface to the bottom between 60°N and 90°N. It is estimated from ORAS4 (blue), GLORYS2V3 (red), SODA3 (green) and the OGCM hindcast (yellow). The shades represent the confidence intervals with one standard deviation.  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

### 2.3.3 MET and the Arctic

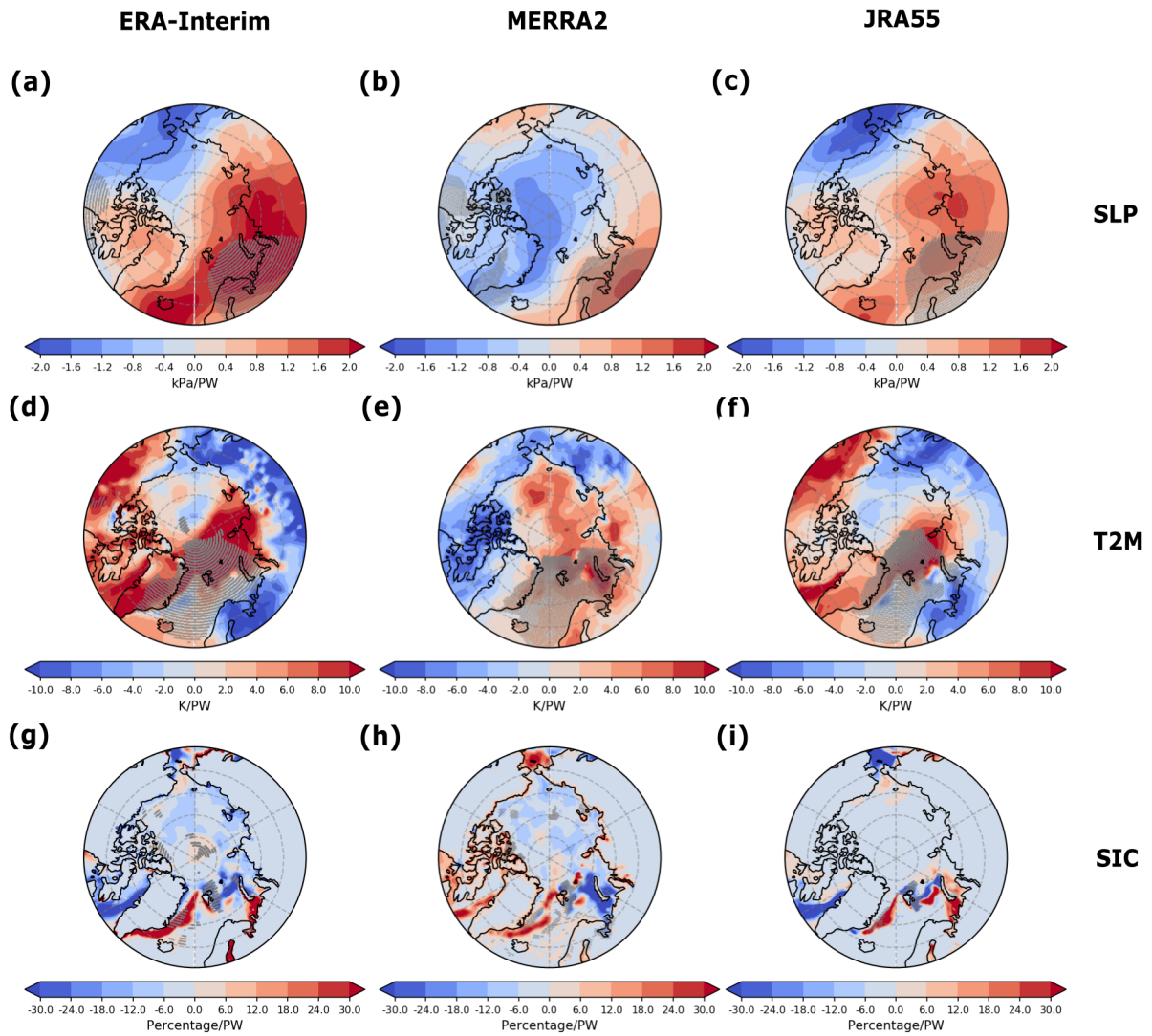
In previous sections, it is found that MET in different reanalysis products at subpolar and subtropical latitudes differ substantially from each other. In order to further evaluate AMET and OMET given by different reanalyses and to provide more insight, we investigate the links between MET and remote regions. We focus on the Arctic because previous studies indicate a strong role for subpolar MET in low frequency variability in the Arctic region. Given the complexity of the interaction between MET and the Arctic, and the short time series available, determining cause-effect relations is out of the scope for this paper. We aim to compare the relation between MET and the Arctic within each reanalysis product to investigate the physical plausibility and compare it with previous studies that use data from one reanalysis product or from coupled climate models [e.g. Graversen, 2006, Van der Swaluw et al., 2007, Graversen et al., 2008, Jungclaus and Koenigk, 2010, Kapsch et al., 2013].

Many of these studies perform linear regressions between a time series of MET and gridpoint values of other physical variables. Here we follow the same procedure and perform linear regressions of sea level pressure (SLP), 2 meter temperature (T2M) and sea ice concentration (SIC) anomalies on AMET and OMET anomalies at  $60^\circ\text{N}$  for the chosen products. We show linear regressions in summer and winter separately in order to account for the seasonal variability of the relationships. It should also be noted that there are strong trends in OMET, T2M and SIC. We removed them by applying a polynomial fit to the time series on each grid point. We find that the second order polynomial fit is able to capture the trend without losing variations at interannual time scales. Hereafter we only address detrended OMET, T2M and SIC. For the sake of consistency, the regressions are carried out on the surface fields included in each respective reanalysis product. For instance, the regression of SLP on AMET estimated from ERA-Interim, involves SLP fields from ERA-Interim itself. For the ocean reanalyses, as they all apply forcing derived from ERA-Interim, the regressions are performed on the fields from ERA-Interim. Note that there is a known issue with the quality of sea ice field close to the north pole in ERA-Interim, which can be inferred from an evaluation of reanalysis data sets concerning near surface fields in Lindsay et al. [2014]. Following the regressions performed by Van der Swaluw et al. [2007] and Jungclaus and Koenigk [2010], we repeated the same procedure here with AMET at interannual scales ( $\sim 5$  year).

First, we investigate the links between MET and the Arctic in winter. The regressions of anomalies of multiple fields on AMET anomalies at  $60^\circ\text{N}$  in each atmospheric product in winter are shown in Figure 2.9. The regression coefficients reach maximum when the regressions are instantaneous with given fields. In ERA-Interim and JRA55, AMET is correlated with SLP over the Greenland, the North Atlantic, the Barents Sea, the Kara Sea and the northern part of the Eurasian continent. It suggests that an increase in subpolar AMET is linked to a northward advection over the Greenland which could bring relatively warm and humid air into the Arctic. Such patterns are consistent with the relatively warm air over the Greenland and part of the Central Arctic close to the Eurasian side shown in Figure 2.9d and f. Using ERA-40, Graversen [2006] found similar correlation between AMET and surface air temperature (SAT) at the Greenland Sea and Barents Sea as Figure 2.9d and f, without time lag. This is also consistent with a model study by Jungclaus and Koenigk [2010]. The decrease of sea ice concentration with



increasing AMET at the Baffin Bay and the northern part of Barents Sea given by Figure 2.9g and i is consistent with the relations between AMET and T2M. A further eddy decomposition of AMET following the method from Peixoto and Oort [1992] indicates that heat transported by standing eddies has the biggest contribution to the total AMET (not shown), which is consistent with Graversen and Burtu [2016]. These patterns are found only in ERA-Interim and JRA55, but not in MERRA2. Such different patterns in MERRA2 likely stem from its shift in AMET around 2000. Hence, there is also large uncertainty in the assertion that heat and humidity transport by stationary eddies contribute to the changes in the subpolar and Arctic regions at interannual time scales.



**Figure 2.9:** Regressions of sea level pressure, 2 meter temperature and sea ice concentration anomalies on AMET anomalies at 60°N in winter (DJF) at interannual time scales with no time lag. The monthly mean fields are used here after taking a running mean of 5 year. Both the 2 meter temperature and sea ice concentration are detrended. From left to right, they are the regressions on AMET of (a, d, g) ERA-Interim, (b, e, h) MERRA2 and (c, f, i) JRA55. The stippling indicates a significance level of 95%.



Moreover, similar to Van der Swaluw et al. [2007] and Jungclaus and Koenigk [2010], we investigate the links between the variability of OMET and variations of multiple fields at interannual time scales. The regressions of anomalies of multiple fields on detrended OMET anomalies at  $60^{\circ}\text{N}$  in winter are shown in Figure 2.10 with OMET leading by 1 month. It is noteworthy that it takes around 1 to 2 years for the OMET anomalies to propagate from the North Atlantic to the Barents Sea [Årthun et al., 2012]. However, the regression coefficients are maximal when the OMET leads by 1 month, which could be attributed to the implementation of the low-pass filter. In ORAS4 and SODA3, increasing OMET can lead to a decrease in SLP in the Arctic, while in ORAS4 this polar-low is much stronger. This seems to indicate that an increase in OMET is related to sea ice melt and increase in T2M around the Nordic seas. There is an AO/NAO like SLP anomaly with the associated large scale temperature pattern. However, GLORYS2V3 tells an entirely different story. This is mainly due to the difference between OMET in this dataset compared to the other ocean data sets during the 1990s as shown in Figure 2.4.

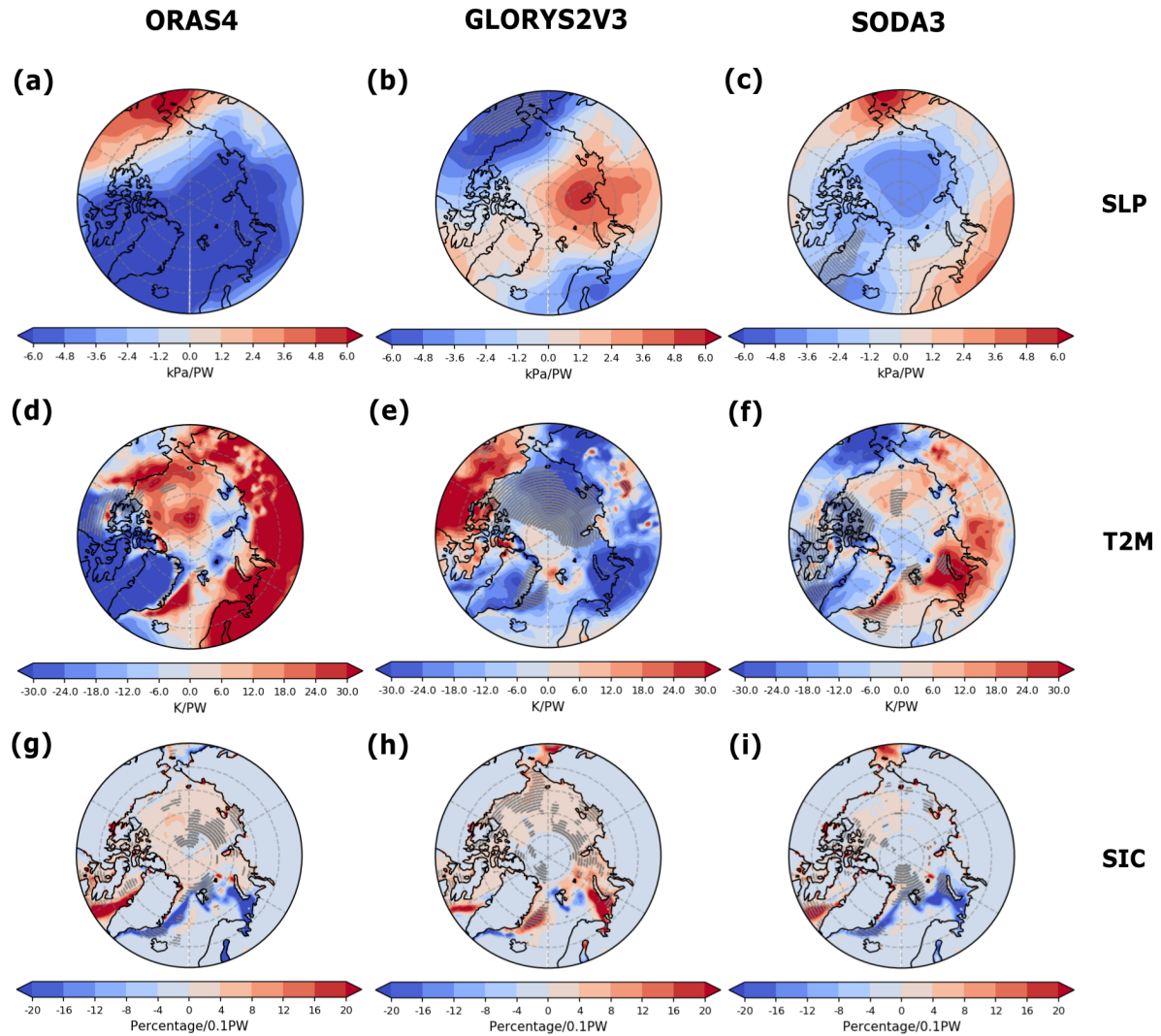
In general, the decrease of OMET leads to an increase in the growth rate of SIC, which is consistent with studies performed with global climate models at decadal to inter-decadal time scales [e.g. Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, van der Linden et al., 2016]. Studies with observations of sea ice at the Barents Sea and OMET across Barents Sea Opening (BSO) also confirm the strong correlation between the OMET and sea ice variation over the Barents Sea [Årthun et al., 2012, Onarheim et al., 2015]. However, note that some discussed regions are below the significance of 95%.

In summer, the situation becomes more intricate and unclear. The same regressions of anomalies of multiple fields on AMET and OMET anomalies at  $60^{\circ}\text{N}$  in each reanalysis product in summer were performed (not shown). It is noticed that the consistency of associations between AMET, OMET and multiple fields is better in winter than that in summer within the chosen products. Atmospheric dynamical processes are more dominant in winter, which is also reflected in large scale patterns of variability such as the AO and NAO which are more pronounced in winter than in summer [e.g. Lian and Cess, 1977, Curry et al., 1995, Goosse et al., 2018]. Therefore the regressions of SLP, T2M and SIC on AMET in winter are easier to understand than those in summer.

In this section we compared the reanalysis data with findings from previous studies. We found that ERA-Interim and JRA55 are most consistent with the results given by coupled numerical models in winter, while MERRA2 does not corroborate model studies. For the ocean, results from ORAS4 and SODA3 are more consistent with literature in winter. However, given the low statistical significance and the difference among chosen products, it is still hard to determine which atmospheric product provides a more convincing plausible interannual variations in AMET.

## 2.4 Discussion

In this study, we found substantial differences between reanalysis products with respect to MET. In order to improve the accuracy of the variability of AMET and OMET estimated from reanalyses, one needs more observations to constrain the models. Vertical profiles differ



**Figure 2.10:** Regressions of sea level pressure, 2 meter temperature and sea ice concentration anomalies on OMET anomalies at 60°N in winter (DJF) at interannual time scales. OMET leads the fields by one month. The 2 meter temperature, sea ice concentration and OMET are detrended. From left to right, they are the regressions on OMET of (a, d, g) ORAS4, (b, e, h) GLORYS2V3 and (c, f, i) SODA3. The stippling indicates a significance level of 95%.

substantially between products, and surface and top of the atmosphere radiation budget are too uncertain to constrain variability in the different products. Note that reanalyses do not assimilate direct observation of the TOA and surface energy budgets. Climate models already provide information on the interaction between atmosphere and ocean and connections provided by the energy transport from mid to high latitudes [Shaffrey and Sutton, 2006, Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010]. This can potentially sketch the mechanism of the interaction between energy transport and the Arctic climate change. Moreover, some studies point out that the latent heat is more influential on the Arctic sea ice rather than the dry static energy [Kapsch et al., 2013, Graverson and Burtu, 2016]. With improved reanalysis

products and independent observations, such as ocean mooring arrays and atmospheric in-situ and satellite observations, to validate the reanalyses, the validity of these mechanisms can be further studied.

The regression of SIC on OMET suggests that sea ice variations are sensitive to changes of meridional energy transport at subpolar latitudes, which is noticed by other studies on SIC and MET as well [Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, van der Linden et al., 2016]. ORAS4 and SODA3 show a large anticorrelation between SIC and OMET in winter around the Greenland Sea and the Barents Sea. However, GLORYS2V3 does not show this relation. The differences in OMET are reflected in the regressions on sea ice. The strong connection between OMET from mid-to-high latitudes and the Arctic sea ice indicates an indirect link between midlatitudes and the Arctic. Many studies that explored these remote links found large scale “horseshoe” and dipole patterns over the Atlantic [Czaja and Frankignoul, 2002, Gastineau and Frankignoul, 2015, Delworth et al., 2017]. However, the physical mechanism remains disputable. Overland et al. [2015] and Overland [2016] propose that the multiple linkages between the Arctic and midlatitudes are based on the amplification of existing jet stream wave patterns, which might also be driven by tropical and midlatitudes SST anomalies [Screen and Francis, 2016, Svendsen et al., 2018]. Cohen et al. [2014] lists possible pathways for the teleconnection between the Arctic and midlatitudes, including changes in storm tracks, the jet stream, and planetary waves and their associated energy propagation. However, due to the shortness of time series, a small signal-to-noise ratio, uncertain external forcing, and the internal atmospheric variability [Overland, 2016, Barnes and Screen, 2015], this question has no easy answer.

Previous studies have shown that the variations of total OMET are very sensitive to the changes of its overturning component [e.g. McCarthy et al., 2015, Lozier et al., 2019]. Hence, the AMOC may serve as an indicator of the changes of OMET. In our case, a quantitative estimation of the difference in the AMOC among the chosen data sets is beyond the scope of this paper. However, the downward trend of AMOC, which has been reported by several studies [Smeed et al., 2014, McCarthy et al., 2015, Oltmanns et al., 2018], is consistent the downward trend observed in OMET at 60°N in our chosen oceanic reanalyses (see Figure 2.4). After analyzing six oceanic reanalysis data sets, Karspeck et al. [2017] find the reanalysis products are not consistent in their year-to-year AMOC variations. The discrepancy between AMOC represented by each reanalysis product may explain the differences in OMET in each reanalysis dataset.

## 2.5 Conclusions

This study aimed to quantify and intercompare AMET and OMET variability from 3 atmospheric and 3 oceanic reanalysis data sets at subpolar latitudes. It also serves to illustrate the relation between AMET and OMET with high latitude climate characteristics. The study is motivated by previous studies with coupled models that show a strong relation between meridional energy transport and sea ice. It is also motivated by previous studies with reanalysis data, where generally only one reanalysis data set is considered, and which includes mostly only oceanic or atmospheric analysis.

All selected data sets agree on the mean AMET and OMET in the Northern Hemisphere. The

results are consistent with those achieved over the previous 20 years [Trenberth and Caron, 2001, Fasullo and Trenberth, 2008, Mayer and Haimberger, 2012]. However, when it comes to anomalies at interannual time scales, they differ from each other both spatially and temporally. The variations between ERA-Interim and JRA55 are small, while MERRA2 is very different from them. Although there is an overlap of observational data assimilated by different reanalysis products, large deviations still exist in many fields, especially for the vertical profiles of temperature and velocity in atmospheric reanalyses, which were also reported by some reanalysis quality reports [Simmons et al., 2014, 2017, Uotila et al., 2018]. A further investigation of the relations between multiple fields in the Arctic and meridional energy transport shows that the Arctic climate is sensitive to the variations of AMET and OMET in winter. The patterns in ERA-Interim and JRA55 are more consistent in winter. For the ocean, ORAS4 and SODA3 provide similar patterns in winter. Based on our results, it seems that the interannual variability of AMET and OMET cannot be constrained by the available observations. The existence of sources and sinks in reanalysis data sets introduces large uncertainties in the computation of energy transport [Trenberth, 1991, Trenberth and Solomon, 1994]. Although the reanalysis data sets are not specifically designed for the studies on energy transport, given the good agreements on mean AMET and OMET and their annual cycles among assessed reanalysis products, we still recommend to use these reanalysis products for the energy transport diagnostics. However, much care should be taken when adopting reanalyses for the examination of energy transport at relatively long time scales. The robustness of those results based on the AMET and OMET estimated from reanalyses should be further assessed.

## Code and data availability

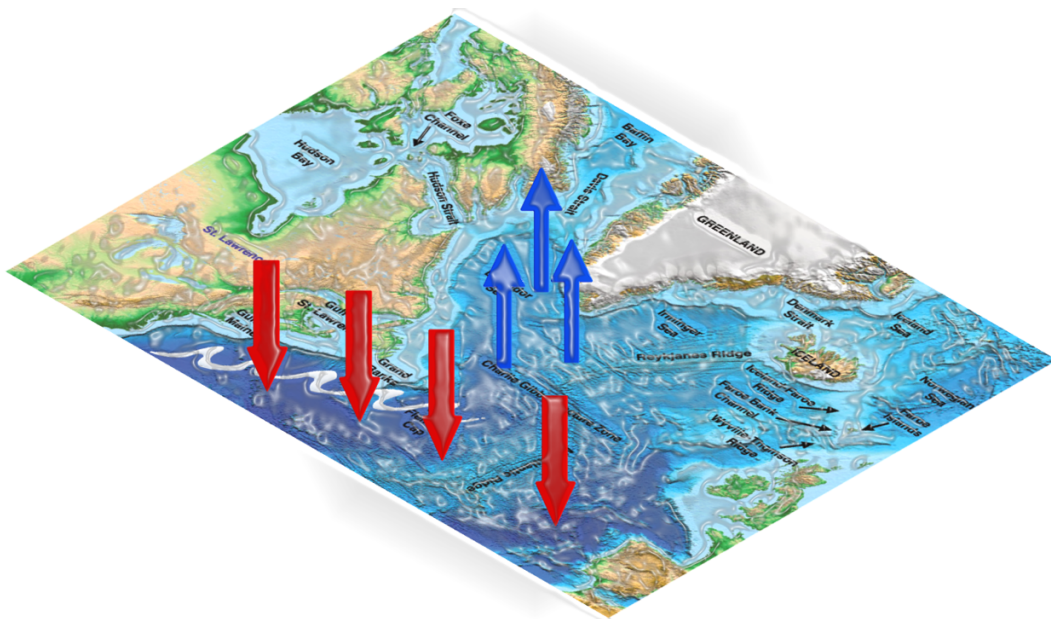
The reanalysis products used in this study are open to the public and they are available online. The computation and post-processing were carried out with the python package META (<https://doi.org/10.5281/zenodo.3609505>).

## Acknowledgements

The research was supported by the Netherlands eScience Center, Wageningen University, the National Oceanography Center in UK, and Blue Action project (European Union's Horizon 2020 research and innovation programme, grant number: 727852). The high resolution NEMO ORCA hindcast was complete in the project North Atlantic Climate System: Integrated Study (ACSIS) [grant number NE/N018044/1]. The authors are grateful for the high performance computational infrastructure (HPC cloud and Cartesius) provided by SURFsara in the Netherlands. We would like to express our gratitude to all the researchers working on the reanalysis data sets and making the data open to the public. We also want to thank the OSNAP (Overturning in the Subpolar North Atlantic Program, doi:10.7924/r4z60gf0f) project and the RAPID-AMOC program (doi:10/bkzc) for making the observation data in the North Atlantic freely available. We also acknowledge the editor Prof. Dr. Gerrit Lohmann and our two anonymous reviewers for their help to improve the manuscript.

## Chapter 3

# Atmosphere-ocean interactions and their footprint on heat transport variability in the Northern Hemisphere



This chapter is based on:

Yang Liu, Jisk Attema, and Wilco Hazeleger. Atmosphere-ocean interactions and their footprint on heat transport variability in the northern hemisphere. *Journal of Climate*, 33(9): 3691–3710, 2020a

## Abstract

Interactions between the atmosphere and ocean play a crucial role in redistributing energy, thereby maintaining the energy balance of the climate system. Here, we examine the compensation between the atmosphere and ocean's heat transport variations. Motivated by previous studies with mostly numerical climate models, this so-called Bjerknes compensation is studied using reanalysis datasets. We find that atmospheric energy transport (AMET) and oceanic energy transport (OMET) variability generally agree well among the reanalysis datasets. With multiple reanalysis products, we show that Bjerknes compensation is present at almost all latitudes from 40°N to 70°N in the Northern Hemisphere from interannual to decadal time scales. The compensation rates peak at different latitudes across different time scales, but they are always located in the subtropical and subpolar regions. Unlike some experiments with numerical climate models, which attribute the compensation to the variation of transient eddy transports in response to the changes of OMET at multi-decadal time scales, we find that the response of mean flow to the OMET variability leads to the Bjerknes compensation, thus the shift of Ferrel cell at midlatitudes at decadal time scales in winter. This cell itself is driven by the eddy momentum flux. The oceanic response to AMET variations is primarily wind driven. In summer, there is hardly any compensation and the proposed mechanism is not applicable. Given the short historical records, we cannot determine whether the ocean drives the atmospheric variations or it is the reverse.

### 3.1 Introduction

Interactions between the atmosphere and ocean play a crucial role in redistributing energy, thereby maintaining the energy balance of the climate system. Here, we study the compensation between the atmosphere and ocean's heat transport variations. It influences large scale dynamics by reshaping the energy budget. Many efforts have been made to determine heat transport. In the 1960s, Jacob Bjerknes suggested that energy transport in the atmosphere and ocean should compensate each other if the variations of heat fluxes at the top of the atmosphere (TOAFlux) and the ocean heat content (OHC) are small [Bjerknes, 1964]. This process has come to be known as Bjerknes compensation.

However, since the variations of total energy are small from the interannual time scales onwards [Shaffrey and Sutton, 2006], Bjerknes compensation is mostly studied using numerical climate models, which allows generating relatively long time series [Shaffrey and Sutton, 2004, 2006, Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, Farneti and Vallis, 2013, van der Linden et al., 2016, Outten and Esau, 2017]. Shaffrey and Sutton [2004] reported an anti-correlation between the atmospheric meridional energy transport (AMET) and oceanic energy transport (OMET) variations in the North Atlantic ocean at midlatitudes with the HadCM3 model. They found that variability of OMET in the North Atlantic is dominated by the wind driven Ekman processes, while AMET in the Northern Hemisphere appears to be governed primarily by changes in stationary waves at midlatitudes across interannual time scales. In their following work with the same model, Shaffrey and Sutton [2006] further claimed that AMET and OMET anomalies are significantly anticorrelated in the extratropical Northern Hemisphere at decadal time scales. They concluded that positive OMET anomalies could reduce atmospheric baroclinicity and hence lead to weakened AMET transport, primarily by transient eddies. It also causes weakened stationary eddy transport, but this is less pronounced than the variations caused by transient eddies.

Those findings are further confirmed by Van der Swaluw et al. [2007] as they also found that compensation between atmosphere and ocean can be attributed to changes in transient eddies with varying OMET at multidecadal time scales in their numerical experiments with the HadCM3 climate model. They emphasized the coupling between OMET in the Atlantic Ocean and the Atlantic Meridional Overturning Circulation (AMOC). Moreover, they argued that there is a link between compensation and the variability of the Arctic sea ice. van der Linden et al. [2016] demonstrated that the link between sea ice cover variations in the Barents Sea and ocean-atmosphere heat exchanges originates from the low frequency changes in the ocean, which supports the statement from Van der Swaluw et al. [2007] about the link between sea ice and Bjerknes compensation. Jungclaus and Koenigk [2010] investigated Bjerknes compensation at decadal time scales with long time series from simulations with the ECHAM5/MPIOM model. They also found that in their model the compensation peaks at 70°N. Both Van der Swaluw et al. [2007] and Jungclaus and Koenigk [2010] agree that the compensation reaches a maximum when the ocean leads the atmosphere. With the GFDL climate model, Farneti and Vallis [2013] detected a strong compensation between AMET and OMET in the North Atlantic at decadal time scales. They confirmed the crucial role of AMOC in generating OMET anomalies as well as the Bjerknes compensation.

Apart from studies using fully coupled general circulation models, some researchers also investigated the mechanism of compensation using simplified mathematical models [Yang et al., 2013, 2018]. By conducting experiments with the Fast Ocean-Atmosphere Model, Yang et al. [2013] found that AMET and OMET compensate very well in the tropics and extratropics driven by the freshwater input in the high latitude Atlantic. Similarly, Yang et al. [2018] found a stabilizing role of Bjerknes compensation caused by the freshwater influx in the North Atlantic using a simple box model.

Recently an intercomparison of Bjerknes compensation found in different numerical models shows that the results are model-dependent [Outten et al., 2018]. Given the differences in representing Bjerknes compensation in the numerical models, we are interested in the interaction between atmosphere and ocean, and Bjerknes feedback in the historical climate record. Serving as a good representation of the historical climate, reanalyses allow us to look for the Bjerknes compensation in a state that is close to reality. Note that these reanalysis products incorporate the numerical models that fail to capture the robust compensation response [Outten et al., 2018], but they are constrained by observations. Nowadays, the time span covered by most reanalysis products allows examining the response and feedback from atmosphere to ocean at interannual to decadal time scales [Dee et al., 2011, Ferry et al., 2012b, Balmaseda et al., 2013, Harada et al., 2016, Gelaro et al., 2017, Carton et al., 2018]. But the time series are not long enough to account for the multidecadal variability of energy transports, which are of interest to many studies with numerical models.

In this paper, for the first time, Bjerknes compensation is studied using multiple atmospheric and oceanic reanalysis datasets. We quantify and intercompare meridional energy transport (MET) in the atmosphere and ocean with six state-of-the-art reanalysis products. With these reanalysis products, Bjerknes compensation is investigated from interannual to decadal time scales. Moreover, we are interested in the energy transport variability and their influence on the circulation of the atmosphere and ocean. The atmospheric response and potential feedback to the variability of the ocean are illustrated in detail. We examine the role of the ocean in the air-sea interaction and explore the mechanism of Bjerknes compensation. The large scale climate variability associated with the interactions between the atmosphere and ocean is also examined in this study.

The paper is organized as follows: The reanalysis datasets used in this study are listed in the section Data and Methodology. In this section, we elaborate on the methods for the quantification of AMET and OMET. A description of the statistical tests is also included in this section. The calculated AMET and OMET, and an overview of Bjerknes Compensation are shown in the section Results, together with our exploration of the physical mechanism of compensation and the atmospheric response and feedback to the ocean. In the section Conclusion and Discussion, we summarize this study and provide suggestions and a perspective for future work.

## 3.2 Data and Methodology

In this section, we introduce six reanalysis datasets used in this study. In addition, we elaborate on our methodology for the computation of AMET and OMET, and the statistical analysis



performed in this work.

### 3.2.1 Reanalysis Datasets

Unlike most of the previous studies that took output from climate models, we investigate Bjerknes compensation within reanalysis datasets. Six reanalysis products are included in this paper: 3 atmospheric reanalysis products (ERA-Interim, MERRA2, and JRA55) and 3 oceanic reanalysis products (ORAS4, GLORYS2V3, and SODA3). A synthesis of the basic specification of the reanalysis products used in this study is provided in Table 2.1. We chose these six reanalysis products because they are the state-of-the-art atmospheric and oceanic reanalyses incorporating the latest weather forecast models and data assimilation schemes. Although there is overlap between the assimilated observations in the chosen products, the intercomparison between them is still fruitful as most of them were built upon different numerical models and data assimilation techniques. For instance, ERA-Interim is based on the ECMWF's Integrated Forecast System (IFS) [Dee et al., 2011], while MERRA2 is based on the Goddard Earth Observing System (GEOS) [Gelaro et al., 2017]. JRA55 generates data with an advanced weather prediction system from JMA [Kobayashi et al., 2015, Harada et al., 2016]. The numerical schemes and data assimilation systems are also different among the chosen oceanic reanalysis products. More details about these reanalysis products can be found in Chapter 2.

### 3.2.2 Methodology

The methods for the computation of AMET and OMET are given in this subsection, together with the statistical analysis performed in this study.

#### *Energy Transport in the Atmosphere and Eddy Decomposition*

Following Magnusdottir and Saravannan [1999], Shaffrey and Sutton [2006] and Van der Swaluw et al. [2007], we calculate the zonal integral of AMET at each latitude as the residual between the zonally integrated net surface flux (SFlux) from the ocean into the atmosphere ( $F_{surface}$ ), and the downward net radiation flux at the top of the atmosphere ( $F_{TOA}$ ). The equation is shown below:

$$\frac{1}{a \cos \theta} \frac{\partial}{\partial \theta} (\cos \theta E) = \int_0^{2\pi} (F_{surface} + F_{TOA}) a \cos \theta d\phi \quad (3.1)$$

where  $\frac{\partial}{\partial \theta} (\cos \theta E)$  is the divergence of global AMET ( $PW$ ),  $a$  is the earth radius ( $m$ ),  $\theta$  denotes the latitude and  $\phi$  indicates the longitude. The calculation of AMET was performed on monthly mean fields with chosen reanalysis products. Note that the calculation of AMET is based on the inferred method which is different from the direct calculation given in Chapter 2.

For the eddy decomposition, we follow the mathematical definition given by Peixoto and Oort [1992] (see chapter 4.1 and equation 4.9). We compute the transient eddies and stationary eddies on selected pressure levels with daily mean fields. The eddy decomposition is performed on each component of atmospheric heat transport separately in order to calculate the total heat

transport by eddies. Taking the internal energy transport as an example, the decomposition is as follows [Peixoto and Oort, 1992] (see equation 4.10):

$$c_p[\overline{vT}] = c_p[\overline{v}][\overline{T}] + c_p[\overline{v}^*][\overline{T}^*] + c_p[\overline{v'T'}] \quad (3.2)$$

where  $[v]$  is the zonal average of meridional wind velocity,  $\overline{v}$  the temporal average,  $v^*$  the departure from zonal mean,  $v'$  is the departure from temporal mean, and similar for temperature  $T$ . The total internal energy is now decomposed into three components: temperature transport by (1) steady mean circulation  $c_p[\overline{v}][\overline{T}]$ , (2) stationary eddies  $c_p[\overline{v}^*][\overline{T}^*]$  and (3) transient eddies  $c_p[\overline{v'T'}]$ . In this case, we assume that cross terms, which indicate the temperature transport carried by instantaneous and asymmetric flow, are small compared to the other components.

After applying the decomposition to all the components in AMET, we can calculate the total heat transport by transient eddies and stationary eddies, respectively. Note that the calculation suffers from mass imbalance reported by previous studies [e.g. Trenberth, 1991, Graversen, 2006, Liu et al., 2020b]. We have examined the mass budget correction methods in a previous paper [Liu et al., 2020b]. In general, barotropic mass correction is the most practical method used by many studies to calculate mass residuals [Trenberth, 1991, Graversen, 2006, Fasullo and Trenberth, 2008, Mayer and Haimberger, 2012, Liu et al., 2020b]. Unfortunately, such barotropic correction is a rough mass budget correction. It is assumed that the barotropic wind causes the mass imbalance and introduces corrections to barotropic wind fields, which only works for the vertical integral of an air column. It cannot quantify reliably the mass residuals on each pressure level. Liang et al. [2018] proposed a new mass budget correction method for the variability of the total heat transport at short timescales. However, it cannot be applied to a decomposition of heat transport driven by transient or stationary eddies.

It is still necessary to have an estimate of the energy transported by mass residuals. Therefore, we performed barotropic mass correction and show the energy transported by mass residuals. For instance, we calculated the zonal integrals of total energy transported by mass residuals at 60°N in the chosen atmospheric reanalysis products (not shown). The values are smaller than the total energy transport at that latitude, but much larger than the anomalies. Although the impact of mass residuals on each layer should be much smaller than the total integral, the results should be interpreted with care.

### *Energy Transport in the Ocean*

Different from the atmosphere, we calculate the total energy transport in the ocean by integrating the divergence of OMET over the entire water column. The mathematical expression of the calculation of OMET at each latitude is given in the methodology section in Chapter 2.

### *Statistical Analysis*

Intending to explore the relations between different variables, we performed linear regressions on various fields. We implemented the Student's t-test to check the significance of the regressions. Before performing the linear regression, all the time series were detrended. We determined the

trends by applying a polynomial fitting to the time series. We found that the second-order polynomial fitting can capture the multidecadal trend without losing variations at decadal time scales, especially for OMET (not shown). The t-tests were performed on detrended monthly fields without a low pass filter.

Note that some of the results discussed in the following sections are below the 99% confidence level. This is mainly due to the short time series provided by the reanalysis data (no more than 456 months at monthly time scales, see Table 2.1). However, compared to the outputs from numerical models, we believe the results are still insightful as these reanalyses can represent AMET and OMET in the real world.

### 3.3 Results

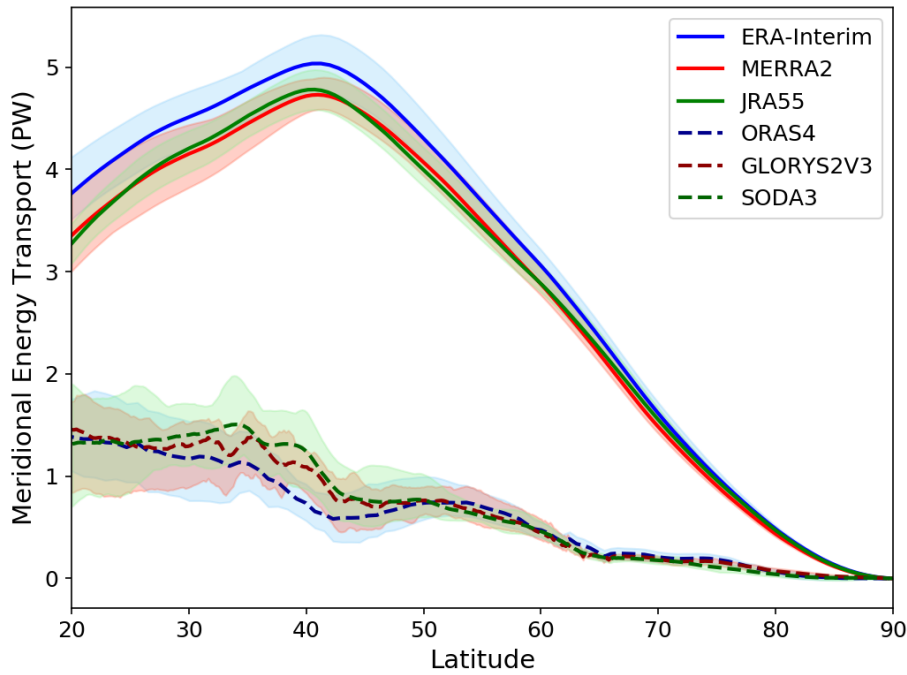
The results shown in this section are based on monthly mean fields. Results across larger time scales are obtained with low pass filters (running mean from 1 to 10 years) applied to the monthly mean fields.

#### 3.3.1 Meridional Energy Transport

To investigate the compensation between the atmosphere and ocean, we start with the quantification of AMET and OMET. Estimated from the chosen reanalysis products, the mean AMET and OMET as a function of latitude in the Northern Hemisphere are shown in Figure 3.1. Generally, AMET between the chosen atmospheric products agrees well. Starting from 20°N, AMET increases with latitude and peaks at 41°N. The peak of AMET in ERA-Interim is about  $5PW$ , which is consistent with previous studies [e.g. Trenberth and Caron, 2001, Fasullo and Trenberth, 2008]. The AMET from MERRA2 and JRA55 are a bit lower than that in ERA-Interim. The meridional variations of mean AMET agree well with literature [Trenberth and Caron, 2001, Fasullo and Trenberth, 2008].

Furthermore, we inspect the temporal evolution of AMET at specific latitudes. Our interest is in meridional variations in energy transport associated with midlatitude-Arctic interactions. The time series of AMET and their anomalies at decadal time scales at 60°N is given in Figure 3.2. The monthly averaged AMET provided by each atmospheric reanalysis product can be found in Figure 3.2a. The variations of AMET are similar among the chosen products, but they differ slightly in amplitudes. The average of AMET in ERA-Interim is  $3.06 \pm 1.33PW$ , while in MERRA2 it is  $2.88 \pm 1.20PW$  and in JRA55 it is  $2.90 \pm 1.20PW$ . After removing the seasonal cycle and implementing a low pass filter of 10 years running mean to each time series, we obtain the anomalies of AMET from each reanalysis (Figure 3.2b). MERRA2 and JRA55 have similar AMET anomalies in terms of the variations and amplitudes, while ERA-Interim differs from them in amplitude. This is also reflected in Figure 3.2a, as ERA-Interim gives a higher mean AMET at each latitude. In addition, ERA-Interim has a larger standard deviation (std) compared to the other two products.

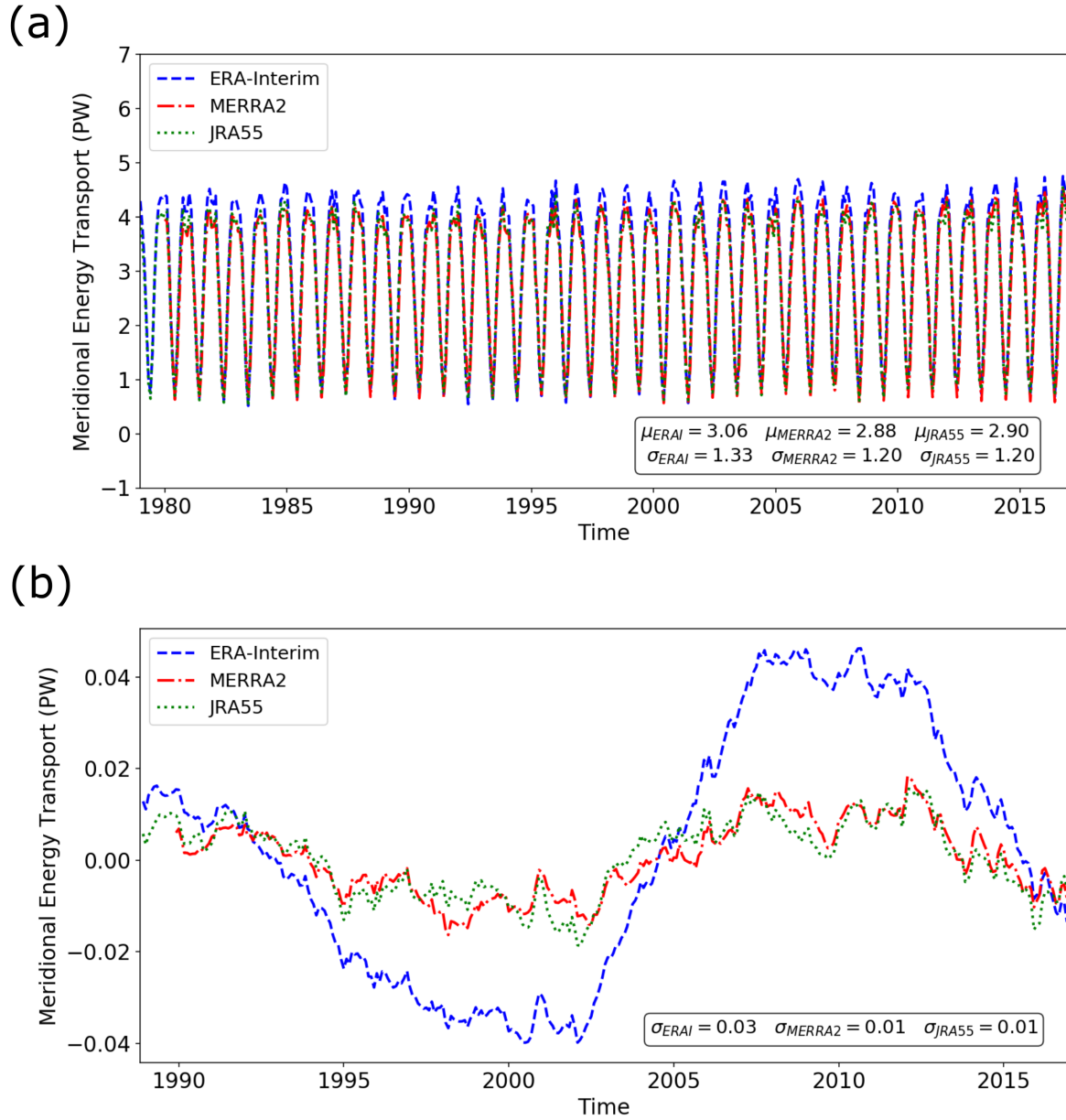
For the ocean, the mean OMET at each latitude estimated from chosen oceanic reanalyses are shown in Figure 3.1. Compared to AMET, OMET is relatively small. The mean OMET in each oceanic reanalysis agrees well at almost all latitudes except for OMET between 30°N and 40°N.



**Figure 3.1:** Mean AMET and OMET over the entire time span of each product as a function of latitude in the Northern Hemisphere. AMET is shown with solid lines and OMET with dash lines. The shaded regions represent the full range of MET across the entire time series at each latitude. The time span of each product used in this study is given in Table 1.

This could be due to the difference in the model’s capability of representing mesoscale eddies in the Gulf Streams among the chosen oceanic reanalyses. We have explored this in our previous work [Liu et al., 2020b]. Both GLORYS2V3 and SODA3 have a high spatial resolution, employ eddy-permitting models, and simulate the large scale eddy variability, but this is not the case for ORAS4 [Ferry et al., 2012b,a, Balmaseda et al., 2013, Carton et al., 2018].

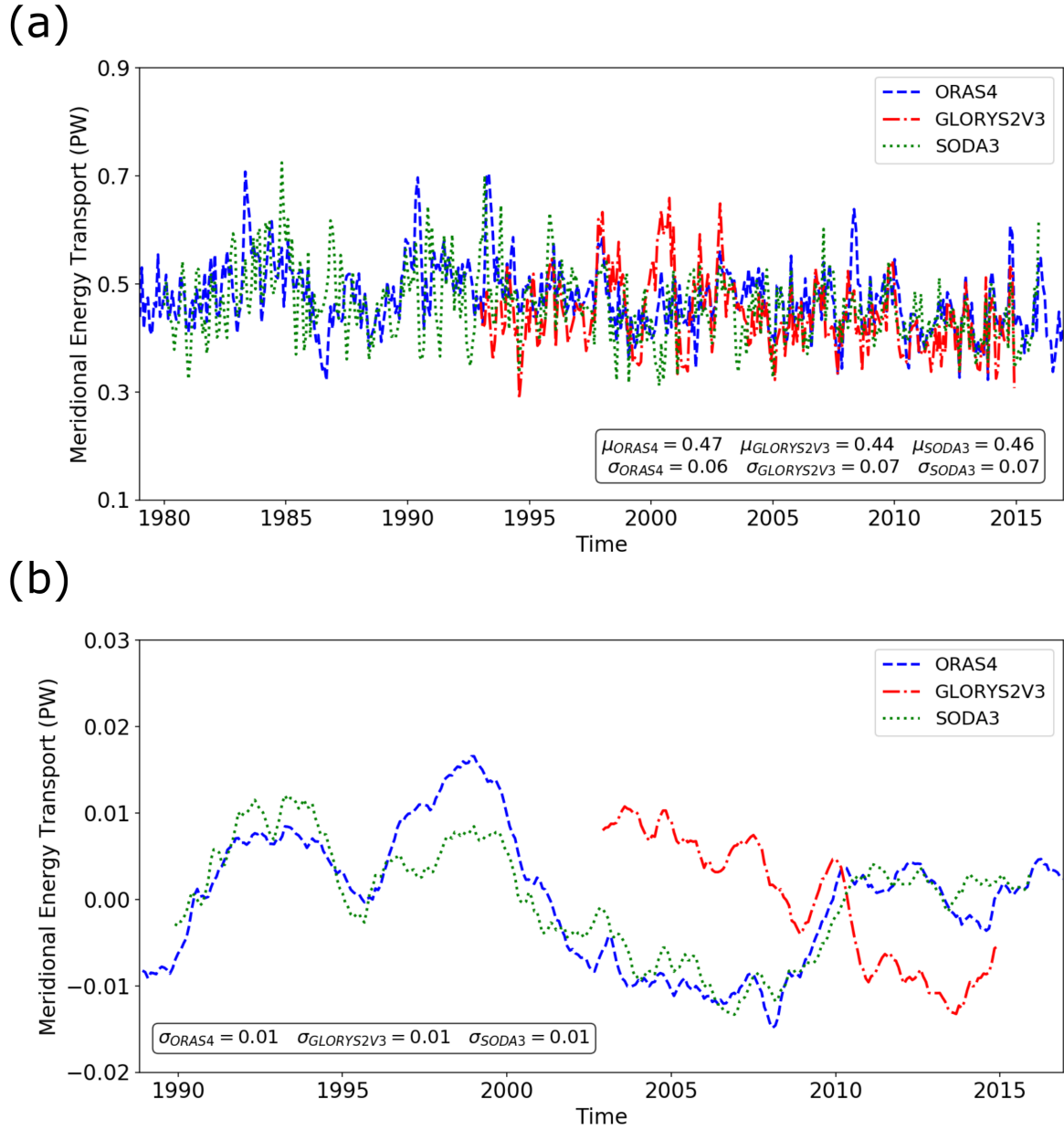
Moreover, we are interested in the temporal evolution of OMET. The time series of OMET and their anomalies at decadal time scales at 60°N after detrending are given in Figure 3.3. Differences in both the amplitude and variations can be observed in all three time series (Figure 3.3a). The differences are consistent with several recent studies. Uotila et al. [2018] demonstrate the large difference in ocean heat transport between several different oceanic reanalyses in their Polar ORA Intercomparison Project, especially for the Arctic region. Karspeck et al. [2017] report that with the same surface forcing (CORE-II forcing), ocean models can behave very differently without further constraints from observations. They find large differences in volume transports between six reanalysis products (also including ORAS4 and SODA) in their tests. The OMET time series considered here do not differ much from each other statistically. The mean OMET at 60°N in ORAS4 is  $0.47 \pm 0.06PW$ , while in GLORYS2V3 it is  $0.44 \pm 0.07PW$  and in SODA3 it is  $0.46 \pm 0.07PW$ . After removing the climatology of OMET and implementing a low pass filter of 10 years, the differences become clear (e.g. Figure 3.3b). It can be noticed



**Figure 3.2:** Time series of the zonal integral of AMET at 60°N without/with a low pass filter. (a) The original time series and (b) the anomalies with a low pass filter after detrending. For the detrended low pass filtered ones, we take a running mean of 10 years.  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

that ORAS4 resembles SODA3 but differs much from GLORYS2V3. It was explained in our previous work that the first 10 years in GLORYS2V3 are not reliable [Liu et al., 2020b]. Given the absolute value of OMET anomalies ( $\sim 0.01PW$ ), the standard deviations of OMET anomalies ( $\sim 0.01PW$ ) from these three oceanic reanalyses are large.

To conclude, all the chosen atmospheric reanalyses agree on mean AMET at almost every lati-



**Figure 3.3:** Time series of the zonal integral of OMET at 60°N. (a) The original time series and (b) the anomalies with a low pass filter after detrending. For the detrended low pass filtered ones, we take a running mean of 10 years.  $\sigma$  is the standard deviation and  $\mu$  is the mean of the entire time series.

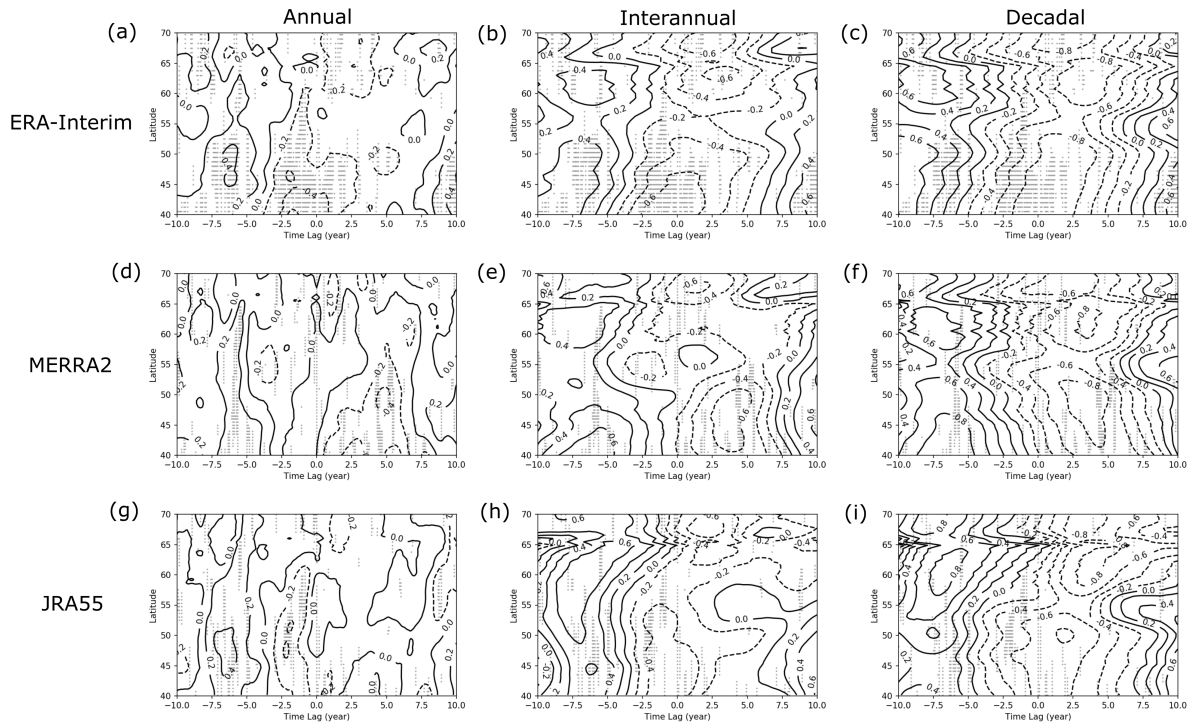
tude in the Northern Hemisphere. The low frequency AMET anomalies in ERA-Interim differ a bit from those in MERRA2 and JRA55, but the differences in variability are relatively small compared to the differences between low frequency AMET anomalies calculated through integration of energy transports over the entire air column [Liu et al., 2020b]. Similarly, the mean OMET given by all three oceanic reanalyses agree at almost all latitudes in the Northern Hemisphere except for around the Gulf Stream. The OMET anomalies given by ORAS4 resembles

that provided by SODA3, especially for the OMET variability. However, in GLORYS2V3 the OMET anomalies are not so reliable. Liu et al. [2020b] found that ORAS4 is more consistent with the direct heat transport observations in the Atlantic Ocean (OSNAP and RAPID array) than SODA3 and GLORYS2V3 with aspect to OMET. Given the similarities in the variability of OMET anomalies in all the chosen oceanic reanalysis products excluding GLORYS2V3, and the consistency of OMET given by ORAS4 and the observations [Liu et al., 2020b], we further investigate the Bjerknes compensation with ORAS4 only. Three atmospheric reanalyses are all included in the following analysis to test the robustness of our study among atmospheric reanalysis products.

### 3.3.2 Bjerknes Compensation and Atmospheric Response to Ocean Variability

According to Bjerknes [1964], compensation between atmosphere and ocean is expected to take place when the net solar radiation flux at TOA and the heat content in the ocean remain relatively stable. This suggests that the Bjerknes compensation should be investigated across large time scales, at least at interannual ( $\sim 5\text{yrs}$ ) time scales. In search of the Bjerknes compensation, we illustrate the correlation between AMET and OMET with time lags up to 10 years from interannual to decadal time scales in Figure 3.4. All months are included in the lag regressions. We also investigated the compensation between AMET and OMET anomalies by season and they will be shown in the following analyses of the mechanism of compensation. Results at annual time scales are included for comparison. At annual time scales, the anticorrelation is small in the chosen atmospheric reanalysis products, which is consistent with results shown by Shaffrey and Sutton [2004] using a numerical climate model. At longer time scales, compensation between AMET and OMET is found at almost all latitudes from  $40^\circ\text{N}$  to  $70^\circ\text{N}$ . The maximal compensation rates are reached when the ocean leads the atmosphere for the chosen atmospheric reanalyses from interannual to decadal time scales. However, the time lag between OMET and AMET for the maximum compensation varies across time scales at different latitudes among chosen atmospheric products. At interannual time scales, compensation reaches a maximum when there is no lag at midlatitudes around  $40^\circ\text{N}$  in ERA-Interim and JRA55, while close to  $70^\circ\text{N}$  the maximum compensation occurs when the ocean leads by 2.5 years. MERRA2 differs slightly. The compensation peaks when the ocean leads by 2.5 years around  $40^\circ\text{N}$ , while between  $65^\circ\text{N}$  and  $70^\circ\text{N}$  the compensation reaches the maximum when the ocean leads the atmosphere by several months. Similarly, in ERA-Interim and JRA55, at decadal time scales, the maximal compensation rate is found at around  $40^\circ\text{N}$  to  $50^\circ\text{N}$ , and  $60^\circ\text{N}$  to  $65^\circ\text{N}$  when the ocean leads by 2 months and 2.5 years, respectively. While in MERRA2 the compensation peaks around the same latitudes but always with ocean leading by 3 years.

Focusing on multi-decadal time scales, previous studies with numerical climate models can provide more insight into the compensation because of longer time series. After analyzing 93 decades of model output from HadCM3, Shaffrey and Sutton [2006] found that at decadal time scales, compensation between atmosphere and ocean occurs in the Northern Hemisphere and reaches maximum only around  $70^\circ\text{N}$ . Similarly, Van der Waluw et al. [2007] investigated Bjerknes compensation with 341 years of data from a preindustrial run of HadCM3. They found a strong anticorrelation between AMET and OMET at around  $70^\circ\text{N}$  at interdecadal time scales



**Figure 3.4:** Lag regression of OMET anomalies on AMET anomalies at different latitudes in the Northern Hemisphere at annual, interannual and decadal time scales. The correlation coefficients are shown by contour lines. Positive time lag indicates that ocean leads the atmosphere and vice versa. The regression was performed on monthly mean series of AMET from (a,b,c) ERA-Interim, (d,e,f) MERRA2 and (g,h,i) JRA55 and OMET integrated over each latitude after taking running means of 1 year, 5 years and 10 years, including all months. The stippling indicates a significance level of 95% based on t-test with unfiltered time series.

and the compensation rate peaks when the ocean leads the atmosphere by 1 year. Using 505 years of the IPCC AR4 preindustrial experiment with the ECHAM5/MPIOM model, Jungclauss and Koenigk [2010] found Bjerknes compensation around 70°N and the compensation became stronger when the ocean leads by 1-2 years. In summary, these studies using numerical climate models all suggest that the compensation between atmosphere and ocean occurs and peaks around 70°N when the ocean leads at decadal and multi-decadal time scales.

With reanalysis data, we find that at decadal time scales the Bjerknes compensation occurs and peaks not only at subpolar latitudes but also at mid-to-low latitudes. Specifically, we notice the highest compensation rate is found at 40°N and 60°N at decadal time scales when the ocean leads. However, we should note that our time series are very short. Recently, Outten and Esau [2017] examined the output from the Bergen Climate Model at decadal time scales using 600 years data and found the strongest compensation at 0 years lag between the oceans and the atmosphere. Their results are supported by other CMIP5 models examined in Outten et al. [2018], which is a synthesis of the diagnostic results using model outputs from 15 CMIP5 models. The finding that Bjerknes compensation appears at mid-latitudes, as well as high latitudes, was put forward in Outten et al. [2018]. They found that peak compensations occurs

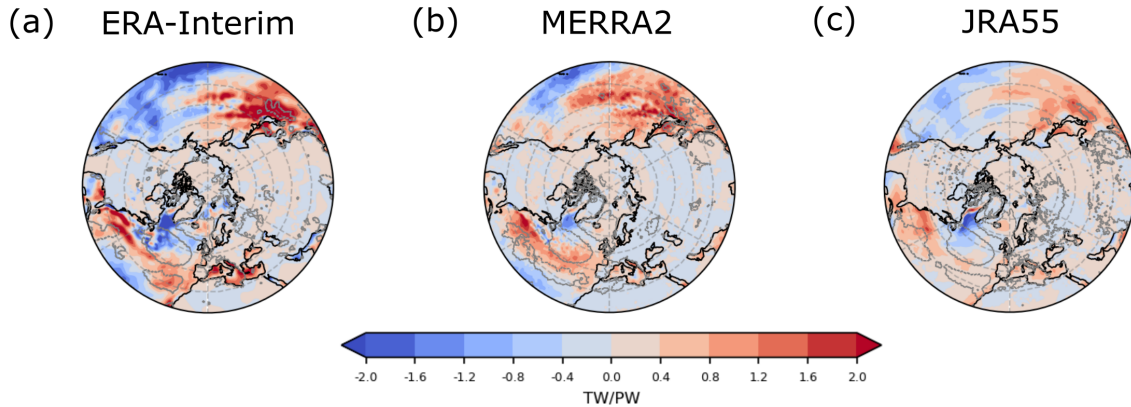


around the latitude 70N (near the sea ice edge) and around 45N (near the tracks of transient eddies). Those results are similar to our findings with reanalysis data. However, Outten et al. [2018] also show that numerical climate models do not agree on the locations and magnitudes of Bjerknes compensation. Given the large uncertainties and differences between the presence of the compensation in numerical models and reanalysis datasets, we further explore the physical mechanism of the compensation between the atmosphere and ocean, so as to obtain a physically consistent picture of Bjerknes compensation.

Since the compensation between the atmosphere and ocean becomes more persistent at decadal time scales in three atmospheric reanalyses and ORAS4, for the rest of this paper we focus on decadal time scales and apply a low-pass filter of 10 years to the monthly fields, unless specifically noted. In order to compare the results given by reanalyses and models from literature, we emphasize on the compensation and its physical mechanism in the subpolar region. Consequently, most of the regressions shown below were conducted on the AMET and OMET anomalies at 60°N. We do realize that the time series are very short for this, but we are looking for various strands of evidence for Bjerknes compensation.

A first impression of the physical picture about the compensation between ocean and atmosphere can be obtained by looking at the surface flux. The SFlux anomalies, which consist of the net turbulent flux and net radiation flux at the surface, regressed on OMET anomalies at 60°N at decadal time scales, are shown in Figure 3.5. There is no time lag between these two variables. Positive values indicate SFlux towards the ocean, and vice versa. The chosen atmospheric reanalysis products provide similar results. It can be observed that the atmosphere provides energy to the ocean in subtropical regions in the Atlantic ocean. In the Pacific Ocean close to the Asian continent, SFlux variations associated with AMET are also directed from the atmosphere to the ocean. This means that in the places where the strongest northward flow is situated in the Atlantic (the Gulf Stream) and in the Pacific (the Kuroshio Current), the ocean always receives heat from the atmosphere. In the subpolar Atlantic, especially in the Labrador Sea, the ocean releases energy back to the atmosphere. Jungclaus and Koenigk [2010] found similar results, but the locations where the air-sea interaction is most active are different. Our results suggest that the change of OMET at 60°N is linked to the changes of SFlux from subtropical to subpolar regions. The cause-effect relation between the atmosphere and ocean, in general, cannot be resolved due to the short timeseries of the analyzed reanalysis products. But the strong correlation between SFlux and OMET at subtropical and subpolar Atlantic is consistent with the anticorrelation in Figure 3.4c, f and i at decadal time scales.

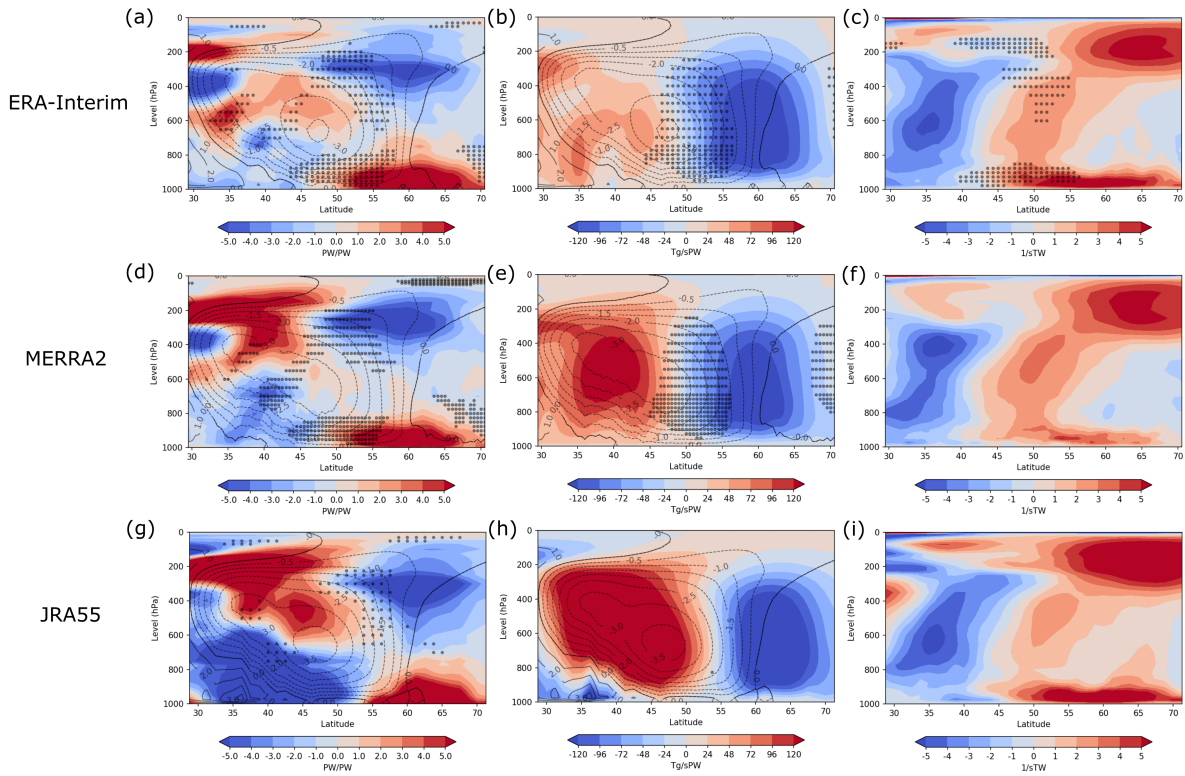
A further look into the mechanism of compensation requires more insight into the physical processes in the atmosphere. We investigated the air-sea interactions in both summer and winter. Since the atmospheric dynamic processes are more dominant in winter than in summer [e.g. Gastineau and Frankignoul, 2015, van der Linden et al., 2016], we found clearer signals of compensation and more active interaction between ocean and atmosphere in winter than in summer. Therefore, we mainly elaborate on processes in winter and will briefly mention the results from summer. The correlations between OMET anomalies at 60°N and AMET anomalies on each pressure level in winter are shown in Figure 3.6a,d and g. In this case, the ocean leads by 1 month. The mean meridional mass streamfunction of the atmosphere is also shown. At



**Figure 3.5:** Regression of net surface turbulent and radiation flux anomalies on vertically integrated OMET anomalies at 60°N at decadal time scales. The regression was performed on monthly fields from (a) ERA-Interim, (b) MERRA2 and (c) JRA55 after taking a running mean of 10 years, including all seasons. The regression coefficients with a unit of Tera Watt per Peta Watt are shown spatially, with positive sign indicating fluxes directed towards the ocean. The gray contour lines indicate a significance level of 99% based on t-test with unfiltered time series.

lower layers, a strong anticorrelation between AMET and OMET in the subtropical region and a strong correlation at the subpolar region is consistent with the air-sea interaction shown in Figure 3.5. However, it is not consistent with Bjerknes compensation. A strong compensation around 60°N at high altitudes is observed, and this should be responsible for the strong compensation illustrated in Figure 3.4c around 60°N. Given the mean meridional atmospheric circulation, the region contributing to the compensation is located at the ascending branch and overturning part of the Ferrel cell. This motivates us to investigate the relation between the Ferrel Cell and OMET variability. We regressed the anomalies of the Stokes streamfunction on OMET at 60°N in winter (DJF) when ocean leads by one month in Figure 3.6b, e and h. A strong anticorrelation is observed between 50°N and 70°N. Together with the climatology of the Stokes streamfunction shown in this figure, we observe that the changes in meridional mean flow is primarily due to a shift rather than a change in intensity of the Ferrel cell. The result is consistent with the observed anticorrelations provided in Figure 3.6a,d and g. Although the magnitudes of regression coefficients are different among the chosen atmospheric reanalysis products in Figure 3.6, the patterns are quite robust among these datasets. In addition, we also notice a strong coupling between the Hadley Cell and OMET (not shown). This suggests that a low latitude driver may contribute to the compensation at mid-to-high latitudes. However, such a teleconnection is beyond the scope of this paper.

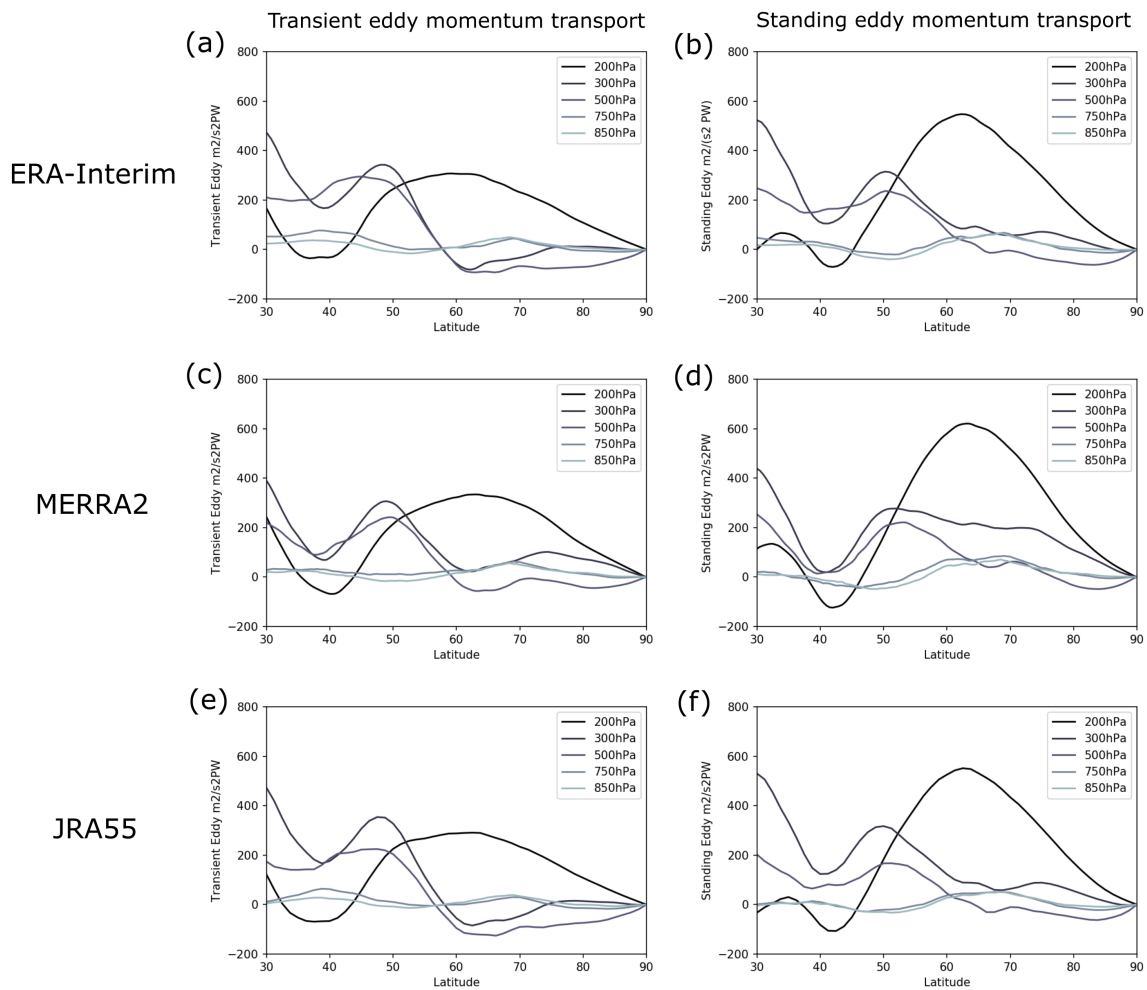
Since the mean meridional cells are primarily driven by the eddy forcing at midlatitudes [e.g. Phillips, 1956, Lorenz and Lorenz, 1967, Holton, 1973, Andrews et al., 1987, Trenberth and Stepaniak, 2003], we expect the shift of the Ferrel cell to be related with changes of transient eddy momentum fluxes and thus the baroclinicity of the atmosphere. In order to inspect the baroclinicity, we regress  $du/dz$  anomalies on OMET anomalies at 60°N in winter (DJF) at decadal time scales when ocean leads by one month in Figure 3.6 c, f and i. The results are



**Figure 3.6:** Regression of AMET anomalies (a, d, g), Stokes streamfunction anomalies (b, e, h), and  $du/dz$  anomalies (c, f, i) on each pressure level on vertically integrated OMET anomalies at  $60^\circ\text{N}$  in winter (DJF) at decadal time scales when OMET leads by one month. The climatology of the Stokes streamfunction of the atmosphere is placed on top of the shades as contour lines with unit of  $10^{10} \text{ kg/s}$ .  $du/dz$  is the change of zonal wind velocity divided by the change of height, which indicates the vertical shear of zonal wind, with positive sign indicates enhancement of baroclinicity. The regression was performed on monthly fields from (a, b, c) ERA-Interim, (d, e, f) MERRA2 and (g, h, i) JRA55 after taking a running mean of 10 years. The stippling indicates a significance level of 99% based on t-test with unfiltered time series.

very similar for different atmospheric reanalysis products. A direct look at the momentum transported by eddies at multiple pressure levels is given in Figure 3.7. Around  $50^\circ\text{N}$  from the surface to 400 hPa and  $70^\circ\text{N}$  above 300 hPa, with increased OMET we find increased baroclinicity in the atmosphere, which implies that more transient eddy activity can be anticipated. Following Peixoto and Oort [1992] we decomposed the total meridional eddy momentum transport anomalies into transient eddy momentum transport and standing eddy momentum transport on multiple pressure levels, and then regressed them on OMET anomalies at  $60^\circ\text{N}$  in winter (DJF). The OMET leads the atmosphere by 1 month. The results are similar among the chosen atmospheric reanalysis products. Between  $50^\circ\text{N}$  and  $70^\circ\text{N}$  at 200 hPa, transient eddies become more active when OMET increases. The standing eddies are correlated with OMET from  $50^\circ\text{N}$  and  $70^\circ\text{N}$  at almost all the chosen pressure levels. Physically, it seems heat converges due to increased OMET and the convergence could lead to increased baroclinicity. A change in gradient will occur as a consequence and hence stronger transient eddy activity in the upper tropopause.

The response of transient eddies and standing eddies are comparable, in terms of the magnitudes of eddy momentum in response to the variability of OMET (Figure 3.7). The mechanisms for changes in standing eddies are unclear. It might be that the processes associated with the increased baroclinicity involve large scale orographic forcing of the stationary waves [Williams et al., 2007]. Due to an interaction with transient eddies and changes in the vertical structure of the atmosphere, stronger standing eddy activity is therefore expected [Held et al., 2002]. The result is consistent with the strong correlation in Figure 3.6 c, f and i between 50°N and 70°N. Consequently, with increased OMET the shift in the Ferrel Cell is likely to be driven by eddy momentum fluxes due to changes in baroclinicity, which explains the anticorrelation between OMET and the Ferrel Cell around 60° in Figure 3.6b, e and h.



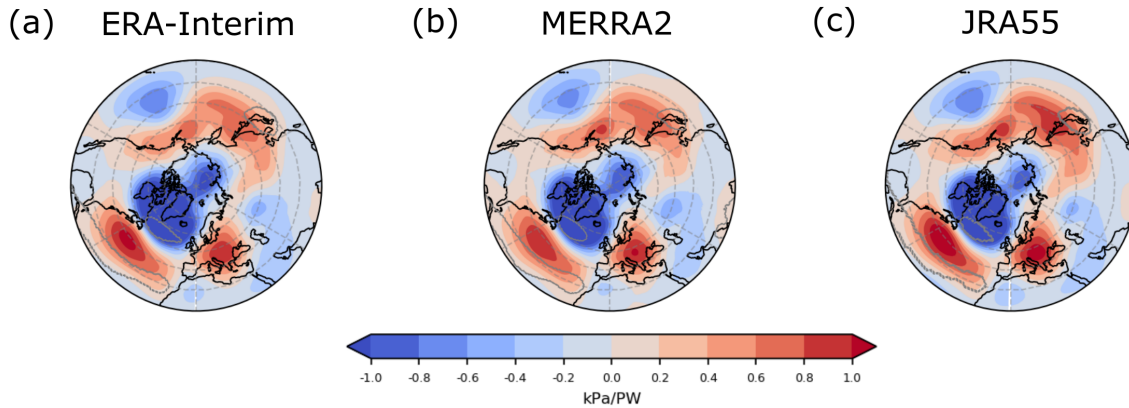
**Figure 3.7:** Regression of meridional component of eddy momentum anomalies on vertically integrated OMET anomalies at 60°N in winter (DJF) at decadal time scales when OMET leads by one month. The eddy momentum  $v \cdot v$  was decomposed into transient eddies and standing eddies. The regression was performed on monthly fields from (a,b) ERA-Interim, (c,d) MERRA2 and (e,f) JRA55 after taking a running mean of 10 years, with positive sign indicating enhancement of eddy transport. The regression coefficients are shown with a unit of  $m^2/s^2PW$ .

In summer, compensation is much weaker than that in winter. Similar to our study in winter months, we examined the compensation and its mechanism with the same methods. We calculated the correlations between OMET anomalies at 60°N and AMET anomalies on each pressure level in summer (JJA) (not shown). We observed that there is a weak compensation at the surface around 60°N. We regressed the anomalies of the Stokes streamfunction on OMET at 60°N in summer (JJA) when ocean leads by one month (not shown). No strong correlation between OMET and the mean flow is found. Similarly, we regressed  $du/dz$  anomalies on OMET anomalies at 60°N in summer (JJA) at decadal time scales when ocean leads by one month (not shown). There is no sign about the coupling between OMET and baroclinicity in subpolar regions. The patterns given by ERA-interim, MERRA2 and JRA55 are slightly different. In summary, the air-sea interaction is weak in summer and it hardly leads to Bjerknes compensation. Given the fact that both the mean flow and the baroclinicity of the atmosphere are not strongly linked to the variability of OMET in summer, the physical processes are also very different than those in winter.

Apart from the zonal mean response, the atmosphere also shows spatial patterns in the horizontal plane in response to OMET variations. The atmospheric response to the changes of OMET in winter at decadal time scales is given in Figure 3.8. It illustrates the relation between geopotential height anomalies at 500 hPa and OMET anomalies at 60°N when OMET leads by 1 month. Both the patterns and magnitudes are very consistent among three atmospheric reanalysis datasets. The variations of OMET are correlated with a low pressure system over the Greenland, subpolar North Atlantic and central Arctic ocean, and a high pressure system over the subtropical Atlantic. In other words, the atmospheric response to positive OMET anomalies is characterized by an AO/NAO like pattern in the North Atlantic and the polar low pressure system is shifted southward. Such an AO/NAO like pattern was found by Jungclaus and Koenigk [2010] too. In addition, they also noticed a high pressure center over the North Pacific (see their Figure 6a) and they claimed that the Bjerknes compensation is likely to be influenced by the changes in the North Pacific sector after inspecting the longitudinal variations of AMET (see their Figure 7). On the other hand, Van der Swaluw et al. [2007] reported a low pressure center driven by positive OMET anomalies at the Greenland Sea and no AO/NAO like pattern was observed at the North Atlantic. Although the chosen reanalyses and models do not agree well on the atmospheric response to OMET anomalies in the North Pacific and the subtropical North Atlantic regions, a strong polar low is found by all of them. This also implies that OMET has a strong impact on the Arctic climate, which has been reported by many studies [Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, Koenigk and Brodeau, 2014, van der Linden et al., 2016].

To conclude, the atmosphere responds to OMET variations with a shift in the Ferrel Cell in winter. This shift is likely to be driven by eddy momentum fluxes due to changes in baroclinicity. Horizontally, this shift shows as an AO/NAO like response.

So far, we discussed the atmospheric response to the variations of OMET. We demonstrated that the atmosphere responds to the OMET variations with a shift of mean flow driven by eddy momentum transport in winter. This was also shown by Outten and Esau [2017] with Bergen Climate Model, where they regressed the OMET variations onto surface fluxes and showed a

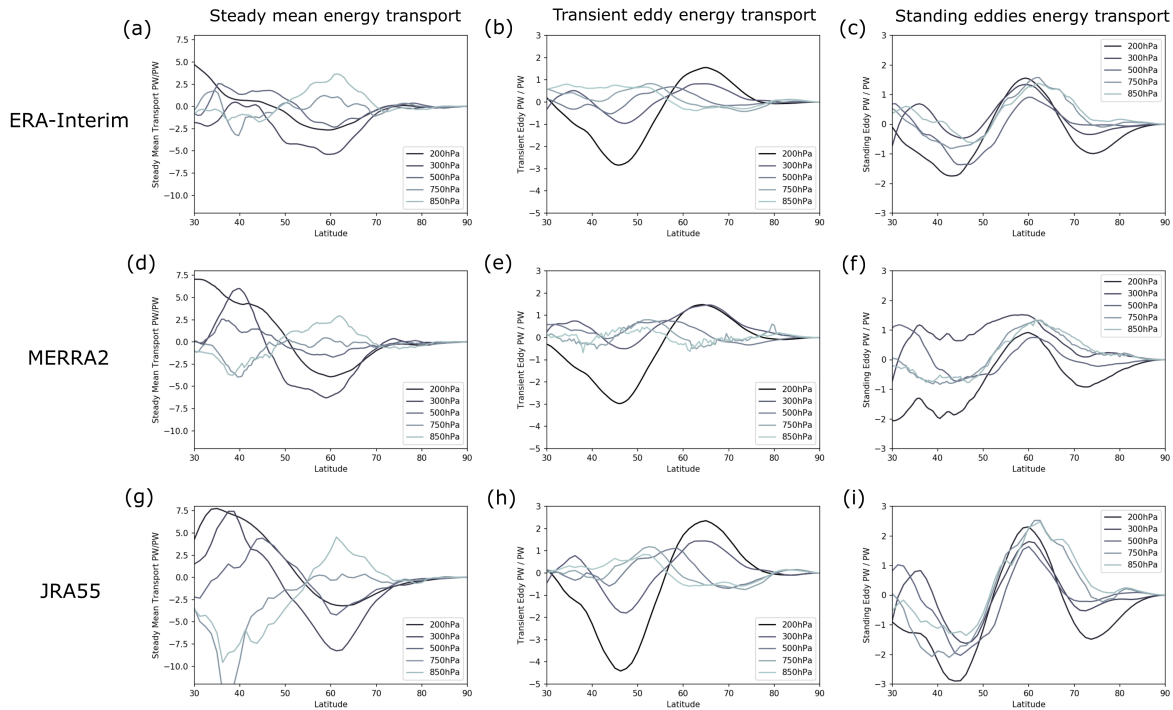


**Figure 3.8:** Regression of geopotential height anomalies at 500hPa on vertically integrated OMET anomalies at 60°N in winter (DJF) at decadal time scales when OMET leads by one month. The regression was performed on monthly fields from (a) ERA-Interim, (b) MERRA2 and (c) JRA55 after taking a running mean of 10 years. The gray contour lines indicate a significance level of 99% based on t-test with unfiltered time series.

shift in the transient eddy storm tracks over the North Atlantic and Pacific, located over the Gulf Stream and Kuroshio Current. The corresponding AMET variations, which links to the compensation directly, contain additional information about the occurrence of Bjerknes compensation as shown in Figure 3.4c, f and i, and 3.6a,d and g. Regarding the AMET response to the variations of OMET in the subpolar Atlantic, there are multiple explanations provided by early studies with numerical climate models. Shaffrey and Sutton [2006] found that a stronger OMET in the North Atlantic Ocean leads to a weakened atmospheric transient energy transport, and thus the Bjerknes compensation. Van der Swaluw et al. [2007] also claimed that the compensation is due to the variations of eddy components in their experiments, especially for the transient synoptic eddies. They found that the vertical shear of zonal wind decreases when OMET increases (see Figure 7a and b in their paper), and this implies a decrease of baroclinicity. Consequently, the decreasing baroclinicity results in less eddy activity and this will lead to the changes in AMET. However, our findings about the changes of eddies are opposite to those from Van der Swaluw et al. [2007] (see Figure 3.7). Moreover, we also notice that there is a shift of Ferrel Cell due to the OMET variations. This encourages us to perform a decomposition of eddy components with AMET and compare them with the energy transported by steady mean flow. The regression of AMET anomalies transported by mean flow and eddy components on OMET on multiple pressure levels (from 850 hPa to 200 hPa) at 60°N in winter (DJF) is given in Figure 3.9. Note that the regression was performed at decadal time scales and the ocean leads by 1 month. In general, the chosen reanalyses products agree well on the energy transport by eddies, especially around 60°N. The magnitudes of regression coefficients differ from product to product, though. Note that in MERRA2 at lower levels (850hPa) there is strong latitudinal variations in Figure 3.9e and f. This is due to the fact that GEOS data assimilation system used to produce MERRA2 does not extrapolate data to pressure levels greater than the surface pressure [Gelaro et al., 2017]. Similarly, the latitudinal variations of steady mean energy transport are consistent



in the chosen atmospheric reanalyses around  $60^\circ\text{N}$ . The magnitudes of regression coefficients are different, though.

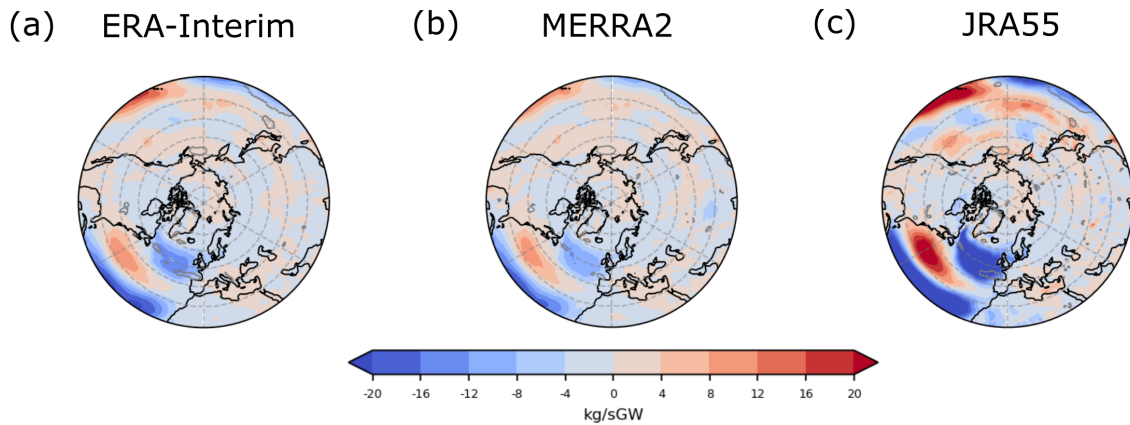


**Figure 3.9:** Regression of meridional energy anomalies transported by steady mean flow (a, d, g), transient eddies (b, e, h) and standing eddies (c, f, i) on multiple pressure levels on vertically integrated OMET anomalies at  $60^\circ\text{N}$  in winter (DJF) at decadal time scales when OMET leads by one month. The regression was performed on monthly fields from (a,b,c) ERA-Interim, (d,e,f) MERRA2 and (g,h,i) JRA55 after taking a running mean of 10 years, with positive sign indicating a northward flow.

To conclude, across interannual to decadal time scales we find that Bjerknes compensation occurs in both subtropical and subpolar regions in reanalysis data. We also investigated the mechanism behind the compensation in the subpolar region. Unlike the numerical experiments which attribute the Bjerknes compensation to the adjustment by the eddy heat transport components [Shaffrey and Sutton, 2006, Van der Waluw et al., 2007], we find that the compensation is primarily achieved by the changes of mean flow in the atmosphere in response to the OMET variability, thus a shift of the Ferrel cell, driven itself by eddy momentum fluxes around  $60^\circ\text{N}$  due to changes in baroclinicity in winter. Horizontally, the shift of the Ferrel Cell leads to a shift of AO/NAO like pattern, which confirms the early results with numerical model by Jungclaus and Koenigk [2010], as they also found a shift of patterns in atmospheric circulations due to compensation. In summer, there is hardly any compensation between the atmosphere and ocean.

### 3.3.3 Drivers for the OMET variations

In the last section, we explored the atmospheric response to the OMET variations with respect to atmospheric circulation and heat transport variations. Now we turn to the drivers of OMET variations. In this section, we study wind and buoyancy forcing of the ocean by the atmosphere. It is well established that the atmosphere can force the ocean at the surface by wind stress and buoyancy flux, and the latter is related to the thermohaline circulation, which is closely linked to variations of AMOC [e.g. Kuhlbrodt et al., 2007, Thomas et al., 2014]. At short time scales, a shallow upper ocean layer adjusts quickly to wind stress variations, which causes the anomalous wind driven Ekman transports. The relation between variations of meridional mass transport in this Ekman layer and changes of OMET anomalies at 60°N in winter (DJF) is illustrated in Figure 3.10. The atmosphere (i.e. surface wind stress) leads by one month. Again, similar results are obtained with the chosen atmospheric reanalysis products, the surface wind stress is larger in JRA55 than the other two products, though. Around 60°N in the North Atlantic, where the subpolar gyre resides, OMET is correlated to the Ekman transports as expected. It indicates that atmosphere variability can influence the OMET variations through wind stress anomalies. This is consistent with Shaffrey and Sutton [2004], as they also found the variability of OMET in the North Atlantic is dominated by Ekman processes. Outten and Esau [2017] noticed that the variations in the strength of the subpolar gyre were well correlated to AMET but not well correlated to OMET, despite AMET and OMET being well correlated to one another. This indicates that the atmosphere could influence the ocean at decadal time scales by causing a spin up or spin down of the subpolar gyre. However, this relationship is not robust when they examined it in many CMIP5 models [Outten et al., 2018].

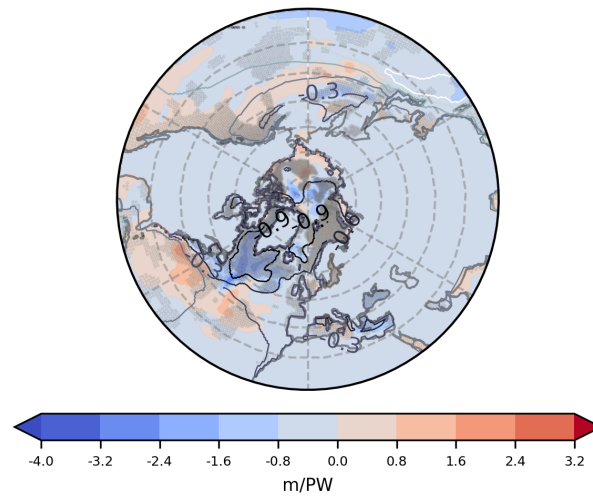


**Figure 3.10:** Regression of meridional Ekman transport anomalies on vertically integrated OMET anomalies at 60°N in winter (DJF) at decadal time scales when atmosphere leads by one month. The meridional Ekman transport anomalies are calculated as  $-\frac{\tau_x}{f}$  with  $\tau_x$  the zonal wind stress and  $f$  the coriolis parameter. The regression was performed on monthly fields from (a) ERA-Interim, (b) MERRA2 and (c) JRA55 after taking a running mean of 1 year. The gray contour lines indicate a significance level of 99% based on t-test with unfiltered time series.

The atmosphere can further affect OMET through modifying the intensity of gyres or initiating a shift of the location of gyres and the compensating western boundary current strength through



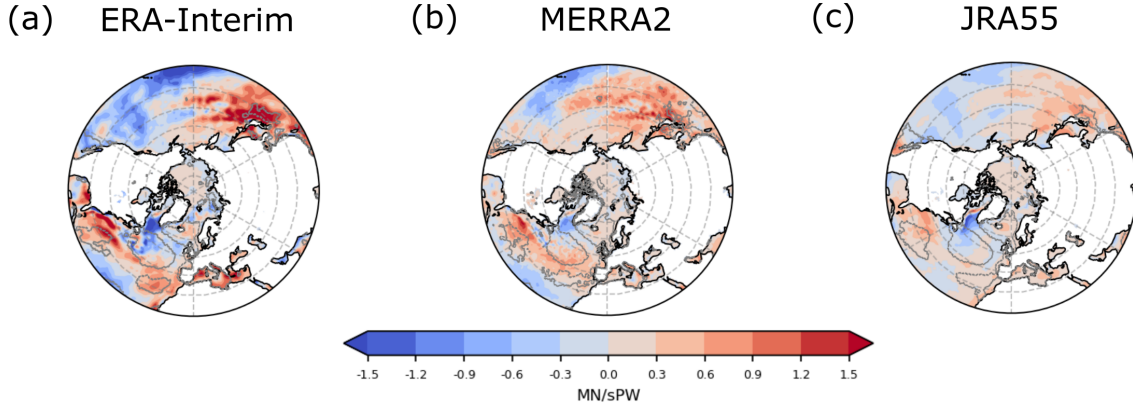
wind stress curl variations [e.g. Sverdrup et al., 1942, Thomas et al., 2014]. Since the impact of wind stress curl on gyres are reflected in sea surface height (SSH), by exploring the relation between SSH anomalies and OMET variations, we can further investigate the drivers for wind-driven OMET variations. We performed regression of SSH anomalies on OMET anomalies at 60°N (see Figure 3.11). A correlation pattern and an anticorrelation pattern are observed at the Gulf Stream extension and the Labrador Sea, respectively. This is consistent with Figure 3.10. We can conclude that the atmospheric winds influence OMET variations through anomalous Ekman transport and variations in the strength of gyres in the North Atlantic, which is consistent with early studies [e.g. Kuhlbrodt et al., 2007].



**Figure 3.11:** Regression of sea surface height anomalies on vertically integrated OMET anomalies at 60°N at decadal time scales. The regression was performed on monthly fields from ORAS4 after taking a running mean of 10 years, including all seasons. The mean sea surface height is illustrated by contour lines with an unit of meter. The stippling indicates a significance level of 99% based on t-test with unfiltered time series.

Moreover, the atmosphere can influence the thermohaline circulation of the ocean through surface buoyancy fluxes. The thermohaline processes affect the AMOC, and thus the variations of OMET [e.g. Kuhlbrodt et al., 2007, Smeed et al., 2014]. The total buoyancy flux consists of a thermal flux and a haline flux. We have already elucidated the influence of SFlux on OMET in Figure 3.5. To provide a complete picture, here we investigate the linkage between anomalous buoyancy flux variations and OMET anomalies. The regression of buoyancy flux anomalies on OMET anomalies in winter (DJF) at decadal time scales is shown in Figure 3.12. Among the chosen atmospheric reanalysis products, a strong coupling between OMET and buoyancy forcing is observed in the subtropical Atlantic, while an anticorrelation between them is found in the subpolar Atlantic. The patterns resemble the correlation map in Figure 3.5, which indicates that the buoyancy forcing is dominated by its thermal component. The anticorrelation between OMET and buoyancy forcing in the Labrador Sea and the Irminger Sea indicates that weaker buoyancy fluxes are associated with stronger OMET in these regions. This means that it is unlikely that the OMET variations are driven by the AMOC associated with convection in this region at these short time scales. In that case, one would have expected the reverse sign.

The patterns seem consistent with the wind variations associated with the OMET variability and therefore the confounded Ekman-driven OMET variations may cause the high correlations. However, it should be noted that large uncertainties have been found by Karspeck et al. [2017] regarding the location and intensity of deep convection in different reanalysis products. Also, advective feedbacks associated with anomalous temperature and salinity transports in the ocean can drive AMOC and OMET variations [e.g. Jackson, 2013].



**Figure 3.12:** Regression of buoyancy forcing anomalies on vertically integrated OMET anomalies at 60°N at decadal time scales without time lags. The buoyancy forcing is calculated as  $\frac{g\alpha}{c_{p0}}Q_s + gS\beta(P-E)$  with  $g$  the gravitational acceleration,  $c_{p0}$  the specific heat capacity of sea water,  $Q_s$  the net sea surface heat flux,  $S$  the surface layer salinity,  $\alpha$  and  $\beta$  the thermal expansion coefficient and saline expansion coefficient,  $P$  the precipitation and  $E$  the evaporation. For simplification, we use constant  $c_{p0} = 3987 J/kg^\circ C$ ,  $S = 35 psu$ ,  $\alpha = 250 \times 10^{-6} K^{-1}$  and  $\beta = 7.5 \times 10^{-4} psu^{-1}$ . The regression was performed on monthly fields from (a) ERA-Interim, (b) MERRA2 and (c) JRA55 after taking a running mean of 10 year. The regression coefficients with a unit of  $N/sPW$  are shown spatially. The gray contour lines indicate a significance level of 99% based on t-test with unfiltered time series.

In summary, the oceanic response to the variations in the atmosphere at the time scales considered here is primarily wind-driven. Together with the atmospheric response to the variations in the ocean, they form a physical picture of the air-sea interaction at subpolar Atlantic which can lead to Bjerknes compensation in reanalysis datasets. It should be noted that remote processes that could affect the ocean have not been considered here.

### 3.4 Conclusion and Discussion

In this paper, we studied the interactions between atmosphere and ocean, and their relation to meridional energy transport. We quantified AMET and OMET using six reanalysis products and explored the Bjerknes compensation within the chosen reanalysis datasets. Moreover, we wanted to understand the physical mechanism of the compensation and we revisit the roles of atmosphere and ocean in terms of their forcing and response. Our work is motivated by previous studies on Bjerknes compensation based in numerical climate models only.

We find that the chosen reanalysis datasets agree well on the mean AMET and OMET in the Northern Hemisphere. Our findings are consistent with the results provided by previous studies

[e.g. Trenberth and Caron, 2001, Fasullo and Trenberth, 2008]. The differences between AMET anomalies given by the chosen atmospheric reanalyses are small. This is the same for OMET anomalies in the chosen oceanic reanalysis products, except for GLORYS2V3. Consequently, we chose ERA-Interim, MERRA2, JRA55 and ORAS4 as our benchmark to further investigate the interaction between atmosphere and ocean. From interannual to decadal time scales, we demonstrate that the Bjerknes compensation occurs in the Northern Hemisphere at almost all the latitudes from 40°N to 70°N in reanalysis data. The intensity of compensation correlates to the length of time scale, which is consistent with Shaffrey and Sutton [2006]. The compensation shows clearly in the surface flux, which is necessary because net radiation flux variations at TOA are small.

Further inspection of the mechanism of Bjerknes compensation at decadal time scales in winter shows that the atmosphere responds to OMET variations with a shift in the Ferrel Cell. This shift is likely to be driven by eddy momentum fluxes due to changes in baroclinicity. Horizontally, this shift shows an AO/NAO like pattern, which is also found by Jungclaus and Koenigk [2010] in their numerical model. The corresponding AMET variations are primarily found in the zonal mean part and less in the eddy heat transport part. This is different from previous studies with numerical models [Shaffrey and Sutton, 2006, Van der Swaluw et al., 2007], but they address larger time scales. This proposed mechanism is robust across the different atmospheric reanalysis data sets used in this study. It is noteworthy that this mechanism does not apply to summer, as in summer there is hardly any compensation between the atmosphere and ocean meridional heat transports. It should be noted that there is a big caveat due to mass budget imbalances [Trenberth, 1991, Graversen, 2006, Trenberth et al., 2014].

Moreover, the oceanic response to the changes in the atmosphere is illustrated. The atmosphere forces the surface ocean through wind stress and buoyancy fluxes [e.g. Kuhlbrodt et al., 2007, Gregory and Tailleux, 2011, Thomas et al., 2014]. The surface wind anomalies, which leads to the Ekman transport and Sverdrup transport variations, modify the OMET through an intensification of gyre transports and anomalous Ekman transport. On the other hand, the thermal flux and the saline flux, which constitute a buoyancy flux, could influence the thermohaline processes in the ocean and their influence is then reflected in the variations of AMOC [Timmermann and Goosse, 2004, Gregory and Tailleux, 2011, Yang et al., 2016]. However, at decadal time scales the oceanic response to the variations in the atmosphere appears to be primarily wind-driven. It should be noted that the time series are too short to unequivocally determine forcing-response relations. With our analysis we can not decide whether the ocean drives the atmospheric MET variations, or it is the reverse. The question of causality is beyond the scope of this study.

Note that different methods for the quantification of AMET can potentially lead to different results Armour et al. [2019]. We are aware of the two practical ways to calculate AMET with atmospheric reanalysis datasets. Apart from the implied method, which takes AMET as the residual between SFlux and TOAFlux as shown in this paper, a direct method that is used by many studies is to integrate the divergence of AMET over the entire air column [Trenberth, 1991, Graversen, 2006, Fasullo and Trenberth, 2008, Mayer and Haimberger, 2012]. Mayer and Haimberger [2012] and Armour et al. [2019] elaborate on the difference between AMET computed using these two different methods. The method based on vertical integration suffers

from problems like mass imbalance, unrealistic moisture budget, and sparseness of observations [Trenberth, 1991]. We also computed AMET as the vertical integral of AMET divergence over the air column in a previous paper [Liu et al., 2020b], but the variability of AMET anomalies given by these three atmospheric reanalysis products are very different, which indicates a lack of constraints on AMET in reanalysis products [Liu et al., 2020b]. A comparison of low frequency AMET anomalies at 60°N with implied and direct methods using all three chosen atmospheric reanalysis products was performed (not shown). With the direct method, the correlation coefficients of low frequency AMET anomalies at 60°N between ERA-Interim and JRA55 is 0.82, and between ERA-Interim and MERRA2 is -0.53. While those correlation coefficients are 0.94 and 0.97 with implied methods. Moreover, with the AMET estimated as vertical integrals, a consistent picture of Bjerknes compensation was not observed and the energy flow between atmosphere and ocean even contradicts the surface and TOA energy budget (not shown). Hence we chose to compute AMET as the residual between SFlux and TOAFlux as in this way we can obtain a physically consistent picture.

Nevertheless, uncertainties are large in the SFlux and TOAFlux fields. Huang et al. [2017] compared the radiation flux at TOA and surface given by ERA-Interim, MERRA2 and JRA55 with NASA CERES. He found that the radiation fluxes at TOA given by the chosen datasets agree well in winter but deviate against the observation in summer. For the surface flux, short wave radiation differs more in summer while longwave radiation differs more in winter. It seems the energy fluxes at TOA is more trustworthy than the energy fluxes at the surface [Huang et al., 2017]. In addition, due to a lack of turbulent fluxes observations with global coverage, currently reanalysis products cannot assimilate turbulent fluxes at the surface. This also leads to large uncertainties in SFlux in reanalyses, which is confirmed by Liu et al. [2017] in their study.

In this paper, we demonstrate that the response of Ferrel Cell contributes to the Bjerknes compensation in reanalysis datasets at mid-high latitudes in the atmosphere. It is also interesting to investigate the other processes that are related to this change. For instance, the coupling between Ferrel Cell and Hadley Cell suggests that there might be a low latitude driver for the changes of mean flow in Ferrel Cell. Also, variations can be driven by the tropical ocean or the AMOC [Trenberth and Stepaniak, 2003]. In this case, we need to explore the teleconnections between low latitudes drivers (e.g. ENSO) and mid-high latitudes weather and climate. Much longer time series are needed to obtain statistically significant results and avoid uncertainties. Here, we find that at large time scales the Bjerknes compensation occurs and peaks not only at subpolar latitudes but also at mid-to-low latitudes. Specifically, we notice the highest compensation rate is present at 40°N at decadal time scales at the subtropics when the ocean leads, which is worth a visit for future work. Also, analysis can be extended to the Southern Hemisphere and the relation between global warming and variability of Bjerknes compensation is of interest for further studies.

## Acknowledgements

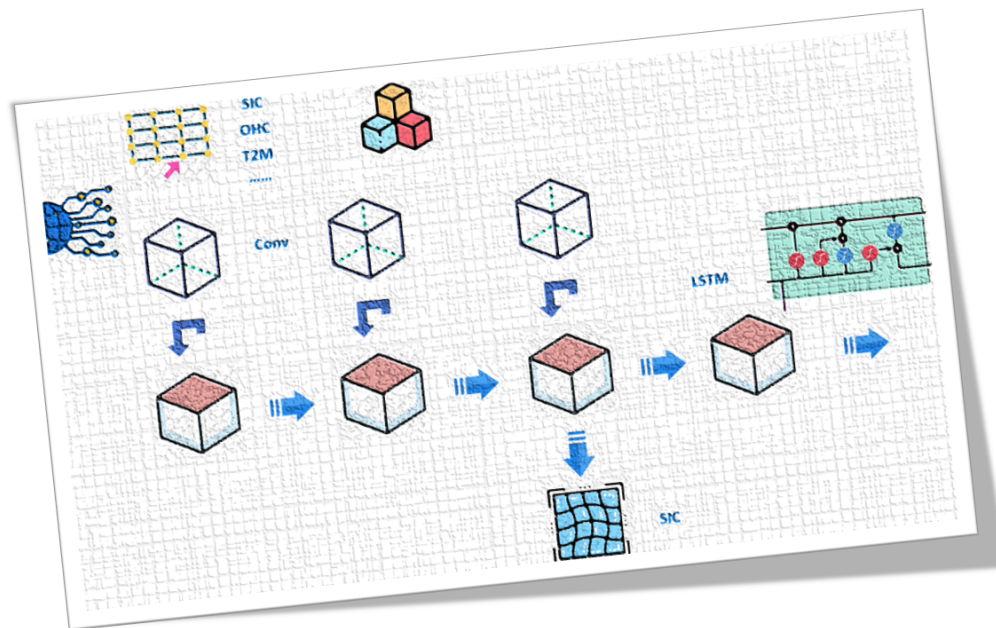
This research was supported by the Blue Action project (European Union's Horizon 2020 research and innovation programme, grant number: 727852). The authors also gratefully acknowledge the support by the Netherlands eScience Center and Wageningen University. We would like

to thank SURFsara (Netherlands) for providing us their super computing infrastructure for our project. We also want to express our gratitude to the reanalysis data sets provided by ECMWF (EU), JMA (Japan), NASA (USA), Mercator Ocean (France) and University of Maryland. We also acknowledge the editor Dr. Isaac Held and our three anonymous reviewers for their help to improve the manuscript.



## Chapter 4

# Extended Range Arctic Sea Ice Forecast with Convolutional Long-Short Term Memory Networks



This chapter is based on:

Yang Liu, Laurens Bogaardt, Jisk Attema, and Wilco Hazeleger. Extended range arctic sea ice forecast with convolutional long-short term memory networks. *Monthly Weather Review*, 149(6):1673–1693, 2021b

## Abstract

Operational Arctic sea ice forecasts are of crucial importance to science and to society in the Arctic region. Currently, statistical and numerical climate models are widely used to generate the Arctic sea ice forecasts at weather time-scales. Numerical models require near real-time input of relevant environmental conditions consistent with the model equations and they are computationally expensive. In this study, we propose a deep learning approach, namely Convolutional Long Short Term Memory Networks (ConvLSTM), to forecast sea ice in the Barents Sea at weather to sub-seasonal time scales. This is an unsupervised learning approach. It makes use of historical records and it exploits the covariances between different variables, including spatial and temporal relations. With input fields from reanalysis data, we demonstrate that ConvLSTM is able to learn the variability of the Arctic sea ice and can forecast regional sea ice concentration skillfully at weekly to monthly time scales. It preserves the physical consistency between predictors and predictands, and generally outperforms forecasts with climatology, persistence and a statistical model. Based on the known sources of predictability, sensitivity tests with different climate fields as input for learning were performed. The impact of different predictors on the quality of the forecasts are evaluated and we demonstrate that the surface energy budget components have a large impact on the predictability of sea ice at weather time scales. This method is promising to enhance operational Arctic sea ice forecasting in the near future.



## 4.1 Introduction

As one of the most noticeable frontiers with visible changes due to global warming, the Arctic has received more and more attention in recent decades. This is accompanied with increased commercial and scientific activities as a result of sea ice melting. This drives a demand for reliable operational sea ice forecasts, especially for shipping companies and related stakeholders [Gascard et al., 2017, Stephenson and Pincus, 2018]. Therefore, it is of crucial importance to improve operational Arctic sea ice forecasts at weather time scales.

The physical interactions between atmospheric and oceanic conditions and Arctic sea ice provide a basis for forecasting sea ice characteristics. Predictability of Arctic sea ice at different time scales in different seasons has been explored extensively in many studies. Blanchard-Wrigglesworth et al. [2011a] investigated the temporal evolution of Arctic sea ice in observations and in ensemble climate model output. They found a summer to summer and a melt season to growth season reemergence effect in sea ice which potentially serves as a good predictor for Arctic sea ice forecasts at monthly to annual time scales. Mohammadi-Aragh et al. [2018] studied the potential predictability of Arctic sea ice in winter, including the deformation and concentration of sea ice at weather time scales. They noticed that the sea ice concentration (SIC) is predictable throughout a 10-day forecast period.

In addition there are studies on longer range forecasts. Guemas et al. [2016] reviewed progress on sea ice forecasts and showed that predictability of Arctic sea ice at seasonal to decadal time scales mainly originates from persistence or advection of sea ice anomalies, air-sea interaction, and changes in radiative forcing. Krikken and Hazeleger [2015] analysed the natural variability of Arctic sea ice from an energy budget perspective. They found strong correlations between the Arctic energy balance components and the reemergence of sea ice anomalies from the melt season to the growth season, which extends the theory proposed by Blanchard-Wrigglesworth et al. [2011a] and further confirms the essential role of the energy budget in sea ice forecasts. Another key element to the predictability of sea ice is the sea ice thickness (SIT). Bonan et al. [2019] explore the role of SIT on the summer predictability of sea ice and found that similar skill of SIT forecasts can be obtained as a perfect model experiment. They also discussed the predictability barrier in late spring and made suggestions for the initialization of forecasts.

Cruz-García et al. [2019] examined seasonal-to-interannual sea ice predictability with multiple climate models and revisited the essential role of the reemergence effect of sea ice anomalies. They observed that SIC anomalies in the Barents Sea have a strong negative correlation with the local sea surface temperature (SST) anomalies. Moreover, Onarheim et al. [2015] emphasized that ocean heat transport (OHT) variations play an important role in the observed winter sea ice variance in the Barents Sea. They claimed an increase in forecast skill at annual times scales up to 2 years using OHT. The knowledge of sea ice predictability associated with different physical processes provided by these studies underlines the importance of choosing relevant predictors related to the target location and time scale.

Currently, many operational Arctic sea ice forecasts are produced by numerical climate models [e.g. Van Woert et al., 2004, Metzger et al., 2014, Hebert et al., 2015, Smith et al., 2013, 2016]. These numerical models are built upon physical linkages in the climate system and are able to

generate accurate sea ice forecasts, but they are computationally expensive, due to the need for relatively high spatial and temporal resolutions, the implementation of ensemble approaches to address uncertainties, their dependency on the real-time input of observed conditions for the data assimilation processes, and the calibration of model output. Moreover, dynamical models are imperfect and many processes have to be parameterized. Specifically for sea ice there are several modelling challenges. For instance, rheology, ice thickness distribution, wave-ice interaction, landfast ice, melt ponding, and floe size distribution [Leppäranta et al., 2020]. Many studies have shown that the forecast skill strongly relies on the initialization [Blanchard-Wrigglesworth et al., 2011b, Goessling et al., 2016], target location and time scales [Cruz-García et al., 2019].

Some operational sea ice forecasts are generated by statistical models [Howell et al., 2015, Yuan et al., 2016, Wang et al., 2019]. Most of these statistical models are linear models, thus they are not suited to learn nonlinear relations between variables in the Arctic climate system. Given the importance of nonlinear feedback mechanisms in the atmosphere, ocean and sea ice coupled system in the Arctic, we may need nonlinear approaches to forecast Arctic sea ice with a statistical model. This brings contemporary machine learning techniques into scope.

Machine learning approaches, especially deep learning, are widely embraced by many fields and are increasingly used to deal with problems like clustering, classification and regression [LeCun et al., 2015]. Benefiting from large volumes of data of the Earth system [Knüsel et al., 2019], those deep learning methods may be appropriate for the weather and climate domain [Reichstein et al., 2019]. Although these applications still have limitations, for example, they rely on the data from numerical weather forecast or reanalysis for the training process and this could be computationally very expensive depending on the configuration as well as the tuning procedure, there are many successful use cases. These cases are, for instance, the representation of physical processes [e.g. Rasp et al., 2018], weather and climate forecasts [e.g. Salman et al., 2015, Ham et al., 2019], and extreme events detection [e.g. Gope et al., 2016]. Deep learning based techniques could potentially be used as an alternative method or an auxiliary approach for the current state-of-the-art forecast systems, or even as a preliminary and fast forecast system. It can be viewed as an enhancement or supplement to our existing tools.

In this study, we consider sea ice forecasts at weekly time scales and we perform extended range sea ice forecasts in the Barents Sea with a complex deep neural network (DNN), namely the Convolutional Long-Short Term Memory Networks. These intricate neural networks are built on top of the basic structures, like convolutional neural networks (CNN) and recurrent neural networks (RNN). Early studies have shown that even these basic neural networks (NN) are able to reproduce both the short-term evolution and the long-term statistics of dynamical systems like a Lorenz system [e.g. Hassanzadeh et al., 2019], which provides a basis for learning the nonlinear relations between meteorological fields and predicting the evolution of a weather-like system in a chaotic regime.

To work with such complex spatial-temporal sequence problems, ConvLSTM are useful [Xingjian et al., 2015]. As a novel combination of CNN and Long-Short Term Memory (LSTM) networks [Fukushima, 1980, Hochreiter and Schmidhuber, 1997], ConvLSTM were first introduced by Xingjian et al. [2015] when dealing with precipitation nowcasting. Until now, many studies

have shown that ConvLSTM are suitable for weather forecasts at different time scales, like precipitation forecasts [Xingjian et al., 2015, Kim et al., 2017], hurricane tracking and forecasting [Kim et al., 2019], sea ice concentration [Kim et al., 2020] and sea ice motion forecasts [Petrou and Tian, 2019]. However, most of those studies only incorporate a few variables and were mostly data-driven without physical insights. In this paper, except for the sea ice forecasts with ConvLSTM using multiple predictors, we also conducted a sensitivity analysis of predictors and focus on the physical consistency between sea ice and other meteorological fields in the trained and forecasted output.

The paper is organized as follows: The methodology and the data sets used in this study are described in section 2. The results are shown in section 3, including constrained forecasts, sensitivity tests of predictors and operational forecasts with ConvLSTM. The discussion and a brief summary of this study are given in section 4 and 5, respectively.

## 4.2 Data and Methodology

A detailed elaboration on the deep neural networks and data sets used in this study is given in this section. In addition, a brief summary about the hyper-parameter tuning of our neural networks and an overview of the evaluation metrics is included at the end of this section.

### 4.2.1 Convolutional Long-Short Term Memory Networks

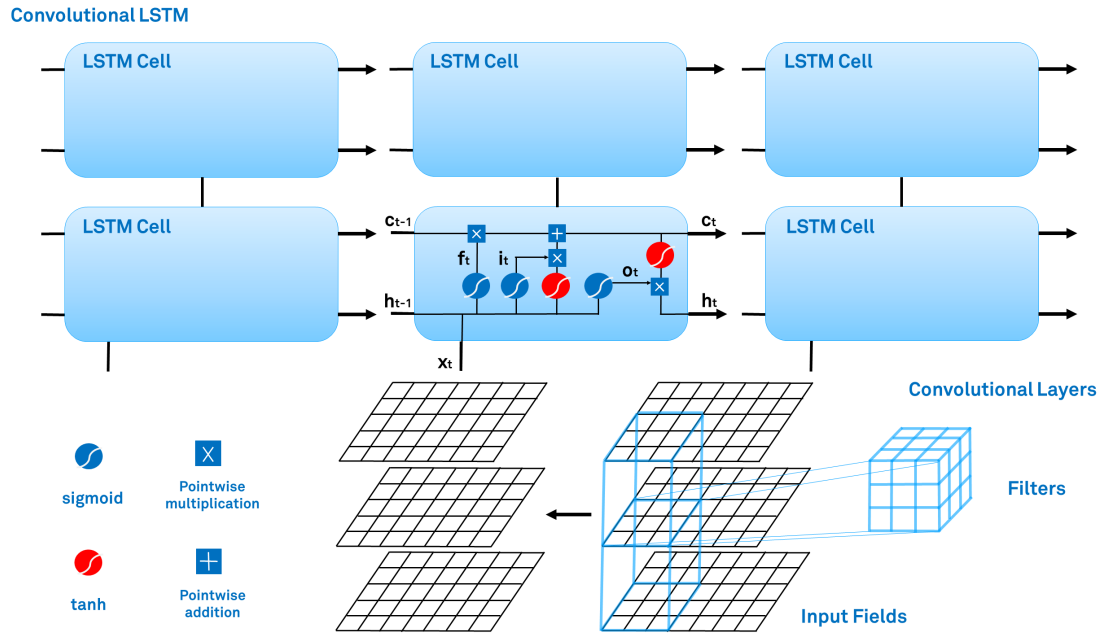
To enhance LSTM networks to include learning and forecasting spatial information, Xingjian et al. [2015] embedded convolutional cells into LSTM cells and created a new neural network structure coined ConvLSTM. Consequently, ConvLSTM inherits the ability of LSTM to “remember” and “forget”, which is achieved by the design of memory cells and multiple gates that control the flow of information [Hochreiter and Schmidhuber, 1997]. Also, the spatial awareness of a convolutional network is added to a LSTM. These aspects of the structure are relevant for weather and climate problems. The structure of ConvLSTM can be defined and explained by the following equations [Xingjian et al., 2015]:

$$\begin{aligned}
 i_t &= \sigma(W_{xi} * x_t + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \\
 c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o) \\
 h_t &= o_t \circ \tanh(c_t)
 \end{aligned} \tag{4.1}$$

With  $i_t$  the input gate,  $f_t$  the forget gate,  $c_t$  the cell state,  $o_t$  the output gate,  $h_t$  the hidden state,  $W$  the weight matrix,  $x$  the input,  $b$  the bias,  $*$  the convolutional operation,  $\circ$  the element-wise product,  $\sigma$  the sigmoid function and  $\tanh$  the hyperbolic tangent function. The subscripts describe the correspondence of the weight matrix to different gates and states. For instance,  $W_{xi}$  indicates the weight matrix of input values related to the input gate, while  $W_{hf}$  represents

the weight matrix of hidden states corresponded to the forget gate. The subscript  $t$  indicates the time step and will be elucidated in section 2.c.

The structure of the ConvLSTM network is illustrated in Figure 4.1. At each time step, convolutions over the input fields (e.g. data of the Arctic climate system) are performed. Then, at each grid point, those values are fed into an LSTM cell. The LSTM cells only differ in their input, and hence their memory, but share all other parameters. Multiple layers can be stacked to further increase the complexity of the network if needed.



**Figure 4.1:** Structure of the convolutional long short term memory neural networks.

ConvLSTM networks are powerful tools for intricate spatial-temporal sequence prediction problems. They are likely suitable for sea ice forecasts. Physically, the use of filters inside convolutional layers accounts for the local interactions between multiple fields (e.g. temperature, wind, etc) which affect the formation of sea ice, and the advance and retreat of sea ice at neighbouring grid points. The temporal evolution of sea ice, including the communication of neighbour points within the convolutional cells, are tracked by the LSTM structure of the network through its recurrence feature. Moreover, this approach is unsupervised learning and it can make use of historical records of weather and climate states.

In this study, we perform many-to-one prediction, which means sequences with spatial structure are taken as input and spatial maps of one time step ahead will be the output by the networks. The numerical processes, including training and testing of ConvLSTM, are elaborated upon in detail in the appendix. The ConvLSTM used in this study are constructed on top of the Pytorch library, and our script is published on Github (<https://github.com/geek-yang/DLACs>).

### 4.2.2 Reanalysis Data Sets

We train ConvLSTM and evaluate its capability to forecast sea ice using reanalysis data sets, namely ERA-Interim and ORAS4.

ERA-Interim is a global atmospheric reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [Dee et al., 2011], which covers the data-rich period since 1979. It employs the cycle 31r2 of ECMWF’s Integrated Forecast System (IFS) and generates atmospheric state estimates using 4D-Var data assimilation with a T255 ( $\sim 79$ km) horizontal resolution on 60 vertical levels [Berrisford et al., 2009]. We use the surface fields, including SIC, 2 meter temperature (T2M), sea level pressure (SLP), net surface turbulent and radiation flux (SFlux), 10 meter zonal and meridional wind (UV10m), and geopotential height at 850hPa (Z850) and 500hPa (Z500), with a  $0.75^\circ \times 0.75^\circ$  horizontal resolution ( $\sim 28.6$  km  $\times$  28.6 km, at  $70^\circ$ N). We take 6-hourly data with a range from 1979 to 2016 and the data is averaged to weekly time scales.

Given a lack of detailed verification studies of the SIC quality in ERA-Interim, we further verify the SIC field in ERA-Interim with a satellite-based product, namely the NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration (version 3) [Peng et al., 2013]. A point-to-point comparison of the SIC in the Barents Sea between ERA-Interim and the chosen satellite product was performed (not shown). They look similar in most of the areas, except for some places close to the continent as a result of land-sea mask and interpolation since the native grids are different. To ensure the robustness of our results, we also perform our training and testing with the SIC field from the chosen satellite-derived product and the results are almost the same as those based on the SIC in ERA-Interim.

The Ocean reanalyses System 4, in short ORAS4, is the replacement of the reanalyses system ORAS3 used by the ECMWF [Balmaseda et al., 2013]. It implements Nucleus for European Modelling of the Ocean (NEMO) as ocean model [Madec, 2008, Ferry et al., 2012a] and uses NEMOVAR as the data assimilation system [Mogensen et al., 2012]. The model is forced by atmosphere-derived daily surface fluxes, from ERA-40 from 1957 to 1989 and ERA-Interim from 1989 onwards. ORAS4 produces analyses with a 3D-Var FGAT assimilation scheme and spans from 1958 to the present. ORAS4 runs on the ORCA1 grid, which is associated with a horizontal resolution of  $1^\circ$  in the extratropics and a refined meridional resolution up to  $0.3^\circ$  in the tropics. It has 42 vertical levels, 18 of which are located in the upper 200m. We use the monthly mean temperature on the native model grid from 1979 to 2016 to calculate the ocean heat content (OHC) from the sea surface to 300m. Given the long memory effect in the ocean, OHC at weekly time scales is interpolated from monthly fields.

We chose this combination because both data sets are ECMWF reanalysis products and ORAS4 takes surface forcing from ERA-Interim [Balmaseda et al., 2013]. This combination is promising to provide a physically consistent picture of the interaction between the atmosphere, ocean and sea ice related processes. The selected fields are potential predictors for the sea ice variations due to their physical relationships in the Arctic [Krikken and Hazeleger, 2015, Guemas et al., 2016].

In this study, we only focus on the Barents Sea. Following the same definition of the Arctic re-

gions as Walsh et al. [2019], our domain is covered by the ERA-Interim grid with 24 (latitudinal) x 56 (longitudinal) points. It is noteworthy that this regional focus has a negative impact on the performance of ConvLSTM. The sampling of convolutional layers is affected by the cut-off of data close to the boundary of the Barents Sea. This partially explains the relatively bad forecast quality in the boundary regions, as we will show, and provides room for improvement if a larger area is included in the future. However, this comes with a computational cost.

### 4.2.3 Evaluation Metrics

Two types of lead time dependent forecast were performed in this study. One is called *constrained forecast*, which takes input fields from the future, excluding SIC only, to test the maximum expected predictability given the chosen forecast methods and input fields. It can be described with the equation shown below:

$$SIC_{pred[t_n+l+1]} = ConvLSTM(SIC_{obs[t_1, t_2, \dots, t_n]} + pred[t_{n+1}, t_{n+2}, \dots, t_{n+l}], OHC_{obs[t_1, t_2, \dots, t_n, t_{n+1}, t_{n+2}, \dots, t_{n+l}], \dots) \quad (4.2)$$

With  $OHC_{observe[t_n+l]}$  the observed OHC for the  $l$  leading week. An extended analysis of the contributions from several predictors was conducted based on the constrained forecast formulation.

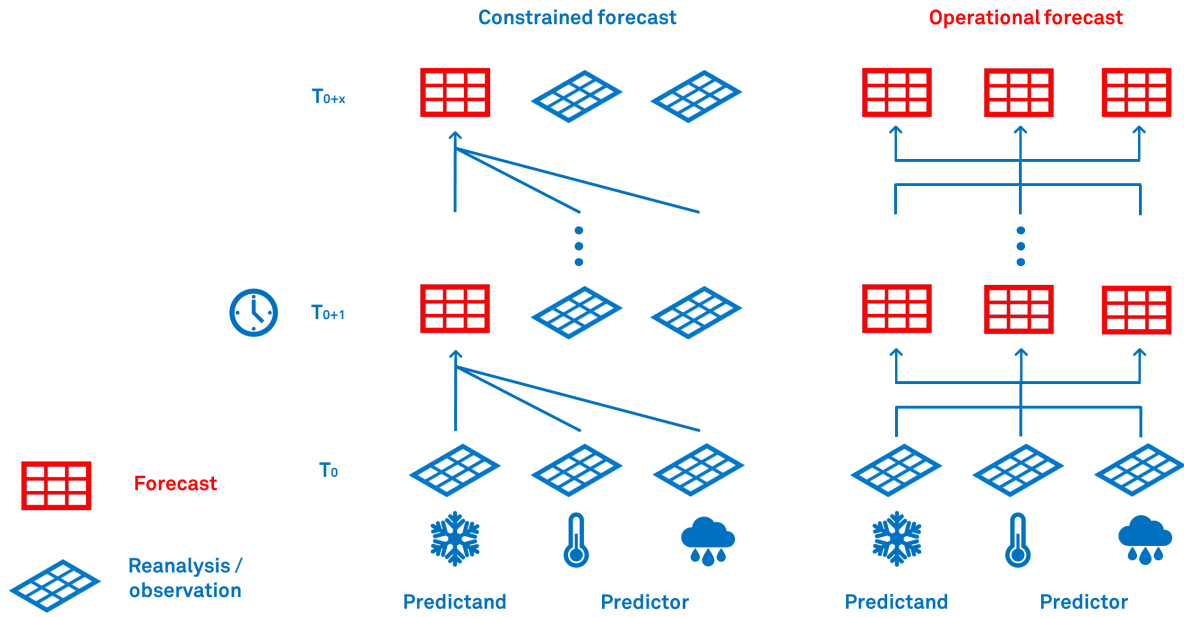
The other setup is called *operational forecast*, which uses only historical records and the forecasts at a specific lead time are based on the predicted fields of all variables (e.g. the week  $n + 2$  forecast is made with all variables of the reanalysis time series considered until the current step  $n$  and the week  $n + 1$  forecast). The procedure can be explained by the equation below:

$$SIC_{pred[t_n+l+1]}, OHC_{pred[t_n+l+1]}, \dots = ConvLSTM(SIC_{obs[t_1, t_2, \dots, t_n]} + pred[t_{n+1}, t_{n+2}, \dots, t_{n+l}], OHC_{obs[t_1, t_2, \dots, t_n] + pred[t_{n+1}, t_{n+2}, \dots, t_{n+l}], \dots) \quad (4.3)$$

With  $SIC_{obs[t_1, t_2, \dots, t_n]}$  and  $OHC_{obs[t_1, t_2, \dots, t_n]}$  the time series of observed SIC and OHC until the current time step,  $SIC_{pred[t_n+l]}$  and  $OHC_{predict[t_n+l]}$  the predicted SIC and OHC for the  $l$  leading week. We can assess the performance of ConvLSTM with re-forecasts, i.e. performing forecasts over the reanalysis period as if they were actual operational forecasts.

The configurations of constrained forecasts and operational forecasts are illustrated in Figure 4.2. The major differences between these two setups are: the constrained forecasts use predictors from the future to predict SIC, whereas the operational forecasts use forecast predictands as predictors to predict future SIC. The value of considering the constrained forecasts is to gain insight into what forecasts of SIC would look like if the ConvLSTM models were able to perfectly forecast the predictors that are then used in the operational forecasts.

In order to evaluate the performance of sea ice forecast by the ConvLSTM, several scores are calculated. The Root Mean Square Error (RMSE) is used to evaluate the sea ice forecast in the



**Figure 4.2:** Configurations of constrained forecast and operational forecast at starting time step  $T_0$  and lead time step  $T_{0+x}$ .

chosen area. To evaluate the predicted data with both temporal and spatial information, we define the RMSE in the following way:

$$RMSE = \frac{1}{N} \sum_{t=1}^N \sqrt{\frac{[\sum_{x=1}^X \sum_{y=1}^Y a_{x,y} [SIC_{x,y,t}^{(predict)} - SIC_{x,y,t}^{(observe)}]]^2}{\sum_{x=1}^X \sum_{y=1}^Y a_{x,y}}} \quad (4.4)$$

With  $x$  the number of points in the longitudinal direction,  $y$  the number of points in latitudinal direction,  $X$  and  $Y$  the total number of grid cell length in longitudinal and latitudinal direction,  $t$  the number of time step,  $N$  the total time steps,  $a_{x,y}$  the area of grid cell denoted by indices  $x$  and  $y$ ,  $SIC_{x,y,t}^{(predict)}$  and  $SIC_{x,y,t}^{(observe)}$  the predicted and observed SIC, respectively.

The Mean Absolute Error (MAE) is used to assess the forecasts on a point-wise basis. The spatial structure is preserved (see Figure 4.4). MAE is defined as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^N |a_{x,y} [SIC_{x,y,t}^{(predict)} - SIC_{x,y,t}^{(observe)}]| \quad (4.5)$$

For several applications, such as shipping and navigation in the Arctic, it is also necessary to know if a certain area is open water or covered by sea ice. This requires a binary forecast. Following Van Woert et al. [2004] and Walsh et al. [2019], the grid boxes with sea ice concentration less than 15% are regarded as open water areas. Based on this criterion, we introduce the

integrated ice-edge error score (IIEE), which is a verification metric first proposed by Goessling et al. [2016], to better represent the performance of binary forecasts. The score is defined as the area where the forecast and the “truth” disagree on the ice concentration being above or below 15% and it can be further decomposed into two components, the overestimated ( $O$ ) and underestimated ( $U$ ) local sea ice extent, respectively:

$$\begin{aligned}
 IIEE &= O + U \\
 O &= \int_A \max(c_f - c_t, 0) dA \\
 U &= \int_A \max(c_t - c_f, 0) dA \\
 c_f, c_t &= \begin{cases} 1, SIC > 15\% \\ 0, SIC < 15\% \end{cases}
 \end{aligned} \tag{4.6}$$

where  $A$  is the area of interest, and subscripts  $f$  and  $t$  denote the forecast and the truth. By definition, overestimated local sea ice extent means the failure of predicting a sea ice free area as ice covered, while underestimated local sea ice extent is the failure of predicting a sea ice covered area as ice free.

Many studies assess their sea ice forecast systems against persistence and climatology at weather time scales [Van Woert et al., 2004, Metzger et al., 2014, Hebert et al., 2015, Smith et al., 2013, 2016]. The reason is that, at weekly to sub-monthly time scales, the persistence of sea ice anomalies is very high [Blanchard-Wrigglesworth et al., 2011a, Guemas et al., 2016]. Therefore, it is very challenging to beat persistence at these time scales. For instance, Van Woert et al. [2004] provide daily ice analyses and 5-day forecasts with their polar ice forecast system but this has almost no skill in winter against persistence. Consequently, in this study, forecasts with the ConvLSTM using different input fields will also be evaluated against persistence and climatology. The persistence is defined as the SIC anomaly at lead time step 0 added to the climatology at each lead time. We use the climatology based on a 10-year sliding window preceding the forecast time [e.g. Zampieri et al., 2018] in order to take changes in the climatology into account.

In order to compare the ConvLSTM with numerical model-based forecasts, we include two ensemble forecast data sets from the sub-seasonal to seasonal prediction project [Vitart et al., 2017]. We use forecasts from the National Centers for Environmental Prediction (NCEP) and the ECMWF ensemble forecasts. These data sets were chosen because of their active sea ice model, available time range (Jan 2015 - Dec 2016), and the forecast frequency.

The NCEP global ensemble forecast system generates real time forecasts using the NCEP Climate Forecast System Version 2 (CFSv2) [Saha et al., 2014]. It consists of 16 ensemble members and the forecast length is 45 days. The atmospheric model has 64 model levels and its horizontal resolution is T126 ( $\sim 100km$ ), which is lower than the ERA-Interim (T255,  $\sim 80km$ ). Its ocean model is GFDL MOM4 [Pacanowski et al., 1991]. It has a spatial resolution in the zonal direction of  $0.5^\circ$  and in the meridional direction,  $0.25^\circ$  from  $10^\circ S$  to  $10^\circ N$ , progressively decreasing to  $0.5^\circ$  from  $10^\circ$  to  $30^\circ$ , and is fixed at  $0.5^\circ$  beyond  $30^\circ$  in both hemispheres. There



are 40 levels in vertical. The system is coupled to an active sea ice model, which is part of the Modular Ocean Model (MOM4) [Pacanowski et al., 1991].

Based on the Integrated Forecasting System (IFS) version CY46R1, the ECMWF global ensemble forecast system has 51 members and it runs twice a week up to day 46. The atmospheric component of the system has a horizontal resolution about  $16km$  up to day 15 and a relatively coarse horizontal resolution about  $32km$  after day 15. Vertically, the atmospheric model has 91 model levels. The ocean model is NEMO3.4.1 with a  $0.25^\circ$  horizontal resolution, 75 vertical levels [Madec, 2008, Ferry et al., 2012a]. The system is coupled to the Louvain-la-Neuve Sea Ice Model (LIM2) [Rousset et al., 2015].

Note that the output from the forecast systems in the S2S project have been regridded to the same model grid. To enable a direct comparison, we interpolate our ConvLSTM results and ERA-Interim sea ice data to the model grid used in the S2S project and weigh the results the same as the area weight applied throughout the paper. More information about these experiments can be found on the homepage of the S2S project (<https://confluence.ecmwf.int/display/S2S/Models>).

Sensitivity tests of the contribution from each predictor to the skill of sea ice forecasts are conducted with the ConvLSTM using SIC plus one extra predictor. To assess the change of forecast skill with different predictors, we introduce a dimensionless score with the definition given below:

$$Relative\ forecast\ skill\ score = \frac{RMSE_{sic} - RMSE_{predictor}}{RMSE_{sic}} \quad (4.7)$$

With  $RMSE_{sic}$  the RMSE of forecast with ConvLSTM using only SIC, and  $RMSE_{predictor}$  the RMSE of forecast with ConvLSTM using SIC and one extra predictor. More details are provided in the section 3b.

#### 4.2.4 Training and Hyperparameter Tuning

The networks are trained with reanalysis data from 1979-2008 (1440 weeks). Given the spatial resolution of input fields, the training set includes  $24 \times 56 \times 1440$  points, thus 1440 points for each node in the convolutional layer. Data from 2009-2012 is used for cross-validation, allowing to implement an early stop module and to avoid overfitting. Data from 2013-2016 is taken as the test set for evaluation. The weight matrix of the ConvLSTM is updated by optimizing the loss function, for which we use mean square error (MSE):

$$MSE = \frac{1}{N} \sum_{t=1}^N \frac{[\sum_{x=1}^X \sum_{y=1}^Y a_{x,y} [SIC_{x,y,t}^{(predict)} - SIC_{x,y,t}^{(observe)}]^2]}{\sum_{x=1}^X \sum_{y=1}^Y a_{x,y}^2} \quad (4.8)$$

Physically, this loss function measures the difference between the actual and forecasted SIC.

The training time varies from 8 hours to 12 hours on a single GPU, depending on the choices of hyperparameters (e.g. number of epochs, filter size, number of layers). With an assessment based

on the RMSE of sea ice forecasts for the first leading week, the results from hyperparameter tuning show that a combination of learning rate equal to 0.01, 3 stacked ConvLSTM layers, filter size 3 x 3 and 1500 epochs is the best. We further inspected the learning curve with the chosen combination of hyperparameters. Although the loss stops decreasing after 600 epochs, the model's skill for anomalies continues to increase.

It is worthwhile emphasizing the importance of the convolutional filter size. In each convolutional layer, the filter size controls the exchange of information between neighbouring points. Physically, it accounts for the influence of for instance SIC, OHC or SLP from adjacent regions on the selected area (node). Given the physical consistency between regional atmospheric and oceanic fields, and the advection of sea ice anomalies [Blanchard-Wrigglesworth et al., 2011a, Guemas et al., 2016], this feature of the convolutional layer should improve the forecast skill. Similarly, the number of stacked ConvLSTM layers also relates to the communication of neighbouring points because of the filtering in each layer with convolutions.

#### 4.2.5 Baseline Statistical Model Analysis

In order to set a baseline for our forecasts with ConvLSTM with less complex statistical models and to provide insight into its ability to account for the non-linearity between multiple physical fields, we also fit a generalized linear model with a “logit” link function. For conciseness, it is referred to as the baseline statistical model in this paper. This method is similar to the logistic regression, which means the non-linear properties of the input fields are covered. In this approach, each spatial location is modelled separately. While the fitted parameters vary over space, the structure of the linear regression is the same everywhere. In particular, the SIC is predicted using an auto-regressive model which includes the value of the previous three weeks at that location, as well as the previous values of the T2M and OHC. In addition, the SIC of all neighbouring locations one week prior is used to establish whether spatial drift is a relevant factor. For all terms, only linear parts are included, although the percentage of sea ice content is first transformed using the “logit” function. Finally, the model is fit using ridge regression, where the optimal amount of regularisation is determined using 5-fold cross-validation over the training set. The predictors are selected in terms of the balance between their expected source of predictability and the cost of training. Similarly, it uses the “constrained forecast” configuration as all the input fields are from the reanalysis. Mathematically, this generalized linear model with a logit link function can be expressed as:

$$\begin{aligned}
SIC(t) &= \beta_t t + \beta_{\sin} \sin(t) + \beta_{\cos} \cos(t) + \overrightarrow{\beta_{SIC}} \cdot \overrightarrow{SIC(t)} + \overrightarrow{\beta_{T2M}} \cdot \overrightarrow{T2M(t)} \\
&\quad + \overrightarrow{\beta_{OHC}} \cdot \overrightarrow{OHC(t)} + \overrightarrow{\beta_{neighbours}} \cdot \overrightarrow{SIC_{neighbours}(t)}
\end{aligned}$$

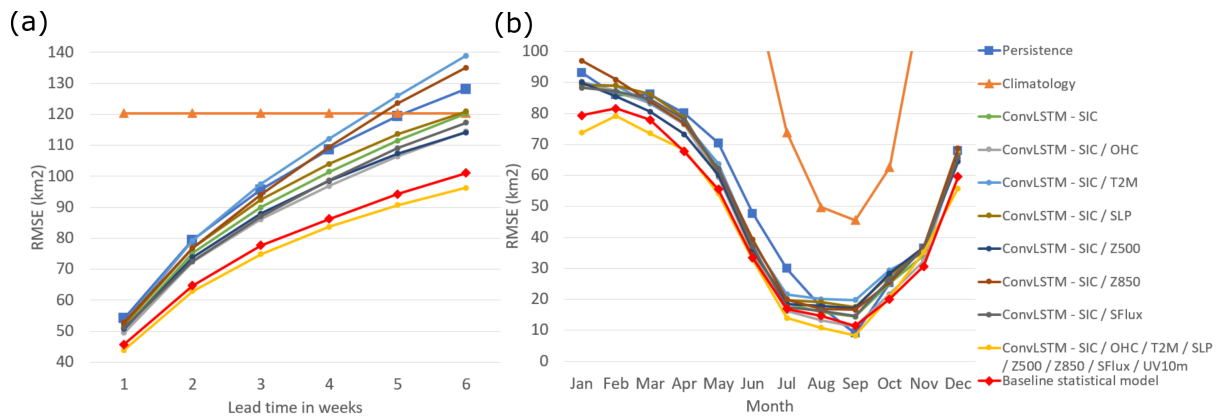
$$\begin{aligned}
\overrightarrow{SIC(t)} &= \begin{pmatrix} \text{logit}(SIC_{t-1, detrend}) \\ \text{logit}(SIC_{t-2, detrend}) \\ \text{logit}(SIC_{t-3, detrend}) \end{pmatrix} & \overrightarrow{T2M(t)} &= \begin{pmatrix} \text{logit}(T2M_{t-1, detrend}) \\ \text{logit}(T2M_{t-2, detrend}) \\ \text{logit}(T2M_{t-3, detrend}) \end{pmatrix} \\
\overrightarrow{OHC(t)} &= \begin{pmatrix} \text{logit}(OHC_{t-1, detrend}) \\ \text{logit}(OHC_{t-2, detrend}) \\ \text{logit}(OHC_{t-3, detrend}) \end{pmatrix} & \overrightarrow{SIC_{neighbours}(t)} &= \begin{pmatrix} \text{logit}(SIC_{t-1, detrend}^N) \\ \text{logit}(SIC_{t-1, detrend}^{NE}) \\ \text{logit}(SIC_{t-1, detrend}^E) \\ \text{logit}(SIC_{t-1, detrend}^{SE}) \\ \text{logit}(SIC_{t-1, detrend}^S) \\ \text{logit}(SIC_{t-1, detrend}^{SW}) \\ \text{logit}(SIC_{t-1, detrend}^W) \\ \text{logit}(SIC_{t-1, detrend}^{NW}) \end{pmatrix} \quad (4.9)
\end{aligned}$$

with  $t$  the current step,  $SIC(t)$  the sea ice forecast of current step,  $\sin(t)$  and  $\cos(t)$  cycles with one year period,  $\beta$  the trainable weight for each term,  $SIC_{t-x, detrend}$ ,  $T2M_{t-x, detrend}$  and  $OHC_{t-x, detrend}$  the detrended SIC, T2M and OHC at previous  $x$  time step,  $SIC_{t-1, detrend}^{direction}$  the detrended SIC of neighbouring point at previous time step at certain location (e.g. N indicates north, NE indicates north-east, and etc.), excluding land pixels.

## 4.3 Results

### 4.3.1 Constrained Predictability

Before implementing the novel deep neural network for sea ice forecasts and performing retrospective analysis of skill, it is worthwhile examining its capability in an ideal set-up. We performed lead-time dependent constrained forecasts of SIC with the ConvLSTM network using different combinations of input fields. The RMSE of these forecasts against those given by persistence and climatology with a lead time up to 6 weeks is shown in Figure 4.3a (details in Table S2). For all the forecasts (except for climatology as it does not vary over time, by definition), RMSE increases with increased lead time. It is observed that most of the forecasts with the ConvLSTM using SIC and one extra field, such as OHC, Z500, SLP and SFlux, can outperform the forecasts with persistence and climatology. The increase of RMSE as a function of lead time from these forecasts with the ConvLSTM is smaller than that with persistence and climatology. However, this is not true for some combinations of input fields, for instance, SIC with T2M or Z850. To obtain more insight on the impact of several predictors, an extended constrained SIC forecast with lead time up to 16 weeks is shown in Figure 4.7. It is found that with multiple input fields, the performance of ConvLSTM is also stable at long lead times. This time scale is beyond the scope of this study though.



**Figure 4.3:** RMSE of (a) the constrained forecast of SIC with a lead time up to 6 weeks and (b) the constrained forecast of SIC for the first week in each month with ConvLSTM using different predictors against persistence, climatology and the baseline statistical model. The unit is square kilometer per grid cell.

Forecasts with a ConvLSTM using only SIC also provides better results than forecasts from persistence and climatology. The non-linearity introduced by the ConvLSTM effectively contributes to the skill of the forecast. A comparison between forecasts with our baseline statistical model and forecasts using the ConvLSTM shows that the ConvLSTM produces slightly better forecasts. They both significantly outperform a persistence forecast.

As SIC has a strong seasonal cycle and the predictability is known to be strongly seasonal dependent, we further evaluate the forecasts by examining the error in each month. The RMSE of the constrained forecast of SIC for the first week in each month with different predictors and methods is given in Figure 4.3b (details in Table S3). In general, the forecast error is larger in winter than that in summer, which is similar to operational forecasts with numerical models [Smith et al., 2016]. This can be explained by the large year-to-year variations of sea ice in winter and a large open water area in summer [Perovich and Richter-Menge, 2009]. Compared to the sea ice forecasts with persistence, the ConvLSTM provides more skillful forecasts in winter and spring, but not for summer and early autumn. Similarly, it was demonstrated by Goessling et al. [2016] that the predictability of sea ice edge is relatively low from summer to autumn at monthly time scales over the whole Arctic. However, for the Barents Sea, it was shown by some studies that the initialized seasonal forecasting models could show monthly skill in the autumn (NDJ) up to 9 months in advance [e.g. Bushuk et al., 2017]. The better performance of the sea ice forecast with the ConvLSTM in winter and during the transition seasons shows that the ConvLSTM is able to capture the intricate sea ice variability in these seasons. This is not so surprising given the significant role of SSTs on sea ice predictability in this region in fall and winter [e.g. Guemas et al., 2016]. However, given the large ice free area in summer from 2013 to 2016 in the Barents Sea, the worse performance of the ConvLSTM than persistence seems to imply that ConvLSTM tend to over-predict sea ice in summer. Similar results are provided by forecast with the baseline statistical model and forecasts using the ConvLSTM. Considering the large difference between RMSE in each calendar month, it is not surprising that the spread of

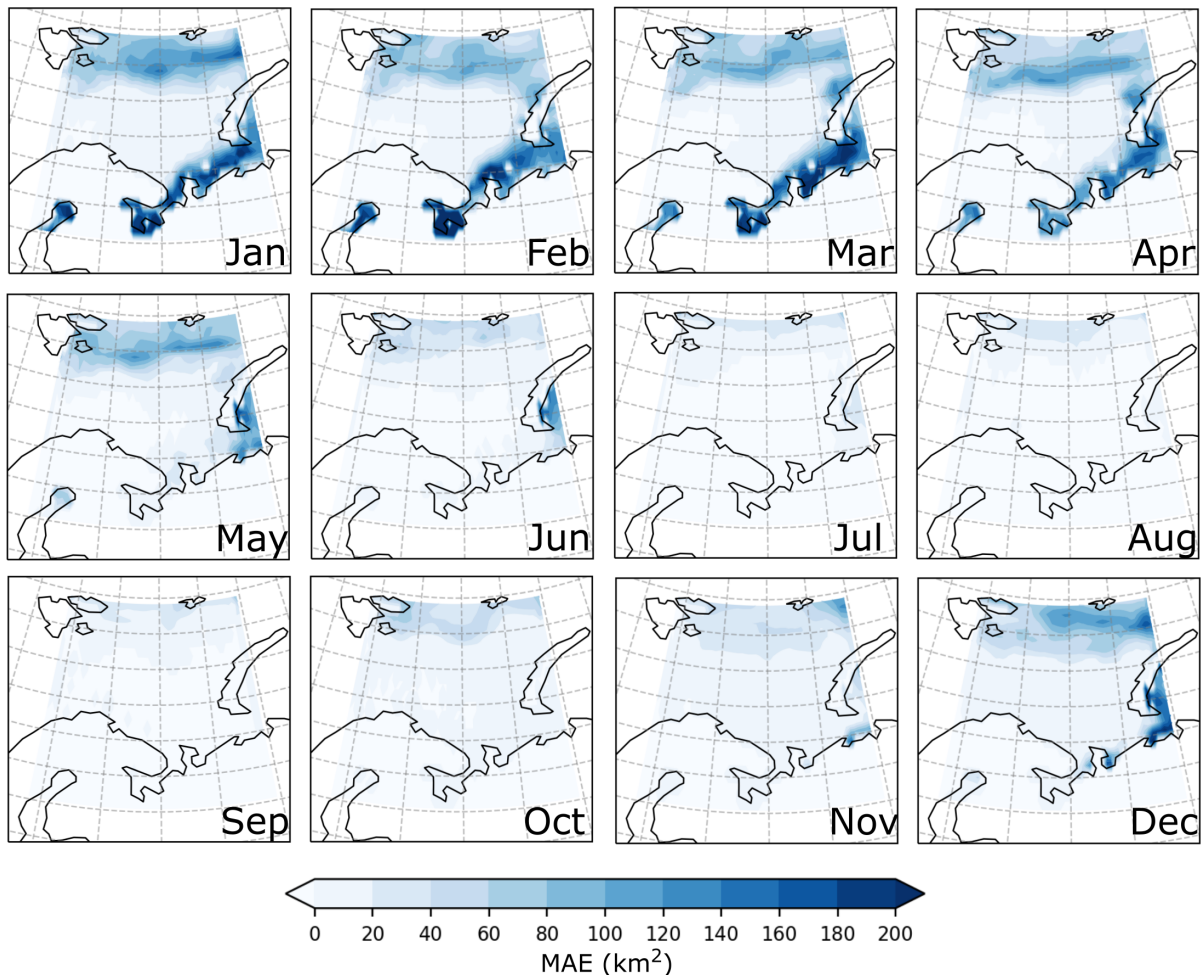
RMSE in Figure 4.3a is very large.

We investigated the seasonal differences in forecast skill at more lead times for the constrained forecasts (not shown). Similar to Figure 4.3b but with different lead times, the error increases with the increase of lead time, especially for the early spring around March. For the ConvLSTM forecast with all the chosen predictors, the reduction of skill is relatively smaller than the others. Forecasts with the ConvLSTM cannot beat persistence in autumn starting from October regardless of the lead time. But they are always better than the persistence in the transition, from spring to summer. At lead time step 6, forecasts with the ConvLSTM behave similar as the climatology in terms of the loss.

In order to understand the source of forecast error in each month, it is insightful to examine the spatial structure of the error. The spatial distribution of the forecast error with the ConvLSTM using SIC and OHC is plotted in Figure 4.4. In winter, the error mainly comes from the coastal area close to the Eurasian continent. Since the boundary sea ice dynamics is relatively complex, it is difficult to forecast, which is also a challenge for most of the operational numerical sea ice forecast systems [e.g. Smith et al., 2013, 2016]. In almost all seasons there is also a contribution to the total error from regions with rapid SIC variations near the northern boundary. Physically, processes like sea ice advection, sea ice advance and retreat, oceanic heat transport, and polar air-sea interaction, make forecasting in this area extremely difficult [Årthun et al., 2012, Smith et al., 2016]. Note that the cut-off of data around the selected boundary also influences the performance of the ConvLSTM, which was discussed in section 2b.

In practice, it is useful to know whether a region is ice free or not, for instance for shipping and related activities. A binary evaluation was carried out based on the forecast made by the ConvLSTM, based on persistence and climatology. We computed the IIEE score of these forecasts and found that the results are comparable among the forecasts with different methods, except for that with climatology. In order to assess the forecast skill location-wise, we plot the difference of IIEE score of the SIC constrained forecast between the ConvLSTM (using SIC and OHC) and persistence in Figure 4.5. Since the IIEE indicates either an overestimation or underestimation of sea ice forecast, areas with lower scores have better forecast skill. In general, the ConvLSTM has better forecast skill than persistence in almost all months. During the transition time from summer to winter, the ConvLSTM gives more skillful forecasts than persistence, especially around the northern boundary of the Barents Sea. In winter and spring, these two methods have skill in different regions and the ConvLSTM is slightly better. In summer, the differences are small.

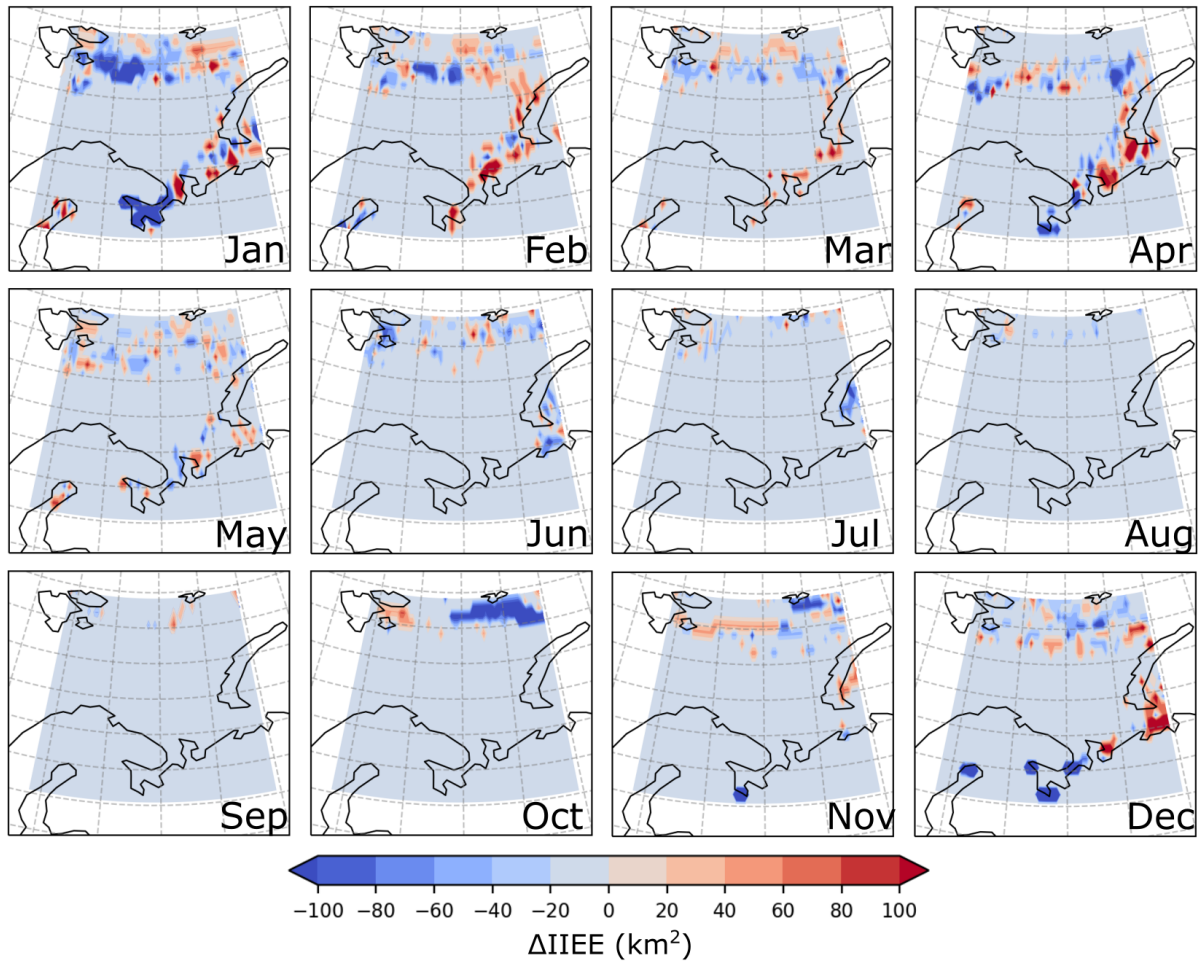
We can learn more about the forecast skill of each method by analysing the overestimated and underestimated sea ice extent separately. The overestimated and underestimated sea ice extent of the constrained forecasts of SIC with the ConvLSTM using different predictors against persistence, and climatology are shown in Figure 4.6. Given the definition of these two components of IIEE (section 2c), in combination they can be interpreted as a trade-off between over-predicting and under-predicting. Climatology tends to over-predict the sea ice in this area ( $\sim 120 \times 10^3 km^2$ ), but at the same time it has the least underestimated sea ice extent ( $\sim 5 \times 10^3 km^2$ ) than the other methods. Almost all the forecasts with the ConvLSTM provide better overestimated component of IIEE score but slightly worse underestimated component of



**Figure 4.4:** MAE of the constrained forecast of SIC for the first week in each month with ConvLSTM using SIC and OHC.

IIEE score than persistence. It reflects that the ConvLSTM learns the temporal evolution of sea ice, especially from an ice free to ice covered period.

To conclude, the ConvLSTM is able to outperform persistence and climatology when conducting sea ice forecasts at weekly to sub-monthly time scales. However, a careful selection of input fields should be made in terms of the physical consistency between predictors and sea ice at chosen time scales. An irrelevant predictor could “confuse” the neural network and hence reduce the performance. It is noteworthy that with more training data, the forecast skill of the ConvLSTM can be improved dramatically. This might indicate that ConvLSTM is good at finding nonlinear relations between variables which could eventually contribute to the predictability of the entire system, as long as the network is complicated enough, and the training data is sufficient.

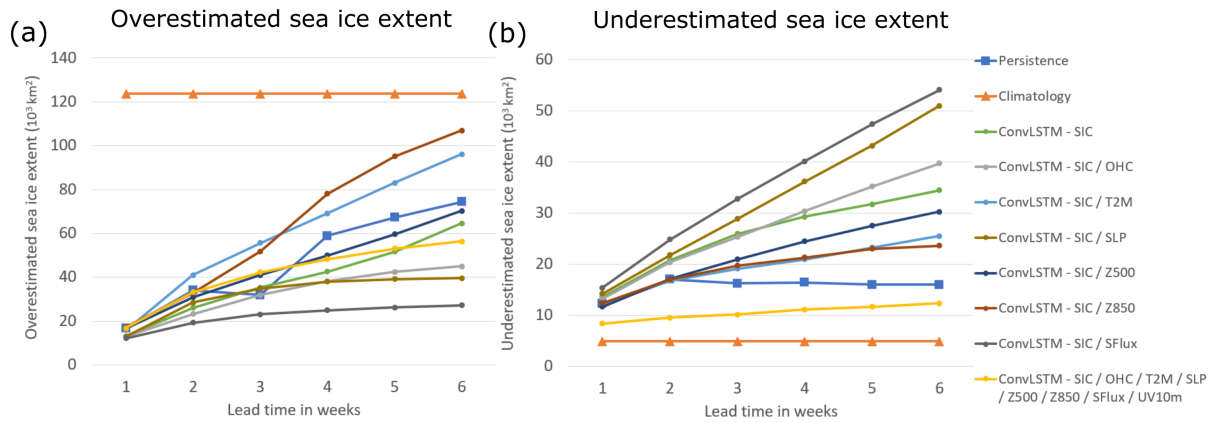


**Figure 4.5:** Difference of the IIEE score of the constrained forecast of SIC for the first week in each month between ConvLSTM and persistence ( $IIEE_{ConvLSTM} - IIEE_{persistence}$ ). The SIC forecast with ConvLSTM uses SIC and OHC fields.

### 4.3.2 Sensitivity Analysis of Predictors

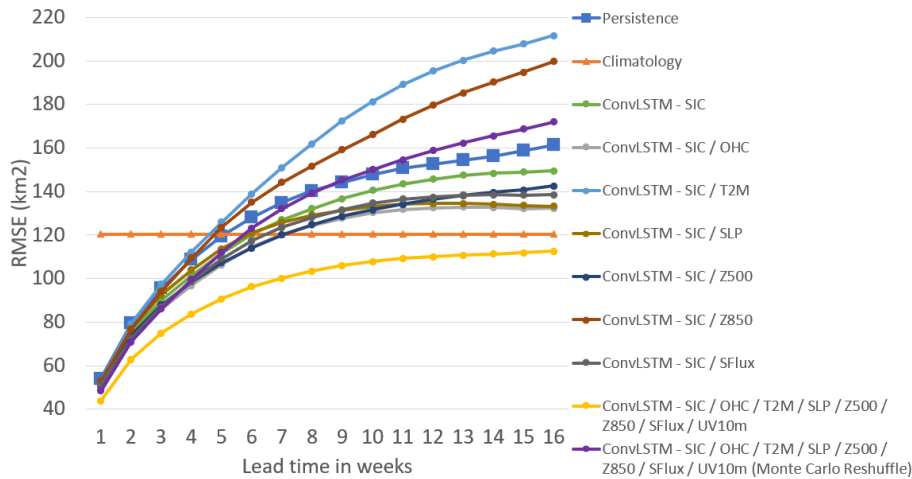
In the previous section, we found that the ConvLSTM is skillful for SIC forecasts using multiple climate fields. The generally better performance of the ConvLSTM than the baseline statistical model indicates that ConvLSTM learns the nonlinear relations between input fields, and sea ice forecast systems for weekly time scales can benefit from such representation of non-linearity. However, the performance could also be attributed to the change of signal to noise ratio and there is a possibility that the model only learns the noise from these predictors. To understand whether these predictors contribute to the forecast or they only pollute the prediction, we apply Monte Carlo reshuffling to all the chosen predictors with respect to the time sequences and use these reshuffled predictors to forecast SIC. Note that the input SIC sequence is not reshuffled. The result is shown in Figure 4.7. It can be noticed that forecasts with reshuffled predictors are much worse than those without reshuffling, especially for a long lead time. This indicates that these predictors provide useful information for the forecast. For a short period, for instance,





**Figure 4.6:** (a) Overestimated and (b) underestimated local sea ice extent of the constrained forecast of SIC with ConvLSTM using different predictors against persistence and climatology. The unit is square kilometer per grid cell.

up to lead week 1 and 2, forecasts with randomly reshuffled series of chosen predictors do not differ much compared to those without reshuffling. This indicates that, in the first two weeks, forecasts mainly rely on the memory of sea ice. This also explains why persistence is difficult to beat at weekly time scales. Unfortunately, the information we have is not sufficient to prove whether the improved forecasts benefit from the nonlinear relations between these variables or an increase of signal to noise ratio. Given that these predictors contribute to the predictability of SIC, we can further evaluate the contribution from different predictors to sea ice forecasts by comparing the forecasts using the same structure of the ConvLSTM, but trained with different input fields.

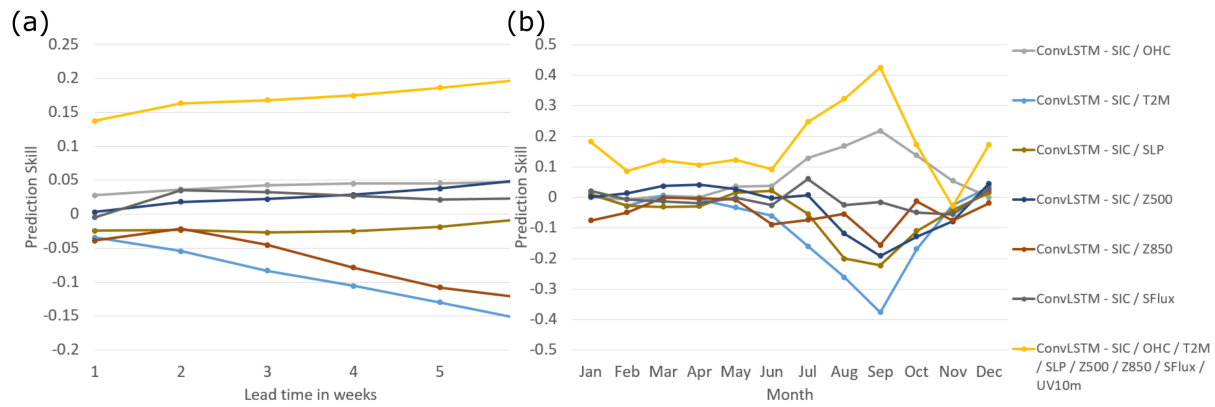


**Figure 4.7:** RMSE of the constrained forecast of SIC with a lead time up to 16 weeks with ConvLSTM using different predictors against persistence and climatology. A ConvLSTM forecast with Monte Carlo reshuffled predictors is included for comparison. The unit is square kilometer per grid cell.

Following the definition of our dimensionless relative forecast skill scores in section 2c, the



sensitivity analysis of predictors is shown in Figure 4.8. Similar to section 3a, we analyze the forecast skill based on lead-time dependent constrained forecasts up to 6 weeks, and for the first week in each month. Positive scores indicate an improvement in forecast skill. Note that the skill is relative to the ConvLSTM model that only uses SIC as a predictor. In Figure 4.8a, it can be observed that OHC, SFlux and Z500 add skill to the sea ice forecasts at chosen time scales, while SLP, T2M and Z850 reduce the skill. Predictability from OHC can be attributed to the long memory of the ocean and the crucial role of OHC in the energy budget, which is also claimed by Guemas et al. [2016], Cruz-García et al. [2019]. Furthermore, it is shown in Figure 4.8b that OHC has a significant contribution to the forecast skill in summer. Another component of the energy budget, SFlux, also plays an important role here since it has a direct relation to sea ice, that is, the surface energy balance between open ocean and sea ice covered ocean is very different. We notice that it contributes to the predictability with a lead-time more than a week, which is consistent with the lead-lag relations between sea ice melting and the variability of surface fluxes found by Krikken and Hazeleger [2015].



**Figure 4.8:** RMSE based relative forecast skill improvement of (a) the constrained forecast of SIC with a lead time up to 6 weeks and (b) the constrained forecast of SIC for the first week in each month with ConvLSTM using different predictors.

At weekly to sub-monthly time scales, Figure 4.8 shows that some surface fields, like SLP and T2M, and some near-surface atmospheric fields, do not contribute to improving sea ice forecasts with the ConvLSTM. It was reported by Onarheim et al. [2015] and Mohammadi-Aragh et al. [2018] that the chaotic behavior of the atmosphere causes the low predictability of the near surface wind divergence and vorticity, which explains the weak relationship between sea ice variability and SLP. The bad performance of sea ice forecast with the ConvLSTM using SIC and T2M suggests that the surface temperature field is not directly related to the variation of sea ice, especially in summer as shown in Figure 4.8b. This is the same for Z850.

In summary, for the extended range sea ice forecasts in the Barents Sea, meteorological and oceanic fields involved in the energy budget (e.g. SFlux, OHC) can enhance the forecast skill of the sea ice forecasts made by deep neural networks. The same holds for the fields representing the free troposphere (e.g. Z500). However, some surface fields (e.g. T2M, SLP) and lower atmospheric fields (e.g. Z850) do not improve the forecast quality in the selected region. Use of the ConvLSTM can potentially help us understand the nonlinear relationships between multiple

selected variables, but limited physical information will be provided by this method. Note that these conclusions about the predictability of sea ice are drawn for specific time scales and region. From this study it is unclear whether it can be generalized to other time scales and locations.

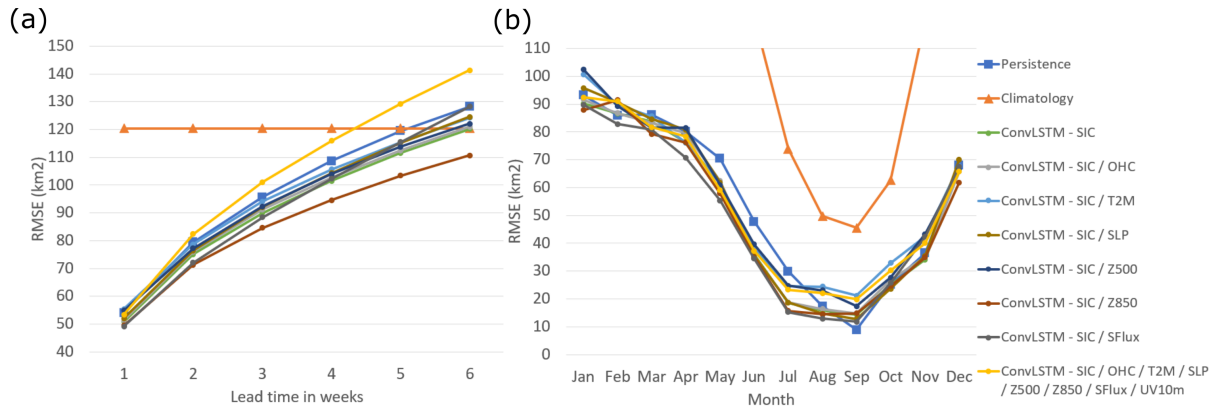
### 4.3.3 Lead-Time Dependent Operational Forecasts

In contrast to constrained forecasts, operational weather forecasts can only proceed from a known state to the future state with predicted fields. Therefore, to achieve a “more than one step” sea ice forecast, we have to predict all the input fields and use the predicted fields for further forecasts, specifically with the ConvLSTM. Similar to evaluations of the constrained forecasts given in section 3a, an assessment of the lead-time dependent forecasts with the ConvLSTM using different combinations of input fields against persistence and climatology are presented in this section. We emphasize that from now on the presented forecasts with the ConvLSTM generate all fields (the same variables as the input) and the loss function also includes these fields. The presented forecasts are re-forecasts and these retrospective forecasts can be analyzed using observations.

We first show the RMSE of lead-time dependent forecasts with different predictors and methods in Figure 4.9a (details in Table S4). In general, RMSE increases with the increased lead-time, which is similar to Figure 4.3a. However, this time, almost all the forecasts with the ConvLSTM outperform the forecasts with persistence, except for the ConvLSTM forecast with all the selected input fields. The ConvLSTM forecast are better than the climatology with lead time from week 1 to 5, in general. The best forecast is given by ConvLSTM with SIC and Z850, which is very different from the constrained forecast (Figure 4.3a). Considering the RMSE in each month (Figure 4.9b, details in Table S5), most of the forecasts with the ConvLSTM are better than the persistence forecast in spring and autumn. All the forecasts are comparable in winter but persistence is much better in summer. Also, the forecasts with the ConvLSTM using all given input fields are worse than the ConvLSTM forecasts using fewer variables. We further inspected the seasonal differences in forecast skill at more lead times for the operational forecasts (not shown). The results are analogous to those given by the constrained forecasts. The only difference is that the ConvLSTM forecasts with one extra predictor (e.g. Z850, SFlux) show better skill in sea ice forecasting for most of the time compared to that with all chosen predictors, which is consistent with our analysis regarding Figure 4.9.

The decrease in performance of the ConvLSTM with more input variables than that with fewer variables can be attributed to the training process of deep neural networks and the setup of operational forecast. Based on this configuration, the number of input fields are equal to the number of output fields. More input variables mean more output variables and therefore more model parameters, which in turn requires more training data. Consequently, the operational forecast with ConvLSTM is a trade-off between skill gain from predictors and skill loss due to the difficulty in learning. Therefore, it is necessary to choose the right combination of input fields for the ConvLSTM, instead of using all the data in a purely data-driven manner.

The spatial distribution of the forecast error for the first week with the ConvLSTM using SIC and OHC is plotted in Figure 4.10. It is very similar to the result in Figure 4.4. It is difficult

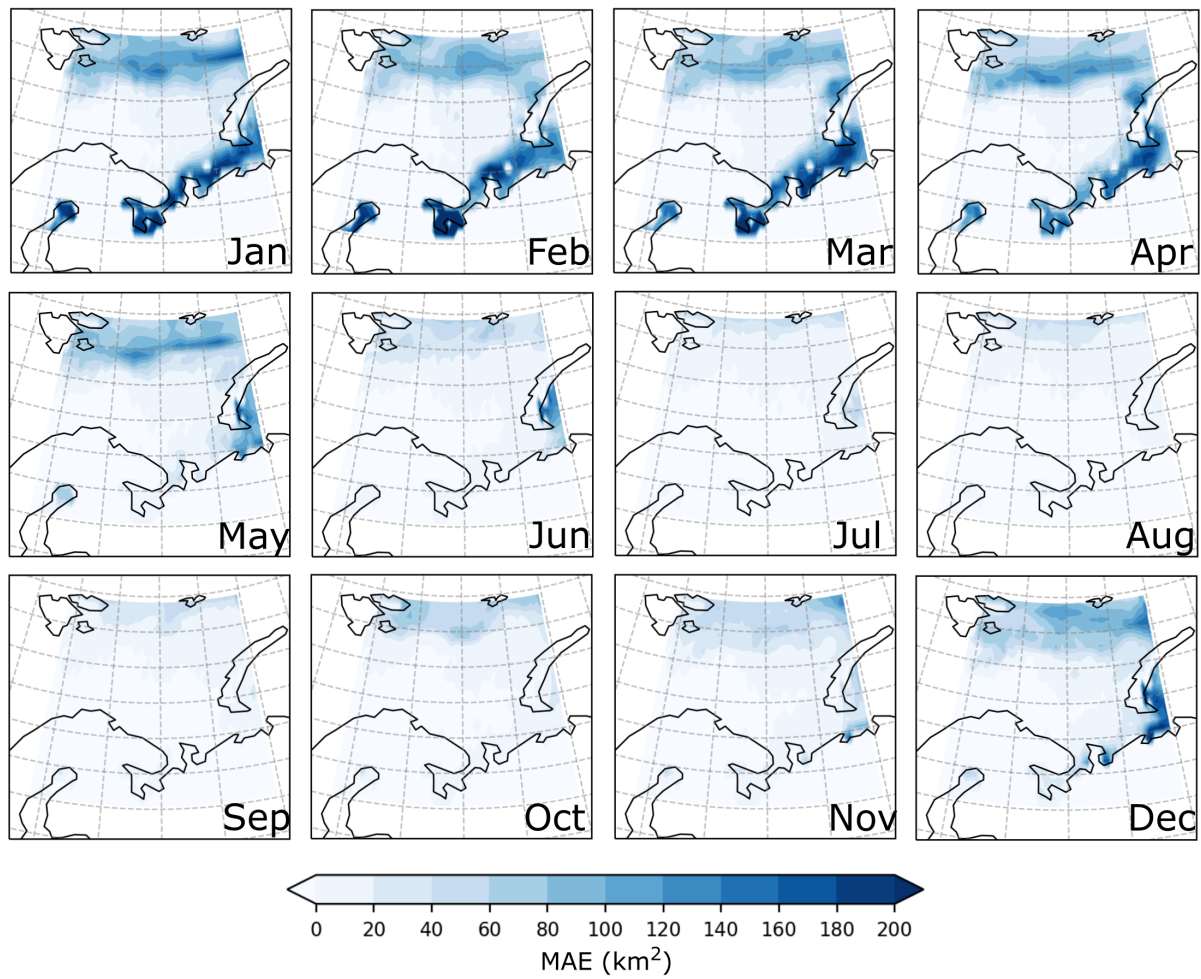


**Figure 4.9:** RMSE of (a) the operational forecast of SIC with a lead time up to 6 weeks and (b) the operational forecast of SIC for the first week in each month with ConvLSTM using different predictors against persistence and climatology. The unit is square kilometer per grid cell.

for the system to predict sea ice in winter and spring, and large forecast errors are mainly found in the coastal area and the northern boundary. In summer and autumn, the forecast errors are relatively small. We now consider the binary forecast of sea ice. The difference of the IIEE score of SIC forecasts between the ConvLSTM (using SIC and OHC) and persistence is shown in Figure 4.11. Given the similarity of the spatial distribution of the forecast errors between the constrained (Figure 4.4) and operational (Figure 4.10) forecasts, it is not surprising to find that the ConvLSTM has better forecast skill than persistence in most of the regions in the Barents Sea in almost all months. More explanation on this can be found in the section 3a. Note that starting from the second week the spatial distribution of the forecast errors (RMSE and binary) becomes different for the constrained and operational forecasts with ConvLSTM, but the regions with high forecast skill with the ConvLSTM in these two cases are still the same (not shown).

Compared to the constrained forecast, small differences are found in the overestimated and underestimated components of the IIEE score for the operational forecasts with the ConvLSTM. The overestimated and underestimated local sea ice extent of the lead-time dependent operational SIC forecast with the ConvLSTM using different predictors against persistence and climatology are shown in Figure 4.12. Still, most of the forecasts with the ConvLSTM provide better overestimated component of the IIEE score than persistence, except for the ConvLSTM forecast using SFlux and all available predictors. However, for the underestimated component of the IIEE score, persistence has always more skill starting from the third week. Considering both components of the IIEE score, it can be noticed that the forecast with the ConvLSTM using all the input fields significantly over-predicts the sea ice. Again, it shows a caveat that rather than blindly using all the available data, a smart selection of input fields is necessary to improve the sea ice forecast with the current structure of the ConvLSTM.

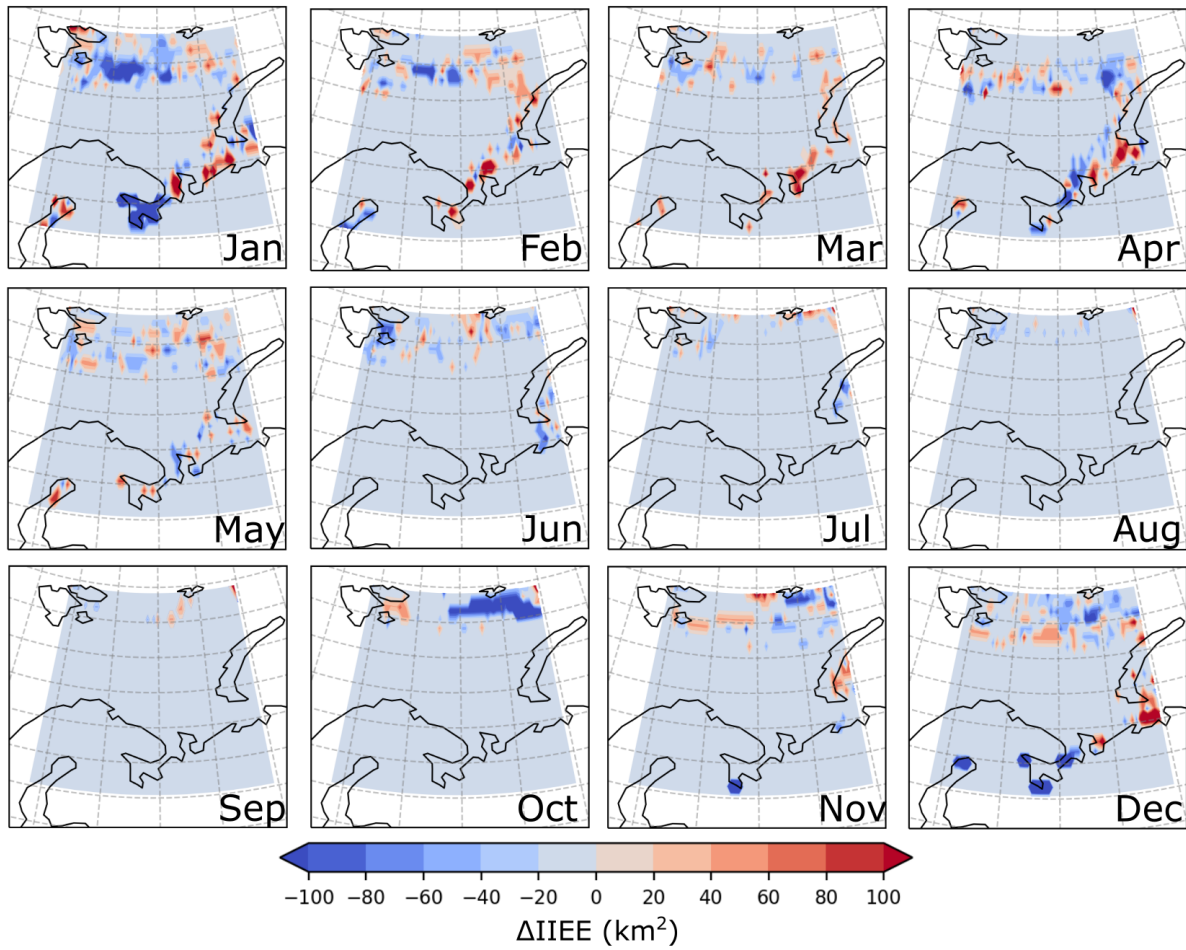
In addition, we compare the ConvLSTM forecasts with the NCEP and ECMWF ensemble forecasts considering period 2015-2016. The results are shown in Figure 4.13. It can be noticed that the NCEP ensemble forecast is worse than most of the ConvLSTM forecasts within 4 lead



**Figure 4.10:** MAE of the operational forecast of SIC for the first week in each month with ConvLSTM using SIC and OHC.

weeks. This may originate from the initialization since the NCEP ensemble forecast is initialized by sea ice conditions from Climate Forecast System Reanalysis (CFSR), and not from the ERA-Interim or any satellite-based observations (e.g. NSIDC/NOAA Passive Microwave SIC). The ECMWF ensemble forecast performs much better than all the ConvLSTM forecasts for more than 2 weeks ahead. Note that since the forecasts are evaluated against ERA-Interim reanalysis and ECMWF ensemble forecasts use the same IFS and similar configurations as ERA-Interim, it is not completely a fair comparison. The forecast error grows slower for the ECMWF ensemble forecast than the ConvLSTM forecasts. Since for the first two weeks memory in sea ice has significant impact on its variations, it reflects that forecasts with ConvLSTM considerably rely on the memory of sea ice, and the chosen numerical model can preserve the physical consistency and therefore may provide better forecasts even for large lead times. In general, at the lead time up to 2 weeks, the forecasts with ConvLSTM are comparable to the best NWP-based forecasts in S2S project.

So far we have not discussed the forecast quality for various start dates. It is useful to know

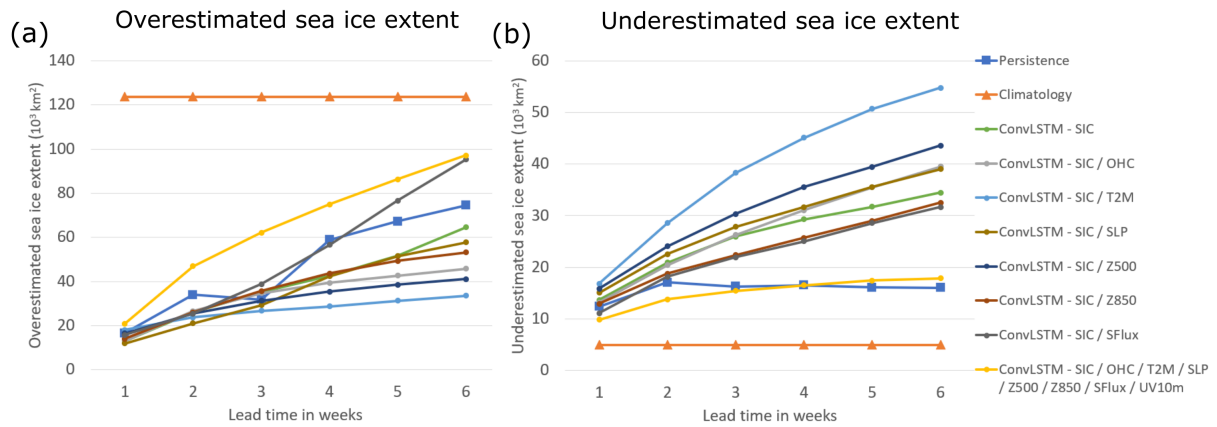


**Figure 4.11:** Difference of the IIEE score of the operational forecast of SIC for the first week in each month between ConvLSTM and persistence ( $IIEE_{ConvLSTM} - IIEE_{persistence}$ ). The SIC forecast with ConvLSTM uses SIC and OHC fields. The unit is square kilometer per grid cell.

if either there is a substantial dependence of forecast errors on climate variability and change, or whether the variations between forecast cases (start dates) are large in comparison to the differences between forecast models (ConvLSTM forecasts with different combination of input fields). The former one can also be interpreted as whether the whole temporal consistency is needed or not for a ConvLSTM.

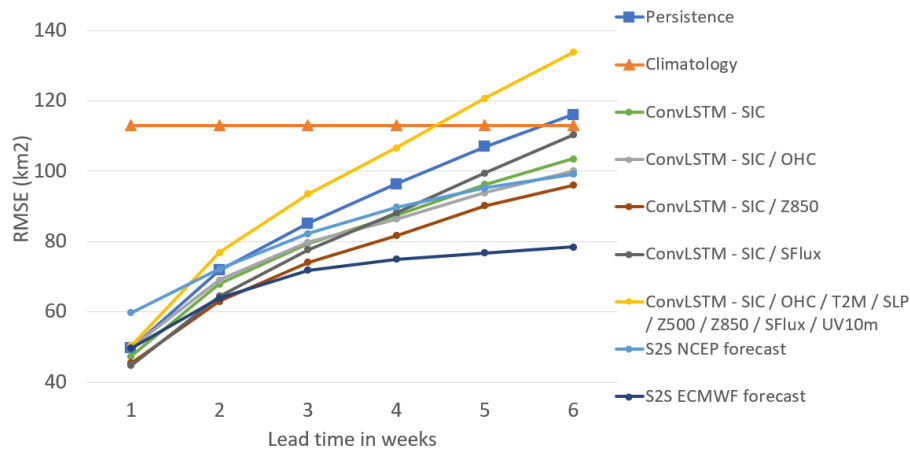
To address this we conducted an extra experiment in which we performed Monte-Carlo sub-sampling of the reanalysis data with a 4-year period. We randomly selected 5 periods (20 years data) to train the network and 4 periods (4 years for each) as valid date. The results indicate that, compared to the climatology and persistence, these ConvLSTM forecasts do not show any skill considering each start date at all lead weeks (not shown). This reflects that this particular ConvLSTM was not able to determine the state of the system and therefore could not provide reliable forecasts. Indeed, this is not so surprising as forecasts with the ConvLSTM are strongly dependent on memory and the temporal order in the data must be preserved during training and predicting. To gain more insight into the dependency of forecasting on the temporal





**Figure 4.12:** (a) Overestimated and (b) underestimated local sea ice extent of the operational forecast of SIC with ConvLSTM using different predictors against persistence and climatology. The unit is square kilometer per grid cell.

consistency, we performed a follow-up test in which we provided the model with 20 years data following temporal order and then applied Monte-Carlo sub-sampling with the same setup. Also, this model gained no skill (not shown). Only when the full time series preceding the valid date were fed to the model, the ConvLSTM started to generate meaningful forecasts. This again demonstrates that forecasts using ConvLSTM rely on the correct temporal order to reproduce and skillfully forecast the state of the system. This indicates that the common workflow used by NWP systems to determine skill for a range of start dates, and therefore an error estimate of forecast quality for a specific forecast system, cannot be transferred meaningfully to ConvLSTM, given the insufficient length of training data. This would only be possible with sufficient length of training data such that the memory can be maintained in the samples.

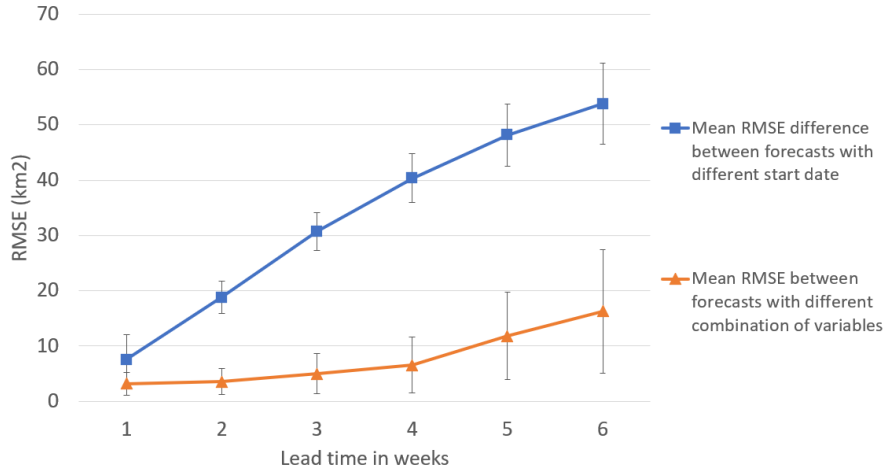


**Figure 4.13:** RMSE of the operational forecast of SIC between 2015 to 2016 with a lead time up to 6 weeks with ConvLSTM using different predictors against persistence, climatology, NCEP ensemble forecast and ECMWF ensemble forecast. The NCEP real-time forecast and ECMWF ensemble forecast are provided by the sub-seasonal to seasonal prediction project (S2S archive). The unit is square kilometer per grid cell.

There is another way to test the state-dependency and address the significance of the conclusions. Given the susceptibility of forecasts to the length of the training set, and a strict requirement on the temporal order of input fields during training and forecasting, we launched another experiment with a reversed time series of the reanalysis data, and chose data from 1979-1982 as testing set, 1983-1986 as cross-validation set and the rest as training set (not shown). Despite slightly different skill compared to those shown in Figure 4.9a, ConvLSTM forecasts with different combinations of input fields exhibit similar errors. From now on we will refer to this experiment as “reversed operational forecast 1979-1982”, and the former experiment with time series following the correct temporal order as “operational forecast 2013-2016” (Figure 4.9a).

By comparing these two operational forecasts, we can evaluate the skill of ConvLSTM for different start dates (between 1979-1982 and 2013-2016) and compare it to the skill of ConvLSTM related to the choices of input fields (e.g. between ConvLSTM-SIC/OHC and ConvLSTM-SIC in Figure 4.9a) in a relatively fair manner, as we can use the same amount of training data and both systems show comparable forecast skill, although the setups are slightly different. We highlight this comparison quantitatively in Figure 4.14. The mean difference of errors between forecasts with different start dates were computed as the mean absolute difference of RMSE between reversed operational forecast 1979-1982 and operational forecast 2013-2016 with the same input fields ( $\frac{1}{7} \sum_{n=1}^7 (|RMSE_{ConvLSTM-SIC/field(2013-2016)} - RMSE_{ConvLSTM-SIC/field(1979-1982)}|)$ , with *SIC/field* indicating different combinations of input fields shown in Figure 4.9a and S11), while the mean difference of RMSE between forecasts with different input fields were obtained from reversed operational forecast 1979-1982 and operational forecast 2013-2016 ( $\frac{1}{14} \sum_{n=1}^{14} (|RMSE_{ConvLSTM-SIC/field} - RMSE_{ConvLSTM-SIC}|)$ , concerning the forecasts in Figure 4.9a and S11). These two cases have errors in skill that are similar. Therefore, these experiments indicate that the ConvLSTM forecasts when properly trained are skillful and their forecast quality is state-dependent just like forecasts with numerical models. The mean differences between forecasts with different input fields are smaller than those related to different start dates, but in general they are of the same amplitude. Note that the differences in skills between the forecasts of period 1979-1982 and 2013-2016 may also originate from the data richness and quality linked to these two different eras.

To conclude, similar to constrained forecasts, the operational forecasts with the ConvLSTM provide better results than persistence in most of the cases. However, unlike the constrained forecasts, the coherence between the way of learning (loss function and an increase in the number of trainable parameters) and the way of predicting (forecast based on predicted fields) with a neural network places a requirement for appropriate choices of input variables, which may a priori not be clear. However, in practice, it is natural to select predictors related to their physical consistency with the predicted variables. The results confirm this, with higher contributions from OHC and SFlux and less clear contributions from more atmospheric characteristics. It is noteworthy that forecasts with ConvLSTM shows state-dependency and the start dates have impact on the forecast quality, just like forecasts using numerical models.



**Figure 4.14:** Comparison between the mean difference of errors among forecasts with different start dates and that among forecasts with different combination of input variables based on reversed operational forecast 1979-1982 and operational forecast 2013-2016, as a function of lead time. The mean difference of errors between forecasts with different start dates were computed as the mean absolute difference of RMSE between reversed operational forecast 1979-1982 and operational forecast 2013-2016 with the same input fields ( $\frac{1}{7} \sum_{n=1}^7 (|RMSE_{ConvLSTM-SIC/field(2013-2016)} - RMSE_{ConvLSTM-SIC/field(1979-1982)}|)$ , with  $SIC/field$  indicating different combinations of input fields shown in Figure 4.9 and S11), while the mean difference of RMSE between forecasts with different input fields were obtained from reversed operational forecast 1979-1982 and operational forecast 2013-2016 ( $\frac{1}{14} \sum_{n=1}^{14} (|RMSE_{ConvLSTM-SIC/field} - RMSE_{ConvLSTM-SIC}|)$ , concerning the forecasts in Figure 4.9a and S11). The standard deviation at each lead week is included.

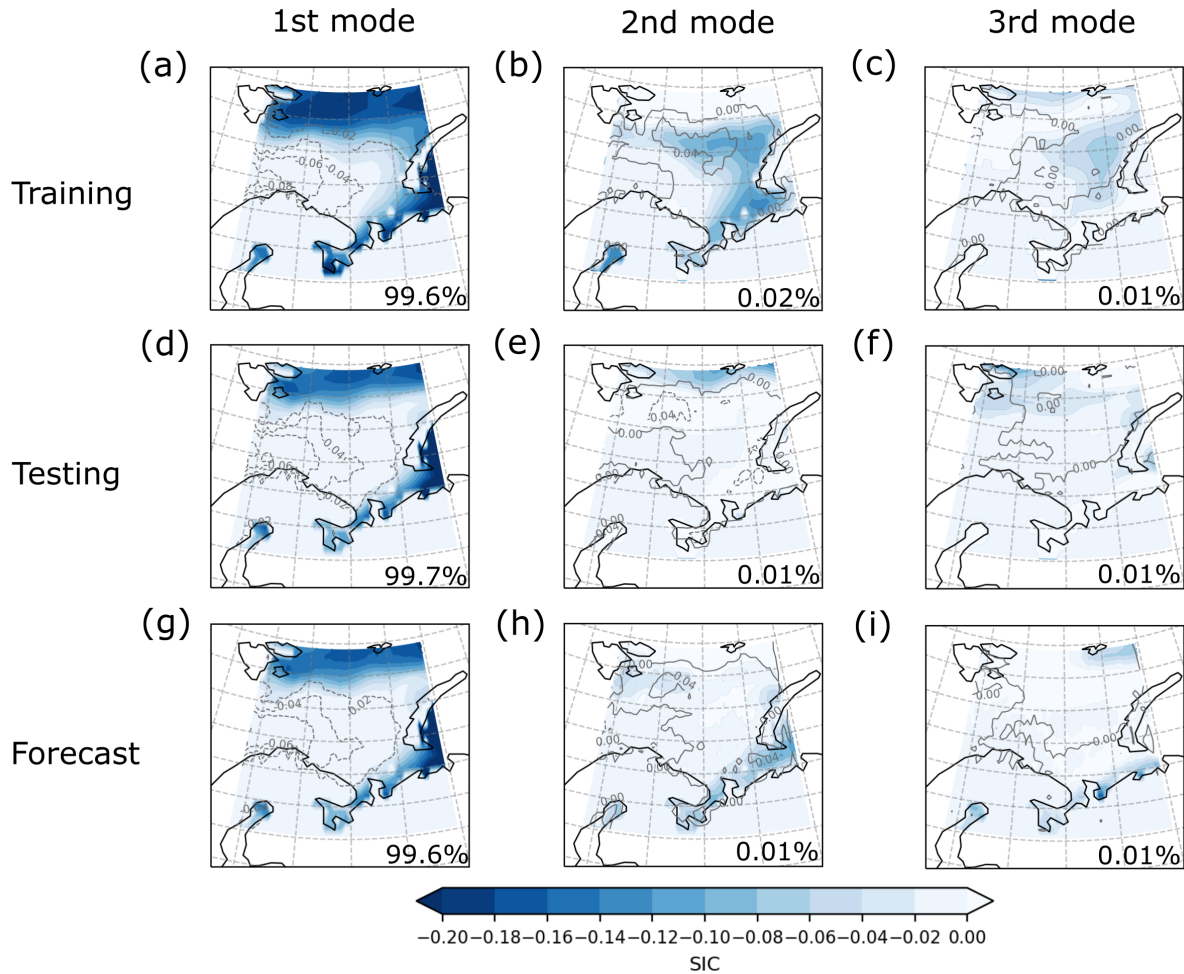
#### 4.3.4 Physical Consistency of ConvLSTM Forecasts

Deep learning techniques are often considered as brute force approaches or black-box methods. There is a valid concern that physical laws may be violated and physical relationships may not hold in the neural network. There is no constraint from first principle in its formulation. With this in mind, we are interested in the physical interpretation of our forecasts with the ConvLSTM. Our first impression comes from the mathematical description of the ConvLSTM (equation 4.1), which includes both linear and non-linear operations. The convolutional operations are all linear (e.g.  $W_{xi} * x_t$ ), and this will maintain the physical consistency of input fields, at least their linear relations. However, the non-linear behaviour caused by the use of sigmoid functions and hyperbolic tangent functions could potentially introduce further physical inconsistency in the forecasts.

In order to evaluate whether the physical links between predicted fields are preserved by the forecasts with the ConvLSTM, we performed singular value decomposition (SVD) on the covariance map of two predicted fields (e.g. SIC and OHC) within training sets (1979-2008), testing sets (2013-2016) and forecast data (2013-2016) [Bretherton et al., 1992]. Since this method searches for the maximum covariance between given variables, it is also known as the maximum covariance analysis (MCA) [Frankignoul et al., 2011]. We first show three SVD modes of the covariance map between SIC and OHC in Figure 4.15. These fields are expected to be related and



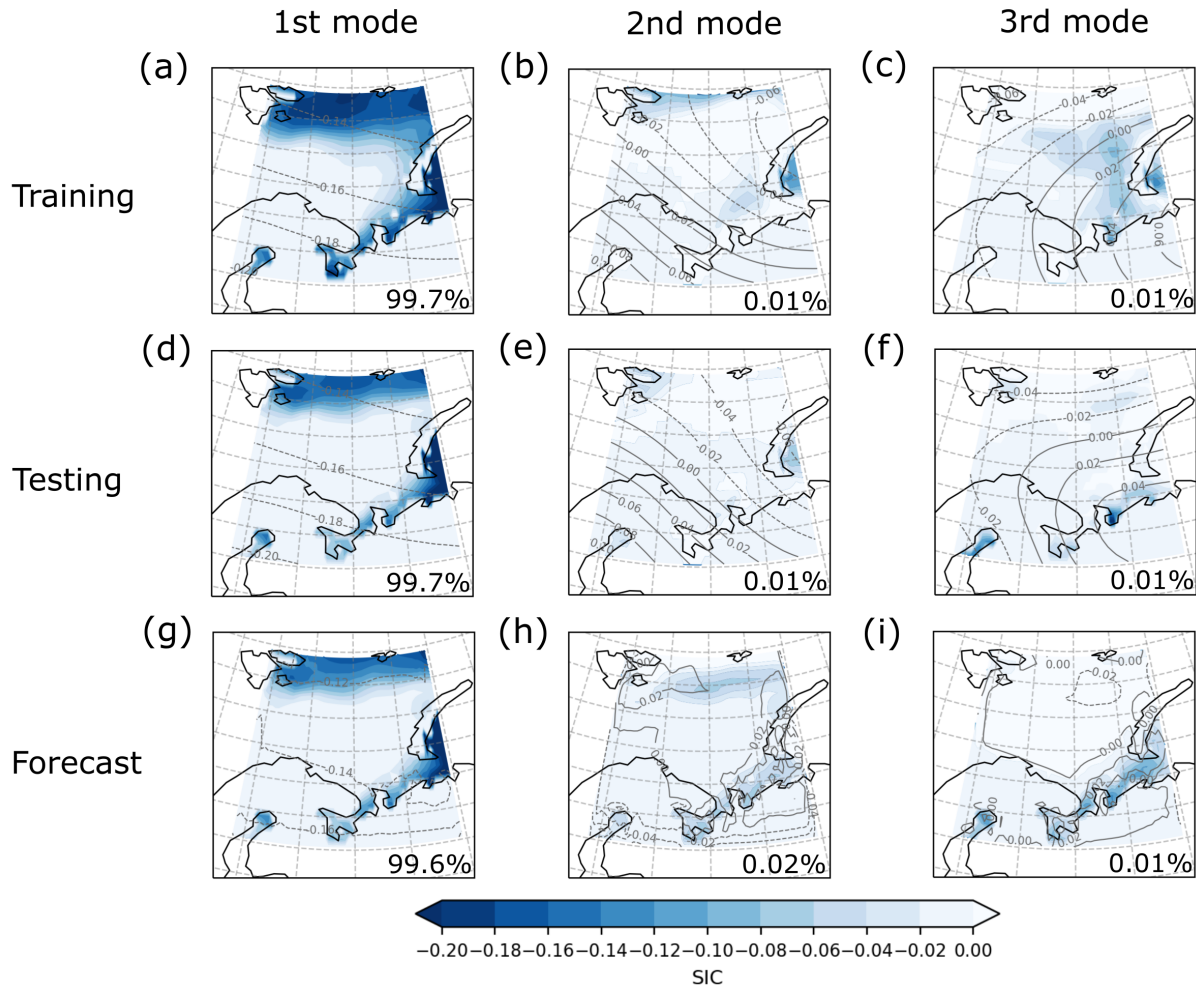
OHC varies relatively slowly. The first modes explain over 99% of covariance, and they mainly represent the trend and climatology, which can be regarded as fairly trivial. Therefore, it is worthwhile to evaluate the second and third modes. In general, the forecast data show similar patterns for both SIC and OHC as training sets and testing sets in all three SVD modes. This indicates that forecasts with ConvLSTM are able to preserve the physical links between given fields, which reflects the source of predictability within chosen fields and the forecast skill of chosen neural networks. This is further confirmed by the projection of SVD modes of the covariance map between SIC and OHC on the actual time series of SIC and OHC (not shown).



**Figure 4.15:** Covariance map of SIC and OHC for the (a, d, g) first, (b, e, h) second and (c, f, i) third SVD modes in (a, b, c) training (d, e, f) testing and (g, h, i) forecast data for the first week, with shades the dimensionless SIC and contour lines the dimensionless OHC. The SVD was performed on the covariance matrix of normalized SIC and OHC.

Similar to the SVD of covariance map between SIC and OHC, we also inspected other variables, for instance, the SVD and SVD projection of the covariance map between SIC and Z500 (Figure 4.16 and S13). In this case, physically interpretable results are obtained in the training and testing data sets. However, the forecasts with the ConvLSTM fails to provide similar coupled patterns of Z500 and SIC. In particular Z500 seems like a spurious pattern. Given the chaotic

behaviour of the atmospheric circulation and the related lack of predictability at extended range, it is not so surprising that the predictability of Z500 at weekly time scales is relatively low in the ConvLSTM [e.g. Hohenegger and Schar, 2007]. Also, the distorted shape of the patterns suggests that the cut-off of input fields around the boundary also influences the physical consistency learned by the ConvLSTM (see the SVD mode patterns of Z500 around the boundary in Figure 4.16 and S13).



**Figure 4.16:** Covariance map of SIC and Z500 for the (a, d, g) first, (b, e, h) second and (c, f, i) third SVD modes in (a, b, c) training (d, e, f) testing and (g, h, i) forecast data for the first week, with shades the dimensionless SIC and contour lines the dimensionless Z500. The SVD was performed on the covariance matrix of normalized SIC and Z500.

We also checked the SVD of the covariance map between SIC and SFlux, and that between SIC and Z850 (not shown). Since these decompositions were all based on the forecast of lead week 1, the results do not differ substantially.

To conclude, depending on the physically interpretable predictability of chosen meteorological and oceanic fields, the ConvLSTM is able to preserve a realistic physical consistency between various predictors and predictands during forecasts. It should be noted that such analysis based

on SVD can only account for the linear relations of the covariance between given fields. To further evaluate the physical consistency during forecast with deep neural networks, a choice of meteorological and oceanic fields could be made based on some lead-lag analysis between potential fields, similar to Krikken and Hazeleger [2015], but at shorter time scales than they considered. Also, other clustering techniques can be used that are not limited to linear relationships.

## 4.4 Discussion

In this study, we introduce a deep neural network to predict the sea ice in the Barents Sea. With the intention to adapt this approach to the Arctic sea ice forecast problem and further assess its capabilities, all the computations were performed with reanalysis data sets. It is noteworthy that our results with the ConvLSTM are comparable to those with the state-of-the-art numerical climate models [Van Woert et al., 2004, Metzger et al., 2014, Hebert et al., 2015, Smith et al., 2013, 2016]. For instance, our forecast errors with the ConvLSTM using SIC and OHC shown in Figure 4.10 are comparable to those with the latest Global Ice Ocean Prediction System (GIOPS) employed by the Canadian Meteorological Centre (see their Figure 4-7) [Smith et al., 2016]. We also inspect the errors based on the normalized sea ice concentration and find a competitive result. However, we only include a brief assessment of the forecast skill between ConvLSTM and numerical weather forecast systems (NCEP and ECMWF ensemble forecast in S2S project) and an extended quantitative comparison between them is beyond our scope. For future work, it is recommended to evaluate ConvLSTM forecasts against numerical weather and seasonal forecast systems consistently using the same observation dataset and an extensive range of metrics.

Typically, operational weather forecast systems are ensemble forecast systems in order to sample several sources of uncertainty [Gneiting et al., 2007]. It is also possible for the deep neural networks to model the uncertainty by either employing the deep learning based ensemble approach (e.g. an ensemble of deep neural networks) [Zaier et al., 2010, Wang et al., 2017], or implementing probabilistic deep neural networks [e.g. Vandal et al., 2018, McDermott and Wikle, 2019, with Bayesian deep learning]. For the deep learning based ensemble approach, the ensemble is generated through perturbing the structure of neural networks (e.g. number of hidden layers, filter size) and the size of input sequences, and therefore very difficult to control. The latter is more common in practice and this technique is generally known as Bayesian deep learning (BDL) [Blundell et al., 2015, Fortunato et al., 2017, Kendall and Gal, 2017, Shridhar et al., 2019]. With Bayesian deep learning, a deterministic NN can easily be transformed into a probabilistic NN, namely a Bayesian neural network (BNN), by replacing the weight with a distribution. Through sampling the distribution, an ensemble forecast can be generated to represent the uncertainty in the forecast to avoid over-confident forecasting. These techniques are able to capture both aleatoric and epistemic uncertainty, but they are very expensive [Kendall and Gal, 2017].

This study is primarily a proof of concept of the capabilities of ConvLSTMs, which goes beyond the extensive use of LSTMs in modern meteorological forecast literature. For this study, incorporating an ensemble or Bayesian network was beyond scope, but we will study it in our following work.

Due to our limited access to the computational resources, the experiments with the ConvLSTM were carried out with a relatively simple network structure and a small dataset. Given the nature of deep neural networks, there is still a lot of room for improvement by increasing the complexity of the neural network (e.g. add more LSTM layers), including more input fields which are physically consistent with sea ice variability (e.g. separate terms in the surface energy budget), enlarging the input area to reduce the boundary effect, training separate models for the forecasts with different lead-time to avoid the accumulation of forecast errors and account for the teleconnections between remote areas. Theoretically, a more complicated deep neural network serves as a better representation of non-linearity between variables, thus the nonlinear relationship between all the meteorological fields in a chaotic climate system. Our study has shown that this is far from straightforward as the inclusion of noisy atmospheric fields deteriorated the forecast skill.

Nevertheless, it should be noted that, compared to numerical weather forecast systems, the deep learning based methods still have drawbacks. For instance, these methods always have a very limited and specialised forecast target and they are not flexible after expensive training and tuning processes. Moreover, such data driven approaches normally require a large training set and therefore the left validation data may not be enough to allow for any statistical significance tests. However, in terms of the fast development of deep learning techniques and the corresponding evolution of hardware, these deep learning approaches could potentially enhance weather forecasts in the near future.

## 4.5 Conclusion

A ConvLSTM is a useful tool to incorporate both the spatial and temporal information within meteorological fields in a model. In this work, we demonstrate that this deep neural network approach can effectively be used to predict sea ice characteristics in the Barents Sea at weekly to sub-monthly time scales. Different combinations of meteorological fields were tested as input variables to train the neural networks and make forecasts. Sea ice forecasts were evaluated against climatology, persistence and the baseline statistical model. They were also compared to operational sub-seasonal to seasonal forecast systems. It is found that in most cases, a ConvLSTM can outperform persistence and climatology for the extended range sea ice forecast at lead week up to 5. However, the choices of input meteorological fields should be made in a smart way based on the physical consistency between sea ice variation and the variability of predictors. Being dependent on the predictability of chosen meteorological fields, an interpretation of forecasts with the ConvLSTM in terms of observed linear relationships indicates that the ConvLSTM is able to preserve the physical consistency between various predictors and predictands. In addition, it should be noted that forecasts with ConvLSTM shows state-dependency and the start dates have impact on the forecast quality, just like forecasts using numerical models.

Both lead-time dependent constrained and retrospective forecasts were performed in this study and their results are slightly different due to the criterion in the loss function. Moreover, sensitivity tests were conducted based on the lead-time dependent constrained forecasts and we notice that energy budget related fields can add skill to the sea ice forecasts in the Barents Sea at weekly to sub-monthly time scales. Krikken and Hazeleger [2015] also found that OHC

and SFlux have a strong impact on the predictability of the Arctic sea ice. This indicates the important role of the energy budget components in the variability of Arctic sea ice and poses a request for accurate and reliable quantification of the energy budget [Liu et al., 2020b] and a deep understanding of the energy balance in the climate system [Liu et al., 2020a]. Moreover, the fields representing the mid-troposphere also have an impact on the predictability of sea ice in the Barents Sea. In contrast, the atmospheric circulation fields close to the surface seem to deteriorate the performance of the deep neural network, thus reducing the forecast skill. Although such deep neural networks are “black boxes”, they can potentially help us gain more knowledge about the nonlinear relationship between multiple meteorological fields.

So far, our experiments with ConvLSTM are limited to simple network structures and small training data sets. It is possible to improve the performance of ConvLSTM by increasing the complexity of the network and including large data sets for training. Fortunately, such attempts will hardly influence the time to obtain a forecast, but the training cost will increase. Compared to the relatively expensive operational numerical forecast systems based on numerical weather models, this method with the ConvLSTM could be suitable for forecasts where time to produce the forecast is limited. In addition, the forecasts made by the ConvLSTM can also potentially be assimilated by the operational weather forecast systems to improve their robustness and reliability. Given its advantages over the conventional way of weather forecast, as well as the fast development of deep learning techniques and the corresponding evolution of hardware, this method is promising to serve as an additional fast and cost-efficient operational sea ice forecast component of a forecast system in the future.

## Acknowledgements

The authors gratefully acknowledge the support by the Netherlands eScience Center and Wageningen University. This study is supported by Blue Action project (European Union’s Horizon 2020 research and innovation programme, grant number: 727852). We would like to thank SURFsara (Netherlands) for providing us their super computing infrastructure for our project. We also acknowledge the editor Dr. Josh Hacker, reviewer Dr. Steffen Tietsche and another anonymous reviewer for their help to improve the manuscript.

## Appendix - Numerical configurations of ConvLSTM

The whole numerical processes of training the ConvLSTM and making forecasts with ConvLSTM are illustrated in details in this part. We train ConvLSTM with the entire time series of the training set (1979-2008), which is a matrix with a dimension of  $X \times 24 \times 56 \times 1440$  ( $X$  indicates the number of input fields). The whole time series are fed into the ConvLSTM time step by time step. Each time step contains  $X \times 24 \times 56$  points and the convolution takes place in these input layers (e.g.  $W_{xi} * x_t$  in equation 4.1). During the convolution, the spatial structure ( $24 \times 56$ ) is preserved by using paddings. After the convolutional processes, the output (dimension  $Y \times 24 \times 56$  with  $Y$  indicates number of channels) is fed into the LSTM layer and it will pass through the input gate ( $i_t$ ) in LSTM (see equation 4.1). The cell state ( $c_t$ ) will be updated if this input gate is activated. At the same time, the forget gate ( $f_t$ ) will decide if the previous cell

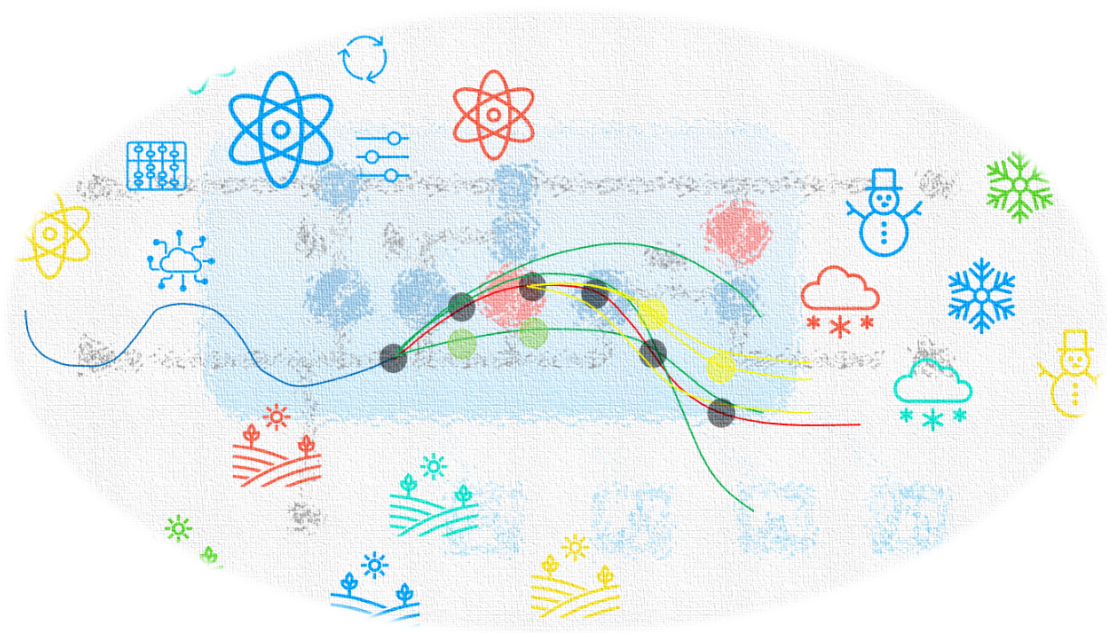
state ( $c_{t-1}$ ) need to be forgotten or not. Finally, the output gate ( $o_t$ ) will determine whether the hidden state ( $h_t$ ) shall include contributions from the current cell state ( $c_t$ ). It will provide a forecast for the next time step, and the hidden state ( $h_t$ ) and cell state ( $c_t$ ) will be passed to the next time step for LSTM related computations. The forecast will be evaluated using the chosen loss function (MSE in this case, see section 2.d) and it will provide a training error for this time step.

This process will be repeated until every time step in the time series of the train set has been fed into the ConvLSTM. Then the aggregation of the training error from each time step will be used to perform the back-propagation and one epoch of training is complete by now. The whole training procedure will be repeated until the required number of epoch is reached.

The convolutions take place through the implementation of filters inside convolutional cells before the start of LSTM processes. Therefore, the convolutions are included in the recurrence. By introducing convolutional layers, communications between adjacent cells are enabled and flow of spatial information is allowed between time steps. In other words, the cross correlations between neighbouring nodes are learnt by the neural network, both spatially and temporally. Multiple LSTM layers can be stacked to improve the complexity of the network.

The forecast (from 2013-2016 with testing set  $X \times 24 \times 56 \times 192$ ) procedure is the same as above, except for the back-propagation processes.

# Exploring Bayesian deep learning for weather forecasting with the Lorenz 84 system



Yang Liu, Jisk Attema, and Wilco Hazeleger. Exploring bayesian deep learning for weather forecasting with the lorenz 84 system. *Geophysical Research Letters*, 2021a. Under review

## Abstract

Recent developments in deep learning have led to many new neural networks and methods potentially applicable to weather forecasting and climate science. These techniques are nearly always based on deterministic deep neural networks (DNNs) and therefore prone to over-confident forecasts. The need for uncertainty quantification placed by weather forecasting makes Bayesian deep learning (BDL) a suited candidate for data-driven weather forecasting. Unlike feed-forward DNN, BDL builds probabilistic DNN by replacing fixed weights with distributions and thus incorporating uncertainty. Here, we aim to understand characteristics BDL in a weather application. In this study, we use Bayesian Long-Short Term Memory neural networks (BayesLSTMs) to forecast output from the Lorenz 84 system with seasonal forcing. The latter represents the dynamics of large scale eddies, or Rossby waves, on a westerly jet. We show that forecasts with the BayesLSTM can stay close to the attractor of the Lorenz model and conclude that they represent the nonlinear relations between each component in this simplified atmospheric circulation system. The forecasts are evaluated against persistence and a baseline statistical model, namely the Vector Autoregressive Model (VAR). We demonstrate that the BayesLSTMs are able to produce reliable probabilistic forecasts and address uncertainties relevant to weather forecasting. Our study indicates that BDL is an easy and fast solution for probabilistic weather forecast and is promising to enhance weather forecasting capabilities at short to medium-range timescales.



## 5.1 Introduction

Deep neural networks (DNNs) are capable of representing intricate features of data and have been proven to be useful for many scientific disciplines [e.g., LeCun et al., 2015], including weather forecasting and climate science [Reichstein et al., 2019]. It has been demonstrated by recent studies that typical DNN are able to mimic and predict the behavior of chaotic systems [e.g., Hochreiter and Schmidhuber, 1997, Chattopadhyay et al., 2019] and therefore they are potentially applicable to weather forecasting. However, mostly deterministic DNNs are considered and these are prone to overfitting and this can result in over-confident forecasts [Shridhar et al., 2019].

Due to the chaotic nature of the atmospheric dynamics and uncertainties in both initial conditions and models representing the atmosphere, weather forecasts are of probabilistic nature. In general, uncertainty estimation is achieved via an ensemble approach within trustworthy Numerical Weather Forecast systems (NWP) [Gneiting et al., 2007, Leutbecher and Palmer, 2008]. However, this strategy is computationally expensive for NWP-based weather forecasts. In order to meet the requirement for uncertainty quantification, many attempts have been made to adapt deterministic DNN to weather forecasting [e.g., Scher and Messori, 2018]. These efforts mainly involve generating a DNN-based ensemble through perturbing either the training data or the structure of DNN [e.g., Zaier et al., 2010, Wang et al., 2017]. However, in practice, this technique is computationally expensive due to multiple training cycles that are needed and it is often difficult to manually select proper perturbations which can approximate the error growth of a real dynamical system. Fortunately, recent developments in deep learning have led to a branch of DNN to cope with overfitting and address uncertainties, which is known as Bayesian deep learning (BDL).

Unlike feed-forward DNN, BDL is constructed by replacing fixed weights with distributions and therefore are designed to represent uncertainties [Blundell et al., 2015]. With a well-defined likelihood function, BDL is able to capture both the aleatoric and epistemic uncertainty [Kendall and Gal, 2017, Shridhar et al., 2018, 2019]. They can avoid making over-confident decisions and incorporate regularization naturally by implementing the variational approaches [Shridhar et al., 2019]. Together with the simplicity of implementing BDL on an already defined deep neural network, these make BDL an attractive approach for representing atmospheric dynamics and the practice of weather forecasting [Vandal et al., 2018].

An operational numerical weather forecast system is very complex. Here, we want to understand the characteristics of BDL within a simplified dynamical system that represents the essence of midlatitude atmospheric dynamics and explore the types of uncertainties addressed by BDL. In particular we examine how BDL can replicate the phase and amplitude of midlatitude Rossby waves on a jet as represented in a Lorenz 84 model [Lorenz, 1984, Wang et al., 2014]. The predictive nature and time scale of propagation and development of Rossby waves form the basis of short to medium-range weather forecasting. We will assess whether BDL can represent the predictability of this simplified atmospheric circulation system. We notice that the concept of BDL in the perspective of weather forecasting is quite similar to the implementation of the Bayesian theorem in data assimilation [e.g., Ghil and Malanotte-Rizzoli, 1991, Navon, 2009,

Bannister, 2017].

Long-Short Term Memory neural networks (LSTMs) have a network structure and characteristics that are found to be suitable to represent fluids in environmental studies [Liu et al., 2021b]. In this study, we explore BDL by turning LSTMs into Bayesian LSTMs (BayesLSTMs). We will use the BayesLSTMs to forecast the Lorenz 84 model and assess the forecast quality in the spatial and temporal space at different lead times. The probabilistic forecasts produced by the BayesLSTM will be evaluated against those with persistence of initial conditions and a baseline statistical model. An emphasis is placed on the uncertainties represented by the BayesLSTM and its capacity in preserving the physical consistency in a simplified atmospheric circulation system.

The paper is organized as follows: we elaborate on the concept of BDL and Lorenz 84 model with seasonal forcing in the section Methodology. An analysis of uncertainty estimation with BDL, and the procedure of sampling the BayesLSTM and generating ensemble forecasts are also provided in this section. The probabilistic forecasts of the Lorenz 84 system using the BayesLSTM are elucidated and analyzed in the section Results. This section also includes forecasts with persistence and a baseline statistical model for comparison and evaluation. Finally, in the section Conclusion and Discussion, we summarize this study and provide our perspective for future work.

## 5.2 Methodology

In this section, we briefly introduce the Lorenz 84 model with seasonal forcing and elaborate upon the concept of BDL as well as how an LSTM network is transformed into a BayesLSTM. Based on the characteristics of BDL, the procedure of producing ensemble forecasts and a description of uncertainty estimation with BayesLSTM is presented in this section.

### 5.2.1 Lorenz 84 Model with Seasonal Forcing

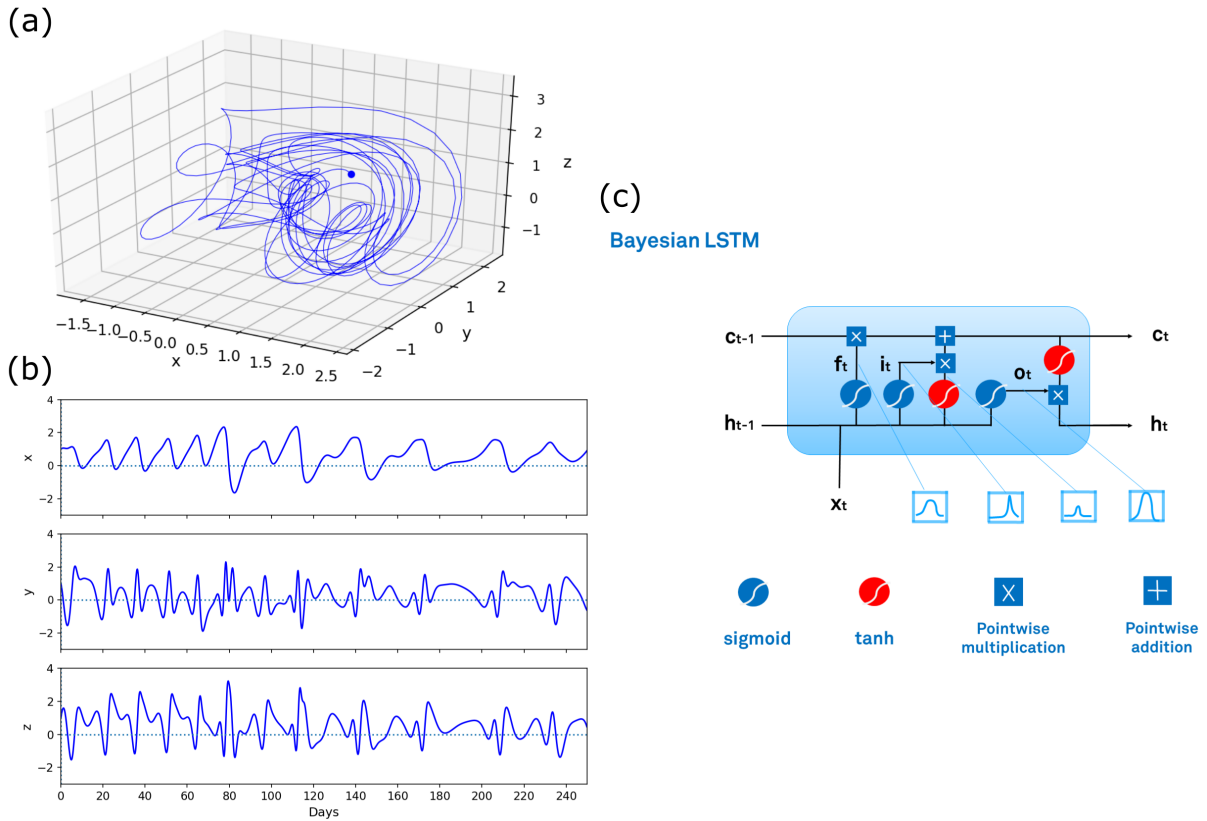
The Lorenz 84 system represents the general circulation of the atmosphere in a low dimensional space and therefore it is useful as a baseline model for exploring BayesLSTMs in weather forecasting [Lorenz, 1984]. To incorporate more realistic features into the simple Rossby wave evolution system, we add a seasonal forcing to the classical Lorenz 84 model. The dynamical system is formulated as follows:

$$\begin{aligned}\frac{dX}{dT} &= -Y^2 - Z^2 - aX + aF(1 + \epsilon \cos(\omega T)) \\ \frac{dY}{dT} &= XY - bXZ - Y + G(1 + \epsilon \sin(\omega T)) \\ \frac{dZ}{dT} &= bXY + XZ - Z\end{aligned}\tag{5.1}$$

where  $X$  represents the intensity of the westerly wind circulating around the globe,  $Y$  and  $Z$  represent the cosine and sine phases of a chain of superimposed large-scale eddies,  $T$  is the time,

$a$  and  $b$  indicate mechanical and thermal damping,  $F$  and  $G$  the symmetric and asymmetric thermal forcing,  $\epsilon$  the intensity of seasonal forcing, and  $\omega$  the angular frequency of seasonality [Freire et al., 2008]. In this study, we mainly focus on the sensitivity of the forecast quality to variations in the initial condition  $X$  and model parameter  $a$ .

To obtain a chaotic system that is suitable for the assessment of the BayesLSTM forecast, we chose the model parameters to be  $a = 0.25$ ,  $b = 4.0$ ,  $F = 8.0$ ,  $G = 1.0$ ,  $\epsilon = 0.4$ , and the initial conditions as  $X, Y, Z = 1.0$ . One unit of time in the Lorenz model corresponds to 5 days. The damping time of the wave is about 5 days [Lorenz, 1984]. We sample the system with a temporal resolution equal to  $1/30$  unit time, which is 4 hours. The period of seasonal forcing is taken as 73 unit time steps and then the period of the entire system is equivalent to 356 days. With this configuration, the trajectories and the time series of each variable are shown in Figure 5.1. Unless specifically noted, the time step and lead time steps in this paper are based on the sampling interval, which is 4 hours.



**Figure 5.1:** (a) Trajectory and (b) time series of each variable of the Lorenz 84 model with seasonal forcing. The sequences contain 250 days (1500 time steps) and the starting point is marked with a blue dot. (c) Structure of the Bayesian Long-Short Term Memory neural networks Fortunato et al. [2017].

### 5.2.2 BayesLSTM and Bayes by Backprop

Our aim is to investigate whether the BayesLSTM can represent the Lorenz 84 model described above. We can add Bayesian inference to an existing neural network by replacing fixed weights

with distributions [e.g. see Figure 1 in Blundell et al., 2015]. Given the structure of an LSTM network [Hochreiter and Schmidhuber, 1997], the Bayesian form of an LSTM network can be represented by equation 5.2:

$$\begin{aligned}
i_t &= \sigma(W_{xi}^s \circ x_t + W_{hi}^s \circ h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \\
f_t &= \sigma(W_{xf}^s \circ x_t + W_{hf}^s \circ h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}^s \circ x_t + W_{hc}^s \circ h_{t-1} + b_c) \\
o_t &= \sigma(W_{xo}^s \circ x_t + W_{ho}^s \circ h_{t-1} + W_{co} \circ c_t + b_o) \\
h_t &= o_t \circ \tanh(c_t)
\end{aligned} \tag{5.2}$$

with  $i_t$  the input gate,  $f_t$  the forget gate,  $c_t$  the cell state,  $o_t$  the output gate,  $h_t$  the hidden state,  $W^s$  the weight distribution,  $x_t$  the input,  $b$  the bias,  $\circ$  the element-wise product,  $\sigma$  the sigmoid function and  $\tanh$  the hyperbolic tangent function. The subscripts describe the corresponding weight matrix to different gates and states.  $W_{xi}^s$  indicates the weight matrix of input values related to the input gate, while  $W_{hf}^s$  represents the weight matrix of hidden states corresponding to the forget gate. The subscript  $t$  indicates the time step. The structure of a BayesLSTM is illustrated in Figure 5.1c.

We need to search for the weight distribution  $W^s$ , thus the posterior  $p(w|D)$  where  $w$  denotes the weight and  $D = (x_j, y_j)_j$  indicates the training set. As the true posterior probability distribution is intractable (because of the marginal likelihood), we use a variational inference scheme, namely the Bayes by Backprop approach, to approximate it [Blundell et al., 2015, Shridhar et al., 2018, 2019]. The reason for choosing this method is elaborated upon in detail in the appendix. A simple variational distribution  $q(w|\theta)$  (where  $\theta$  is the variational posterior parameter), such as a Gaussian distribution, or a lognormal distribution is often chosen [Blundell et al., 2015, Shridhar et al., 2018, Vandal et al., 2018]. Here we approximate the posterior  $p(w|D)$  with a Gaussian distribution  $q(w|\theta)$ , which consists of two trainable parameters  $\mu \in \mathbb{R}^d$  and  $\sigma \in \mathbb{R}^d$ . As a result,  $\theta$  in the assumed variational distribution  $q(w|\theta)$  can be denoted by  $\mathcal{N}(\theta|\mu, \sigma^2)$ .

The gap between the chosen variational distribution and the exact posterior distribution is reduced using the Kullback-Leibler (KL) divergence between  $p(w|D)$  and  $q(w|\theta)$  [Graves, 2011, Blundell et al., 2015]. KL divergence measures the similarity between two distributions and in this we define the optimal parameters  $\theta^*$  as:

$$\begin{aligned}
\theta^* &= \arg \min_{\theta} [q(w|\theta) || p(w|D)] \\
&= \arg \min_{\theta} KL[q(w|\theta) || p(w)] - \mathbb{E}_{q(w|\theta)} [\log p(D|w)] + \log p(D)
\end{aligned} \tag{5.3}$$

where  $KL$  indicates the full KL divergence operation and  $\mathbb{E}$  represents the expectation. This equation includes a data dependent part  $\mathbb{E}_{q(w|\theta)} [\log p(D|w)]$  and a prior dependent part  $KL[q(w|\theta) || p(w)]$  [Neal and Hinton, 1998, Blundell et al., 2015, Shridhar et al., 2019]. We sample the weight  $w$  from  $q(w|\theta)$  and the cost function that we optimize is:

$$\mathcal{F}(D, \theta) = \sum_{n=1}^N \log q(w^{(n)}|\theta) - \log p(w^{(n)}) - \log p(D|w^{(n)}) \quad (5.4)$$

where  $w^{(n)}$  denotes the  $n$ th Monte Carlo sampling from the variational posterior  $q(w^{(n)}|\theta)$ .

Together with the local reparameterization method, which translates the global uncertainty in the weights into a form of local uncertainty [Kingma et al., 2015, Shridhar et al., 2019], our BayesLSTMs are ready for training and back-propagation. We constructed the networks using the Pytorch library, and our code is published on Github (<https://github.com/geek-yang/DLACs>).

### 5.2.3 Ensemble Forecasting with BDL and Numerical Configurations

The ensemble method is generally used for uncertainty assessment in weather forecasting [Gneiting et al., 2005, Buizza et al., 2008, Leutbecher et al., 2017]. In numerical weather prediction systems (NWP), uncertainties in the initial conditions and model parameters are projected by ensemble forecasts with perturbations in the initial conditions and model formulations [Palmer, 2002, Milinski et al., 2020]. It has been explained in many previous studies that BDL is able to address the uncertainties in initial conditions and model parameters [e.g., Kendall and Gal, 2017]. This characteristic is fundamental for any probabilistic forecast and therefore makes BDL a candidate for weather forecasting. However, they are treated differently than in operational NWP approaches.

The ability of BayesLSTM to characterize uncertainty is reflected in its forecasting procedure. During a prediction process, the whole time series preceding the forecast date ( $t < t_0$ ) will be fed to the model to initialize the memory and position the state of the network. Therefore the model itself is constrained by the past and this is similar to an NWP-based forecast. When producing a forecast that takes uncertainties into account for a next time step, the BayesLSTM will first sample the weight distributions multiple times to build an ensemble and then use the sampled weight matrix to generate the predictions for the target time step ( $t = t_1$ ). The ensemble forecast can be extended to more time steps ahead ( $t = t_n > t_1$ ) by continuing with each individual LSTM.

In order to evaluate ensemble forecasts with the BayesLSTM, several scores are calculated, including continuous ranked probability score (CRPS), root mean square error (RMSE) and Euclidean distance (EuD). The mathematical expressions of these scores can be found in the appendix.

For all the experiments in this paper, we generate sequences including 1500 time steps (250 days) with the Lorenz 84 model. The training set contains 1300 time steps (about 216 days) and the validation set consists of 200 time steps (about 33 days). The optimization is based on the minimization of training loss, which consists of likelihood cost (data-dependent) and complexity cost (prior dependent) [Shridhar et al., 2019]. A scaling factor between these two sources of loss should be tuned, since it accounts for the trade-off between the width of ensemble spread in terms of uncertainty estimation and saturation of forecasts around the variance displayed in the

observations. Note that the scaling factor is related to the normalization of the distributions and cannot be calculated exactly. The training time is about 20 hours on a single GPU (Nvidia Tesla K40m). The hyperparameters like the learning rate, number of epochs and number of layers, were tested and determined in terms of the EuD error. It shows that a combination of a learning rate equal to 0.01, a single BayesLSTM layer and 3000 epochs is sufficient to achieve satisfying results. More details about the numerical configurations are shown in the appendix.

#### 5.2.4 Vector Autoregressive Model

The VAR model is used as a baseline method to assess the probabilistic forecast skill of BayesLSTM. As a variant of the autoregressive model (AR), the VAR model generalizes univariate AR by allowing for multivariate time series and therefore can capture the relation between multiple variables. The VAR model and many variants belonging to the VAR family have shown skill in many weather forecast applications [e.g., Gneiting et al., 2006, Wang et al., 2016a, and many others]. To expand its forecast capacity from the deterministic domain to the probabilistic domain, we replaced the Gaussian noise term ( $\epsilon_t$ ) with Gaussian distributed variations based on the variance of input time series from the chosen lag step to the current step. The optimal number of the lag to be included in the model is determined based on the auto-correlation of each variable of the Lorenz 84 model output, and tests of forecast quality in terms of the CRPS score. In our case, the VAR model with a lag equal to 3 provides the best probabilistic forecast. Mathematically, our modified VAR model can be expressed as:

$$\begin{aligned} X_t &= \alpha_1 + \sum_{l=1}^{Lag} (\beta_{11,l}X_{t-l} + \beta_{12,l}Y_{t-l} + \beta_{13,l}Z_{t-l}) + \epsilon_{1,t} \\ Y_t &= \alpha_2 + \sum_{l=1}^{Lag} (\beta_{21,l}X_{t-l} + \beta_{22,l}Y_{t-l} + \beta_{23,l}Z_{t-l}) + \epsilon_{2,t} \\ Z_t &= \alpha_3 + \sum_{l=1}^{Lag} (\beta_{31,l}X_{t-l} + \beta_{32,l}Y_{t-l} + \beta_{33,l}Z_{t-l}) + \epsilon_{3,t} \end{aligned} \quad (5.5)$$

with

$$\begin{aligned} \epsilon_{1,t} &= \mathcal{N}[0, \sigma(X_{t-1}, X_{t-2}, \dots, X_{t-l})^2] \\ \epsilon_{2,t} &= \mathcal{N}[0, \sigma(Y_{t-1}, Y_{t-2}, \dots, Y_{t-l})^2] \\ \epsilon_{3,t} &= \mathcal{N}[0, \sigma(Z_{t-1}, Z_{t-2}, \dots, Z_{t-l})^2] \end{aligned}$$

Where  $\alpha$  and  $\beta$  are trainable parameters in the model,  $\epsilon_t$  the Gaussian distributed variations, and  $X_{t-l}$ ,  $Y_{t-l}$  and  $Z_{t-l}$  the Lorenz model output at time lag  $l$ . The parameters were updated by fitting the model to the time series of Lorenz model output using maximum likelihood.

### 5.3 Results

We evaluate the capacity of BDL in representing the dynamics of Rossby wave propagation on a westerly jet by investigating the forecasts in the spatial and temporal domains. Based on

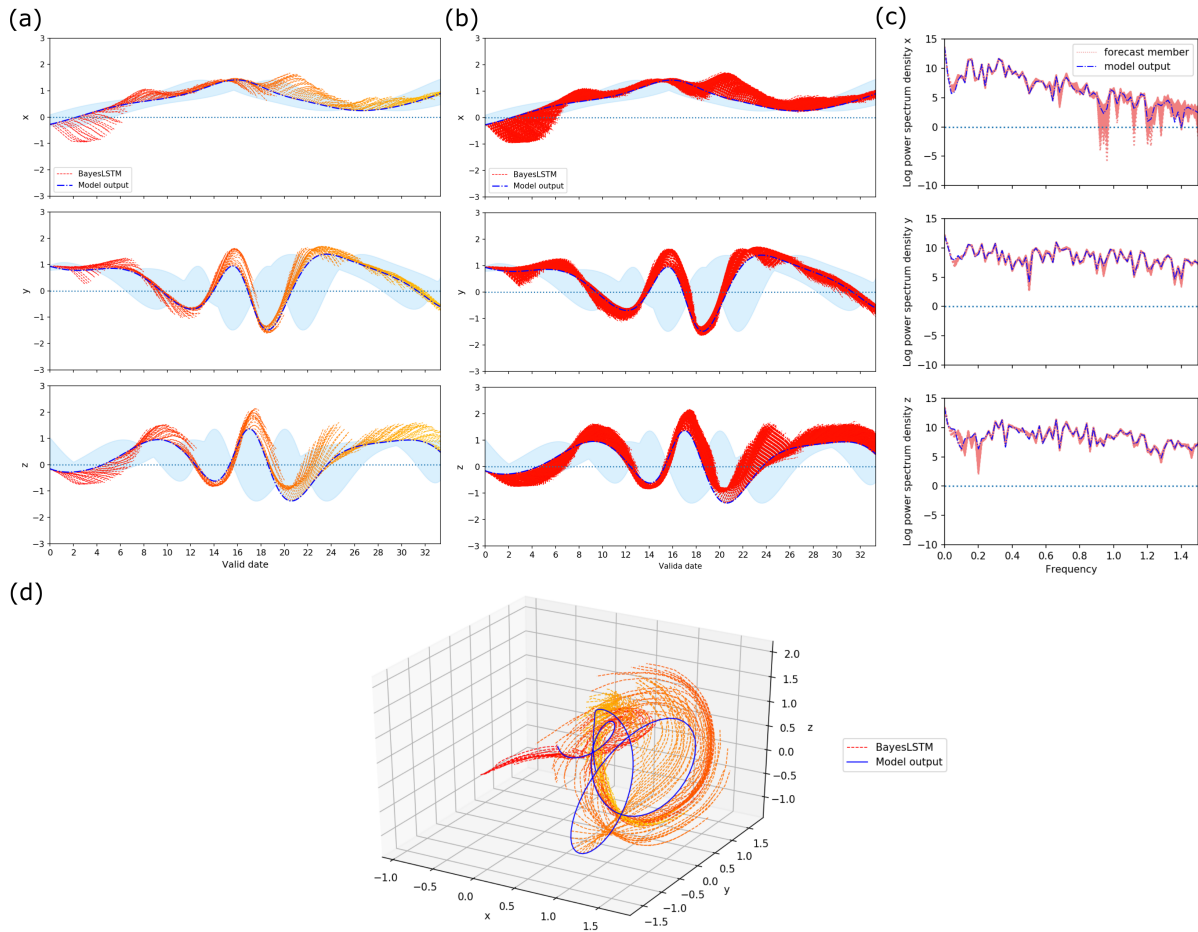
the selected scoring metrics, we further assess the forecast quality of BayesLSTM against the forecasts with persistence and a VAR model.

### 5.3.1 Representing the Evolution of Lorenz 84 Model

A retrospective forecast of the Lorenz 84 system with the BayesLSTM is shown in Figure 5.2. The forecasts start every time step (4 hours) and each has been extended to a lead time of 3 days. Given the time series of the BayesLSTM forecasts in Figure 5.2a, it can be observed that in general the forecasts are close to the time series of the Lorenz 84 model output, which is considered to be the “truth”. Although the forecast quality drops down with the increase of lead time as expected, the BayesLSTM shows good skill in replicating the variations of the Lorenz 84 model, especially for the state-transitions of the Lorenz 84 system and the sinusoidal patterns of the eddy components, like the forecast of  $X$  around valid date 14 and the forecast of  $Y$  around valid date 16. This indicates that the BayesLSTM learns to predict the state of the Lorenz system. Considering the typical predicting process of an LSTM network, in which the whole time series of the Lorenz 84 system preceding the forecast time should be fed to the system, it implies that our BayesLSTM is well constrained by the Lorenz 84 model output. Given the fact that the learning process of a deep neural network is characterized by the relationship between input fields, it further indicates that the non-linear relations between the variables in this Lorenz 84 system, the westerly  $X$  and the large scale eddies  $Y$  and  $Z$ , were addressed by the BayesLSTM.

In addition, we plot the forecast trajectory in Figure 5.2d and compare it with the Lorenz model output to further evaluate the performance of BayesLSTM. It can be noticed that the forecast trajectory is close to the attractor and the “behavior” of the forecast trajectory as a function of lead time resembles the evolution of the Lorenz 84 model. The result is consistent with the assessment based on the time series of each component as shown in Figure 5.2a. As a follow-up check, we investigate the physical consistency of BayesLSTM forecasts via the log power spectrum density of forecast time series, which is shown in Figure 5.2c. Only the high frequency components of  $X$  (with the frequency between 0.9 and 1.5) differ from the Lorenz model output. In general, the power spectrum density of the BayesLSTM forecasts is similar to that of the Lorenz 84 model. This indicates that the phases of the waves simulated by BayesLSTM do not differ much from the Rossby waves in the Lorenz 84 model. Considering the time step (4 hours) and the damping time of the Lorenz system (5 days), such similarity over the whole frequency space reflects that the BayesLSTM can account for the dynamics of this Rossby wave system across different time scales, which potentially benefits from its ability of multiple-level information abstraction. Together with the similar amplitudes of waves displayed in Figure 5.2a, it implies that the BayesLSTM manages to learn the Rossby wave propagation. The interaction between the jet and eddy components in this simplified atmospheric circulation system and the forecasts are physically realistic.

In order to evaluate the probabilistic forecast skill of the BayesLSTM, we generated a 20-member ensemble by sampling the BayesLSTM network and the time series of these retrospective forecasts up to 3 lead days are shown in Figure 5.2b. The blue shades serve to approximate the error growth of the Lorenz 84 system, which are selected as the range between the current



**Figure 5.2:** BayesLSTM retrospective forecasts up to a lead time of 3 days (18 time steps), with forecasts starting every time step (every 4 hours). (a) Time series of each variable (b) time series of a 20-member ensemble (c) logarithmic power spectrum and (d) trajectory in phase space. Except for (b) all the figures contain the results from a single BayesLSTM retrospective forecast. The Lorenz model output is included as reference (blue, labelled as “model output”) and the blue shades indicate the range between the Lorenz model output persisting for 3 days, both lead (3 days forward) and lag (3 days backward).

Lorenz model series persisting for 3 lead and lag days. Note that this selection is made based on the auto-correlation and it aims to assist the evaluation of the probabilistic forecasts, specifically for the uncertainty estimation. It is observed that the forecast members are located around the Lorenz model output and the spread is comparable to the error growth of this Rossby wave system. This indicates that the spread of the BayesLSTM ensemble is neither over-dispersive nor under-dispersive. The probabilistic forecasts therefore address uncertainties in a reasonable way. Collectively, the development of these forecasts as a function of lead time in 5.2b are similar to the single forecast in 5.2a. This means almost all the ensemble members capture the properties of the propagating waves and the jet strength while allowing for the occurrence of uncertainty. Consequently, the probabilistic forecasts generated by sampling the BayesLSTM are physically plausible.



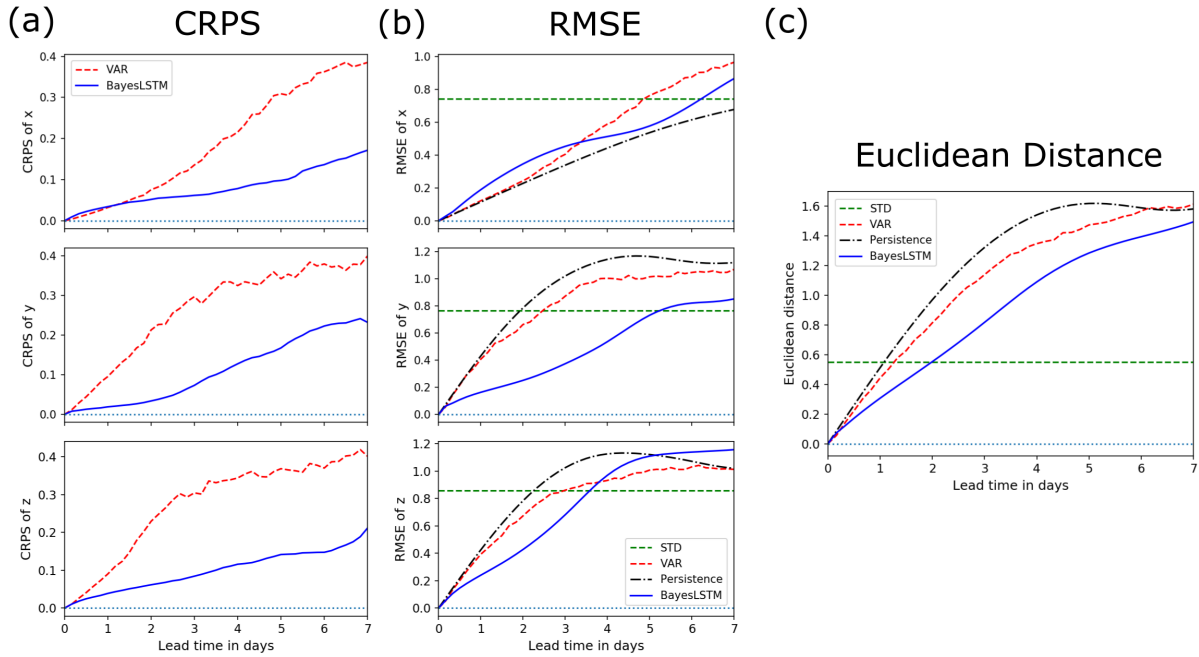
Nevertheless, the BayesLSTM forecasts may lose skill at certain valid time. For instance, in Figure 5.2a and b between valid date 0 to 6, forecasts of  $X$  drift away unrealistically. This might result from the state-dependency of the BayesLSTM, or in general the state-dependency of any deep learning approaches. For a numerical model, it is common to have state-dependency, for example, the prediction of NAO/blocking events in medium-range forecasts [e.g., Parker et al., 2018]. This may also apply to the deep learning approaches if the training data fails to provide adequate information for forecasting at some points.

### 5.3.2 Evaluate the BayesLSTM ensemble forecasts

A reliability assessment of probabilistic forecasts with the BayesLSTM ensemble was performed using the chosen metrics. The BayesLSTM ensemble consists of 20 members and they are evaluated against a deterministic forecast with persistence and a probabilistic forecast with the VAR model, which is also a 20-member ensemble. The results are shown in Figure 5.3. Regarding the CRPS score, in general the BayesLSTM ensemble forecast is better than the VAR ensemble forecast considering all the variables for almost all lead days. Only around day 1 for predictand  $X$ , the VAR ensemble forecast shows slightly better skill. The error growth of the BayesLSTM ensemble forecast is much slower than that of the VAR ensemble forecast. Given the definition of CRPS score, which provides a quadratic measure of discrepancy between the forecast cumulative density function (CDF) and the empirical CDF of the scalar observation [Gneiting et al., 2005], this indicates that the forecast CDF with the BayesLSTM centered around the Lorenz model output, while the forecast CDF with the VAR is relatively over-dispersive.

Regarding the RMSE shown in Figure 5.3b, forecasts with persistence are better than that with BayesLSTM and VAR ensemble concerning only  $X$ . This is consistent with the high auto-correlation of the zonal wind  $X$ . While for the eddy components  $Y$  and  $Z$ , the BayesLSTM provides much better forecasts within 3 lead days, with the averaged RMSE error smaller than the standard deviation of the full time series of the Lorenz 84 model output. Considering the nonlinear relation between the westerly  $X$  and large scales eddies  $Y$  and  $Z$ , this means that the BayesLSTM is able to preserve the physical consistency between the zonal wind and the propagation of large scale eddies in this atmospheric circulation system, and therefore produces better probabilistic forecasts. It is evident by analyzing the time series in Figure 5.2a, that the variations of  $Y$  and  $Z$  are well represented by the BayesLSTM forecasts up to a lead time of 3 days.

More information about forecast quality in terms of the trajectories, which intrinsically embody the properties of Rossby waves and jet strength, is reflected by the EuD in Figure 5.3c. Starting from the first forecast time step (4 hours), the BayesLSTM shows better forecast skill concerning the EuD. Although the EuD error grows with the increase of lead time for all the forecast methods, the BayesLSTM forecasts are better than the others for the whole inspected lead time range. Note that within 2 lead days, the EuD error of BayesLSTM is smaller than the standard deviation of the Lorenz 84 model output, which is about 0.6. Since the EuD of BayesLSTM ensemble forecast shown in Figure 5.3c is the average of 20 members with forecasts starting every time step, this implies that these ensemble members are able to replicate the patterns of the attractor and the spread of the ensemble is properly distributed around the target Lorenz



**Figure 5.3:** Skill evaluation of the BayesLSTM ensemble forecasts against VAR and persistence with (a) CRPS and (b) RMSE and (c) EuD, which are averaged over 200 forecasts starting every time step (4 hours). The standard deviation of the full time series of Lorenz model output (based on 250 days data) is included in (b) and (c) (green, labelled as “STD”).

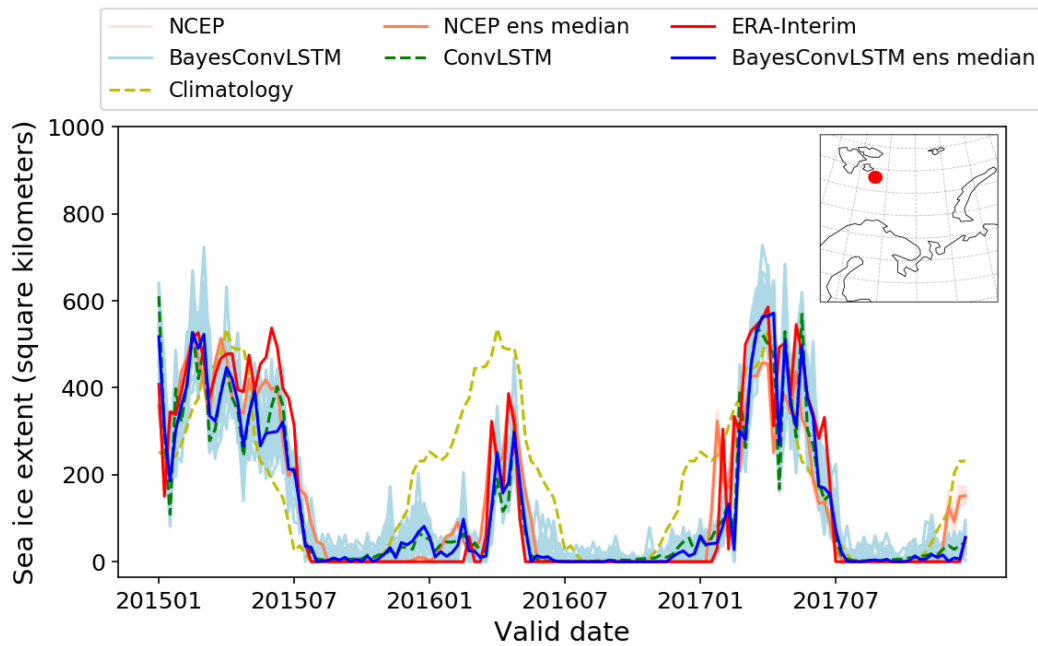
84 model trajectory. It further demonstrates that probabilistic forecast with BayesLSTM can address uncertainties adequately.

### 5.3.3 Probabilistic Sea Ice Forecasting with BDL

To better assess the probabilistic forecast skill of the BDL and provide more insights into its capability with the climatological fields from the real climate system, we revisited the topic of sea ice forecasting in the Barents Sea in Chapter 4 with the BDL. Similar to the transformation of the LSTMs into the BayesLSTMs, we built the Bayesian ConvLSTM (BayesConvLSTM) on the basis of the ConvLSTM and used it to perform a retrospective sea ice forecast with the operational set-up, as explained in Chapter 4. Concerning the training process, the input fields include T2M, SLP, Z850, SFlux, and SIC. The numerical and experimental configurations are exactly the same as those in Chapter 4, except that we chose 2015 to 2017 as the new valid dates. This choice was made because of the data availability of NCEP ensemble forecast from S2S project, which is used for evaluation (the ECMWF forecast system from S2S project only produced deterministic forecasts of sea ice for the inspected period and therefore it is not included here). A brief introduction about this numerical weather and seasonal forecast system is available in Chapter 4. Note that the NCEP ensemble forecast is initialized by sea ice conditions from the NCEP climate forecast system reanalysis (CFSR), and not from the ERA-Interim. Although previous study has shown that the sea ice fields in the CFSR are similar to those in the ERA-40 [Wu et al., 2010], they still differ from the ERA-Interim. As a consequence,

the comparison should be interpreted with cautions.

To begin with, we investigate the time series of sea ice forecasts at different locations. For instance, the retrospective operational sea ice forecast at  $76.75^\circ$  East and  $25.5^\circ$  North in the Barents Sea is shown in Figure 5.4. This example was chosen because of the rapid sea ice advance and retreat at this location. Forecasts from the ConvLSTM and climatology are also included for comparison. In general, the observation (ERA-Interim) falls into the spread of the BayesConvLSTM forecasts in the inspected period. This shows that the uncertainty addressed by the BayesConvLSTM is relevant to the variability of sea ice in this area, as well as to weather forecasting. The ensemble mean of the BayesConvLSTM forecasts resembles the ConvLSTM forecast and the ensemble mean of the NCEP ensemble forecast, but they all differ from the ERA-Interim at certain periods, for instance, around June in 2015 and around February in 2017. For the rest of the inspected time, the forecasts from the BayesConvLSTM, the ConvLSTM and the NCEP ensemble are similar to the observations and they are much better than the climatology. This indicates that forecasts by BDL show strong state-dependency like the numerical weather forecast system. It is consistent with our findings from the forecasting of Lorenz 84 system by the BDL, as elaborated upon in the section 5.3.1.

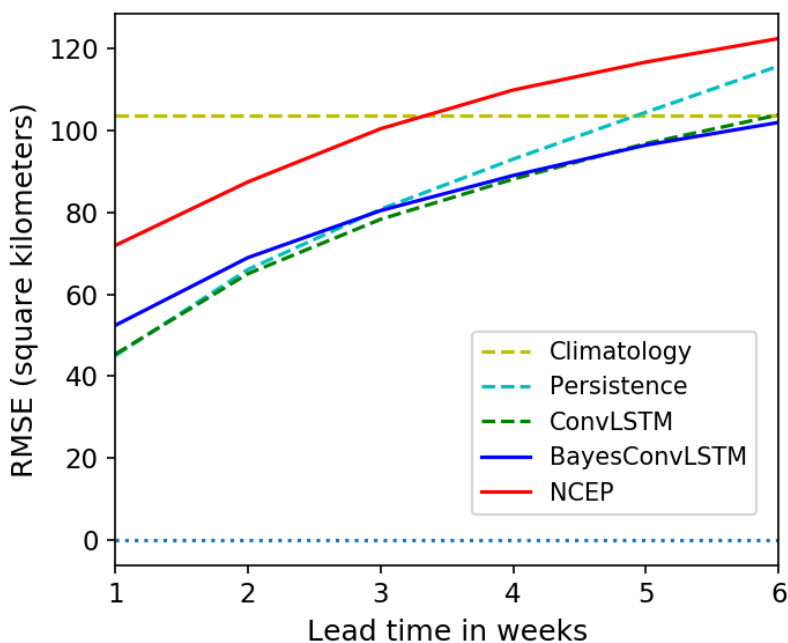


**Figure 5.4:** Time series of the sea ice forecast with BayesConvLSTM, NCEP ensemble forecast, ConvLSTM (green) and climatology (yellow) at  $76.75^\circ$  East and  $25.5^\circ$  North for the first week from 2015 to 2017. Both the members (BayesConvLSTM, light blue) and the ensemble median (BayesConvLSTM ens median, blue) from the BayesConvLSTM forecasts are included. The figure also contains the member (NCEP, light orange) and the ensemble median (NCEP ens median, orange) from the NCEP ensemble forecast. The ground truth (ERA-Interim) is illustrated in red. The results are weighted by the size of the grid and the unit is square kilometers.

It can be observed that the spread of the NCEP ensemble forecast is much smaller compared to the BayesConvLSTM forecasts and it is not sufficient to cover the observed variations of sea ice

in the ERA-Interim. This implies that the NCEP ensemble forecast is under-dispersive and the BayesConvLSTM forecasts perform in the perspective of uncertainty estimation, though it might be relatively over-dispersive and this needs further inspection. Note that concerning the ice-free seasons, some members of the BayesConvLSTM forecasts erroneously predicted the presence of sea ice and the ensemble median is similar to the ConvLSTM forecast. This shows that the bad performance of the BayesConvLSTM in these ice-free periods are associated with both the uncertainty quantification and the representation of the physical processes by the neural network. There is still room for improvement and it is worth further investigation.

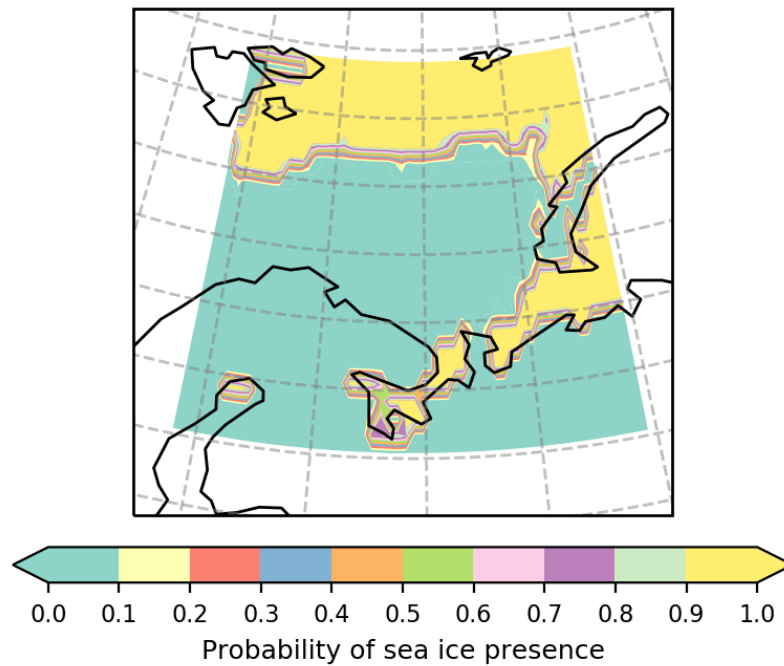
The forecast quality with the deep neural networks and numerical climate models deteriorates with the increase of lead time, which is illustrated by the mean RMSE of the sea ice forecast in the Barents sea for the inspected period with a lead time up to 6 weeks in Figure 5.5. Note that the RMSE errors of the BayesConvLSTM forecast and the NCEP ensemble forecast are the RMSEs averaged over the ensemble members. The performance of the BayesConvLSTM forecast is slightly worse than persistence within a lead time up to 3 weeks and becomes better after the 3rd lead week. The error growth of the BayesConvLSTM forecasts resembles that of the ConvLSTM forecast and both of them show better skill than the NCEP ensemble forecast. As mentioned before, this might relate to the initialization by sea ice conditions.



**Figure 5.5:** Mean RMSE of the operational forecast of SIC in the Barents Sea with a lead time up to 6 weeks from the forecasts with BayesConvLSTM, NCEP ensemble forecast system, ConvLSTM, persistence and climatology, concerning all the valid dates from 2015 to 2017. For the BayesConvLSTM forecasts and the NCEP ensemble forecast, the RMSEs are averaged over all members.

These probabilistic forecasts are useful to determine the probability of the presence of sea ice, which are of interest to shipping and expeditions in the Arctic. Same as the criterion adopted in Chapter 4, the grid boxes with sea ice concentration less than 15% are regarded as open water

areas [Van Woert et al., 2004, Walsh et al., 2019]. An example of the probability of sea ice presence based on the binary forecasts from 20 members with the BayesConvLSTM is shown in Figure 5.6. This figure illustrates whether a certain area in the Barents Sea is open water or covered by sea ice for the first week in March 2015, as predicted by the BayesConvLSTM. It is noted that the disagreement between forecast members mainly lies in the areas close to the sea ice edge. This is physically consistent with the features of sea ice variations in this area, and further reflects that the BayesConvLSTM is able to account for the features of sea ice variability. Moreover, it also shows that the uncertainty produced by the BayesConvLSTM is relevant to the evolution of this physical system.



**Figure 5.6:** Probability of the presence of sea ice for the first week in March 2015, based on 20 members from the BayesConvLSTM forecasts for the first lead week. The grid boxes with sea ice concentration less than 15% are regarded as open water areas.

To conclude, the BayesConvLSTM shows skill in operational sea ice forecasts and it can generate uncertainties relevant to the natural variability of sea ice. Concerning its probabilistic features that similar to the combination of ensemble approach and numerical weather prediction systems, it can be an useful tool for operational weather forecasting. The potential of the BDL in forecasting the evolution of a real climate system is not fully explored here due to the limited time and a different scope of this study, but it is worth more efforts in the future.

## 5.4 Discussion

We demonstrate the capability of BayesLSTM in probabilistic weather forecasting. Intuitively, by perturbing the Lorenz 84 model, it seems possible to compare the BayesLSTM forecasts to the perturbed Lorenz model output and check if the BayesLSTMs are able to address uncertainties in the initial conditions and model formulation, respectively. However, there is no objective way

to determine the amplitude of the perturbation which can appropriately approximate the error growth in the Lorenz system that analogue to a realistic dynamical system of Rossby waves on a jet. So this experiment is not feasible at the moment.

In addition, we extended the ensemble forecasts to more than 60 lead days and noticed that after 20 days the forecast errors increase dramatically with the increase of lead time (not shown). From this point, it seems that the BayesLSTM is useful for medium-range forecasts and it is not suitable for seasonal forecast and climate change predictions. The outcome of this study is insufficient to prove that, either the Bayesian deep neural networks can mathematically represent the differential equations which depict the Lorenz 84 system with seasonal forcing (note that due to the features of deep learning and the nature of deep neural networks, there is no direct mapping between weight matrix in a trained BayesLSTM and Lorenz model parameters), or BayesLSTMs only abstract and store the physical linkages in a latent space and use them to produce memory-based forecasts at relatively short time scales. This can be explored in the future.

Although not the main topic of this paper, we note that the formulations of BDL are very similar to data assimilation, specifically the Bayesian data assimilation, which is extensively used in weather forecasting to combine the knowledge from observations and models, and deal with the uncertainty in the initial conditions [Evensen, 1994, Houtekamer and Mitchell, 1998, Houtekamer and Zhang, 2016]. Based on the Bayes' theorem, it incorporates model knowledge into the prior and corresponding observations as likelihood, and treats the observation involving uncertainty estimation as posterior. Given the large dimensional systems, in reality approximate solutions are always made based on different methods, like variational methods, Kalman-based methods and particle filters [Navon, 2009].

Given the fact that forecasts with BayesLSTM stay close to the Lorenz 84 attractor, the BayesLSTM may be also chaotic. This question can be answered by the chaotic system diagnostics, for instance, with the Lyapunov spectrum [Broer et al., 2002, Freire et al., 2008]. However, this is beyond our scope now but worth the effort in the future. So far, it can be concluded that BayesLSTM is a useful candidate for weather forecasts, at relatively small lead times up to several days. For a long term climate forecast, the BayesLSTM may not be a good choice in terms of the error accumulation and the lack of skill in physical model representation. Also, for simulating and forecasting changes in the climate system boundary condition uncertainty will need to be taken into account. This can be further tested by studies using observational data and climate model ensembles in the future.

## 5.5 Conclusion

In this study, we explored the potential of BDL for weather forecasting using the modified Lorenz 84 model as a model for the atmosphere. The probabilistic character of the BDL is addressed and assessed using the chaotic nature of the Lorenz 84 system with seasonal forcing as "truth". Specifically, we chose BayesLSTM as an example of BDL to forecast the Lorenz 84 model and evaluate its forecast skill. It was observed that the retrospective forecasts are similar to those of the Lorenz model output in the spatial and temporal domain. The forecast trajectories are close

to the attractor. This indicates that BayesLSTM is able to learn the propagation of Rossby waves in this atmospheric system, in terms of both the amplitude and phase. It further demonstrates that the BayesLSTM is able to replicate the interaction between the jet stream and large-scale eddies and thus the evolution of Rossby waves on a midlatitude jet. The forecasts get worse with increasing lead times due to the accumulation of errors, as expected.

The probabilistic forecast skill of BayesLSTM was analyzed and evaluated against persistence and a VAR model. We found that the BayesLSTM forecasts saturate around the model output considering both the sequences of each variable and the trajectory. In terms of the scores in the chosen metrics, the BayesLSTM shows better probabilistic forecast skill than persistence and the VAR model in the inspected lead days. It shows that the BayesLSTM is able to account for uncertainties relevant to the evolution of this simplified atmospheric circulation system, though the procedure differs from well-known NWP based approaches. Given the relatively low cost of ensemble forecasts compared to deterministic DNN and NWP systems, and the capacity in probabilistic forecasting, BayesLSTM, or in general BDL, is useful to produce fast and reliable probabilistic weather forecast and therefore is promising to enhance weather forecasting capabilities at short to medium-range timescales.

Moreover, we assessed the probabilistic forecast skill of the BDL in the real climate system by revisiting the case study of sea ice forecasting in the Barents Sea. We built BayesConvLSTM by applying the BDL to the ConvLSTM and used it to forecast sea ice with the same setup as that given in Chapter 4. The probabilistic forecasts with the BayesConvLSTM were evaluated against the NCEP ensemble forecast, as well as deterministic forecasts by climatology, persistence and the ConvLSTM. We found that, in general, the BayesConvLSTM forecasts are better than the NCEP ensemble forecast and forecast by climatology at a lead time up to 6 weeks. They are comparable to the ConvLSTM forecasts and better than the forecast by persistence with a lead time of more than 3 weeks.

Considering the binary forecast with the BayesConvLSTM, which provides insight into the probability of the presence of sea ice in the inspected area, we demonstrate that the BayesConvLSTM can quantify uncertainties relevant to the natural variability of sea ice. Our exploration suggests that the BayesConvLSTM, or generally the BDL, is useful for the operational probabilistic forecasts of sea ice. However, it should be noted that this is only a preliminary check and a deep look at the potential of the BDL in forecasting the real climate system is necessary for future studies.

## Acknowledgements

The authors gratefully acknowledge the support by the Netherlands eScience Center and Wageningen University. This study is supported by Blue Action project (European Union's Horizon 2020 research and innovation programme, grant number: 727852). We would like to thank SURFsara (Netherlands) for providing us their super computing infrastructure for our project.

## Appendix

This appendix includes additional information regarding BDL and the variational inference method for training the Bayesian neural networks, namely the Bayes by Backprop. It also includes a brief introduction of the local re-parameterization method and the evaluation metrics. A detailed explanation of the lead time dependent forecasts and numerical configurations of training and forecasting with BayesLSTM is provided.

**A1. BDL and Bayes by Backprop** In this section, we elaborate on BDL and Bayes by Backprop in detail. Deep neural networks are equipped with multiple layers of fixed weights to extract the multiple levels of abstraction in the training data LeCun et al. [2015]. These neural networks are deterministic and prone to overfitting and forecasts with over-confidence Blundell et al. [2015]. In order to incorporate uncertainty estimation during training and forecasting, we need to transform the deterministic neural network into a probabilistic structure and therefore we seek help from the Bayes' theorem.

To begin with, we have a training set  $D = (x_i, y_i)_i$  which consists of input data  $x \in \mathbb{R}^d$  and output data  $y \in \mathbb{R}^d$ . We aim for a probabilistic neural network  $p(y|x, w)$  with a distributed weight  $w$ . To obtain the weight distribution, we apply the Bayesian inference and we will get the expression:

$$p(w|x, y) = \frac{p(y|x, w)p(w)}{p(x, y)} \quad (5.6)$$

By substituting  $p(x, y)$  according to the law of total probability, we can obtain:

$$p(w|x, y) = \frac{p(y|x, w)p(w)}{\int p(y|x, w)p(w)dw} \quad (5.7)$$

The integral in the denominator requires coverage of all possible values of  $w$ , which is computationally prohibitive. This makes the true posterior probability distribution intractable Blundell et al. [2015], Shridhar et al. [2018, 2019]. To solve this, various approximation methods (e.g. maximum-a-posteriori scheme, variational inference schemes) were studied in the past Graves [2011], Shridhar et al. [2018]. Among all of them, variational inference schemes are widely embraced and they work well for a wide range of BDL applications. So far, three popular variational inference schemes are always used to approximate the intractable posterior: Monte Carlo Markov Chain (MCMC), Monte Carlo dropout (MCD) and Bayes by Backprop (BBB).

MCMC is a family of methods that combines stochastic variational inference and Monte Carlo approaches Salimans et al. [2015]. Unlike variational inference in general which takes a random draw from a simple variational distribution and keeps optimizing the distribution, MCMC methods subsequently apply a stochastic transition operator to the random draw, so as to cover the exact posterior distribution. As long as the iterations are sufficient, MCMC can approximate the exact posterior well. However, we do not know if the iterations are sufficient and this procedure is always very costly.



An alternative is MCD, which uses dropout to perform variational inference where the variational distribution comes from a Bernoulli distribution. Hence it is also known as Variational inference with Bernoulli Distribution Gal and Ghahramani [2015, 2016]. This method makes use of dropout in each layer of neural network during training as well as testing and thus the process is equivalent to sampling from a Bernoulli distribution and provides a measure of uncertainty. Nevertheless, it provides a rough approximation of the target posterior and the control of uncertainty is limited by the complexity of the neural network.

Given the drawback of MCMC and MCD, in this study, we choose a more functional and affordable approach, BBB, to approximate the posterior. This method is a back-propagation (BP) compatible variational inference approach that estimates a density function with a known distribution and progressively update it with BP Blundell et al. [2015], Shridhar et al. [2018]. More details about BBB can be found in the section Methodology in the main body of the paper.

**A2. Local re-parameterization method** With the implementation of BBB and optimization function, it seems our BayesLSTM is ready for training. However, the whole procedure is not ready for back-propagation (BP) due to the stochastic node including  $\mathcal{N}(\theta|\mu, \sigma^2)$  in our chosen variational distribution. In order to enable the optimization of the parameters of Gaussian posterior with BP, we introduce the local reparameterization method, which translates the global uncertainty in the weights into a form of local uncertainty Kingma et al. [2015], Shridhar et al. [2019]. In this case, we replace  $\mathcal{N}(\theta|\mu, \sigma^2)$  with  $w = \mu + \sigma * \epsilon$  where  $\epsilon \sim \mathcal{N}(0, 1)$ . More information about the local reparameterization method can be found in the related literature Kingma and Welling [2013], Kingma et al. [2015], Shridhar et al. [2018, 2019].

**A3. Estimate uncertainty with BDL** Typically, three types of uncertainties are considered: (1) Uncertainties in the initial conditions (2) Necessary approximations and corresponding uncertainties in the construction of a numerical model of the real atmosphere (3) Uncertainties posed by external forcing and boundaries. The last one is often ignored in an operational weather forecast system, but relevant for climate studies and local forecasts. [Kendall and Gal, 2017] explained how the model uncertainty and initial condition uncertainty are addressed by BDL respectively. Given the nature of the forward process of BDL, model uncertainty is addressed by the BDL via placing a prior distribution over a model's weight, which is already explained by the prior dependent part  $KL[q(w|\theta)||p(w)]$  during the training of BDL when minimizing the KL divergence [Kendall and Gal, 2017].

The representation of initial condition uncertainty with BDL is less straightforward. During training, the BayesLSTMs approach output  $y_j \in \mathbb{R}^d$  with an input  $x_j \in \mathbb{R}^d$  and weight distribution  $w_j$ , as  $P(y_j|x_j, w_j)$ . However, there is always noise  $\epsilon_j$  in the input fields (initial conditions) and this propagates through the network and affects the output, thus  $P(y_j|x_j, w_j, \epsilon_j)$ . With the help of Bayes rule, we bring in the pre-knowledge of model weight distribution as prior and search for the best weight distribution as posterior that is able to achieve the testing set  $\hat{x}_j$  and  $\hat{y}_j$  following  $P(\hat{y}_j|\hat{x}_j) = \mathbb{E}_{P(w_j|D)}[P(\hat{y}_j|\hat{x}_j, w_j)]$ . To simplify the problem, we assume a Gaussian distribution for the prior and the weight distribution then can be solved via maximum a posteriori (MAP). Due to the use of the BBB approach, we need to solve the KL divergence and the knowledge of observation is brought in via the data-dependent term  $\mathbb{E}_{q(w|\theta)}[\log p(D|w)]$  in equa-

tion 3 in the paper, and finally the  $-\log p(D|w^{(j)})$  in equation 4. It then can be reformulated as:

$$\mathbb{L}(\theta, D) = -\frac{1}{N} \sum_{j=1}^N \log p(y_j | f^{\hat{w}_j}(x_j, \epsilon_j)) \quad (5.8)$$

If we assume the error in the output is Gaussian, then  $-\log p(y_j | f^{\hat{w}_j}(x_j, \epsilon_j)) \propto \frac{1}{2\sigma^2} \|y_j - f^{\hat{w}_j}(x_j)\|^2 + \frac{1}{2} \log \sigma^2$ . The shape of this Gaussian probability is determined by the variance  $\sigma$ , which is in this case the deviation between the ground truth in reality and the uncertainty corresponding to the noise in the input fields  $\epsilon_j$ . As a result, the data dependent loss evolves as:

$$\mathbb{L}_{NN}(\theta, D) = \frac{1}{N} \sum_{j=1}^N \frac{1}{2\sigma(\epsilon_j)^2} \|y_j - f^{\hat{w}_j}(x_j)\|^2 + \frac{1}{2} \log \sigma(\epsilon_j)^2 \quad (5.9)$$

with  $\sigma(\epsilon_i)$  indicating the variance as a function of noise in the input fields, which represents the uncertainty in the initial conditions. In other words, uncertainty in the initial condition is modeled through the neural networks by generating a distribution over the output of the model. Note that this is recognized as heteroscedastic and aleatoric uncertainty in machine learning and it does not exist in a non-Bayesian (deterministic) neural network since this observation noise parameter  $\epsilon$  is fixed as part of the model's weight decay and therefore neglected after training Kendall and Gal [2017]. So far, our analysis is based on the hypothesis of linear propagation of expectation with variational inference in the LSTM Fortunato et al. [2017]. However, it should be noted that this becomes different when forecasting more than one step ahead since there is also uncertainty coming from the forecasts of previous time steps.

**A4. Lead time dependent forecast and numerical configurations** The whole numerical processes of training the BayesLSTM and making lead time dependent forecasts with BayesLSTM are described in details in this section. We generate time sequences of variable X, Y and Z with modified Lorenz 84 model, which including 1500 time steps in total. We perform sequence-to-one forecast with BayesLSTM and the first 1300 time steps are used for training. During training, in each epoch we pass 3 x 1300 points to the BayesLSTM model. The model samples its weight distribution at each time step (with 3 points X, Y and Z passed to the model) and after 1300 iterations we perform back-propagation to update the BayesLSTM model. The procedure will be repeated until the maximum epochs are reached (or the early stop module is called, depending on the setup).

The validation set consists of 200 time steps. The forecasting procedure is the same as the training process, except that there is no back-propagation and we perform lead time dependent forecasts Liu et al. [2021b]. At each forecast time step  $t_n$ , we use the Lorenz model output (observation) to initialize the model and make forecasts, which can be described by the equation below:

$$(X, Y, Z)_{pred[t_n]} = BayesLSTM((X, Y, Z)_{obs[t_0, t_1, t_2, \dots, t_{n-1}]}) \quad (5.10)$$

with  $(X, Y, Z)_{pred[t_n]}$  denotes the forecast  $X$ ,  $Y$  and  $Z$  at time step  $t_n$ , and  $(X, Y, Z)_{obs[t_m]}$  represents the Lorenz model output of  $X$ ,  $Y$  and  $Z$  at time step  $t_m$  with  $m < n$ . It is still sequence-to-one prediction, which means data at time step  $t_0$  to  $t_{n-1}$  will be passed to the BayesLSTM model to make the forecast for the time step  $t_n$ .

For one more lead time step  $t_{n+1}$ , the forecast is based on both the model output from  $t_0$  to  $t_{n-1}$  and the forecast of time step  $t_n$ , and therefore it can be expressed as:

$$(X, Y, Z)_{pred[t_{n+1}]} = BayesLSTM((X, Y, Z)_{obs[t_0, t_1, t_2, \dots, t_{n-1}]} + (X, Y, Z)_{pred[t_n]}) \quad (5.11)$$

It can be noticed that starting from the time step  $t_{n+1}$  (the second lead time step), the forecast quality also depends on the forecasts of previous time steps. This helps us explore how far the BayesLSTM can predict the Lorenz system, which corresponds to the actual predictability of a weather and climate system.

**A5. Evaluation metrics** In order to evaluate these ensemble forecasts, we include continuous ranked probability score (CRPS), root mean square error (RMSE) and Euclidean distance.

CRPS is a popular verification tool for the assessment of probabilistic forecast systems. It is used to evaluate an ensemble forecast against a single deterministic observation and it has the following form Gneiting et al. [2005]:

$$CRPS(F, v) = \int_{-\inf}^{\inf} [F(r) - H(r - v)]^2 dr \quad (5.12)$$

with  $F$  the predictive cumulative density function (CDF),  $v$  the verifying observation,  $r$  the threshold value, and  $H(r - v)$  the Heaviside function which takes the value 0 when  $r < v$  and otherwise 1.

RMSE score is also included in this study and it is defined in the following way so as to work with the probabilistic forecast results:

$$RMSE = \frac{1}{N} \sum_{n=1}^N \frac{1}{T} \sum_{t=1}^T \sqrt{(x_{pred} - x_{obs})^2} \quad (5.13)$$

with  $N$  the number of ensemble members,  $T$  the number of time steps,  $x_{pred}$  and  $x_{obs}$  the predicted and observed value.

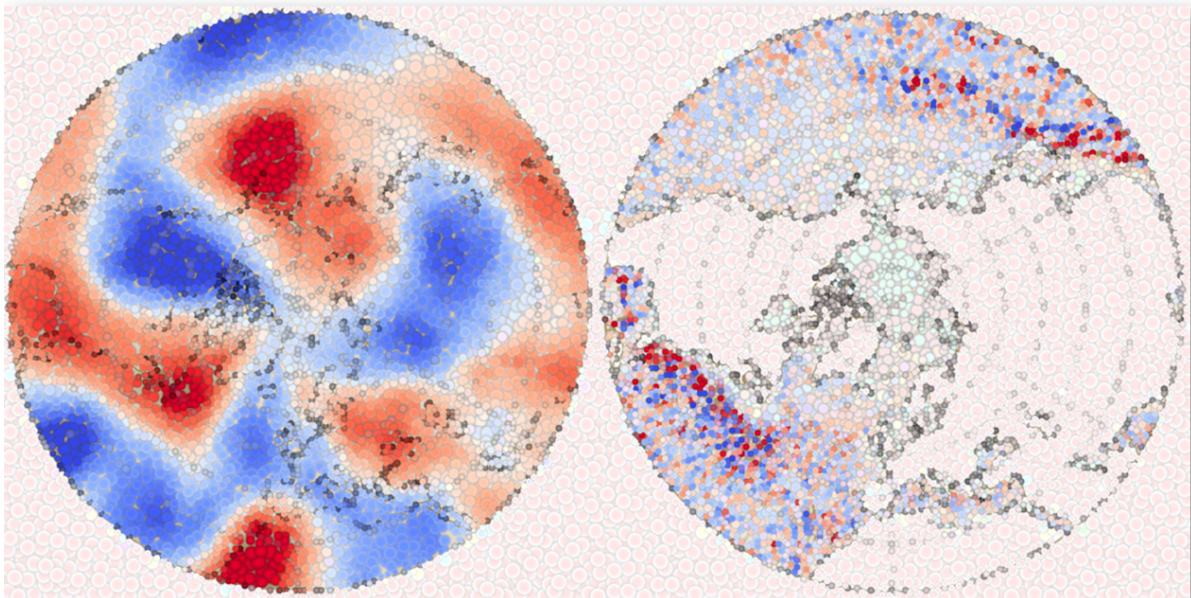
Euclidean distance (EuD) is used to evaluate the similarity between trajectories, which is expressed as:

$$EuD = \frac{1}{N} \sum_{N=1}^N \frac{1}{T} \sum_{T=1}^T \sqrt{(x_{pred} - x_{obs})^2 + (y_{pred} - y_{obs})^2 + (z_{pred} - z_{obs})^2} \quad (5.14)$$



## Chapter 6

# Synthesis and Outlook



Given the essential role of the meridional energy transport in regulating the variability of the Arctic climate, I focus on the quantification of AMET and OMET, and the atmosphere-ocean interactions in connection to the energy transport. The strong link between the energy transport and sea ice variations was underlined in this thesis and this further motivates our explorations of improving the Arctic sea ice forecasts with deep learning techniques. In this synthesis chapter, I will address the questions raised in the first chapter. The results shown in the previous chapters will be summarized and discussed in the following sections and the thesis will be closed with an outlook.

## 6.1 Physical insights into Arctic climate

Meridional energy transport closely relates to high-latitude climate characteristics and therefore is recognized as a key mechanism to the climate variability in the Arctic. We provide physical insights into the Arctic climate by inspecting the energy transport towards the Arctic in historical records and addressing the following question: *Can we obtain a reliable and consistent estimation of meridional energy transport in the atmosphere and ocean using multiple modern reanalysis data sets, and are the results coherent with the Arctic climate variability at different time scales?* This challenging research question was addressed by many previous studies, but most of them only use coupled models or a single reanalysis dataset. This is not sufficient to account for the uncertainties associated with data sources. In chapter 2, I elaborated upon the quantification and evaluation of AMET and OMET from three atmospheric and three oceanic reanalysis products. AMET and OMET are calculated as the vertical integral of the MET divergence over the air and water columns (the so-called direct method), respectively.

We found that all the selected data sets agree on the mean AMET from midlatitudes towards the Arctic. For the atmosphere, the peak of AMET occurs around  $41^{\circ}\text{N}$  at 4.45 PW to 4.5 PW in the analyzed products. Considering the variations of AMET from midlatitudes towards the Arctic, I inspected the zonally integrated AMET over  $60^{\circ}\text{N}$  and found that all the chosen atmospheric reanalysis data sets provide similar results, in terms of the seasonal cycling and the phases of the time series. The amplitudes are slightly different as a function of time, but on average the difference is smaller than 0.1 PW, excluding some outliers. Our results are similar to those from the studies in the past two decades [e.g., Trenberth and Fasullo, 2008, Mayer and Haimberger, 2012, and many others].

However, the low frequency variations of the anomalies of AMET are different within the chosen reanalysis products. Concerning the AMET, differences between ERA-Interim and JRA55 are relatively small, which is no more than 0.05 PW. And their signals are strongly correlated with each other, with a correlation coefficient 0.82. While MERRA2 provides very different results and the difference can exceed 0.1 PW for most of the time in the considered period. We investigated the vertical profile of AMET from ERA-interim, JRA55 and MERRA2 and identified that differences from the temperature and meridional wind fields. This is consistent with some early assessment reports of reanalysis products [e.g., Simmons et al., 2017, Uotila et al., 2018]. Our study can only provide insight qualitatively as the comparison was carried out on pressure levels and mass conservation was not ensured. But this analysis indicates that

uncertainties in both the temperature and meridional velocity fields are too large to constrain the AMET. Given the difference in the numerical schemes among the chosen reanalysis products, like the horizontal advection schemes, this analysis is not sufficient to judge their similarity to actual energy transport due to a lack of independent observations in the atmosphere.

For the energy transport in the ocean, the analyzed oceanic reanalysis products agree well on the mean OMET from midlatitudes to the Arctic. Between 30°N and 40°N, the latitude of the Gulf Stream, their differences are large (up to 0.5 PW), which might be associated with the differences in eddy representation approaches. An emphasis was placed on the poleward energy transport at higher latitudes. Taking the zonal integral of OMET at 60°N as an example, the mean OMET given by these three products varies from 0.44 PW to 0.47 PW and the variations of OMET are very similar. However, large differences were observed in the low frequency signals. We found that the low frequency signals of OMET anomalies in ORAS4 and SODA3 are consistent with each other and the difference is no larger than 0.02 PW. While GLORYS2V3 differs much from them and the differences is more than 0.05 PW between 2000 and 2005. Note that starting from 2007, all the oceanic reanalysis products agree well on OMET considering the mean and anomalies, which might be associated with the increase of observations due to the increasing number of Argo floats in use [Riser et al., 2016]. Only the OMET from ORAS4 resembles the observations in the Atlantic (RAPID array and OSNAP), albeit ORAS4 benefits from assimilating the observations from RAPID array [Balmaseda et al., 2013].

The study was further extended to analyze the relations between the climatological fields and MET in the Arctic. Considering the surface fields, we observed physically consistent patterns in ERA-Interim and JRA55 in winter. For the ocean, similar patterns were found in ORAS4 and SODA3 in winter. In summer, the relation between MET and the surface fields become unclear, which reflects that the large-scale dynamics processes are more pronounced and dominant in winter [Curry et al., 1995, Goosse et al., 2018]. These strong links all indicate that the Arctic climate variability is very sensitive to changes of AMET and OMET at subpolar latitudes, which is consistent with other studies as well [Van der Swaluw et al., 2007, Jungclauss and Koenigk, 2010, van der Linden et al., 2016].

Our findings indicate that the interannual variability of AMET and OMET calculated via the direct method cannot be constrained by the available observations. The presence of sources and sinks in the reanalysis products introduce large uncertainties in the quantification of MET, which confirms the caveats raised by Trenberth [1991], Trenberth and Solomon [1994]. Nevertheless, given the good agreement on mean AMET and OMET and their annual cycles, these reanalysis products are still useful and reliable tools for estimating and assessing the meridional energy transport in the atmosphere and ocean. Most of them can provide trustworthy energy transport diagnostics coherent with the Arctic climate variability. Note that when adopting these reanalysis products for the examination of energy transport at relatively long time scales (e.g. interannual to decadal time scales), the results should be interpreted with caution and the robustness of these results must be further evaluated.

Another physical mechanism that influences the Arctic climate variability is the variability in atmosphere-ocean interactions. This involves energy compensation. The compensation between AMET and OMET variations redistributes the energy between the atmosphere and ocean, and

it contributes to the energy balance in the climate system. By reshaping the energy budget, it leaves its footprint in the general circulations in the atmosphere and ocean and influences the large-scale dynamics of the whole system. It is investigated in chapter 3 and here I will summarize and discuss my findings by addressing the question ***Does the compensation between energy transport variations in the atmosphere and ocean in the Northern Hemisphere occur in the historical records and is there a robust mechanism found in multiple modern reanalysis products?*** On the basis of our outcome from chapter 2, we approached this question by investigating the AMET and OMET using these six state-of-the-art reanalysis products again and explored the Bjerknes compensation within the chosen reanalysis data sets. We demonstrated that the Bjerknes compensation occurs in the Northern Hemisphere at almost all latitudes from 40°N to 70°N from interannual to decadal time scales within the selected reanalysis products. The results are consistent with some previous work [e.g., Shaffrey and Sutton, 2006, and many others].

In order to understand the physical mechanism of the compensation, we revisited the roles of atmosphere and ocean in terms of their forcing and response, including the driven force for the compensation, the relevant variations in the general circulation of the atmosphere and ocean, and the corresponding response characterized by large-scale patterns like AO/NAO. With ERA-Interim, MERRA2, JRA55 and ORAS4, we demonstrated that the Bjerknes compensation occurs in the Northern Hemisphere at almost all the latitudes from 40°N to 70°N and the intensity varies as a function of the length of time scales. By investigating the flow of the anomalies of net surface flux, we observed that the atmosphere provides energy to the ocean in subtropical regions in the Atlantic Ocean and in the Pacific Ocean close to the Asian continent. In the subpolar Atlantic, the ocean releases energy back to the atmosphere. The compensation is shown clearly in the surface flux as a result of small net radiation flux variations at TOA. Our results are similar to the findings from [Jungclaus and Koenigk, 2010].

We further noticed that the atmosphere responds to OMET variations with a shift in the Ferrel cell at decadal time scales in winter. Horizontally, this shift shows patterns comparable to AO/NAO. This has also been reported by Jungclaus and Koenigk [2010] in their numerical model simulation. The associated variations in AMET are mainly observed in zonal mean part of heat transport and less in the eddy heat transport part. This is different from previous studies with numerical models [Shaffrey and Sutton, 2006, Van der Waluw et al., 2007], but they concerned larger time scales. This proposed mechanism is robust across the different atmospheric reanalysis data sets used in this study.

Moreover, we found that the oceanic response to the changes in the atmosphere is primarily wind-driven at decadal time scales. The ocean adjusts to surface wind anomalies and this modifies the OMET through an intensification of gyre transports and anomalous Ekman transport. The buoyancy fluxes, which could influence the thermohaline circulation, also play a role in this process. The buoyancy flux consists of a thermal flux and a haline flux. By inspecting each component respectively, we identify that the buoyancy forcing is dominated by its thermal component. Through linear regressions, we showed that the anomalous buoyancy flux variations are closely linked to the variability of OMET, especially in the subpolar Atlantic and in the Labrador Sea and the Irminger Sea.



When interpreting the results given by chapter 3, we must note that the mass budget imbalance suffered by the chosen reanalysis products might influence the conclusions. We also noticed that different methods for the quantification of AMET can potentially lead to different results [Armour et al., 2019]. Our results are based on AMET calculated using indirect method, which takes AMET as the residual between net surface flux and net radiation flux at TOA [e.g., Mayer and Haimberger, 2012]. With AMET computed as the vertical integral of AMET divergence over the air column [Liu et al., 2020b], a consistent picture of Bjerknes compensation was not observed and the energy flow between atmosphere and ocean even contradicts the surface and TOA energy budget. In addition, the time series in the historical records are too short to unequivocally determine forcing–response relations. So, with our analysis we cannot prove whether the ocean drives the atmospheric MET variations or the reverse. This question could be addressed with longer historical records in the future, which is beyond the scope of this thesis.

Chapter 3 provides new insight into the Bjerknes compensation in the Northern Hemisphere by identifying compensations between energy transport variations in the atmosphere and ocean in multiple modern reanalysis products. A robust physical mechanism was found among all the selected atmospheric reanalysis data sets. In contrast, previous studies on Bjerknes compensation are mostly based in numerical climate models only. Different from the previous work, we demonstrated that the response of the Ferrel cell contributes to the Bjerknes compensation in reanalysis data sets at middle to high latitudes in the atmosphere. Other processes that are related to this change, for instance, the coupling between the Ferrel cell and the Hadley cell, and teleconnections between low-latitude drivers and middle- to high-latitude weather and climate, are also interesting to investigate if much longer time series can be available in the future. This poses scientific challenges and is worth a visit for further studies.

## 6.2 Improving weather forecast in the Arctic

The strong demand for reliable weather forecasts in the Arctic poses a need for improved weather forecast skill in the polar region. Motivated by the potential of deep neural networks for addressing complicated nonlinear relationships between multiple variables during the evolution of a complex dynamical system, in chapter 4 and 5 we explored the deep learning techniques for weather forecasting. To begin with, we employed a deterministic neural network to forecast sea ice in the Barents Sea at weather to sub-seasonal time scales. The corresponding research question reads: *Can we produce skillful Arctic sea ice forecasts at weekly time scales using deep neural networks, and is the physical consistency preserved in these forecasts?* We address this question by predicting sea ice with a special neural network, namely the Convolutional Long Short Term Memory Network. The architecture of this neural network is designed to incorporate both the spatial and temporal information in meteorological fields. Sea ice forecasts were evaluated against climatology, persistence, a baseline statistical model and operational sub-seasonal to seasonal forecast systems. In chapter 4, we demonstrated that this deep neural network can effectively be used to predict sea ice characteristics in the Barents Sea at weekly to sub-monthly time scales. Two types of lead time dependent forecast were conducted in this work, which are called the constrained and retrospective forecasts. The former

set-up is used to gain insight into the sea ice forecast skill of the applied deep neural networks with perfect predictors. In other words, it shows the forecast “ceiling” of the chosen approach. The latter follows a configuration generally taken by the operational weather forecasting. The results from these two set-ups differ slightly, but in most cases, ConvLSTMs can outperform persistence and climatology for the extended range sea ice forecast at a lead time up to 5 weeks, with proper choices of predictors. Forecasts from the ConvLSTM are comparable to those provided by the operational sub-seasonal to seasonal forecast systems, namely the NCEP global ensemble forecast system and the ECMWF ensemble forecast system.

We noticed that the forecast quality has strong dependency on the input meteorological fields and for the selection of predictors we must consider the physical consistency between sea ice variation and the variability of predictors. This further motivates our interest in the contribution from different predictors to the sea ice forecasting. Note that in general such inspection can hardly be performed in an experiment with numerical climate models. On the basis of known sources of predictability, we performed sensitivity tests with different climate fields as input for learning and found that energy budget related fields can add skill to the sea ice forecasts in the Barents Sea at weekly to sub-monthly time scales. This indicates the important role of the energy budget components in the variability of Arctic sea ice, which is consistent with our findings in chapter 2 and 3, and the findings from many other related work [e.g., Krikken and Hazeleger, 2015].

To conclude, by choosing suitable deep neural networks and the proper combination of predictors, we can produce skillful extended-range Arctic sea ice forecasts comparable to the operational sub-seasonal to seasonal forecast systems. The physical consistency can also be preserved in the forecasts. It should be noted that other research also points to the potential value of deep neural networks for weather forecasting, and in general the climate science [e.g., Salman et al., 2015, Gope et al., 2016, Kim et al., 2017, Ham et al., 2019, Kim et al., 2019, 2020, and many others]. However, most of these studies only focus on deep learning in the deterministic domain. And, these techniques can not address the probabilistic nature of weather forecasts due to the chaotic nature of the atmospheric dynamics and uncertainties in both initial conditions and models representing the atmosphere. Driven by the need uncertainty quantification, we extended our study to the probabilistic domain and introduced the Bayesian deep learning to weather forecast.

Here we will summarize and discuss our findings of chapter 5 and answer the research question *Can probabilistic deep neural networks learn to represent the evolution of weather and climate system with an appropriate estimation of uncertainty, and can we use them to generate reliable weather forecast?* In this chapter, we built Bayesian Long-Short Term Memory neural networks by replacing fixed weights of normal LSTMs with distributions, and used them to forecast the evolution of a modified Lorenz 84 system (with a damping time equal to 5 days). we tried to understand the characteristics of BDL within this simplified dynamical system that represents the essence of midlatitude atmospheric dynamics, and explore the types of uncertainties addressed by BDL. We found that, despite the forecasts get worse with increasing lead times, the retrospective forecasts from BayesLSTM are similar to those of the Lorenz model output in the spatial and temporal domain up to a lead time of 3 days. The

forecast trajectories are close to the attractor. Considering both the amplitude and phase, this indicates that BayesLSTM is able to learn the propagation of Rossby waves on a midlatitude jet in this atmospheric system, which reflects that the interaction between the jet stream and large-scale eddies is replicated by the BayesLSTM.

We further assessed the probabilistic forecast skill of BayesLSTM against persistence of initial conditions and a vector autoregressive model. We found that the BayesLSTM forecasts resemble the model output. With evaluation metrics including CRPS, RMSE, and Euclidean distance scores, we observed that the BayesLSTM showed better probabilistic forecast skill than persistence and the VAR model up to a lead time of 3 days. We further explored the probabilistic forecast skill of the BDL through the sea ice forecasting case study in the Barents Sea, just like the experiments shown in chapter 4. Our preliminary results suggest that the BDL shows skill in operational sea ice forecasts and it can generate uncertainties relevant to the natural variability of sea ice. It can be useful tool for operational weather forecasting, though its potential was not fully explored here due to the limited time and a different scope of this study.

It is fundamental for any probabilistic weather forecast to address uncertainties in the initial conditions and model parameters. We found that, compared to the NWP-based approaches, the BDL addresses these uncertainties in a different manner. For NWP systems, the aleatoric uncertainties are reflected by the perturbations in the initial conditions and the chaotic nature of the system, while BDL addresses it by placing a distribution over the weight. Moreover, NWP based approaches represent epistemic uncertainties by perturbing model parameters, parameterization, or including multiple models, while BDL deals with it by placing a distribution over the weights. Such differences, as well as the data driven aspects, make the BDL a method compared to the classical approaches in climate science.

Based on these results, we concluded that the BayesLSTM is useful for estimating uncertainties relevant to the evolution of this simplified atmospheric circulation system, though the procedure differs from the general NWP based approaches. This implies that the BDL can be added to the toolkit of weather forecasting at relatively small lead times. However, for a long term climate forecast or climate projections, the BayesLSTM may not be a good candidate in terms of the error accumulation and the lack of skill in physical model representation. Our attempt only serves as a pilot study for adapting the BDL to climate science. There is still a lot of room to improve the performance of BDL, which can be explored in the future.

## 6.3 Outlook

The threats from adverse effects of climate change and global warming raise our awareness about the rapid changes occurring in the Arctic. The growing interests in the Arctic lead to a growing number of publications related to the Arctic climate change, which reveals our eagerness of the knowledge about the Arctic. This thesis concerns the physical aspects of the Arctic and contributes to the research on Arctic sea ice forecasting with data-driven techniques. It not only adds new content to our knowledge base, but also points the way for future work, to some extent.

In chapter 2 and 3, we studied the meridional energy transport using reanalysis data sets and

analyzed the variability of the energy transport and compensation in the Northern Hemisphere at interannual to decadal time scales. However, these studies are limited by the shortness of historical records and the conclusions are therefore prone to the uncertainties associated with the insufficiency of samples available. As a consequence, some results show relatively low statistical significance and this hampers the corresponding interpretation of our findings. Similar problems were encountered when testing the forecast quality for various start dates in chapter 4. Since the neural networks were trained with reanalysis data and the chosen neural networks require about 30 years of training data strictly following the temporal order to ensure the quality of learning, the test for multiple start days can not be performed. For meteorology, the data-rich decades originate from the start of satellite era in 1979. Hence we only have about 40 years historical records, but their lengths are still increasing over time. This explains why most of the previous work choose to study the interannual variability of energy transport and the Bjerknes compensation using the long time series generated by numerical climate models [e.g., Van der Swaluw et al., 2007, Jungclaus and Koenigk, 2010, and many others]. In the future, with relatively long time series, we will be able to revisit these topics with much convincing analysis and the outcome could be more trustworthy.

The sparseness of observations is a big obstacle for studying the meridional energy transport. In chapter 2, we noticed that the interannual variability of AMET and OMET cannot be constrained by the available observations, which explains the difference of our results among chosen analysis products. The strong correlation between the estimation accuracy of energy budget components, and the number and coverage of observations available has been revealed by many studies. For instance, the advent of the global Argo array of profiling floats (see Figure 2 in Riser et al. [2016]) and the decrease of uncertainties in ocean heat budget (see Figure 2 in Cheng et al. [2020]). In addition, with independent observations, we will be able to identify the sources of uncertainty related to the energy budget at TOA and sea surface, and provide more insights into the energy transport diagnostics.

Fortunately, there are many ongoing projects aiming at documenting the variability of changes in the atmosphere and ocean. For the atmosphere, there are many efforts to provide better monitoring of the atmosphere from both satellite and in-situ observations, for instance, the European Union's Earth observation programme, namely the Copernicus program, and the Clouds and the Earth's Radiant Energy System (CERES) project [Loeb et al., 2018]. Similarly, there are also many on-going observation programs for the ocean, like RAPID-MOCHA-WBTS program [Johns et al., 2011, McCarthy et al., 2015], Overturning in the Subpolar North Atlantic Program (OSNAP) [Susan Lozier et al., 2017, Lozier et al., 2019], Observatoire de la variabilité interannuelle et décennale en Atlantique Nord project (OVIDE) [Mercier et al., 2008], and Greenland Scotland Ridge observation project (GSR) [e.g., Berx et al., 2013, and many others in GSR project]. Given the fact that the reconstruction of historical records relies on both the model output and assimilated observations, we need more observation data to better constrain the models and improve the reanalysis products. In addition, considering the demand of independent observations to validate the reanalyses, at this moment the amount of observations available are far from sufficient. With improved numerical models and independent observations, the validity of these mechanisms in chapter 2 and 3 can be further studied.

In chapter 4 and 5, we explored two types of deep neural networks for weather forecasting. So far, our experiments with ConvLSTM and BayesLSTM are limited to simple network structures and small training data sets. It is possible to improve the performance of these neural networks by increasing the complexity of the network and including large data sets for training. When exploiting the BDL with Lorenz 84 model, our experiments contribute in a way as a proof of concept. These experiments can be further extended to address a real weather forecast case in the future and the body of the research can be expanded with more training data, as well as more output from numerical models for intercomparison and evaluation. This will also allow for different kinds of sensitivity tests, in terms of different predictors and various start days, and statistical tests for the validity of lessons learnt with the analyzed deep learning techniques.

Inspired by recent developments in artificial intelligence, many studies have introduced advanced machine learning techniques to tackle the challenges in meteorology and oceanography [Reichstein et al., 2019, Knüsel et al., 2019]. Until now, most of these studies have provided positive feedbacks about the implementation of machine learning approaches, especially for deep neural networks, to revisit common research topics in climate science [e.g., Vandal et al., 2018, Kim et al., 2020, and many others]. However, these studies still struggle with the interpretability of their results from a physical perspective and these deep learning techniques are often considered as brute force approaches or black-box methods. In chapter 4, we performed singular value decomposition on the covariance map of the forecasts from ConvLSTM, with a valid concern that physical laws may be violated and physical relationships may not hold in the neural networks. Although we demonstrated that the ConvLSTM is able to preserve the physical consistency between various predictors and predictands, our method is limited to linear relationships. With the fast development of machine learning techniques, the community is dedicated to disentangle the multi-layer representations of abstracted information from training data and provide more physical explanations [e.g., Chakraborty et al., 2017, Zhang and Zhu, 2018, and many others]. Such explorations are promising for the “decryption” of deep learning in climate science, and therefore are worth the effort in the future.

As a closing remark, I would like to emphasize the value of multifaceted analysis and multidisciplinary approaches to the climate science. This thesis contributes to improve our understanding about the Arctic, in terms of the physical processes and forecasting practices. Apart from the scientific content, it also acts as a good example of embracing new approaches when revisiting old issues. As a practitioner, our endeavor involves the employment of new combinations of data sets, the implementation of new approaches, the utilization of new tools, and finally the development of new ideas. Although there are always some defects related to the novelty, for instance, incomplete knowledge, lack of benchmarks, and inadequacy of consideration, unexpected breakthroughs can often be expected, as “the role of pain” from time to time.



# References

- Knut Aagaard and EC 1994 Carmack. The arctic ocean and climate: A perspective. *GMS*, 85: 5–20, 1994.
- David G Andrews, Conway B Leovy, and James R Holton. *Middle atmosphere dynamics*, volume 40. Academic press, 1987.
- Kyle C Armour, Nicholas Siler, Aaron Donohoe, and Gerard H Roe. Meridional atmospheric heat transport constrained by energetics and mediated by large-scale diffusion. *Journal of Climate*, 32(12):3655–3680, 2019.
- M Årthun, T Eldevik, LH Smedsrud, Ø Skagseth, and RB Ingvaldsen. Quantifying the influence of atlantic heat on barents sea ice variability and retreat. *Journal of Climate*, 25(13):4736–4743, 2012.
- Marius Årthun, Tor Eldevik, and Lars H Smedsrud. The role of atlantic heat transport in future arctic winter sea ice loss. *Journal of Climate*, 32(11):3327–3341, 2019.
- Jürgen Bader, Michel DS Mesquita, Kevin I Hodges, Noel Keenlyside, Svein Østerhus, and Martin Miles. A review on northern hemisphere sea-ice, storminess and the north atlantic oscillation: Observations and projected changes. *Atmospheric Research*, 101(4):809–834, 2011.
- Magdalena Alonso Balmaseda, Kristian Mogensen, and Anthony T Weaver. Evaluation of the ecmwf ocean reanalysis system oras4. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1132–1161, 2013.
- RN Bannister. A review of operational methods of variational and ensemble-variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143(703):607–633, 2017.
- Elizabeth A Barnes and James A Screen. The impact of arctic warming on the midlatitude jet-stream: Can it? has it? will it? *Wiley Interdisciplinary Reviews: Climate Change*, 6(3): 277–286, 2015.
- Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.
- Paul Berrisford, DPKF Dee, Keith Fielding, Manuel Fuentes, P Kallberg, Shinya Kobayashi, and Sakari Uppala. The era-interim archive. *ERA report series*, (1):1–16, 2009.
- Paul Berrisford, P Kållberg, S Kobayashi, D Dee, S Uppala, AJ Simmons, P Poli, and H Sato. Atmospheric conservation properties in era-interim. *Quarterly Journal of the Royal Meteorological Society*, 137(659):1381–1399, 2011.

- B Berx, B Hansen, S Østerhus, KM Larsen, T Sherwin, and K Jochumsen. Combining in situ measurements and altimetry to estimate volume, heat and salt transport variability through the faroe–shetland channel. *Ocean Science*, 9(4):639–654, 2013.
- Jf Bjerknes. Atlantic air-sea interaction. In *Advances in geophysics*, volume 10, pages 1–82. Elsevier, 1964.
- Eduardo Blanchard-Wrigglesworth and Mitch Bushuk. Robustness of arctic sea-ice predictability in gcms. *Climate Dynamics*, 52(9-10):5555–5566, 2019.
- Edward Blanchard-Wrigglesworth, Kyle C Armour, Cecilia M Bitz, and Eric DeWeaver. Persistence and inherent predictability of arctic sea ice in a gcm ensemble and observations. *Journal of Climate*, 24(1):231–250, 2011a.
- Edward Blanchard-Wrigglesworth, CM Bitz, and MM Holland. Influence of initial conditions and climate forcing on predicting arctic sea ice. *Geophysical Research Letters*, 38(18), 2011b.
- Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. *arXiv preprint arXiv:1505.05424*, 2015.
- David B Bonan, Mitchell Bushuk, and Michael Winton. A spring barrier for regional predictions of summer arctic sea ice. *Geophysical Research Letters*, 46(11):5937–5947, 2019.
- Christopher S Bretherton, Catherine Smith, and John M Wallace. An intercomparison of methods for finding coupled patterns in climate data. *Journal of climate*, 5(6):541–560, 1992.
- Laurent Brodeau, Bernard Barnier, Anne-Marie Treguier, Thierry Penduff, and Sergei Gulev. An era40-based atmospheric forcing for global ocean circulation models. *Ocean Modelling*, 31(3-4):88–104, 2010.
- Henk Broer, Carles Simó, and Renato Vitolo. Bifurcations and strange attractors in the lorenz-84 climate model with seasonal forcing. *Nonlinearity*, 15(4):1205, 2002.
- Kirk Bryan. Measurements of meridional heat transport by ocean currents. *Journal of Geophysical Research*, 67(9):3403–3414, 1962.
- Roberto Buizza, Martin Leutbecher, and Lars Isaksen. Potential use of an ensemble of analyses in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 134(637):2051–2066, 2008.
- Mitchell Bushuk, Rym Msadek, Michael Winton, Gabriel A Vecchi, Rich Gudgel, Anthony Rosati, and Xiaosong Yang. Skillful regional prediction of arctic sea ice on seasonal timescales. *Geophysical Research Letters*, 44(10):4953–4964, 2017.
- Levke Caesar, Stefan Rahmstorf, Alexander Robinson, G Feulner, and V Saba. Observed fingerprint of a weakening atlantic ocean overturning circulation. *Nature*, 556(7700):191, 2018.
- Eddy Carmack, I Polyakov, Laurie Padman, Ilker Fer, E Hunke, Jennifer Hutchings, Jennifer Jackson, D Kelley, Ronald Kwok, Chantelle Layton, et al. 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, 2015.
- James A Carton, Gennady A Chepurin, and Ligang Chen. Soda3: a new ocean climate reanal-



- ysis. *Journal of Climate*, (2018), 2018.
- Supriyo Chakraborty, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, Moustafa Alzantot, Federico Cerutti, Mani Srivastava, Alun Preece, Simon Julier, Raghuveer M Rao, et al. Interpretability of deep learning models: a survey of results. In *2017 IEEE smart-world, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, Internet of people and smart city innovation (smartworld/SCALCOM/UIC/ATC/CBDcom/IOP/SCI)*, pages 1–6. IEEE, 2017.
- Ashesh Chattopadhyay, Pedram Hassanzadeh, Krishna Palem, and Devika Subramanian. Data-driven prediction of a multi-scale lorenz 96 chaotic system using a hierarchy of deep learning methods: Reservoir computing, ann, and rnn-lstm. *arXiv preprint arXiv:1906.08829*, 2019.
- Lijing Cheng, John Abraham, Jiang Zhu, Kevin E Trenberth, John Fasullo, Tim Boyer, Ricardo Locarnini, Bin Zhang, Fujiang Yu, Liying Wan, et al. Record-setting ocean warmth continued in 2019, 2020.
- Gabriel Chiodo and Leopold Haimberger. Interannual changes in mass consistent energy budgets from era-interim and satellite data. *Journal of Geophysical Research: Atmospheres*, 115(D2), 2010.
- Judah Cohen, James A Screen, Jason C Furtado, Mathew Barlow, David Whittleston, Dim Coumou, Jennifer Francis, Klaus Dethloff, Dara Entekhabi, James Overland, et al. Recent arctic amplification and extreme mid-latitude weather. *Nature geoscience*, 7(9):627, 2014.
- Josefino C Comiso and Dorothy K Hall. Climate trends in the arctic as observed from space. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3):389–409, 2014.
- Rubén Cruz-García, Virginie Guemas, Matthieu Chevallier, and François Massonnet. An assessment of regional sea ice predictability in the arctic ocean. *Climate Dynamics*, 53(1-2): 427–440, 2019.
- Judith A Curry, Julie L Schramm, and Elizabeth E Ebert. Sea ice-albedo climate feedback mechanism. *Journal of Climate*, 8(2):240–247, 1995.
- Arnaud Czaja and Nicholas Blunt. A new mechanism for ocean–atmosphere coupling in mid-latitudes. *Quarterly Journal of the Royal Meteorological Society*, 137(657):1095–1101, 2011.
- Arnaud Czaja and Claude Frankignoul. Observed impact of atlantic sst anomalies on the north atlantic oscillation. *Journal of Climate*, 15(6):606–623, 2002.
- Mladen Dalto, Jadranko Matusko, and Mario Vašak. Deep neural networks for ultra-short-term wind forecasting. In *2015 IEEE International Conference on Industrial Technology (ICIT)*, pages 1657–1663. IEEE, 2015.
- Fabio D’Andrea, Arnaud Czaja, and John Marshall. Impact of anomalous ocean heat transport on the north atlantic oscillation. *Journal of climate*, 18(23):4955–4969, 2005.
- JJ Day, Ed Hawkins, and Steffen Tietsche. Will arctic sea ice thickness initialization improve seasonal forecast skill? *Geophysical Research Letters*, 41(21):7566–7575, 2014a.
- JJ Day, S Tietsche, and E Hawkins. Pan-arctic and regional sea ice predictability: Initialization month dependence. *Journal of Climate*, 27(12):4371–4390, 2014b.

- Dick P Dee, S M Uppala, AJ Simmons, Paul Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, d P Bauer, et al. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656):553–597, 2011.
- Thomas L Delworth and Fanrong Zeng. The impact of the north atlantic oscillation on climate through its influence on the atlantic meridional overturning circulation. *Journal of Climate*, 29(3):941–962, 2016.
- Thomas L Delworth, Anthony Rosati, Whit Anderson, Alistair J Adcroft, Venkatramani Balaji, Rusty Benson, Keith Dixon, Stephen M Griffies, Hyun-Chul Lee, Ronald C Pacanowski, et al. Simulated climate and climate change in the gfdl cm2. 5 high-resolution coupled climate model. *Journal of Climate*, 25(8):2755–2781, 2012.
- Thomas L Delworth, Fanrong Zeng, Liping Zhang, Rong Zhang, Gabriel A Vecchi, and Xiaosong Yang. The central role of ocean dynamics in connecting the north atlantic oscillation to the extratropical component of the atlantic multidecadal oscillation. *Journal of Climate*, 30(10):3789–3805, 2017.
- Clara Deser, Robert A Tomas, and Lantao Sun. The role of ocean–atmosphere coupling in the zonal-mean atmospheric response to arctic sea ice loss. *Journal of Climate*, 28(6):2168–2186, 2015.
- Ambroise Dufour, Olga Zolina, and Sergey K Gulev. Atmospheric moisture transport to the arctic: Assessment of reanalyses and analysis of transport components. *Journal of Climate*, 29(14):5061–5081, 2016.
- R Dussin, B Barnier, L Brodeau, and JM Molines. The making of the drakkar forcing set dfs5. *DRAKKAR/MyOcean Rep*, pages 01–04, 2016.
- ECMWF. Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate. *Copernicus Climate Change Service Climate Data Store (CDS)*, 2017.
- Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.
- Riccardo Farneti and Geoffrey K Vallis. Meridional energy transport in the coupled atmosphere–ocean system: Compensation and partitioning. *Journal of Climate*, 26(18):7151–7166, 2013.
- John T Fasullo and Kevin E Trenberth. The annual cycle of the energy budget. part ii: Meridional structures and poleward transports. *Journal of Climate*, 21(10):2313–2325, 2008.
- Nicole Feldl, Stephen Po-Chedley, Hansi KA Singh, Stephanie Hay, and Paul J Kushner. Sea ice and atmospheric circulation shape the high-latitude lapse rate feedback. *npj Climate and Atmospheric Science*, 3(1):1–9, 2020.
- Qing Yi Feng, Ruggero Vasile, Marc Segond, Avi Gozolchiani, Yang Wang, Markus Abel, Shilomo Havlin, Armin Bunde, and Henk A Dijkstra. Climatelearn: A machine-learning approach for climate prediction using network measures. *Geoscientific Model Development*, 2016.

- N Ferry, B Barnier, G Garric, K Haines, S Masina, L Parent, A Storto, M Valdivieso, S Guinehut, and S Mulet. Nemo: the modeling engine of global ocean reanalyses. *Mercator Ocean Quarterly Newsletter*, 46:46–59, 2012a.
- N Ferry, L Parent, G Garric, C Bricaud, CE Testut, O Le Galloudec, JM Lellouche, M Drevillon, E Greiner, B Barnier, et al. Glorys2v1 global ocean reanalysis of the altimetric era (1992–2009) at meso scale. *Mercator Ocean-Quarterly Newsletter*, 44, 2012b.
- Nicolas Ferry, Laurent Parent, Gilles Garric, Bernard Barnier, Nicolas C Jourdain, et al. Mercator global eddy permitting ocean reanalysis glorys1v1: Description and results. *Mercator-Ocean Quarterly Newsletter*, 36:15–27, 2010.
- Meire Fortunato, Charles Blundell, and Oriol Vinyals. Bayesian recurrent neural networks. *arXiv preprint arXiv:1704.02798*, 2017.
- Jennifer A Francis, Stephen J Vavrus, and Judah Cohen. Amplified arctic warming and mid-latitude weather: new perspectives on emerging connections. *Wiley Interdisciplinary Reviews: Climate Change*, 8(5):e474, 2017.
- Claude Frankignoul, Nadine Chouaib, and Zhengyu Liu. Estimating the observed atmospheric response to sst anomalies: maximum covariance analysis, generalized equilibrium feedback assessment, and maximum response estimation. *Journal of climate*, 24(10):2523–2539, 2011.
- Joana G Freire, Cristian Bonatto, Carlos C DaCamara, and Jason AC Gallas. Multistability, phase diagrams, and intransitivity in the lorenz-84 low-order atmospheric circulation model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(3):033121, 2008.
- Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.
- Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2015.
- Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.
- Alexandre Ganachaud and Carl Wunsch. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature*, 408(6811):453, 2000.
- Alexandre Ganachaud and Carl Wunsch. Large-scale ocean heat and freshwater transports during the world ocean circulation experiment. *Journal of Climate*, 16(4):696–705, 2003.
- Yongqi Gao, Richard Davy, Lingling Suo, Guillaume Gastineau, Young-Oh Kwon, Vladimir Semenov, and Yang Liu. Identification of the surface state influence in representing the Arctic warming by coordinated atmosphere-only simulations (D3.1), September 2019. URL <https://doi.org/10.5281/zenodo.3559462>.
- Jean-Claude Gascard, Kathrin Riemann-Campe, Rüdiger Gerdes, Harald Schyberg, Roger Rاندريامپيانينا, Michael Karcher, Jinlun Zhang, and Mehrad Rafizadeh. Future sea ice conditions and weather forecasts in the arctic: Implications for arctic shipping. *Ambio*, 46(3):

- 355–367, 2017.
- Guillaume Gastineau and Claude Frankignoul. Influence of the north atlantic sst variability on the atmospheric circulation during the twentieth century. *Journal of Climate*, 28(4):1396–1416, 2015.
- Guillaume Gastineau, Fabio D’Andrea, and Claude Frankignoul. Atmospheric response to the north atlantic ocean variability on seasonal to decadal time scales. *Climate dynamics*, 40(9-10):2311–2330, 2013.
- Ronald Gelaro, Will McCarty, Max J Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A Randles, Anton Darmenov, Michael G Bosilovich, Rolf Reichle, et al. The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of Climate*, 30(14):5419–5454, 2017.
- Peter R Gent and James C McWilliams. Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, 20(1):150–155, 1990.
- Michael Ghil and Paola Malanotte-Rizzoli. Data assimilation in meteorology and oceanography. In *Advances in geophysics*, volume 33, pages 141–266. Elsevier, 1991.
- Luis Gimeno-Sotelo, Raquel Nieto, Marta Vázquez, and Luis Gimeno. The role of moisture transport for precipitation in the inter-annual and inter-daily fluctuations of the arctic sea ice extension. *Earth System Dynamics*, 10(1):121–133, 2019.
- Tilmann Gneiting and Adrian E Raftery. Weather forecasting with ensemble methods. *Science*, 310(5746):248–249, 2005.
- Tilmann Gneiting, Adrian E Raftery, Anton H Westveld III, and Tom Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum crps estimation. *Monthly Weather Review*, 133(5):1098–1118, 2005.
- Tilmann Gneiting, Kristin Larson, Kenneth Westrick, Marc G Genton, and Eric Aldrich. Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association*, 101(475):968–979, 2006.
- Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.
- Helge F Goessling, Steffen Tietsche, Jonathan J Day, Ed Hawkins, and Thomas Jung. Predictability of the arctic sea ice edge. *Geophysical Research Letters*, 43(4):1642–1650, 2016.
- Hainan Gong, Lin Wang, Wen Chen, and Debashis Nath. Multidecadal fluctuation of the wintertime arctic oscillation pattern and its implication. *Journal of Climate*, 31(14):5595–5608, 2018.
- Hugues Goosse, Jennifer E Kay, Kyle C Armour, Alejandro Bodas-Salcedo, Helene Chepfer, David Docquier, Alexandra Jonko, Paul J Kushner, Olivier Lecomte, François Massonnet, et al. Quantifying climate feedbacks in polar regions. *Nature communications*, 9(1):1919, 2018.
- Sulagna Gope, Sudeshna Sarkar, Pabitra Mitra, and Subimal Ghosh. Early prediction of extreme

- rainfall events: a deep learning approach. In *Industrial Conference on Data Mining*, pages 154–167. Springer, 2016.
- Rune G Graversen and Mattias Burtu. Arctic amplification enhanced by latent energy transport of atmospheric planetary waves. *Quarterly Journal of the Royal Meteorological Society*, 142(698):2046–2054, 2016.
- Rune G Graversen, Thorsten Mauritsen, Michael Tjernström, Erland Källén, and Gunilla Svensson. Vertical structure of recent arctic warming. *Nature*, 451(7174):53, 2008.
- Rune Grand Graversen. Do changes in the midlatitude circulation have any impact on the arctic surface air temperature trend? *Journal of climate*, 19(20):5422–5438, 2006.
- Rune Grand Graversen, Erland Källén, Michael Tjernström, and Heiner Körnich. Atmospheric mass-transport inconsistencies in the era-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 133(624):673–680, 2007.
- Alex Graves. Practical variational inference for neural networks. In *Advances in neural information processing systems*, pages 2348–2356, 2011.
- Jonathan M Gregory and Rémi Tailleux. Kinetic energy analysis of the response of the atlantic meridional overturning circulation to co 2-forced climate change. *Climate dynamics*, 37(5-6):893–914, 2011.
- Virginie Guemas, Francisco J Doblas-Reyes, Kristian Mogensen, Sarah Keeley, and Yongming Tang. Ensemble of sea ice initial conditions for interannual climate predictions. *Climate dynamics*, 43(9-10):2813–2829, 2014.
- Virginie Guemas, Edward Blanchard-Wrigglesworth, Matthieu Chevallier, Jonathan J Day, Michel Déqué, Francisco J Doblas-Reyes, Neven S Fučkar, Agathe Germe, Ed Hawkins, Sarah Keeley, et al. A review on arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 142(695):546–561, 2016.
- Mindy M Hall and Harry L Bryden. Direct estimates and mechanisms of ocean heat transport. *Deep Sea Research Part A. Oceanographic Research Papers*, 29(3):339–359, 1982.
- Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572, 2019.
- Yayoi Harada, Hirotaka Kamahori, Chiaki Kobayashi, Hirokazu Endo, Shinya Kobayashi, Yukinari Ota, Hirokatsu Onoda, Kazutoshi Onogi, Kengo Miyaoka, and Kiyotoshi Takahashi. The jra-55 reanalysis: Representation of atmospheric circulation and climate variability. *Journal of the Meteorological Society of Japan. Ser. II*, 94(3):269–302, 2016.
- Pedram Hassanzadeh, Ashesh Chattopadhyay, Krishna Palem, and Devika Subramanian. Data-driven prediction of a multi-scale lorenz 96 chaotic system using a hierarchy of deep learning methods: Reservoir computing, ann, and rnn-lstm. *Bulletin of the American Physical Society*, 2019.
- Chengfei He, Zhengyu Liu, and Aixue Hu. The transient response of atmospheric and oceanic heat transports to anthropogenic warming. *Nature Climate Change*, 9(3):222–226, 2019.
- David A Hebert, Richard A Allard, E Joseph Metzger, Pamela G Posey, Ruth H Preller, Alan J

- Wallcraft, Michael W Phelps, and Ole Martin Smedstad. Short-term sea ice forecasting: An assessment of ice concentration and ice drift forecasts using the us navy's arctic cap nowcast/forecast system. *Journal of Geophysical Research: Oceans*, 120(12):8327–8345, 2015.
- Isaac M Held, Mingfang Ting, and Hailan Wang. Northern winter stationary waves: Theory and modeling. *Journal of climate*, 15(16):2125–2144, 2002.
- Michael Hilmer and Thomas Jung. Evidence for a recent change in the link between the north atlantic oscillation and arctic sea ice export. *Geophysical Research Letters*, 27(7):989–992, 2000.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- Cathy Hohenegger and Christoph Schar. Atmospheric predictability at synoptic versus cloud-resolving scales. *Bulletin of the American Meteorological Society*, 88(11):1783–1794, 2007.
- James R Holton. An introduction to dynamic meteorology. *American Journal of Physics*, 41(5):752–754, 1973.
- Peter L Houtekamer and Herschel L Mitchell. Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 126(3):796–811, 1998.
- PL Houtekamer and Fuqing Zhang. Review of the ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12):4489–4532, 2016.
- Carl Howell, Martin Richard, Joshua Barnes, and Tony King. Short-term operational sea ice forecasting for arctic shipping. In *ASME 2015 34th International Conference on Ocean, Off-shore and Arctic Engineering*. American Society of Mechanical Engineers Digital Collection, 2015.
- Yiyi Huang, Xiquan Dong, Baike Xi, Erica K Dolinar, Ryan E Stanfield, and Shaoyue Qiu. Quantifying the uncertainties of reanalyzed arctic cloud and radiation properties using satellite surface observations. *Journal of Climate*, 30(19):8007–8029, 2017.
- Yen-Ting Hwang, Dargan MW Frierson, and Jennifer E Kay. Coupling between arctic feedbacks and changes in poleward energy transport. *Geophysical Research Letters*, 38(17), 2011.
- LC Jackson. Shutdown and recovery of the amoc in a coupled global climate model: the role of the advective feedback. *Geophysical Research Letters*, 40(6):1182–1188, 2013.
- Ralf Jaiser, Klaus Dethloff, and Dörthe Handorf. Stratospheric response to arctic sea ice retreat and associated planetary wave propagation changes. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19375, 2013.
- Eystein Jansen, Jens Hesselbjerg Christensen, Trond Dokken, Kerim H Nisancioglu, Bo M Vinther, Emilie Capron, Chuncheng Guo, Mari F Jensen, Peter L Langen, Rasmus A Pedersen, et al. Past perspectives on the present era of abrupt arctic climate change. *Nature Climate Change*, 10(8):714–721, 2020.
- William E Johns, Molly O Baringer, Lisa M Beal, SA Cunningham, Torsten Kanzow, Harry L Bryden, JJM Hirschi, J Marotzke, CS Meinen, B Shaw, et al. Continuous, array-based estimates of atlantic ocean heat transport at 26.5 n. *Journal of Climate*, 24(10):2429–2449,

- 2011.
- Didier Jourdan, Efstathios Balopoulos, M-J Garcia-Fernandez, and Catherine Maillard. Objective analysis of temperature and salinity historical data set over the mediterranean basin. In *OCEANS'98 Conference Proceedings*, volume 1, pages 82–87. IEEE, 1998.
- Johann H Jungclaus and Torben Koenigk. Low-frequency variability of the arctic climate: the role of oceanic and atmospheric heat transport variations. *Climate dynamics*, 34(2-3):265–279, 2010.
- Christopher Kadow, David Matthew Hall, and Uwe Ulbrich. Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, pages 1–6, 2020.
- Marie-Luise Kapsch, Rune Grand Graversen, and Michael Tjernström. Springtime atmospheric energy transport and the control of arctic summer sea-ice extent. *Nature Climate Change*, 3(8):744, 2013.
- AR Karspeck, Detlef Stammer, Armin Köhl, G Danabasoglu, M Balmaseda, DM Smith, Y Fujii, S Zhang, B Giese, H Tsujino, et al. Comparison of the atlantic meridional overturning circulation between 1960 and 2007 in six ocean reanalysis products. *Climate Dynamics*, 49(3):957–982, 2017.
- Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584, 2017.
- Seongchan Kim, Seungkyun Hong, Minsu Joh, and Sa-kwang Song. Deeprain: Convlstm network for precipitation prediction using multichannel radar data. *arXiv preprint arXiv:1711.02316*, 2017.
- Sookyung Kim, Hyojin Kim, Joonseok Lee, Sangwoong Yoon, Samira Ebrahimi Kahou, Karthik Kashinath, and Mr Prabhat. Deep-hurricane-tracker: Tracking and forecasting extreme climate events. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1761–1769. IEEE, 2019.
- Young Jun Kim, Hyun-Cheol Kim, Daehyeon Han, Sanggyun Lee, and Jungho Im. Prediction of monthly arctic sea ice concentrations using satellite and reanalysis data based on convolutional neural networks. 2020.
- Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. In *Advances in neural information processing systems*, pages 2575–2583, 2015.
- Benedikt Knüsel, Marius Zumwald, Christoph Baumberger, Gertrude Hirsch Hadorn, Erich M Fischer, David N Bresch, and Reto Knutti. Applying big data beyond small problems in climate research. *Nature Climate Change*, 9(3):196–202, 2019.
- Shinya Kobayashi, Yukinari Ota, Yayoi Harada, Ayataka Ebita, Masami Moriya, Hirokatsu Onoda, Kazutoshi Onogi, Hirotaka Kamahori, Chiaki Kobayashi, Hirokazu Endo, et al. The

- ja-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1):5–48, 2015.
- Torben Koenigk and Laurent Brodeau. Ocean heat transport into the arctic in the twentieth and twenty-first century in ec-earth. *Climate dynamics*, 42(11-12):3101–3120, 2014.
- Folmer Krikken and Wilco Hazeleger. Arctic energy budget in relation to sea ice variability on monthly-to-annual time scales. *Journal of Climate*, 28(16):6335–6350, 2015.
- Till Kuhlbrodt, Alexa Griesel, Marisa Montoya, Anders Levermann, Matthias Hofmann, and Stefan Rahmstorf. On the driving processes of the atlantic meridional overturning circulation. *Reviews of Geophysics*, 45(2), 2007.
- Rothrock Kwok and DA Rothrock. Decline in arctic sea ice thickness from submarine and icesat records: 1958–2008. *Geophysical Research Letters*, 36(15), 2009.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- Matti Leppäranta. *The drift of sea ice*. Springer Science & Business Media, 2011.
- Matti Leppäranta, Valentin P Meleshko, Petteri Uotila, and Tatiana Pavlova. Sea ice modelling. In *Sea Ice in the Arctic*, pages 315–387. Springer, 2020.
- Martin Leutbecher and Tim N Palmer. Ensemble forecasting. *Journal of computational physics*, 227(7):3515–3539, 2008.
- Martin Leutbecher, Sarah-Jane Lock, Pirkka Ollinaho, Simon TK Lang, Gianpaolo Balsamo, Peter Bechtold, Massimo Bonavita, Hannah M Christensen, Michail Diamantakis, Emanuel Dutra, et al. Stochastic representations of model uncertainties at ecmwf: State of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707):2315–2339, 2017.
- S Levitus, TP Boyer, ME Conkright, T Obrien, J Antonov, C Stephens, L Stathoplos, D Johnson, and R Gelfeld. NOAA atlas nesdis 18, world ocean database 1998: vol. 1: Introduction. *US Government Printing Office, Washington DC*, 346, 1998.
- MS Lian and RD Cess. Energy balance climate models: A reappraisal of ice-albedo feedback. *Journal of the Atmospheric Sciences*, 34(7):1058–1062, 1977.
- Minyi Liang, Arnaud Czaja, Rune Graversen, and Remi Tailleux. Poleward energy transport: is the standard definition physically relevant at all time scales? *Climate dynamics*, 50(5-6):1785–1797, 2018.
- R Lindsay and A Schweiger. Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *Cryosphere*, 9(1), 2015.
- Ron Lindsay, Mark Wensnahan, A Schweiger, and J Zhang. Evaluation of seven different atmospheric reanalysis products in the arctic. *Journal of Climate*, 27(7):2588–2606, 2014.
- Zachary C Lipton, John Berkowitz, and Charles Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.
- Chunlei Liu, Richard P Allan, Michael Mayer, Patrick Hyder, Norman G Loeb, Chris D



- Roberts, Maria Valdivieso, John M Edwards, and Pier-Luigi Vidale. Evaluation of satellite and reanalysis-based global net surface energy flux and uncertainty estimates. *Journal of Geophysical Research: Atmospheres*, 122(12):6250–6272, 2017.
- Wei Liu, Alexey Fedorov, and Florian Sévellec. The mechanisms of the atlantic meridional overturning circulation slowdown induced by arctic sea ice decline. *Journal of Climate*, 32(4): 977–996, 2019.
- Yang Liu and Jisk Attema. Synthesis and dissemination of ocean and atmosphere heat transport to the Arctic (D2.4), November 2019. URL <https://doi.org/10.5281/zenodo.3631084>.
- Yang Liu, Jisk Attema, and Wilco Hazeleger. Atmosphere–ocean interactions and their footprint on heat transport variability in the northern hemisphere. *Journal of Climate*, 33(9):3691–3710, 2020a.
- Yang Liu, Jisk Attema, Ben Moat, and Wilco Hazeleger. Synthesis and evaluation of historical meridional heat transport from midlatitudes towards the arctic. *Earth System Dynamics*, 11(1):77–96, 2020b.
- Yang Liu, Jisk Attema, and Wilco Hazeleger. Exploring bayesian deep learning for weather forecasting with the lorenz 84 system. *Geophysical Research Letters*, 2021a. Under review.
- Yang Liu, Laurens Bogaardt, Jisk Attema, and Wilco Hazeleger. Extended range arctic sea ice forecast with convolutional long-short term memory networks. *Monthly Weather Review*, 149(6):1673–1693, 2021b.
- Zhengyu Liu, Haijun Yang, Chengfei He, and Yingying Zhao. A theory for bjerknes compensation: The role of climate feedback. *Journal of Climate*, 29(1):191–208, 2016.
- Norman G Loeb, John M Lyman, Gregory C Johnson, Richard P Allan, David R Doelling, Takmeng Wong, Brian J Soden, and Graeme L Stephens. Observed changes in top-of-the-atmosphere radiation and upper-ocean heating consistent within uncertainty. *Nature Geoscience*, 5(2):110–113, 2012.
- Norman G Loeb, David R Doelling, Hailan Wang, Wenying Su, Cathy Nguyen, Joseph G Corbett, Lusheng Liang, Cristian Mitrescu, Fred G Rose, and Seiji Kato. Clouds and the earth’s radiant energy system (ceres) energy balanced and filled (ebaf) top-of-atmosphere (toa) edition-4.0 data product. *Journal of Climate*, 31(2):895–918, 2018.
- Edward N Lorenz. Irregularity: A fundamental property of the atmosphere. *Tellus A*, 36(2): 98–110, 1984.
- Edward N Lorenz and Edward N Lorenz. *The nature and theory of the general circulation of the atmosphere*, volume 218. World Meteorological Organization Geneva, 1967.
- MS Lozier, F Li, S Bacon, F Bahr, AS Bower, SA Cunningham, MF de Jong, L de Steur, J Fischer, SF Gary, et al. A sea change in our view of overturning in the subpolar north atlantic. *Science*, 363(6426):516–521, 2019.
- Gurvan Madec. Nemo reference manual, ocean dynamic component: Nemo-opa. *Note du Pôle modélisation, Inst. Pierre Simon Laplace, Fr*, 2008.
- Gudrun Magnusdottir and R Saravannan. The response of atmospheric heat transport to zonally-

- averaged sst trends. *Tellus A: Dynamic Meteorology and Oceanography*, 51(5):815–832, 1999.
- Irina Mahlstein and Reto Knutti. Ocean heat transport as a cause for model uncertainty in projected arctic warming. *Journal of Climate*, 24(5):1451–1460, 2011.
- Alice Marzocchi, Joël J-M Hirschi, N Penny Holliday, Stuart A Cunningham, Adam T Blaker, and Andrew C Coward. The north atlantic subpolar circulation in an eddy-resolving global ocean model. *Journal of Marine Systems*, 142:126–143, 2015.
- Simona Masina, Andrea Storto, Nicolas Ferry, Maria Valdivieso, Keith Haines, Magdalena Balmaseda, Hao Zuo, Marie Drevillon, and Laurent Parent. An ensemble of eddy-permitting global ocean reanalyses from the myocean project. *Climate Dynamics*, 49(3):813–841, 2017.
- Michael Mayer and Leopold Haimberger. Poleward atmospheric energy transports and their variability as evaluated from ecmwf reanalysis data. *Journal of Climate*, 25(2):734–752, 2012.
- Michael Mayer, Leopold Haimberger, Marianne Pietschnig, and Andrea Storto. Facets of arctic energy accumulation based on observations and reanalyses 2000–2015. *Geophysical research letters*, 43(19):10–420, 2016.
- Michael Mayer, Leopold Haimberger, John M Edwards, and Patrick Hyder. Toward consistent diagnostics of the coupled atmosphere and ocean energy budgets. *Journal of Climate*, 30(22):9225–9246, 2017.
- Michael Mayer, Steffen Tietsche, Leopold Haimberger, Takamasa Tsubouchi, Johannes Mayer, and Hao Zuo. An improved estimate of the coupled arctic energy budget. *Journal of Climate*, 32(22):7915–7934, 2019.
- GD McCarthy, DA Smeed, WE Johns, E Frajka-Williams, BI Moat, Darren Rayner, MO Baringer, CS Meinen, J Collins, and HL Bryden. Measuring the atlantic meridional overturning circulation at 26 n. *Progress in Oceanography*, 130:91–111, 2015.
- Patrick L McDermott and Christopher K Wikle. Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data. *Entropy*, 21(2):184, 2019.
- Herle Mercier, Pascale Lherminier, and Claire Gourcuff. North atlantic ocean circulation: the ovide project. *Houille blanche-revue internationale de l’eau*, (2):30–32, 2008.
- E Joseph Metzger, Ole Martin Smedstad, Prasad G Thoppil, Harley E Hurlburt, James A Cummings, Alan J Wallcraft, Luis Zamudio, Deborah S Franklin, Pamela G Posey, Michael W Phelps, et al. Us navy operational global ocean and arctic ice prediction systems. *Oceanography*, 27(3):32–43, 2014.
- Sebastian Milinski, Nicola Maher, and Dirk Olonscheck. How large does a large ensemble need to be. *Earth System Dynamics*, 11:885–901, 2020.
- Gifford H Miller, Richard B Alley, Julie Brigham-Grette, Joan J Fitzpatrick, Leonid Polyak, Mark C Serreze, and James WC White. Arctic amplification: can the past constrain the future? *Quaternary Science Reviews*, 29(15-16):1779–1790, 2010.
- Ben Moat, Christophe Herbaut, Karin Margretha Larsen, Bogi Hansen, Bablu Sinha, Alejandra Sanchez-Franks, Loic Houpert, Yang Liu, Wilco Hazeleger, Jisk Attema, Stephen Yeager, Justin Small, Hedinn Valdimarsson, Barbara Berx, Stuart Cunningham, Loic Houpert,

- Samantha Hallam, Rebecca Woodgate, Craig Lee, Young Oh Kwon, Laura Flemming, Herle Mercier, Kerstin Jochumsen, Jennifer Mecking, Penny Holliday Holliday, and Simon Josey. Model-observation and reanalyses comparison at key locations for heat transport to the Arctic (D2.1), January 2020. URL <https://doi.org/10.5281/zenodo.3631100>.
- BI Moat, SA Josey, B Sinha, AT Blaker, DA Smeed, GD McCarthy, WE Johns, JJ-M Hirschi, E Frajka-Williams, D Rayner, et al. Major variations in subtropical north atlantic heat transport at short (5 day) timescales and their causes. *Journal of Geophysical Research: Oceans*, 121(5):3237–3249, 2016.
- Kristian Mogensen, Magdalena Alonso Balmaseda, and Anthony Weaver. *The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4*. European Centre for Medium-Range Weather Forecasts, 2012.
- Mahdi Mohammadi-Aragh, HF Goessling, Martin Losch, Nils Hutter, and Thomas Jung. Predictability of arctic sea ice on weather time scales. *Scientific reports*, 8(1):6514, 2018.
- Andrea Molod, Lawrence Takacs, Max Suarez, and Julio Bacmeister. Development of the geos-5 atmospheric general circulation model: Evolution from merra to merra2. *Geoscientific Model Development*, 8(5):1339–1356, 2015.
- Ionel M Navon. Data assimilation for numerical weather prediction: a review. In *Data assimilation for atmospheric, oceanic and hydrologic applications*, pages 21–65. Springer, 2009.
- Radford M Neal and Geoffrey E Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pages 355–368. Springer, 1998.
- Matthew Newman, Michael A Alexander, Toby R Ault, Kim M Cobb, Clara Deser, Emanuele Di Lorenzo, Nathan J Mantua, Arthur J Miller, Shoshiro Minobe, Hisashi Nakamura, et al. The pacific decadal oscillation, revisited. *Journal of Climate*, 29(12):4399–4427, 2016.
- Aleksi Nummelin, Camille Li, and Paul J Hezel. Connecting ocean heat transport changes from the midlatitudes to the arctic ocean. *Geophysical Research Letters*, 44(4):1899–1908, 2017.
- Marilena Oltmanns, Johannes Karstensen, and Jürgen Fischer. Increased risk of a shutdown of ocean convection posed by warm north atlantic summers. *Nature Climate Change*, 8(4):300, 2018.
- Ingrid H Onarheim, Tor Eldevik, Marius Årthun, Randi B Ingvaldsen, and Lars H Smedsrud. Skillful prediction of barents sea ice cover. *Geophysical Research Letters*, 42(13):5364–5371, 2015.
- Abraham H Oort and Thomas H Vonder Haar. On the observed annual cycle in the ocean-atmosphere heat balance over the northern hemisphere. *Journal of Physical Oceanography*, 6(6):781–800, 1976.
- Stephen Outten and Igor Esau. Bjerknes compensation in the bergen climate model. *Climate Dynamics*, 49(7-8):2249–2260, 2017.
- Stephen Outten, Igor Esau, and Odd Helge Otterå. Bjerknes compensation in the cmip5 climate models. *Journal of Climate*, 31(21):8745–8760, 2018.
- James Overland, Jennifer A Francis, Richard Hall, Edward Hanna, Seong-Joong Kim, and Timo

- Vihma. The melting arctic and midlatitude weather patterns: Are they connected? *Journal of Climate*, 28(20):7917–7932, 2015.
- James E Overland. A difficult arctic science issue: Midlatitude weather linkages. *Polar Science*, 10(3):210–216, 2016.
- Jonathan Overpeck, Konrad Hughen, D Hardy, Raymond Bradley, R Case, Marianne Douglas, Bruce Finney, Konrad Gajewski, G\_ Jacoby, Anne Jennings, et al. Arctic environmental change of the last four centuries. *Science*, 278(5341):1251–1256, 1997.
- Ronald C Pacanowski, K Dixon, and Anthony Rosati. The gfdl modular ocean model users guide. *GFDL Ocean Group Tech. Rep*, 2:142, 1991.
- TN Palmer. Predicting uncertainty in numerical weather forecasts. In *International Geophysics*, volume 83, pages 3–13. Elsevier, 2002.
- Tess Parker, Tim Woollings, and Antje Weisheimer. Ensemble sensitivity analysis of greenland blocking in medium-range forecasts. *Quarterly Journal of the Royal Meteorological Society*, 144(716):2358–2379, 2018.
- José Pinto Peixoto and Abraham H Oort. *Physics of climate*. 1992.
- G Peng, Walter N Meier, DJ Scott, and MH Savoie. A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring. 2013.
- Donald K Perovich and Jacqueline A Richter-Menge. Loss of sea ice in the arctic. *Annual Review of Marine Science*, 1:417–441, 2009.
- Zisis I Petrou and Yingli Tian. Prediction of sea ice motion with convolutional long short-term memory networks. *IEEE Transactions on Geoscience and Remote Sensing*, 2019.
- Norman A Phillips. The general circulation of the atmosphere: A numerical experiment. *Quarterly Journal of the Royal Meteorological Society*, 82(352):123–164, 1956.
- Felix Pithan and Thorsten Mauritsen. Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, 7(3):181–184, 2014.
- Andrei Y Proshutinsky and Mark A Johnson. Two circulation regimes of the wind-driven arctic ocean. *Journal of Geophysical Research: Oceans*, 102(C6):12493–12514, 1997.
- William M Putman and Shian-Jiann Lin. Finite-volume transport on various cubed-sphere grids. *Journal of Computational Physics*, 227(1):55–78, 2007.
- Stephan Rasp and Sebastian Lerch. Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146(11):3885–3900, 2018.
- Stephan Rasp, Michael S Pritchard, and Pierre Gentine. Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, 115(39):9684–9689, 2018.
- Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9):2352–2449, 2017.
- Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system

- science. *Nature*, 566(7743):195–204, 2019.
- Stephen C Riser, Howard J Freeland, Dean Roemmich, Susan Wijffels, Ariel Troisi, Mathieu Belbéoch, Denis Gilbert, Jianping Xu, Sylvie Pouliquen, Ann Thresher, et al. Fifteen years of ocean observations with the global argo array. *Nature Climate Change*, 6(2):145, 2016.
- Clément Rousset, Martin Vancoppenolle, Gurvan Madec, Thierry Fichefet, Simona Flavoni, Antoine Barthélemy, Rachid Benshila, Jérôme Chanut, Claire Lévy, Sébastien Masson, et al. The louvain-la-neuve sea ice model lim3. 6: global and regional capabilities. 2015.
- Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu-Tai Hou, Hui-ya Chuang, Mark Iredell, et al. The ncep climate forecast system version 2. *Journal of climate*, 27(6):2185–2208, 2014.
- Tim Salimans, Diederik Kingma, and Max Welling. Markov chain monte carlo and variational inference: Bridging the gap. In *International Conference on Machine Learning*, pages 1218–1226, 2015.
- Afan Galih Salman, Bayu Kanigoro, and Yaya Heryadi. Weather forecasting using deep learning techniques. In *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 281–285. IEEE, 2015.
- AB Sandø, Y Gao, and HR Langehaug. Poleward ocean heat transports, sea ice processes, and arctic sea ice variability in noresm1-m simulations. *Journal of Geophysical Research: Oceans*, 119(3):2095–2108, 2014.
- Kazutoshi Sato, Jun Inoue, Akira Yamazaki, Joo-Hong Kim, Marion Maturilli, Klaus Dethloff, Stephen R Hudson, and Mats A Granskog. Improved forecasts of winter weather extremes over midlatitudes with extra arctic observations. *Journal of geophysical research: oceans*, 122(2):775–787, 2017.
- Ursula Schauer and Agnieszka Beszczynska-Möller. Problems with estimation and interpretation of oceanic heat transport—conceptual remarks for the case of fram strait in the arctic ocean. *Ocean Science*, 5(4):487–494, 2009.
- Sebastian Scher and Gabriele Messori. Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society*, 144(717):2830–2841, 2018.
- Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61: 85–117, 2015.
- James A Screen and Jennifer A Francis. Contribution of sea-ice loss to arctic amplification is regulated by pacific ocean decadal variability. *Nature Climate Change*, 6(9):856, 2016.
- Denis Sergeev, Ian A Renfrew, and Thomas Spengler. Modification of polar low development by orography and sea ice. *Monthly Weather Review*, 146(10):3325–3341, 2018.
- Mark C Serreze and Roger G Barry. Processes and impacts of arctic amplification: A research synthesis. *Global and planetary change*, 77(1-2):85–96, 2011.
- Mark C Serreze and Roger G Barry. *The Arctic climate system*. Cambridge University Press, 2014.
- Mark C Serreze, Andrew P Barrett, Andrew G Slater, Michael Steele, Jinlun Zhang, and Kevin E

- Trenberth. The large-scale energy budget of the arctic. *Journal of Geophysical Research: Atmospheres*, 112(D11), 2007.
- Florian Sévellec, Alexey V Fedorov, and Wei Liu. Arctic sea-ice decline weakens the atlantic meridional overturning circulation. *Nature Climate Change*, 7(8):604, 2017.
- Len Shaffrey and Rowan Sutton. The interannual variability of energy transports within and over the atlantic ocean in a coupled climate model. *Journal of climate*, 17(7):1433–1448, 2004.
- Len Shaffrey and Rowan Sutton. Bjerknes compensation and the decadal variability of the energy transports in a coupled climate model. *Journal of Climate*, 19(7):1167–1181, 2006.
- Adam P Showman, Robin D Wordsworth, Timothy M Merlis, and Yohai Kaspi. Atmospheric circulation of terrestrial exoplanets. *Comparative Climatology of Terrestrial Planets*, 1:277–326, 2013.
- Kumar Shridhar, Felix Laumann, and Marcus Liwicki. Uncertainty estimations by softplus normalization in bayesian convolutional neural networks with variational inference. *arXiv preprint arXiv:1806.05978*, 2018.
- Kumar Shridhar, Felix Laumann, and Marcus Liwicki. A comprehensive guide to bayesian convolutional neural network with variational inference. *arXiv preprint arXiv:1901.02731*, 2019.
- AJ Simmons, P Poli, DP Dee, P Berrisford, H Hersbach, S Kobayashi, and C Peubey. Estimating low-frequency variability and trends in atmospheric temperature using era-interim. *Quarterly Journal of the Royal Meteorological Society*, 140(679):329–353, 2014.
- AJ Simmons, Paul Berrisford, DP Dee, H Hersbach, S Hirahara, and J-N Thépaut. A re-assessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quarterly Journal of the Royal Meteorological Society*, 143(702):101–119, 2017.
- Lars H Smedsrud, Igor Esau, Randi B Ingvaldsen, Tor Eldevik, Peter M Haugan, Camille Li, Vidar S Lien, Are Olsen, Abdirahman M Omar, Odd H Otterå, et al. The role of the barents sea in the arctic climate system. *Reviews of Geophysics*, 51(3):415–449, 2013.
- DA Smeed, G McCarthy, SA Cunningham, E Frajka-Williams, D Rayner, WE Johns, CS Meinen, MO Baringer, BI Moat, A Ducheze, et al. Observed decline of the atlantic meridional overturning circulation 2004–2012. *Ocean Science*, 10(1):29–38, 2014.
- Gregory C Smith, François Roy, and Bruce Brasnett. Evaluation of an operational ice–ocean analysis and forecasting system for the gulf of st lawrence. *Quarterly Journal of the Royal Meteorological Society*, 139(671):419–433, 2013.
- Gregory C Smith, François Roy, Mateusz Reszka, Dorina Surcel Colan, Zhongjie He, Daniel Deacu, Jean-Marc Belanger, Sergey Skachko, Yimin Liu, Frédéric Dupont, et al. Sea ice forecast verification in the canadian global ice ocean prediction system. *Quarterly Journal of the Royal Meteorological Society*, 142(695):659–671, 2016.
- Michael Steele, Rebecca Morley, and Wendy Ermold. Phc: A global ocean hydrography with a high-quality arctic ocean. *Journal of Climate*, 14(9):2079–2087, 2001.

- David J Stensrud. *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press, 2009.
- Vladimir Stepanov and Keith Haines. Mechanisms for amoc variability simulated by the nemo model. *Ocean Science*, 10(4):645–656, 2014.
- Scott R Stephenson and Rebecca Pincus. Challenges of sea-ice prediction for arctic marine policy and planning. *Journal of Borderlands Studies*, 33(2):255–272, 2018.
- Julienne Stroeve and Dirk Notz. Changing state of arctic sea ice across all seasons. *Environmental Research Letters*, 13(10):103001, 2018.
- M Susan Lozier, Sheldon Bacon, Amy S Bower, Stuart A Cunningham, M Femke de Jong, Laura De Steur, Brad deYoung, Jürgen Fischer, Stefan F Gary, Blair JW Greenan, et al. Overturning in the subpolar north atlantic program: A new international ocean observing system. *Bulletin of the American Meteorological Society*, 98(4):737–752, 2017.
- Lea Svendsen, Noel Keenlyside, Ingo Bethke, Yongqi Gao, and Nour-Eddine Omrani. Pacific contribution to the early twentieth-century warming in the arctic. *Nature Climate Change*, 8(9):793, 2018.
- Harald Ulrik Sverdrup, Martin Wiggo Johnson, Richard H Fleming, et al. *The Oceans: Their physics, chemistry, and general biology*, volume 7. Prentice-Hall New York, 1942.
- Matthew D Thomas, Agatha M De Boer, Helen L Johnson, and David P Stevens. Spatial and temporal scales of sverdrup balance. *Journal of Physical Oceanography*, 44(10):2644–2660, 2014.
- David WJ Thompson and John M Wallace. The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research letters*, 25(9):1297–1300, 1998.
- Axel Timmermann and Hugues Goosse. Is the wind stress forcing essential for the meridional overturning circulation? *Geophysical Research Letters*, 31(4), 2004.
- Mary-Louise Timmermans and John Marshall. Understanding arctic ocean circulation: A review of ocean dynamics in a changing climate. *Journal of Geophysical Research: Oceans*, 125(4), 2020.
- Kevin E Trenberth. Climate diagnostics from global analyses: Conservation of mass in ecmwf analyses. *Journal of Climate*, 4(7):707–722, 1991.
- Kevin E Trenberth and Julie M Caron. Estimates of meridional atmosphere and ocean heat transports. *Journal of Climate*, 14(16):3433–3443, 2001.
- Kevin E Trenberth and John T Fasullo. An observational estimate of inferred ocean energy divergence. *Journal of Physical Oceanography*, 38(5):984–999, 2008.
- Kevin E Trenberth and John T Fasullo. Atlantic meridional heat transports computed from balancing earth’s energy locally. *Geophysical Research Letters*, 44(4):1919–1927, 2017.
- Kevin E Trenberth and John T Fasullo. Applications of an updated atmospheric energetics formulation. *Journal of Climate*, 31(16):6263–6279, 2018.

- Kevin E Trenberth and Amy Solomon. The global heat balance: Heat transports in the atmosphere and ocean. *Climate Dynamics*, 10(3):107–134, 1994.
- Kevin E Trenberth and David P Stepaniak. Seamless poleward atmospheric energy transports and implications for the hadley circulation. *Journal of Climate*, 16(22):3706–3722, 2003.
- Kevin E Trenberth, David P Stepaniak, and Julie M Caron. Accuracy of atmospheric energy budgets from analyses. *Journal of Climate*, 15(23):3343–3360, 2002.
- Kevin E Trenberth, John T Fasullo, and Magdalena A Balmaseda. Earth’s energy imbalance. *Journal of Climate*, 27(9):3129–3144, 2014.
- Takamasa Tsubouchi, Kjetil Våge, Bogi Hansen, Karin Margretha H Larsen, Svein Østerhus, Clare Johnson, Steingrímur Jónsson, and Hédhinn Valdimarsson. Increased ocean heat transport into the nordic seas and arctic ocean over the period 1993–2016. *Nature Climate Change*, 11(1):21–26, 2021.
- Petteri Uotila, Hugues Goosse, Keith Haines, Matthieu Chevallier, Antoine Barthélemy, Clément Bricaud, Jim Carton, Neven Fučkar, Gilles Garric, Doroteaciro Iovino, et al. An assessment of ten ocean reanalyses in the polar regions. *Climate Dynamics*, pages 1–38, 2018.
- Sakari M Uppala, PW Kållberg, AJ Simmons, U Andrae, V Da Costa Bechtold, M Fiorino, JK Gibson, J Haseler, A Hernandez, GA Kelly, et al. The era-40 re-analysis. *Quarterly Journal of the royal meteorological society*, 131(612):2961–3012, 2005.
- Eveline C van der Linden, Richard Bintanja, Wilco Hazeleger, and Rune G Graversen. Low-frequency variability of surface air temperature over the barents sea: causes and mechanisms. *Climate dynamics*, 47(3-4):1247–1262, 2016.
- E Van der Swaluw, SS Drijfhout, and W Hazeleger. Bjerknes compensation at high northern latitudes: The ocean forcing the atmosphere. *Journal of climate*, 20(24):6023–6032, 2007.
- Michael L Van Woert, Cheng-Zhi Zou, Walter N Meier, Philip D Hovey, Ruth H Preller, and Pamela G Posey. Forecast verification of the polar ice prediction system (pips) sea ice concentration fields. *Journal of Atmospheric and Oceanic Technology*, 21(6):944–957, 2004.
- Thomas Vandal, Evan Kodra, Jennifer Dy, Sangram Ganguly, Ramakrishna Nemani, and Auroop R Ganguly. Quantifying uncertainty in discrete-continuous and skewed data with bayesian deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2377–2386. ACM, 2018.
- Timo Vihma. Effects of arctic sea ice decline on weather and climate: A review. *Surveys in Geophysics*, 35(5):1175–1214, 2014.
- F Vitart, C Ardilouze, A Bonet, A Brookshaw, M Chen, C Codorean, M Déqué, L Ferranti, E Fucile, M Fuentes, et al. The subseasonal to seasonal (s2s) prediction project database. *Bulletin of the American Meteorological Society*, 98(1):163–173, 2017.
- Karina von Schuckmann, Pierre-Yves Le Traon, Neville Smith, Ananda Pascual, Pierre Brasseur, Katja Fennel, Samy Djavidnia, Signe Aaboe, Enrique Alvarez Fanjul, Emmanuelle Autret, et al. Copernicus marine service ocean state report. *Journal of Operational Oceanography*, 11(sup1):S1–S142, 2018.



- Thomas H Vonder Haar and Abraham H Oort. New estimate of annual poleward energy transport by northern hemisphere oceans. *Journal of Physical Oceanography*, 3(2):169–172, 1973.
- John E Walsh, J Scott Stewart, and Florence Fetterer. Benchmark seasonal prediction skill estimates based on regional indices. *The Cryosphere*, 13(4):1073–1088, 2019.
- Hu Wang, Yongguang Yu, and Guoguang Wen. Dynamical analysis of the lorenz-84 atmospheric circulation model. *Journal of Applied Mathematics*, 2014, 2014.
- Huai-zhi Wang, Gang-qiang Li, Gui-bin Wang, Jian-chun Peng, Hui Jiang, and Yi-tao Liu. Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied energy*, 188:56–70, 2017.
- Lei Wang, Xiaojun Yuan, Mingfang Ting, and Cuihua Li. Predicting summer arctic sea ice concentration intraseasonal variability using a vector autoregressive model. *Journal of Climate*, 29(4):1529–1543, 2016a.
- Lei Wang, Xiaojun Yuan, and Cuihua Li. Subseasonal forecast of arctic sea ice concentration via statistical approaches. *Climate Dynamics*, 52(7-8):4953–4971, 2019.
- Xuanji Wang, Jeffrey Key, Ron Kwok, and Jinlun Zhang. Comparison of arctic sea ice thickness from satellites, aircraft, and piomas data. *Remote Sensing*, 8(9):713, 2016b.
- Richard G Williams, Chris Wilson, and Chris W Hughes. Ocean and atmosphere storm tracks: The role of eddy vorticity forcing. *Journal of Physical Oceanography*, 37(9):2267–2289, 2007.
- Xingren Wu, Robert Grumbine, et al. Sea ice in the ncep climate forecast system reanalysis. In *The Twentieth International Offshore and Polar Engineering Conference*. International Society of Offshore and Polar Engineers, 2010.
- Carl Wunsch. The total meridional heat flux and its oceanic and atmospheric partition. *Journal of climate*, 18(21):4374–4380, 2005.
- SHI Xingjian, Zhouong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.
- Haijun Yang, Yuxing Wang, and Zhengyu Liu. A modelling study of the bjercknes compensation in the meridional heat transport in a freshening ocean. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):18480, 2013.
- Haijun Yang, Kun Wang, Haijin Dai, Yuxing Wang, and Qing Li. Wind effect on the atlantic meridional overturning circulation via sea ice and vertical diffusion. *Climate Dynamics*, 46(11-12):3387–3403, 2016.
- Qianzi Yang, Yingying Zhao, Qin Wen, Jie Yao, and Haijun Yang. Understanding bjercknes compensation in meridional heat transports and the role of freshwater in a warming climate. *Journal of Climate*, 31(12):4791–4806, 2018.
- Xiao-Yi Yang, John C Fyfe, and Gregory M Flato. The role of poleward energy transport in arctic temperature evolution. *Geophysical Research Letters*, 37(14), 2010.
- Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, 31(7):1235–1270, 2019.

- 
- Xiaojun Yuan, Dake Chen, Cuihua Li, Lei Wang, and Wanqiu Wang. Arctic sea ice seasonal prediction by a linear markov model. *Journal of Climate*, 29(22):8151–8173, 2016.
- I Zaier, C Shu, TBMJ Ouarda, O Seidou, and F Chebana. Estimation of ice thickness on lakes using artificial neural network ensembles. *Journal of Hydrology*, 383(3-4):330–340, 2010.
- Lorenzo Zampieri, Helge F Goessling, and Thomas Jung. Bright prospects for arctic sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, 45(18):9731–9738, 2018.
- Quan-shi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1):27–39, 2018.
- Rong Zhang. Mechanisms for low-frequency variability of summer arctic sea ice extent. *Proceedings of the National Academy of Sciences*, page 201422296, 2015.
- Yangxing Zheng and Benjamin S Giese. Ocean heat transport in simple ocean data assimilation: Structure and mechanisms. *Journal of Geophysical Research: Oceans*, 114(C11), 2009.

# Acknowledgements

A wonderful journey cannot begin without the company of wonderful people. I would like to thank everyone who encourages me, supports me and helps me during my PhD. There are so many names I would like to mention and so many glowing moments I would wish to share. But then I definitely need to expand these pages into a name book. Therefore, I will try to refrain from losing the control of my gratitudes and touching every piece of precious memory in my brain.

First, I would like to thank my promoter, my boss, my tutor, my guide, Wilco. Thank you so much for your steering! It is so lucky to receive the scientific training and grow up as an independent researcher under your wing. You taught me so many things, about the atmosphere and ocean, about the science community, about the career path as a young scientist, and about the Dutch culture and society. You lead me to the gallery of science and teach me to think critically. I am so impressed by your broad knowledge and your way of thinking. Although your agenda is full all the time, when I need help, you will always be there. Those update meetings on Monday at the Netherlands eScience Center, inspiring talks on Friday in Lumen at Wageningen, and interesting discussions on Friday in Vening Meineszgebouw at Utrecht University, they all recorded your mentoring and my growing. Science is not easy, but with your guide and help, we made it possible!

And, my co-promoter, my teacher, my friend, Jisk, many thanks for your daily-supervising. I'm so happy to spend lots of time with you at the center. You are always prepared to help me, with your multi-disciplinary background and your instinct of science. Without your support I cannot tackle so many technical challenges. Your patience, your care, and your advices pave my way to success. We had many shining moments working together, like the special "baby-sitting" time after each update meeting with Wilco in case I had some confusions (well, I always did), and our wet day tour in Lisbon in an epic heavy rain after the Blue Action assembly. With you company, no matter how challenging the topic is, I will not be afraid.

I kicked the start of my PhD life at the Netherlands eScience Center. The story began with a very warm welcome from everyone in our center. I was so lucky to join this big and warm family. There were so many memorable moments here, which became the main theme of my PhD life. For instance, interesting discussions at the coffee corner, fancy lunch presentations, fantastic SIG meetings, exciting excursions and social activities. It was you who filled these moments with happiness, Faruk, Carlos, William, Patrick, Sacha, Otto, Karima, Dafne, Gijs, Ronald, Erik, Arnold, Maarten, Christiaan, Sonja, Elena, Hanno, Janneke, Stefan, Johan, Lourens,

---

Rena, Noura, Tom, Yifat, Laurens, Lars, Kim, Johann, Frank, Nico, Niels, Rob, Florian, Stijn, Berend, Inti, Sarah, Prem, Marjie, Mariëtte, Mateusz, Cunliang, Pusha, Meiert, Merijn, Adam, Peter, Bouwe, Alessio, Ben, Jason, Jurriaan, Adriënnna, Jesus, Pablo, Jaro, Felipe, Dujra, Sven, Jens, Joris ... (I know when start enumerating names, I will get myself in trouble. But I sincerely want to do it). Thank you all for creating this wonderful place. I am also very happy that I can continue my journey in our center with everyone.

I also want to thank my fellows, peers, colleagues and friends at Wageningen. My first atmospheric course (atmospheric dynamics) was given by Jordi and Gert-Jan. I had my first MAQ group meeting with Chiel in Forum. Imme and Anna-Lena, as my seniors and friends, shared with me their experience about mastering the PhD. Robin, Menno, and I, we together organized a very successful online symposium and overcame many difficulties. Auke, Arjan, Gerband, Stijn, Anja, Aris, Frank, Liesbeth, Ara, Ingrid, Sandra, Wouter and Maarten, many thanks for the interesting talks at MAQ table, during monthly seminars and at BBOS. Also, I am grateful for the help from the staff at MAQ. In addition, I would like to mention two master students I supervised, Daan and Judith. It was nice to work with you and I would also like to offer my gratitudes to you, for the interesting talks between us and the inspirations I received from you. All in all, I feel so lucky to join the MAQ group and spent time with so many nice people at Wageningen.

当然在此，我更要感谢的是，爸妈和所有家人对我的支持！感谢你们对我无微不至的关心和理解。每次遇到挫折和困难，爸妈都会在屏幕前倾听，尽管相隔万里，还是能十分真切地感受到你们对我的牵挂。我开心的时候，你们与我分享喜悦；我失落的时候，你们的鼓励无处不在；我遇到困难的时候，你们又积极的为我出谋划策。正是你们支持与理解让我走完了四年的漫漫征程。此外，还有我的其他亲人，我的爷爷奶奶，姥姥姥爷，我的舅舅阿姨们，我的姑姑，还有各位长辈和兄弟姐妹们，谢谢你们的牵挂。也正是家的陪伴，给了我勇气和毅力，远渡重洋，历经磨难，才终于达成愿望，学业有成。

此外，还要感谢每天陪伴我，分享欢声笑语的小伙伴们。在荷兰各个地方，都有你们的身影。感谢海牙的库欧静雯夫妇，还有尔康。感谢瓦村的彭彭和燕宁师妹。感谢阿姆斯特丹大学还有各个研究所和医院的小伙伴们，感谢天翼，博哥，王教授，璇姐，薛彤师姐，宝烘师弟，晓冉师妹，祯姐，少杰和彩霞。还有活跃在阿姆斯特丹和周边的好友们，感谢郝越意Guido夫妇，康哥小艺姐，培菁，还有静仪。因为有你们的陪伴，单调的学术生活也变得多姿多彩。感谢在代村以及从代村离开的，我们海工的老同学们。当然，还有我可爱的室友，文斐，兴宇，canders。要感谢的，还有我们大工船舶散落在世界各地的同学们，在苏黎世的坤哥，在赫尔辛基的放哥，在广州的海鑫。最后，还要感谢我家乡的朋友们和老同学们，与你们的互动和联系也为我的博士生活注入了诸多的色彩，感谢冯尧，赵昱昊，赵伯洋，周千惠，赵炜光，诗航，常琳，小孟，张苑，还有大乐。

Last but not the least, I'm grateful for the chance to join the EU Blue Action project as a PhD student. I very much enjoyed the collaborations with many organization, universities and centers of excellence around the world. I learnt a lot from the junior and senior scientists in this project.

Limited by the length, I cannot list every name on these pages. But I do appreciate everyone's presence and help during my PhD. I want to let you know that, without you and your help, I

---

cannot graduate on time and kick the start of my career as an independent researcher! Time flies and this four-year journey serves as the prelude of a new movement of my life. Again, I would like to thank you all, for leaving your note on my piece of melody and we, all together, composed fantasy.

Yang



# About the author

Yang Liu was born on February 27th, 1991 in Tianjin, China. With a strong interest in physics and engineering, Yang started his bachelor in 2010 at Dalian University of Technology (DUT), major in Naval Architecture and Ocean Engineering. After obtaining his bachelor degree in 2014, he decided to continue his study in Offshore Engineering and moved to the Netherlands to pursue his master degree at Delft University of Technology (TUD). During his study in Delft, he discovered the beauty of physics and numerical modelling, and further extended his interest in the Computational Fluid Dynamics. This brought him to Wageningen for his master thesis at the Maritime Research Institute Netherlands (MARIN). He obtained his master degree in 2016 and lived in Wageningen for a while. Driven by the broad interest in science, nature and technology, he was lucky to joined the Netherlands eScience Center and was enrolled as a PhD candidate in the Meteorology and Air Quality Group at Wageningen University (WUR). There, Yang worked on the Blue Action project, which was funded by the EU Horizon 2020 Work Programme, and started his adventure in the Arctic weather and climate system. His research concerns atmospheric dynamics, numerical modelling, data mining and artificial intelligence. After wrapping up his scientific contribution to the climate science, Yang was obsessed with digital science and therefore continued his journey as an eScience research engineer in the Netherlands eScience Center.

## Peer-reviewed journal publications

Yang Liu, Jisk Attema, Ben Moat, and Wilco Hazeleger. Synthesis and evaluation of historical meridional heat transport from midlatitudes towards the arctic. *Earth System Dynamics*, 11(1): 77–96, 2020b

Yang Liu, Jisk Attema, and Wilco Hazeleger. Atmosphere–ocean interactions and their footprint on heat transport variability in the northern hemisphere. *Journal of Climate*, 33(9):3691–3710, 2020a

Yang Liu, Laurens Bogaardt, Jisk Attema, and Wilco Hazeleger. Extended range arctic sea ice forecast with convolutional long-short term memory networks. *Monthly Weather Review*, 149(6):1673–1693, 2021b

Yang Liu, Jisk Attema, and Wilco Hazeleger. Exploring bayesian deep learning for weather forecasting with the lorenz 84 system. *Geophysical Research Letters*, 2021a. Under review

---

## Other scientific publications

Ben Moat, Christophe Herbaut, Karin Margretha Larsen, Bogi Hansen, Bablu Sinha, Alejandra Sanchez-Franks, Loic Houpert, Yang Liu, Wilco Hazeleger, Jisk Attema, Stephen Yeager, Justin Small, Hedinn Valdimarsson, Barbara Berx, Stuart Cunningham, Loic Houpert, Samantha Hallam, Rebecca Woodgate, Craig Lee, Young Oh Kwon, Laura Flemming, Herle Mercier, Kerstin Jochumsen, Jennifer Mecking, Penny Holliday Holliday, and Simon Josey. Model-observation and reanalyses comparison at key locations for heat transport to the Arctic (D2.1), January 2020. URL <https://doi.org/10.5281/zenodo.3631100>

Yang Liu and Jisk Attema. Synthesis and dissemination of ocean and atmosphere heat transport to the Arctic (D2.4), November 2019. URL <https://doi.org/10.5281/zenodo.3631084>

Yongqi Gao, Richard Davy, Lingling Suo, Guillaume Gastineau, Young-Oh Kwon, Vladimir Semenov, and Yang Liu. Identification of the surface state influence in representing the Arctic warming by coordinated atmosphere-only simulations (D3.1), September 2019. URL <https://doi.org/10.5281/zenodo.3559462>





*Netherlands Research School for the  
Socio-Economic and Natural Sciences of the Environment*

# D I P L O M A

*for specialised PhD training*

The Netherlands research school for the  
Socio-Economic and Natural Sciences of the Environment  
(SENSE) declares that

***Yang Liu***

born on 27 February 1991 in Tianjin, China

has successfully fulfilled all requirements of the  
educational PhD programme of SENSE.

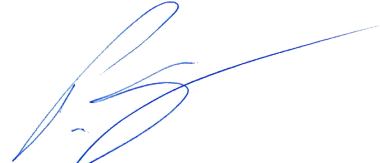
Wageningen, 12 October 2021

Chair of the SENSE board



Prof. dr. Martin Wassen

The SENSE Director



Prof. Philipp Pattberg

*The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)*



K O N I N K L I J K E N E D E R L A N D S E  
A K A D E M I E V A N W E T E N S C H A P P E N



The SENSE Research School declares that **Yang Liu** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 46.6 EC, including the following activities:

#### SENSE PhD Courses

- o Environmental research in context (2017)
- o Research in context activity: 'Organising the symposium "Machine Learning in Meteorology and Climate"' (2020)

#### Other PhD and Advanced MSc Courses

- o Predictability and ocean-atmosphere ensemble forecasting, European Centre for Medium-Range Weather Forecasts (2017)
- o GPU Programming, Netherlands Science Center (2017)
- o Atmospheric Modelling, Wageningen University (2017)
- o Dynamical Oceanography, Utrecht University (2018)
- o Writing a Scientific Article, VU-Amsterdam (2018)
- o Workshop on Parallel Programming in Python, Netherlands Science Center (2019)

#### Management and Didactic Skills Training

- o Supervising two MSc students with thesis entitled 'Improving the predictive skill of Arctic climate predictions' (2019) and 'Influence of melting Arctic sea-ice on the waviness of the summer jet stream' (2020)

#### Oral Presentations

- o *Meridional energy transport from mid-latitudes towards the Arctic*. European Geoscience Union, 9-13 April 2018, Vienna, Austria
- o *Synthesis and evaluation of historical heat transport in the ocean and atmosphere*. Buys Ballot research school (BBOS) autumn symposium, 25-26 October 2018, Utrecht, The Netherlands
- o *Synthesis and evaluation of historical heat transport from midlatitudes towards the Arctic*. Earth Energy Imbalance workshop, 13-16 November 2018, Toulouse, France
- o *Extended range Arctic sea ice forecast with convolutional long-short term memory networks*. European Geoscience Union, 3-8 May 2020, Vienna, Austria

SENSE coordinator PhD education

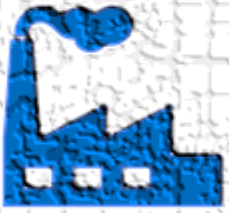
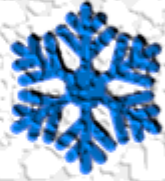
Dr. ir. Peter Vermeulen



The research described in this thesis is supported by the Blue Action project, which received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 727852.

Numerical simulations and data analysis were carried out on the supercomputing facilities of SURFsara, sponsored by the Dutch national e-infrastructure project with the support of SURF Cooperative.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.



1010  
1010

