Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Multi-annual predictability of the Atlantic Meridional Overturning Circulation

Supervisor:
Dr. Paolo Ruggieri

Submitted by:
Chiara Galeotti

Co-supervisors:
Dr. Alessio Bellucci
Dr. Dario Nicolí

Academic Year 2020/2021

Abstract

The North Atlantic is among the few places on Earth where decadal climate variations are considered potentially predictable with an added value of the initial state of the system. This predictability stems also from multi-annual fluctuations of the Atlantic Meridional Overturning Circulation (AMOC). Because the AMOC is responsible for a global-scale energy transport, a correct representation of this process is fundamental to skillfully predict climate variability in the Northern Hemisphere at these timescales. In this thesis, AMOC predictability is investigated in the CMCC-CM2-SR5 (CMCC Coupled Model v2 in standard resolution) decadal predictions system. The ability of the model to forecast the AMOC is evaluated in both a deterministic and probabilistic way, comparing a set of hindcasts initialized between 1960 and 2018 with observations, ocean reconstructions, and a non-initialized historical simulation. Special attention is devoted to the analysis and management of AMOC biases. Indeed, it is documented that predictions suffer from initial shocks and tend to drift towards the model's equilibrium state. We find that the potential predictability of the system is high up to a ten-year forecast range, but this is not reflected in the AMOC transport forecast skill, which undergoes a sudden reduction after the first year. An interesting finding is that the drift of the model is start-date dependent: we leverage on this feature to propose a new post-processing approach for the drift adjustment, different from the usual one in which drifts are treated as stationary. The experimented approach significantly increases the forecast skill. Furthermore, we identify a reduction of convection in the Labrador Sea, a feature that previous studies linked with the model drift of the AMOC. Further research with an increased ensemble size of both initialized and historical simulations and with a multi-model set is envisaged.

Contents

Chapter 1

Introduction

Climate shows variations on a wide range of timescales and our ability to predict it is strongly linked to which timescale we are looking at. Short-term processes, such as atmospheric weather, highly depend on initial conditions and thus result in chaotic motion, limiting the predictability up to a couple of weeks. From a climatic perspective, however, we are interested in larger spatial scales than synoptic weather: hence, we must focus on those processes which have a slower evolution. The natural variability internally generated by the climate system, as opposed to variability forced by human activities, dominates on the interannual-to-decadal scale, representing 80% of the total variance on these periods (Hawkings and Sutton, 2009). We can potentially exploit natural variability to obtain successful predictions of its future evolution, treating climate as an initial value problem where the knowledge of the initial state is lost after a decade or two due to chaotic processes. In recent years, this perspective has prompted the development of decadal climate predictions. These are based on atmosphere-ocean coupled models both forced by radiative scenarios and initialized with an observed state, so that they make it possible to track interannual variations for some years since the starting date.

Oceans play a fundamental role in decadal climate variability because of their slow thermal variations. The heat capacity of the entire atmosphere is equivalent to that of the first 2 m of the ocean surface, so seawater stores the greatest part of energy in the ocean-atmosphere system and acts as a thermal reservoir, delaying atmospheric signals. Oceanic regions in which climate shows the highest potential predictability at decadal scale are the Southern Ocean and, particularly, the North Atlantic (Latif et al., 2006). Coupled general circulation models have shown that, in the Atlantic sector, the climate has strong decadal-to-multidecadal oscillations (the Atlantic Multidecadal Variability, AMV) and these are claimed to be related to Atlantic Meridional Overturning Circulation (AMOC) variations (Zhang et al., 2019). The AMOC is responsible for up to 25% of the northward atmosphere-ocean heat transport necessary to balance the Earth's radiation budget (Bryden et al., 2014). It has important effects on climate and these are not limited just to the Atlantic region (Buckley and Marshall, 2016). Also, since the deployment of the

RAPID-array experiment in 2004, the AMOC has been directly monitored and it resulted in having interannual-to-decadal shifts (Frajka-Williams et al., 2019). With these premises, predicting the AMOC variability is arguably crucial to have good decadal climate predictions in areas such as (but not only) those surrounding the North Atlantic basin. In fact, different modelling studies have shown that the ability to predict temperatures is enhanced in the North Atlantic region for initialized models, and this is related to the initialization of AMOC (e.g. Keenlyside et al., 2008; Pohlmann et al., 2009).

Decadal predictions are very recent (less than fifteen years passed since the first applications) and there are still many challenges on the way for their routine operational production. These issues are related both to the understanding of climate variability and predictability and to modelling aspects such as initial shock, drift and bias, or techniques and data for forecast initialization (Kushnir et al., 2019). Hence, when producing decadal forecasts, these issues are also reflected in the AMOC. In light of this, it is important to have a deeper insight on AMOC variability and predictability in decadal models. In the next section we introduce the AMOC, its relation with Atlantic and global climate and the experimental ways to measure its variability. After a general presentation of decadal predictions (section 1.2), we explain in more detail the present status of the research about AMOC in decadal models (section 1.2.1). Based on these considerations, an overview of this study is provided in section 1.3.

1.1 The Atlantic Meridional Overturning Circulation

The Meridional Overturning Circulation (MOC) is the oceanic vertical circulation associated with sinking of waters mostly in the subpolar and polar regions and a large-scale upwelling elsewhere. These two vertical motions are connected through meridional flows, a northward one in the surface layer and the other southward below the main thermocline, referred to as abyssal circulation. Thus, in a first approximation the MOC can be simply seen as an interhemispheric cell: waters become denser and sink in the Northern polar regions and upwell in the Southern Ocean, with the motion in-between following approximately the isopycnals (Vallis, 2019). In reality, it has a more complex structure, as shown in figure 1.1. The drivers of the circulation are heat and water exchange, i.e. temperature and salinity gradients, at the air-sea interface, but many measurements have suggested that also wind-induced Ekman forcing is involved and that it has a fundamental role in setting surface waters in motion (Kuhlbrodt et al., 2007). Each ocean basin has its own MOC and all of them are connected in a global circulation pattern through the Antarctic Circumpolar Current (ACC) (figure 1.2). Despite this, sinking, also known as *deep water formation*, takes place only in a few areas in the world: the

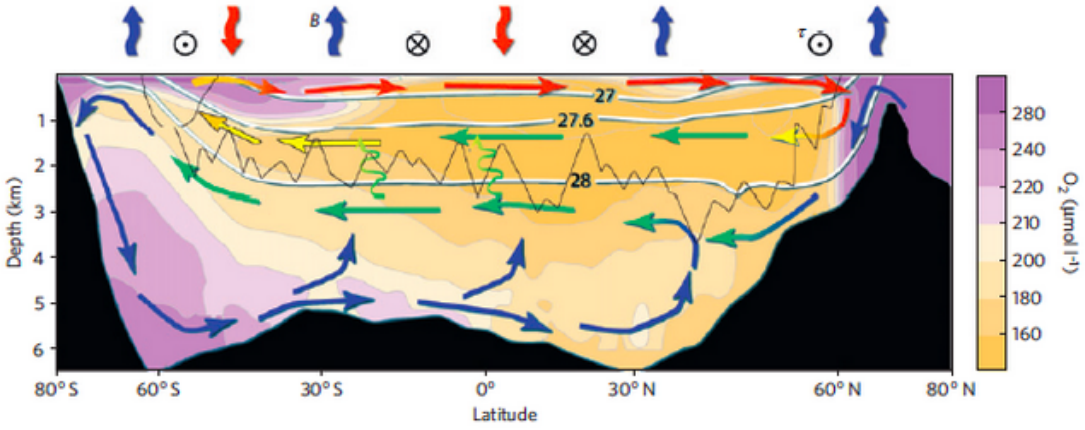

Figure 1.1: Schematic of Upper and Lower Cell of the global MOC (figure from Marshall and Speer, 2012). Thin arrows show the relative density of water masses: surface waters (red), upper deep waters (yellow), deep waters (green), bottom waters (blue). It can be noticed that arrows follow the isopycnals (white lines). The thin black line indicates the depth of the Mid-Atlantic Ridge and Scotia Ridge in the Southern Ocean. B: air-sea surface buoyancy flux, $\boldsymbol{\tau}$: wind stress. Zonally averaged oxygen distribution is superimposed, purple are more ventilated (and hence recent) waters, yellows are low-oxygen older waters.

Weddell and Ross Sea, the Mediterranean Sea and, primarily, in the Nordic and Labrador Seas in the North Atlantic. In the latter is produced the North Atlantic Deep Water (NADW).

The Atlantic Meridional Overturning Circulation (AMOC) often is schematized as an upper northward flow of warm waters and a deeper southward flow returning the cold NADW. Actually, looking at the true structure of AMOC in figures 1.1 and 1.2 we can recognize three levels characterized by peculiar properties.

In the surface layer of the North Atlantic, the northward-flowing waters belong to the Gulf Stream, which separates from the coast at Cape Hatteras and becomes a meandering jet to the northeast. This warm water mixes with the Labrador Current and forms the North Atlantic Current (NAC) east of the Grand Banks (Buckley and Marshall, 2016). When the NAC reaches the subpolar region, it enters the Nordic Seas and interacts with the East Greenland Current, creating a complex system of streams. Warm Atlantic water mixes with fresh polar water and exchanges heat with the atmosphere: the NAC waters undergo a huge heat loss, releasing heat to the air above. This cooling increases the surface water density, leading to buoyancy loss and NADW formation. Waters coming from the Gulf stream system thus cool and sink in the central Nordic Seas between Greenland, Iceland and Norway; the other fundamental region of NADW formation is the central Labrador Sea, where the cold arctic waters arrive, transported by the West Greenland Current (Buckley and Marshall, 2016).

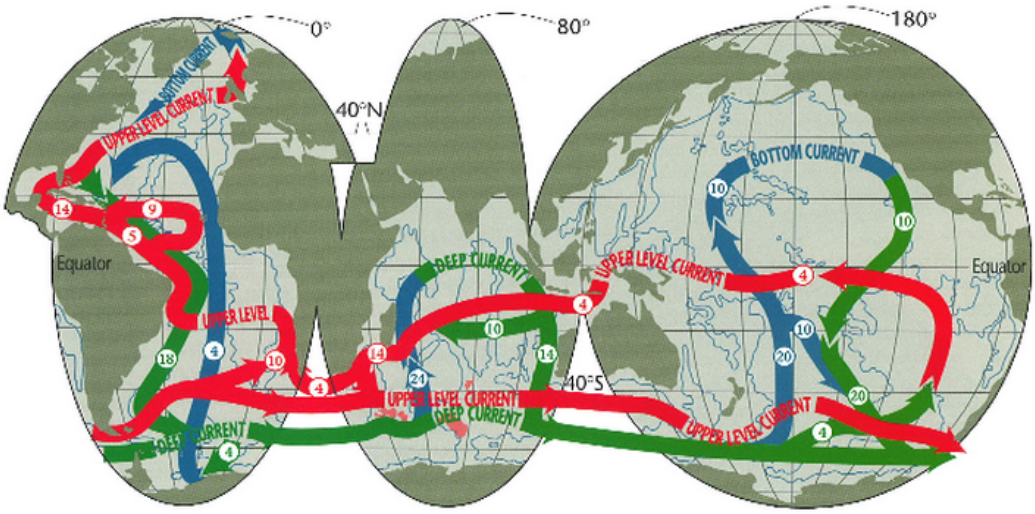

Figure 1.2: Schematic of a three-layer conveyor belt across the main oceanic basins (figure from Schmitz and William, 1995). Red lines are relative to upper layers, blue denote the deepest bottom currents and green represent an intermediate layer of transport. The numbers in circles are the mean values of transport in Sverdrups.

The main Atlantic basin is separated from the Nordic Seas by a shallow ridge from Greenland to Scotland. NADW flows across this ridge by three main routes: through the Denmark Strait, across the broad Shetland Faroe Ridge, and through the deeper Scotland Faroe Channel. The bottom branches of AMOC thus are fed by the *Nordic Seas overflow*, which plays an important -but not yet fully understood- role in the circulation. As it flows down from the North Atlantic Ridge, the Nordic overflow water mixes with Atlantic intermediate water (Eldevik et al., 2009). These deep waters flow southwards mainly in the Deep Western Boundary Current (DWBC), at depths of 2-4000 m. The DWBC is the western intensification of the intermediate AMOC branch. Near 26°N the majority of southward-flowing deep waters rejoin the DWBC. At 8°S the DWBC breaks up and, southward of this latitude, the transport of NADW is accomplished by migrating eddies. Deep waters are exported from the South Atlantic, and the upwelling of dense waters occurs primarily along isopycnals in the Southern Ocean, drawn up by strong winds blowing around Antarctica. An inflow of surface waters into the South Atlantic is needed to balance the export of NADW: waters from the Pacific enter through Drake Passage and waters from the Indian Ocean enter through cape Agulhas (Buckley and Marshall, 2016).

In the Southern Ocean, and more precisely in the Weddel Sea, takes place the formation of the second most important deep waters in the world after NADW: the Antarctic Bottom Water (ABW). Waters belonging to the Antarctic Circumpolar Current (ACC) sink and the ABW is exported northward. Two counter-rotating

meridional cells dominate the overturning circulation and represent distinct circulation regimes. The upper cell is fed both from the northern Atlantic and from below by deep diapycnal upwelling. Hence, a key part of AMOC closure is the return path which connects the interior ocean to the surface through upwelling, induced by the strong, persistent westerly winds that blow over the Southern Ocean. It has become ever more clear that, for our understanding of climate, this upwelling is critical at least as North Atlantic downwelling, because it controls how ocean reservoirs of heat and carbon communicate with the surface (Marshall and Speer, 2012).

Role of AMOC in local and global climate

The AMOC is responsible for the northward transport of a huge heat amount (1.3 PW), with a peak in the tropics at about 25°N. Its contribution is so important that it accounts for about 30% of the total poleward heat transport of atmosphere and ocean at 30°N (Johns et al., 2011). For this reason, north of 40°N the Atlantic thermocline is warmer by about 3°C than the Pacific. The AMOC heat transport has an important role in shifting the Intertropical Convergence Zone (ITCZ) north of the equator, because in this way the atmosphere can compensates the ocean heat transport across the equator with a southern contribution (Yu and Pritchard, 2019).

It is now evident that the AMOC is related to the Atlantic Multidecadal Variability (AMV), an oscillation in the South/North Atlantic bipolar temperature pattern, and hence to the atmospheric conditions over the Atlantic and the surrounding areas. The AMV is related to a range of climate phenomena: for example, it impacts on summer climate over Europe and North America, Sahel summer monsoon rainfall and Atlantic hurricanes (Zhang et al., 2019).

An atmospheric variability mode strictly related to the AMOC is the North Atlantic Oscillation (NAO), which is a fluctuation in the difference of atmospheric pressure at sea level between the Icelandic Low and the Azores High. A positive NAO phase in winter is associated with stronger westerlies and increased ocean heat loss at high latitudes. Hence, while the immediate effect of a positive phase is a reduction in sea surface temperature, the long-term effect of consecutive winter positive phases is an increased production of Labrador Sea Water and a strengthening of MOC and ocean heat transport (OHT), thus resulting in increased temperatures (Ortega et al., 2017). For negative NAO phases the opposite is true. It has been demonstrated by some case studies, such as the AMOC shift in the '90s or the cold blob that formed in the North Atlantic in 2015, that the AMOC state depends on a confluence of positive and negative NAO phases at different time scales (e.g. Robson et al., 2012b; Yeager et al., 2016). Hence, a fraction of AMOC decadal variability can also be inferred by the NAO history. A relevant example was the AMOC decline in winter 2009-2010. The reduction was much greater than the range of interannual change found in climate models used for the IPCC assessments, and that coincided with a negative NAO phase. This event led to an extreme sea level rise at the eastern coast of the US and was associated with a freezing winter in

Western Europe (Bryden et al., 2014).

Of course the AMOC has an impact on -and is impacted by- the North Atlantic climate; however, it has a role also in the global temperature shift. As a first thing, it mediates the transport of anthropogenic heat and carbon signals from the surface to the deep ocean, delaying the surface temperature response and regulating the effective ocean heat capacity under global warming (Kostov et al., 2014). Second, due to anthropogenic causes, AMOC strength has been found to have a negative long-term trend. The rising of global temperatures causes ice melting in the Arctic, and thus a release of fresh (and less dense) water in a critical region for deep water formation. Moreover, higher temperatures cause also a reduction in ocean heat transfer to the atmosphere (Buckley and Marshall, 2016). For these reasons, the AMOC is expected to slow down with associated reduction of OHT and therefore with cooling of the Atlantic subpolar gyre, which has been called the "warming hole". Caesar et al. (2021) have found from a set of proxy records that in the last decades the AMOC state has reached the weakest state of the last millennium.

Experimental measures: the RAPID array

Due to its large scale and its structural complexity, directly measuring the AMOC strength represents a hard experimental challenge. The AMOC state in the past is reconstructed from sparse observations of density, temperature and salinity. The first direct measures have been taken only since 2004, thanks to the RAPID array project located at 26.5°N, that is the latitude of maximum estimated water transport. In 2014 another experiment, the Overturning in the Subpolar North Atlantic Program array (OSNAP, 2016), has begun to take measures in the subpolar region. The conjunction of these two experiments has begun to provide a comprehensive understanding of the AMOC structure and of what drives its variability.

The first RAPID Program lasted from 2002 to 2008 providing the first observations and an unprecedented insight, then measures continued with RAPID-WATCH (2008-2014) and RAPID-AMOC (2014-2020), extending the data sets further. Transport time series now cover the period from April 2004 to September 2018, with values taken twice a day.

RAPID is made out of three interlinked experiments, each one of them studying a particular component of the total overturning circulation (figures 1.3 and 1.4) (Rayner et al., 2011).

- The southward branch of the AMOC is measured by nineteen moorings, deployed in 2004 across the Atlantic basin; these reconstruct the baroclinic circulation related to the difference in sea water density between the American and African continents, so what is called *upper mid-ocean transport* (Hirschi et al., 2003). Barotropic fluctuations are given using bottom pressure recorders.

- Additional moorings have been set on the western and eastern boundaries and on either side of the Mid-Atlantic Ridge, in the framework of the U.S.

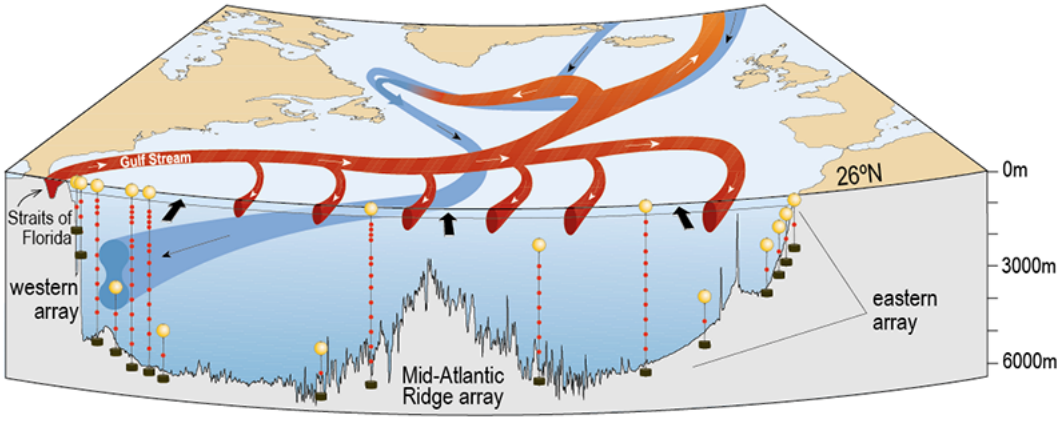

Figure 1.3: Schematic of Rapid array, with respect to the Gulf Stream and the Deep Western-Boundary Current.

Meridional Overturning Circulation and Heatflux Array (MOCHA) project, and allow surface-to-bottom density profiles to be monitored. As a result, the transatlantic pressure gradient can be continuously measured, contributing to reconstruct the upper mid-ocean transport. On the western boundary the Deep Western Boundary Current (DWBC) is detected with direct velocity measures (Johns et al., 2005).

- The northward branch of the MOC is observed using submarine cables in the Florida Straits. In this way is measured the transport due to the Gulf Stream.

A part of transport is forced by the interaction between zonal wind stress and ocean surface. Thus, measures are integrated with wind estimates provided by ship and satellites, helping to reconstruct also the Ekman induced flow. As shown in figure 1.4, the Ekman component at 26.5°N contributes with just 2-4 Sv to northward transport; despite this, it is responsible for the largest short-term (less then 100 days) variability (Kanzow et al., 2010).

Some important results have been discovered since RAPID first measurements. For example, there is much more AMOC variability over a year than previously expected from ship-based observations, with a marked seasonal cycle (Srokosz and Bryden, 2015). Furthermore, over the period of measures the AMOC has been declining at a rate ten times as fast as predicted by climate models. To say whether this is an ongoing trend caused by global warming or it is related to the AMV, we must wait for other observations (Moat et al., 2020).

1.2 The emergence of Decadal Predictions

Decadal climate predictions are a fairly recent but rapidly evolving field of research. Since the late '90s the rate of extreme events increased sufficiently to draw the atten-

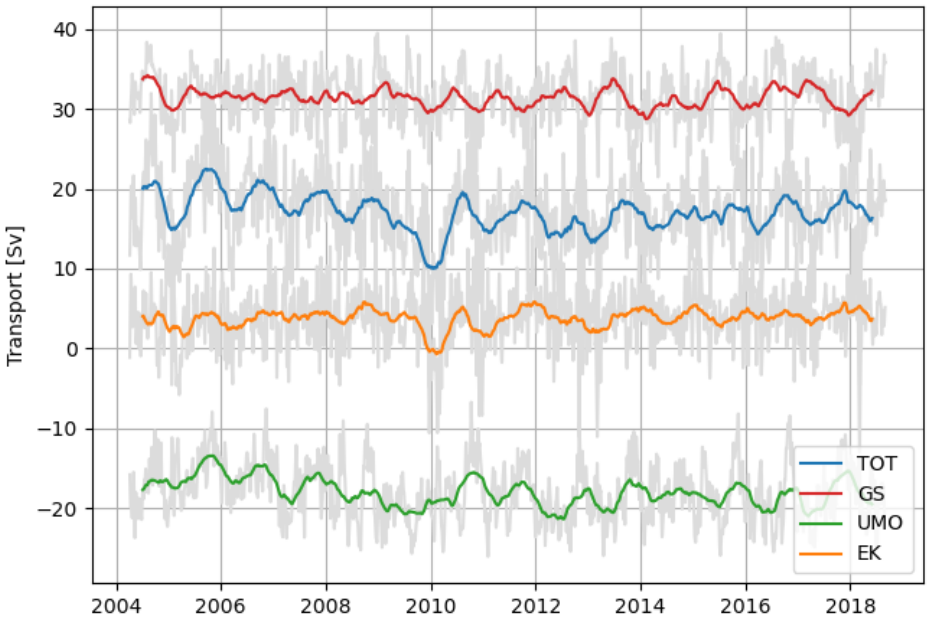

Figure 1.4: Different components of transport measured by the RAPID array, smoothed with a 6-months running mean. Transport is divided in Gulf Stream (GS, red), Upper Mid-Ocean (UMO, green) and Ekman induced flow (EK, yellow). Total transport is obtained from the sum of the three (blue). Raw data are superimposed (grey). Positive transports represent a northward flow. Time series were downloaded from the RAPID website.

tion of the global community: among the others, prolonged drought in the American Southwest, increased hurricane activity in the tropical Atlantic and changes in fisheries regimes happened. This caused concern about climate change on timescales of 10-30 years. Governments and business sectors now need to consider the social, economic and environmental impacts that these changes may have in the near term (Meehl et al., 2009). It is now clear that climate change from anthropogenic greenhouse gas forcing is happening and the rate of shift is higher than anything seen in nature in the past 10,000 years (Guiot and Cramer, 2016). However, this does not mean that future changes will be the same in any place around the globe: anthropogenic forcing is not the only component as it acts together with intrinsic climate variability and natural external forcings, and climate evolution arises from their combination. Given no possibility to foresee the third component, due e.g. to volcanic eruptions or the solar cycle, we must focus our efforts to address the other two. In particular, on the interannual-to-decadal timescale, regional variations in climate patterns and their related impacts are strongly influenced by natural variability (Hurrell et al., 2010; Smith et al., 2018).

Until the beginning of the XXI century the climate information available was of two types. By one side there were short-term seasonal-to-interannual predictions given by initialized models, provided by the National Meteorological and Hydrological Services. By the other, long-term multidecadal projections based on plausible radiative forcing scenarios. Since the end of the '00s, several centers from all around the world have initiated services which could provide predictions between the seasonal and the multidecadal scale, that is precisely decadal predictions (Smith et al., 2007; Keenlyside et al., 2008; Pohlmann et al., 2009). The first systems have been established in the United States from the National Oceanic and Atmospheric Administration (NOAA), in Germany with the Climate Service Center (CSC), in the United Kingdom with the Climate Impacts Programme (UKCP) and in Italy among the Euromediterranean Center for Climate Change (CMCC) (Meehl et al., 2009).

The World Climate Research Programme (WCRP) endorsed the Coupled Model Intercomparison Project version 5 (CMIP5) to, among the other goals, explore the ability of models to predict climate on time horizons of about a decade (Taylor et al., 2012). Near-Term Climate Prediction is consolidated as one of the WCRP Grand Challenges: the objectives it sets are to improve the quality of decadal information, to collect and synthesize prediction output, and to develop processes to assess and communicate the degree of confidence in the predictions[1]. Actually CMIP has reached version 6 (CMIP6) and in this framework was born the Decadal Climate Prediction Project (DCPP) (Boer et al., 2016). The DCPP Panel is under the guidance of the Working Group on Subseasonal to Interdecadal Prediction (WGSIP) and is organized into three components:

(A) *Hindcasts*: the organization of a coordinated experiment of retrospective fore-

[1]From the WCRP Grand Challenge on Near-Term Climate Prediction website.

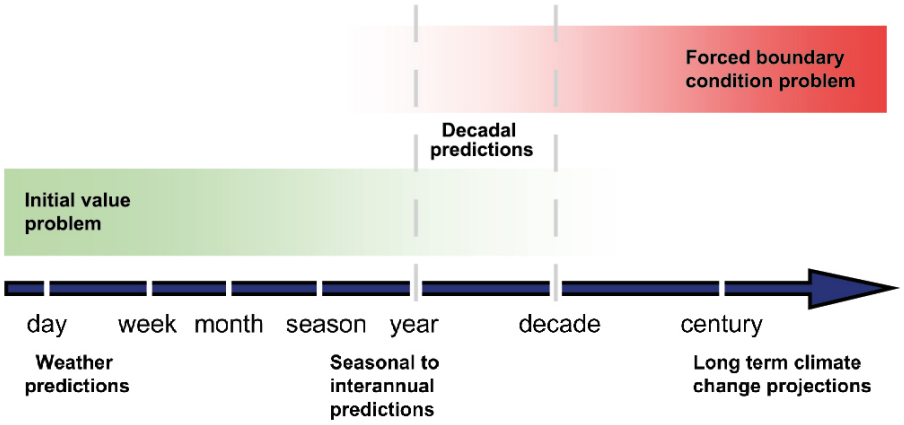

Figure 1.5: Schematic to illustrate the relation between predictions and projections at different timescales, from the daily one of weather forecasts to the multidecadal one of climate projections.

casts production, in conjunction with seasonal-prediction and climate modelling communities, and the collection of results for research and applications;

(B) *Forecasts*: the production of operational decadal climate predictions in support of multimodel annual-to-decadal forecasting and the application of the results to societal needs;

(C) *Predictability, mechanisms and case studies*: the coordination of predictability studies and of case studies of particular climate shifts, and the study of the mechanisms that cause these behaviours.

This large coordinated effort is aimed at understanding if and to which extent decadal predictions are sufficiently reliable to serve as a basis for future decisions. To do so, one of the question that must be answered is whether initial conditions play a significant role in providing more skill than uninitialized projections. The multimodel ensemble produced under DCPP thus represents a major contribution in advancing our understanding of the problem and in making available this coordinated set of initialized predictions to the international science community. Furthermore, analyses have shown that multimodel average outperforms most single models in predictive skill (Smith et al., 2012; Bellucci et al., 2015b). The development of decadal predictions, however, involves not only questions related to the skill of the predictions themselves, but also the science questions that could lead to a better understanding of processes and modeling of climate variability, sensitivity, and predictions at other time scales.

Let us now fix the terminology in a more rigorous way. A *decadal climate prediction* or, equivalently, *decadal forecast* provides information about the future evolution of regional climate from the output of a numerical model, that has been initialized with observations and runs with multiple ensemble members on timescales

of 10 years. This could be done either with a single model or a multimodel ensemble (Meehl et al., 2014). *Climate projections* can be considered "boundary-condition problems". They have a multidecadal timescale and they only provide an estimate of the mean climatic anomalies from today. This estimate depends on the external forcing inherent to a particular emission scenario. Hence, projections are started from randomly selected preindustrial states, and the variability inherent to natural climate change is not synchronized with observations. For this reason, the ensemble generation is a fundamental feature: multimember ensembles allow to average the natural variability out, leaving only a probabilistic estimate of the mean anomalies (Meehl et al., 2007). The difference of projections with *numerical weather forecasts* is that the latter are initialized and their aim is to track the time evolution of individual weather variables. The same can be done for seasonal forecasts, i.e. climate predictions at short time scale, selecting more slowly varying systems. Differently from projections, these can be considered as "initial value problems" (Kalnay, 2009). Decadal predictions rely on both principles: they are initialized from observations, so they track the time evolution of the climatic system, but in combination with forcings and thus exploiting the probabilistic features of the ensemble (see the schematic in figure 1.5). Thus, initialization offers the potential to predict internal variability in addition to externally forced climate change. We will use often the terms *hindcasts* or *retrospective forecasts*, which are decadal predictions initialized in the past.

Predictability, forecast skill and sources of uncertainty

Although the climate system is essentially deterministic in nature, it strongly depends on initial conditions, resulting in a chaotic system. The natural variability of the atmosphere leads to a loss of predictability for deterministic forecasting in a matter of days. In order to predict climate thus we have to leverage on long-timescales "signals", on which the "noise" generated by the atmosphere and other components is superimposed. *Potential predictability* can be seen as the fraction of long-term variability that may be disentangled from noise, thus making the signal usable for forecasting (Boer, 2004). Classical predictability studies perform ensemble experiments in which initial conditions of a coupled model are perturbed, then, the predictability of a variable is given by the ratio of the signal variance to the ensemble (noise) variance (Latif et al., 2006).

The premise for near-term climate predictions hence is that the coupled climate system contains elements and interactions that are predictable on the timescales at interest. An external source of potential predictability are the components of anthropogenic forcing that are also essential to century-timescale projections (Kushnir et al., 2019). On the other side, ocean and coupled atmosphere-ocean systems represent the main internal sources for long term signals, because of the thermal inertia provided by water masses (Cassou et al., 2018). The oceanic climate modes that are most significant at decadal timescales are the Atlantic Multidecadal Variabil-

ity (AMV) in the North Atlantic, related to the Atlantic Meridional Overturning Circulation (AMOC), and the Pacific Decadal Variability (PDV) in the tropical and extratropical Pacific, related to El Niño-Southern Oscillation (ENSO). Coherently with that, the first classical predictability studies identified few oceanic regions where predictability exist at decadal scale, and the most prominent one is the North Atlantic (Latif et al., 2006). Further prediction experiments have confirmed this statement, attributing predictability to the presence of the AMV and the AMOC (Yeager and Robson, 2017).

The *forecast skill* is intended as a measure of how much the predicted state is coherent with the observed one: hence, it quantifies the prediction reliability. Predictability is the theoretical limit for predictive skill in optimum conditions: we can represent it as the climate system "ability to be predicted", both in models and real world, while the forecast skill can be more described as the model's "ability to predict" a certain quantity. We can evaluate the forecasts skill by comparing the probability of an event predicted by retrospective forecasts with that of past observations: a forecast is reliable enough if the prediction probability for an event is the same as the observed frequency of the occurrence (Kumar, 2007). The most widely used metric for this purpose is the correlation between forecasts and observations. There is today enough knowledge to believe that decadal coupled model predictions are able to provide skilful information on the future climate evolution, especially regarding surface air temperature and, to some extent, precipitation and extreme events such as tropical storms or heatwaves (Kushnir et al., 2019).

Decadal forecast skill is frequently compared with that of uninitialized climate simulations. Both are intended to capture the forced response to changing atmospheric composition, but only initialized decadal hindcasts carry potential information on the time evolution of internally generated climate variability (Meehl et al., 2014). Branstator and Teng (2012), using an initialized multi-model ensemble, have found increased skill up to a decade in North Atlantic upper-ocean temperatures: from experiments like that, it is now clear that initialization benefits the simulations skill especially in oceanic regions.

The problem of predictability is linked to the *signal-to-noise ratio*, namely, the extent to which predictable signal variations could rise above the noise produced from uncertainties in the forced response and from unpredictable aspects of internal variability (Meehl et al., 2009). In climate predictions indeed, different sources of uncertainty can be identified (Hawkings and Sutton, 2009):

1. *Internal variability of the climate system*: it is the natural climate fluctuation that arises in the absence of any radiative forcing. It includes processes intrinsic to the ocean, the atmosphere, and the coupled ocean-atmosphere system. The uncertainty arises due to errors in the estimate of the initial conditions.

2. *Model uncertainty* (or *response uncertainty*): models have different structures and parametrizations to mathematically represent the process at interest and

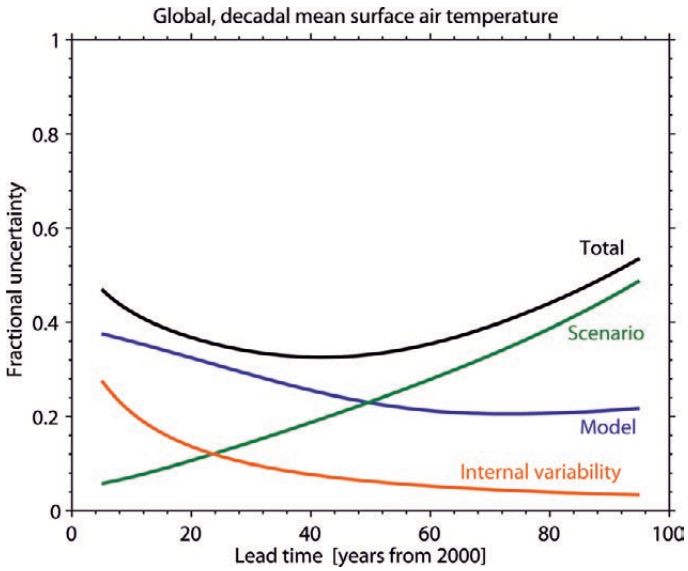

Figure 1.6: Relative importance of error sources at different time scales in IPCC GCM projections of decadal-mean global-mean surface air temperature in the XXI century (figure from Hawkings and Sutton, 2009). What is shown is the fractional uncertainty, i.e. prediction uncertainty divided by the expected mean change, relative to the 1971-2000 mean.

thus, in response to the same radiative forcing, simulate different changes in climate. Moreover, numerical errors must be considered.

3. *Scenario uncertainty*: uncertainty in future emissions of greenhouse gases, for example, causes uncertainty in future radiative forcing and climate change.

Hawkings and Sutton (2009) show that the relative importance of the sources of uncertainty varies with prediction lead time and with the spatial and temporal scale of the simulation (see figure 1.6). Over a timescale of many decades, the dominant sources of uncertainty at regional or larger spatial scales are model and scenario uncertainty. For timescales of a decade or two, model uncertainty increases and becomes the dominant source of uncertainty together with internal variability. As we expect, the importance of internal variability increases as spatial and time scales decrease. In general, the potential to reduce the total uncertainty is associated entirely with the contributions from internal variability and model error: clearly, we cannot have better estimates on future scenarios, but this is not completely a disadvantage because, as said above, in decadal predictions the scenario error contributes only with a little weight on the total. We cannot reduce the contribution from internal variability far ahead but proper initialization of climate predictions with observational data enables some reduction of this contribution for forecasts of the next decade or so. The model uncertainty, instead, has the potential to be reduced at all lead times. The generation of an ensemble of forecasts starting

from a set of similar initial conditions allows to account for the uncertainties associated with initial state and model errors. The ensemble spread is related to the forecast uncertainty, and increasing the ensemble size it is possible to reduce the unpredictable noise (Scaife and Smith, 2018).

Finally, it is important to say that the choice of the initialization method is not a secondary matter, because it is related to model biases and systematic errors, and thus may affect the forecast skill in different ways. Different techniques have been proposed to date but we can group them in two main classes. A possibility is the *full-field initialization*, in which the model state is set close to observations. This technique has the problem that the model drifts towards its systematic error state during the prediction (Yeager et al., 2018; Bilbao et al., 2020). We can minimize the drift with the *anomaly initialization* method, in which the model deviates from its preferred climatology only by the anomalous observed variability. This technique however may lead to mismatches between the observational anomalies and the model climatology (e.g. Keenlyside et al., 2008; Hurrell et al., 2010). The AMOC is not immune from the drifts and biases due to initialization. Let us have now a deeper insight of AMOC in decadal predictions.

1.2.1 AMOC decadal variability and predictability

As we have already introduced, the North Atlantic (NA) is among the few places where the climate is considered potentially predictable at a decadal scale and there is general agreement that this is due to the presence of the Atlantic Meridional Overturning Circulation (AMOC) (e.g. Latif et al., 2006). Bjerknes (1964) indeed advanced the hypothesis that, in the Atlantic sector, whereas the atmosphere drives the ocean at interannual time scales, at decadal scale it is the ocean that drives the atmosphere variability, and many observations agree with this statement (e.g. Delworth et al., 1993). It has been extensively demonstrated that initialization improves temperature predictions in the NA at decadal timescale (Keenlyside et al., 2008; Pohlmann et al., 2009; Doblas-Reyes et al., 2013; Ham et al., 2014). This is expected to be related to skilful predictions, and thus initialization, of the AMOC, because of the link between this circulation and temperatures associated to the Atlantic Multidecadal Variability (AMV) pattern (e.g. Delworth et al., 2013). Anyway, there are still considerable issues in forecasting the AMOC state even a few years in advance. In this section we will provide a picture of the main possibilities and challenges and of the present status of the problem.

Assessing the AMOC decadal predictability

Predicting AMOC evolution is far from straightforward because its variations are not regular and, overall, they are typical of a chaotic system -which means that the circulation state is highly sensitive on initial conditions (Strogatz, 1994). Therefore, before any attempt in predicting the AMOC, the first question that must

be answered is: does AMOC decadal predictability exist? That is: which are the fundamental limits in predicting AMOC at decadal scale?

In order to answer these questions, Sévellec and Sinha (2018) use two conceptual frameworks: on one side we can represent synoptic atmospheric variability with stochastic chaos, on the other, internal variability with deterministic chaos. In real world, both types of chaos are active and they affect each other. Using these conceptual models they demonstrate that internal and external processes are both sources of uncertainty and as such they are a limitation to predictability.

Techniques to theoretically estimate the error growth have led to an estimate of AMOC predictability time of 10-30 years. Besides theoretical investigation it is possible to use more pragmatic approaches, such as estimates of uncertainty in ensemble experiments. In these cases often the underlying assumption is that of a "perfect model", i.e. that the uncertainty only depends on initial conditions and there is no error due to the model or the emission scenario. In ensemble experiments the uncertainty can be estimated sampling the spread of the simulations over the forecast time, so that using a perfect model approach it is possible to study the relation between the ensemble spread and the error linked to initial state (e.g. Meehl et al., 2009). There is a large body of ensemble experiments suggesting the potential decadal predictability of the AMOC, but little quantitative agreement among them, ranging from 5 years to even 50 years under certain conditions (Latif et al., 2006; Sévellec and Sinha, 2018). With the current computational sources it is not easy to go beyond a qualitative evaluation, because, to have an accurate estimate, ensembles composed by much more then 10 members would be required. Multi-model analyses highlighted that the subpolar NA is the region of highest and most robust potential predictability, in terms of sea surface temperatures and ocean circulation (Boer, 2004).

In addition to presenting the factors of error and giving an estimate of the uncertainty, we can also look at the factors contributing to predictability. A first source of predictability are the oceanic internal modes of variability. Since there are still short direct observations of AMOC -at least, not at a multidecadal scale- efforts have been concentrated on the climate variability modes in the Atlantic which may be linked to AMOC: as we have already underlined, the most evident is the AMV. The problem is that the AMV is a mode of variability that shows important differences across models in terms of periodicity, amplitude, spatial structure and climate footprints (Zhang et al., 2019).

Besides the modes of variability we can also recognize some AMOC precursor factors. It is important to distinguish precursors which emerge just from a statistical analysis from those which effectively identify a cause-effect relation. For the former there may not be a causal link: an example have been identified in volcanic eruptions, anticipating AMOC changes by 15 years (Swingedouw et al., 2015). Precursors in a dynamical sense include at interannual scale the wind stress over the subtropical gyre, and at decadal scale the surface heat flux gradient across the North Atlantic Current and temperature and salinity perturbations in regions bordering

Greenland and Eastern Europe (Sévellec et al., 2008). Furthermore, changes in deep buoyancy in the Labrador Sea (LS) have gained increasing attention since the first studies which identified only a statistical relation (Robson et al., 2014). Ortega et al. (2021) have explored the role of Labrador Sea density as a precursor of western boundary density anomalies and thus of the overturning circulation. They found a clear multidecadal variability in Labrador Sea density, with positive density anomalies linked to a strengthening of the AMOC especially in the subpolar region. The models considered in the study show some limitations and an uncertainty remains, but the chain of relationships pictured is consistent with the mechanisms proposed by Yeager and Robson (2017). The latter conclude that the initialization of water mass anomalies driven by the North Atlantic Oscillation (NAO) (and in particular LS water anomalies) is one of the keys to explain high levels of decadal predictive skill in the subpolar NA. Surely, from these studies emerges the importance of convective sites and deep oceanic regions but, in general, there is still not a solid consensus about which are the mechanisms that drive AMOC decadal variability and this is a strong limitation in initializing and predicting it.

The AMOC in current decadal models

The first works demonstrating the potential for skillful prediction of Atlantic decadal climate using suitably initialized coupled general circulation models appeared little more than a decade ago (Keenlyside et al., 2008; Pohlmann et al., 2009). Most experiments participating to CMIP5 have reported substantial improvement in the NA, and especially in the subpolar region, for up to a decade ahead compared to uninitialized simulations, despite the modest improvement in global mean temperatures (García-Serrano and Doblas-Reyes, 2012; Doblas-Reyes et al., 2013; Bellucci et al., 2015b). The improvement seems to be independent on the initialization strategy (Smith et al., 2013).

The ocean has begun to be less sparsely observed only in relatively recent times: the first start dates in CMIP5 and CMIP6 hindcasts are initialized in 1960. This makes the skill assessment for the AMOC forecasts a difficult task. Thus, more efforts have been applied in understanding specific case studies. Examples are the rapid warming in the '90s or the cooling of the '60 in the NA subpolar gyre, which have been found to be predictable in a number of studies. In particular, the ocean heat transport, and thus initialization of AMOC, played an important role (Robson et al., 2012a; Yeager et al., 2012; Robson et al., 2018). The subpolar region consistently stands out as the most improved region for upper ocean heat content and surface temperature in state-of-the-art initialized climate predictions. Furthermore, in some systems, high skill in the SPNA appears to reverberate around the Atlantic sector as improved skill in predicting climate over land in Europe, decadal Arctic winter sea ice trends, upper ocean heat content in the Nordic Seas, and Atlantic tropical cyclone frequency (Yeager and Robson, 2017).

In the late '90s case study, besides the direct role of ocean circulation, also

the NAO has been found to have an important impact on temperature prediction (Robson et al., 2012b; Yeager et al., 2012). However, interannual NAO variations are not well predicted by hindcasts. As we said in section 1.1, the NAO acts on ocean heat fluxes and buoyancy forcing (in particular on the Labrador Sea) at different time scales. The difficulties in forecasting the NAO may explain why, in the '90s, decadal models do not capture the warming with the correct speed or magnitude (Robson et al., 2012a).

Despite the increasing number of studies demonstrating the potential of the NA sector for decadal predictions, there is emerging evidence that current models do not adequately represent all of the mechanisms that give rise to real-world predictability in this region and, as a result, models are too "noisy". What is remarkable is that there is a "signal-to-noise paradox": that is, in the NA models are better at predicting observed climate than we would expect from their low signal-to-noise ratio (Scaife and Smith, 2018). This calls into question the assumption that perfect model predictability reflects an upper limit of real-world predictability. Smith et al. (2020) show that current models underestimate the predictable signal of NA atmospheric oscillations. However, despite the poor skill of raw model outputs, atmospheric and AMV decadal predictions can greatly improve if forecast ensembles are post-processed so as to extract the signal. Also, an underestimated low-frequency AMOC variability is found in CMIP models by Yan et al. (2018), affecting in turn AMV. These findings have led to the conclusion that NA climate is more predictable on decadal timescales than previously believed, because current models, underestimating AMOC and AMV, lead to overly weak linkages between their signals and degrade decadal predictability. Some important questions arise: why is the real world more predictable than climate models suggest? How and to which extent can we improve model outputs in the NA?

Further progress will require more research of Atlantic climate variability and relevant mechanisms, such as the AMV, the NAO, the LS water formation and their related impacts. It is also evident that we need a deeper understanding of the behavior of initialized coupled prediction systems themselves regarding issues such as initialization shock, drift, ensemble generation, optimal ensemble size, and sensitivity to initial conditions. However, in the upcoming years we expect progressive advancements in our ability to predict the AMOC on interannual to decadal timescales as models and initialization methods improve, as computing power increases, and as observational data sets expand.

1.3 This study

At present there are hints that AMOC variations are predictable some year or decade in advance, but models disagree in giving quantitative estimates (Sévellec and Sinha, 2018). This uncertainty also derives from a lack of consensus in the physical mechanisms that affect AMOC decadal variability, so that it is not easy to understand what should we focus on when initializing the North Atlantic.

It is documented that full-field initialized models are subject to biases and drifts from the initial state, but there seems to be a gap in literature about how and why drifts affect the AMOC. A discussion of the problem is provided by the recent work of Bilbao et al. (2020). From their study a clear correlation between the convection in the Subpolar North Atlantic and the AMOC drift emerges. Understanding which are the areas potentially related to initialization shocks could supply indications for the modelling community towards better initialization strategies, but at present the available information is poor, if not totally absent.

Moreover, there is increasing evidence that climate models tend to underestimate the magnitude of the predictable signal (Scaife and Smith, 2018). This is especially true for the Atlantic sector, including the AMOC (Yan et al., 2018). Further research is essential to understand why the signal-to-noise ratio in climatic variables is too small in the current model generation and to develop post-processing strategies that could extract as much signal as possible.

This work focuses on the AMOC simulated in the decadal forecasts system based on CMCC-CM2-SR5, the coupled-climate model developed by the Euro-Mediterranean Center for Climate Change (CMCC). We use a set of hindcasts initialized between 1960 and 2018. First, we build a general picture of how the AMOC is simulated in the model through its climatology and variability and we compare it to observational data, with particular attention to biases and drifts. Then, we assess the ability of the model to skillfully predict the real circulation. This is done by comparing the hindcasts with the RAPID dataset and, given that direct measures began in 2004, with a set of ocean reanalyses spanning over the whole period. Successively, we focus on Labrador Sea convection, trying to get more information about the origin of the drift. In the final part we investigate if there are some features in the model drift that have a particular impact on the forecast skill and we leverage on these features to experiment a new post-processing strategy, in order to obtain more reliable predictions.

Studying whether the AMOC is realistically simulated by the model, whether the forecast skill reflects the potential predictability, and which processes lead to eventual issues, may clear if there is a degradation in the AMOC signal and which may be the causes. Furthermore, today large ensembles are required to reproduce the correct signal. Hence, a more realistic initialization and a targeted post-processing strategy would also enable smaller-size ensembles to be used for prediction, with saving of computational expense.

The document is structured as follows: in Chapter 2 are presented the mathe-

matical concepts to manage the overturning circulation and deep convection. The experimental methodology is presented in Chapter 3, including technical details about decadal predictions generation, the used data sets, the verification theoretical tools and their numerical implementation. In Chapter 4 are shown the main findings about AMOC variability and predictability, with an insight in LS convection and the experimented post-processing strategy. Chapter 5 is devoted to the discussion of results. The conclusion follows in Chapter 6.

Chapter 2

Theoretical background

In this chapter we will set the theoretical framework to manage the overturning circulation in an ocean basin, and present an essential mathematical tool for the current study: the meridional streamfunction (MSF).

Let us consider a fluid element characterized by the velocity field

$$\mathbf{v} = \mathbf{v}(x, y, z, t) = u()\mathbf{i} + v\mathbf{j} + w\mathbf{k} \tag{2.1}$$

where the three components u, v and w are respectively the zonal, meridional and vertical one, and they depend on (x, y, z, t). The rate of change of a property of the fluid element is given by the operator

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \tag{2.2}$$

which is called the *total* (or *material* or *Lagrangian*) *derivative* and takes into account the fluid advection.

The fundamental equations of motion for fluid are the Navier-Stokes equations, which express conservation of momentum and mass for viscous fluids: they come from the application of Newton's second law to fluid motion, where the stress is given by the sum of a pressure term and a diffusing viscous term proportional to the velocity gradient. The equation of momentum in vectorial form is thus

$$\frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho}\nabla p - g\mathbf{k} + \nu\nabla^2\mathbf{v}, \tag{2.3}$$

where p is the pressure, ρ is the density, g the acceleration of gravity and ν the kinematic viscosity. The conservation of mass is expressed by the *continuity equation*

$$\frac{D\rho}{Dt} + \rho\nabla \cdot \mathbf{v} = 0 \tag{2.4}$$

or equivalently, for the definition of total derivative in (2.2),

$$\frac{\partial\rho}{\partial t} + \nabla \cdot (\rho\mathbf{v}) = 0. \tag{2.5}$$

The *primitive equations* of motion for the ocean come from the viscous Navier-Stokes equations in a rotating frame of reference. The system is closed with a thermal energy equation, an equation for tracers conservation and an equation of state which relates density, potential temperature (Θ) and salinity (S). If $\mathbf{\Omega}$ is the Earth's angular velocity, $\mathbf{k}$ is the direction of the rotation axis and θ is the latitude, the Coriolis term in the equations of motion is $\mathbf{f} = f\mathbf{k}$ and $f = 2\Omega sin(\theta)$ is the Coriolis parameter. The primitive equations for a stratified ocean come from the Navier-Stokes equations with the following simplifications:

1. *Incompressibility approximation*: under the approximation that the fluid is incompressible, from the continuity equation in (2.4), we get

$$\nabla \cdot \mathbf{v} = 0. \tag{2.6}$$

2. *Boussinesq approximation*: density variations in liquids are small, thus we take a first-order balance in which density does not change unless it is multiplied by gravity (g). We write density as $\rho = \rho_0 + \delta\rho(x, y, z, t)$ where ρ_0 is a constant and assume $|\delta\rho| \ll \rho_0$. Terms which contain $g\delta\rho$ should not be neglected because, even if we have small density variations, g is of the order of $\sim 10\ m/s^2$.

3. *Hydrostatic balance:* associated with density there is a reference pressure that is defined to be in hydrostatic balance with it. This means that $p = p_0(z) + \delta p(x, y, z, t)$ and, being $|\delta p| \ll p_0$, at first order we write the vertical component of the momentum equations as

$$\frac{dp_0}{dz} = -g\rho_0. \tag{2.7}$$

Hence, the primitive seawater equations for a Boussinesq fluid are:

$$\begin{cases} \dfrac{D\mathbf{v}}{Dt} + 2\mathbf{\Omega} \times \mathbf{v} \ = -\nabla\phi - \dfrac{\rho}{\rho_0} g\mathbf{k} + \nu\nabla^2\mathbf{v} \\ \nabla \cdot \mathbf{v} = 0 \\ \dfrac{D\Theta}{Dt} = k_T \nabla^2 \Theta \\ \dfrac{DS}{Dt} = k_S \nabla^2 S \\ \rho = \rho(\Theta, S, p) \end{cases} \tag{2.8}$$

where $\phi = \delta p/\rho_0$ and k_T, k_S are the thermal and tracers diffusivity. As we said, the vertical component of the momentum equation is given by the hydrostatic balance in equation (2.7).

For our purposes we will take the primitive equations written in buoyancy terms, defined as $b = -g\delta\rho/\rho_0$, as it is a more intuitive physical quantity for treating fluids vertical motions:

$$\begin{cases} \dfrac{D\mathbf{v}}{Dt} + \mathbf{f} \times \mathbf{v} = -\nabla\phi + b\mathbf{k} + \nu\nabla^2\mathbf{v} \\ \nabla \cdot \mathbf{v} = 0 \\ \dfrac{D\Theta}{Dt} = k_T\nabla^2\Theta \\ \dfrac{DS}{Dt} = k_S\nabla^2 S \\ b = b(\Theta, S, z) \end{cases} \tag{2.9}$$

The last three equations of (2.9) can be replaced with the less general form

$$\frac{Db}{Dt} = k\nabla^2 b. \tag{2.10}$$

The vertical component of the momentum equation, that is the hydrostatic balance, in buoyancy terms becomes

$$\frac{\partial\phi}{\partial z} = b. \tag{2.11}$$

The condition for the balance in (2.11) to hold is that vertical accelerations are small compared to $g\delta\rho_0$, and not compared to the gravity acceleration itself.

After this general introduction, we can focus on the overturning dynamics. One of the simplest models to study the overturning circulation is the one called *sideways convection* (Vallis, 2019). Suppose we have a fluid in a container insulated on the sides and on its bottom, so that we can transfer heat only at surface. Boundary conditions hence can be written as

$$b(x, y, 0, t) = g(x, y) \tag{2.12}$$

at the surface, where $g(x, y)$ is a generic field, and

$$\partial_n b = 0 \tag{2.13}$$

on lateral and bottom boundaries, where the subscript n stands for direction normal to the walls. Thus, the fluid exchanges heat from above, while the buoyancy flux normal to the boundary is zero. The thermal forcing in the real ocean is in part due to a radiative transfer and in part to latent and sensible heat flux from the atmosphere.

For simplicity we will describe the overturning circulation in latitude and height coordinates, so that the problem of sideways convection can be restricted to a two-dimensional case. This is a strong idealization of the real overturning circulation,

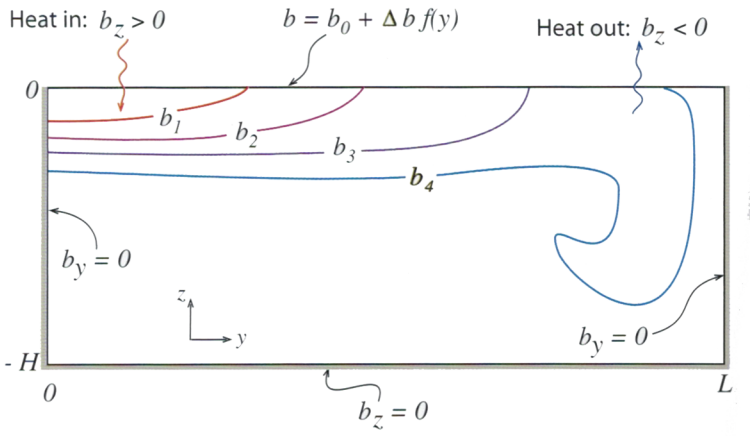

Figure 2.1: Schematic of the sideways convection. The fluid exchanges heat only at surface, while the side and bottom walls are insulated. Where the buoyancy b_z is negative, heat is extracted and the fluid sinks in convective plumes, reaching the depth scale H. Deep convection happens over smaller regions with respect to the basin horizontal scale L (figure from Vallis, 2019).

but the results do not depend on this simplification. From the incompressibility approximation in (2.6), in two dimensions we have that

$$\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{2.14}$$

therefore we can define a *meridional streamfunction*, ψ, such that

$$v = -\frac{\partial \psi}{\partial z}, \qquad w = \frac{\partial \psi}{\partial y}. \tag{2.15}$$

The vorticity ζ in the meridional plane is

$$\zeta = \nabla_x^2 \psi = \left(\frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right). \tag{2.16}$$

For how we defined the MSF in (2.15), it can be calculated at any given latitude y and depth z_0 integrating the meridional velocity $v(x, y, z, t)$ over all longitudes and over depths from z_0 to surface:

$$\psi(y, z_0, t) = \int_{z_0}^{0} \int_{x_W}^{x_E} v(x, y, z, t)\, dx\, dz \tag{2.17}$$

The integration limits x_W and x_E stand for the zonal positions of the western and eastern boundaries of the basin. The MSF represents the total volume transport of water mass in the overturning circulation (m^3/s).

We can then take the curl of the equations of motion and buoyancy diffusivity in (2.9) and (2.10), and obtain:

$$\frac{\partial \nabla^2 \psi}{\partial t} + J(\psi, \nabla^2 \psi) = \frac{\partial b}{\partial y} + \nu \nabla^4 \psi \tag{2.18a}$$

$$\frac{\partial b}{\partial t} + J(\psi, b) = k \nabla^2 b \tag{2.18b}$$

where $J(a,b) \equiv (\partial_y a)(\partial_z b) - (\partial_z a)(\partial_y b)$ is the Jacobian operator. Equations (2.18) can be non-dimensionalized to obtain an estimate of the parameters that govern the process at interest. We set

$$b = \Delta b\, \hat{b}, \qquad \psi = \Psi \hat{\psi}, \qquad y = L\hat{y}, \qquad z = H\hat{z}, \qquad t = \frac{LH}{\Psi}\hat{t}. \tag{2.19}$$

The hatted variables are non-dimensional. Δb is the difference of temperature across the surface, L and H are respectively the horizontal and vertical scale of the domain. Ψ and H are to be determined. Using these parametrizations into (2.18) and choosing Ψ as

$$\Psi = \frac{kL}{H}, \tag{2.20}$$

the equations (2.18) become

$$\frac{\partial \widehat{\nabla}^2 \hat{\psi}}{\partial \hat{t}} + \hat{J}(\hat{\psi}, \widehat{\nabla}^2 \hat{\psi}) = Ra\, \sigma \alpha^5 \frac{\partial \hat{b}}{\partial \hat{y}} + \sigma \widehat{\nabla}^4 \hat{\psi}, \tag{2.21a}$$

$$\frac{\partial \hat{b}}{\partial \hat{t}} + \hat{J}(\hat{\psi}, \hat{b}) = \widehat{\nabla}^2 \hat{b}. \tag{2.21b}$$

The three adimensional parameters that govern the behaviour of the system are

$$Ra = \Delta b L^3 / \nu k \qquad \text{the Rayleigh number,} \tag{2.22a}$$

$$\sigma = \nu / k \qquad \text{the Prandtl number,} \tag{2.22b}$$

$$\alpha = H/L \qquad \text{the aspect ratio.} \tag{2.22c}$$

The Rayleigh number provides an estimate of the strength of the buoyancy forcing with respect to the viscous term: in the ocean is very large, because it is $\sim 10^{24}$. The Prandtl number for seawater is ~ 7, and the aspect ratio in a typical ocean basin is $\sim 10^{-2} - 10^{-3}$. If we want the buoyancy term in (2.21a) to be $\mathcal{O}(1)$, using the parameters (2.22) we find that the most suitable scaling for H and Ψ are

$$H = \left(\frac{k^2 L^2}{\Delta b}\right)^{1/5} = L\,(\sigma Ra)^{-1/5}, \qquad \Psi = \left(k^3 L^3 \Delta b\right)^{1/5} = Ra^{1/5} \sigma^{-4/5} \nu. \tag{2.23}$$

This scaling is supported by numerical experiments and makes emerge some important points for the real ocean. First, the horizontal scale of the overturning

circulation is large, nearly the scale of the basin. Indeed, it is far more efficient to produce a single large overturning cell than a multitude of small ones. Second, the downwelling (convection) regions are of smaller horizontal scale than the upwelling regions. This is because high Rayleigh number convection is more efficient than diffusive upwelling. Hence, the convective buoyancy flux can only match the diffusive flux if convection occurs in a narrower region than diffusion. Let us see now in more detail the aspect of convection.

Deep convection and mixed layer

The North Atlantic deep water is produced in the convective region north of 60°N where convection produces vertical isopycnals, which connect with the flat ones in the basin region spanning from about 50°S to 60°N. Thus, in the northern region (which we denote with the subscript N) the interior flow connects to the surface through convection. The values of buoyancy at the surface, $b_N(y, z = 0)$, are represented by a flat isopycnal, while $b_B(z)$ is the buoyancy in the interior basin region. If the isopycnal connects by way of convection, the surface waters sinking down towards the level of neutral buoyancy produce vertical levels, that is, $\partial b_N/\partial z = 0$. Then, the deep waters return to flow southward in the basin region. The vertical isopycnals generates a zonal flow by thermal wind. This zonal transport is determined by the meridional gradient of temperature and the depth h to which waters convect:

$$u_N(y, z) = -\frac{1}{f}\int_{-h}^{z} \frac{\partial b_N}{\partial y} dz' + C \tag{2.24}$$

where C is a constant determined by the condition

$$\int_{-h}^{0} u_N dz = 0. \tag{2.25}$$

Following (2.17), the total volume transport is represented by the MSF:

$$\int_{x_W}^{x_E} \psi_N dx = \int_{-h}^{z}\int_{L_s}^{L_n} u_N \, dy \, dz' \tag{2.26}$$

where L_s is the latitude of the southern edge of the region (about 60°N), L_n is the northern edge. The zonal extent is thus $L_x = L_n - L_s$. If Δb is the surface buoyancy difference across the region N anf f_N the local Coriolis coefficient, inserting (2.24) in (2.26) we can have an estimate of the MSF:

$$\Psi_N = \frac{\Delta b H^2}{L_x f_N}. \tag{2.27}$$

H is a representative value of h, that is, the length scale of depth reached by convection.

The ocean in normal conditions is a stratified fluid, which means that the average density gradient on the vertical is larger than the horizontal one. When convection happens, the fluid undergoes a turbulent mixing process and stratification is much weaker. The mixed layer thickness is the range of surface depths over which stratification is weak and water properties are homogeneous. To measure the mixed layer depth is often used the σ_T criterion. The latter is based on the quantity σ_T:

$$\sigma_T = \rho(S,T) - 1000\frac{kg}{m^3}. \tag{2.28}$$

The criterion sets the mixed layer thickness as the depth at which a change in σ_T of 0.125 has occurred from the surface (Levitus, 1982).

From a phenomenological point of view, deep water formation takes place in three phases (Marshall and Schott, 1999):

(i) a preconditioning phase with increased stratification at surface. The general circulation (often cyclonic gyre-scale) predisposes some parts of a gyre to locally overturn;

(ii) a violent-mixing phase: if seawater in the surface layer is denser than the water at depth, it is convectively unstable. Deep convection occurs in localized, intense plumes (on scales of $\sim 1km$);

(iii) a sinking and spreading phase, that is, a lateral exchange between the convection site and the ambient fluid through advective processes. In the deep layers, water denser than its surrounding generates a pressure gradient, thus the horizontal domain fills with the densest available fluid.

The last two phases often occur concurrently. After deep convection the surface re-stratifies and a new cycle takes place. This process of downwelling and spreading is continuous: the surface water sinks in plumes, dense waters reach the bottom and spread generating a flow, and in this way the overturning circulation sets in motion. Even for large viscosity and diffusion, the problem has no stationary solution: the fluid begins to circulate as soon as we impose a temperature gradient at the surface, because this generates a density gradient. The latter leads to a pressure gradient, which sets waters in motion.

Chapter 3

Methods

In this chapter are presented the experimental data and methods. The main technical aspects of decadal forecasts, such as initialization, ensemble generation, and bias and drift adjustment, are discussed in section 3.1. In section 3.2 are described the CMCC-CM2-SR5 decadal system and the data used for the predictability verification. In order to study the time variability of the physical systems at interest, some representative indicators have been chosen: they are listed in section 3.3. The actual experimental procedure with details about the chosen predictability-verification metrics is in section 3.4, followed by the relative computational strategies by which they are implemented.

3.1 Technical issues in Decadal Predictions production

Initialization

As anticipated in chapter 1, decadal predictions can be initialized with a full-field method or with an anomaly initialization method. To make a comparison, on seasonal timescales full-field initialization of regional temperature and precipitation leads to significantly more skill on average than anomaly initialization. Differences in skill on multi-year timescales are generally not significant. Anomaly initialization provides a better estimate from a limited hindcast set (Smith et al., 2013). Hence, there is no clear indication of which method is the best as both show pros and cons, and modeling groups are experimenting different strategies (Meehl et al., 2014).

The simplest framework to assimilate data (referred to as *weakly coupled*) consists in producing individual assimilation runs for the single model components (Sanchez-Gomez et al., 2016). Despite individual components may benefit from this strategy, initialisation shocks may occur due to incompatibilities among the initial conditions of ocean and atmosphere. Hence, many forecast centres are moving towards fully coupled (or *strongly coupled*) assimilation, i.e. a joint assimilation

of observations in coupled climate systems (Brune and Baehr, 2020). This is more technically challenging because there are problems such as the treatment of ocean and atmosphere data covariances or the weighting of different observed variables in the various components, but assures physical consistency of the initial conditions (Kushnir et al., 2019).

A matter of discussion is which field variables should be used for initialization. For the timescales at interest, the full ocean should be initialized as realistically as possible (Kushnir et al., 2019). There are also non-oceanic drivers of variability that would be important to extend the temporal range of predictability: primarily, sea ice, snow cover, frozen soil and soil moisture. The contribution of these constituents to the slow climatic variability is controversial in comparison to that of the large heat capacity of the oceans. For this reason, they have received relatively little attention and their initialization and representation still remains a challenge (Bellucci et al., 2015a). In particular, the treatment of sea ice and ocean conditions under sea ice could be a significant issue, but it is also important because dense water formation there have strong climate impacts (Meehl et al., 2009). Regarding the ocean state, historically the subsurface ocean has been sparsely observed and some of the data can be significantly biased (Ishii and Kimoto, 2009). A solution to avoid this difficulty could be to initialize models by assimilating only sea surface temperatures, so that the subsurface ocean is indirectly initialized through ocean transport processes (Keenlyside et al., 2008). In an alternative approach, subsurface ocean temperature and salinity fields are taken from an ocean model forced by atmospheric reanalysis based on observations, and then nudged into a coupled model to produce the initial conditions (e.g. Smith et al., 2007; Pohlmann et al., 2009). Currents can also be included in the initialization.

The direct use of subsurface ocean observations can improve predictions and the experimental progress will make this more possible is the upcoming years. Several reanalyses of historical ocean observations have been constructed to make up for the lack of experimental measures in the past. A range of techniques have been developed to produce the reconstructions, from classical nudging approaches based on Newtonian relaxation to more complex and computationally expensive methods like the Ensemble Kalman Filter approach that take into account aspects of the observational uncertainty (Bilbao et al., 2020). The Climate Variability and Predictability (CLIVAR) Global Synthesis and Observations Panel (GSOP) intercomparison project has been created to collect all the reanalyses, to help understand the differences among these products and to provide insight into why they may disagree (GSOP, 2019). In this way, also, modeling groups without data assimilation schemes have the possibility to perform initialized climate predictions.

Ensemble generation

The ensemble is generated by varying the initial state with small perturbations. Climate at decadal scale is sensitive to these perturbation, so the ensemble members

will spread as run time goes by. The ensemble generation is strictly related to the problem of uncertainties because through the ensemble it is possible to sample the spread of possible outcomes, thus the model error that is consistent with the uncertainty of the initial state. It is thus clear that the way in which we design the ensemble and set the initial conditions is an important issue (Sévellec and Sinha, 2018).

Different strategies are possible to perturb the initial state. Some methods focus on the atmospheric state, adding random perturbations to it or selecting for it different start days around the initialization time (e.g. Yeager et al., 2012). However, some studies highlight the role of uncertainty in ocean state in ensemble experiments, since it modifies the decadal climate predictability and the divergence of the atmospheric state. This is true especially in initialization-sensitive regions such as the Subpolar North Atlantic (Kröger et al., 2018), but also perturbations of the deep ocean are involved: for example, the Atlantic overturning circulation is sensitive on deep density anomalies because of their long residence time (Germe et al., 2017). Moreover, oceanic perturbations obtained from the use of different oceanic reanalyses tend to increase the spread of oceanic variables in comparison to atmospheric-only perturbations (Du et al., 2012).

Other ensemble production methods can be the generation of optimal perturbations through breeding strategies (Ham et al., 2014) or the ensemble Kalman filter-type assimilation, in which reanalyses of observations are created by using the forecast model and observations to update an ensemble of previous reanalyses (Karspeck et al., 2013), or variations on these methods. This diversity of techniques results from the philosophical approaches to the initialization and ensemble generation problem: for this reason, which are the best methods is still an issue under debate.

Bias adjustment

Linked to the management of model errors there is the issue of bias adjustment. It is standard practice to remove biases by an a posteriori empirical correction computed from a series of hindcasts, i.e. post-processing the model output. There is now strong evidence that, for robust estimates of the bias, increasing the number of start dates is essential at least as increasing the ensemble size (Sévellec and Sinha, 2018).

In full-field initialized models, forecasts are initially set near the observed state but they drift back towards the model's preferred state. The *drift* occurs because the model attractor is different from the real world attractor. Thus, when initialized from observations, the model tends to deviate towards its equilibrium state (Sévellec and Sinha, 2018). In this way it re-establishes the bias, which is then removed with an a posteriori lead-time dependent correction diagnosed from a set of hindcasts. The model drift follows an *initialization shock*: these two must not be confused, as the former is a consequence of the latter. Initial shocks happen soon after

initialization and are due to the adjustment of the model to the constrained state, because of structural errors in the model and biases in its climatology. In such models, we can adjust bias and remove the drift by calculating a climatological average over the total period of predictions and considering the forecast anomalies obtained by subtracting the average. The same calculation is done for observations, and the anomalies are compared (Smith et al., 2013). Besides an a posteriori bias adjustment, of course it is important to reduce the systematic errors to minimize the initialization shock. With this objective, several decadal forecast centers are producing in-house assimilation experiments based on the same model components used for the simulations so that, from these experiments, initial states more suitable for the forecasts can be derived (Bilbao et al., 2020). In anomaly-initialized models, instead, there is not such a drift and bias may be removed based on the difference between independent simulations and observations (Smith et al., 2013).

Sources of error which can lead to a bias in the model include, among the other things, the inability to realistically simulate the natural modes of interannual variability, uncertain future levels of radiative forcings and imperfect experimental observations (Meehl et al., 2014).

Finally, an aspect that should be considered is that the model bias may vary not just with forecast time, but also with the model state (Sévellec and Sinha, 2018). Systematic errors in hindcasts may be removed by subtracting their average bias, but it is clear that a mean adjustment does not address issues such as trends or time dependence in the bias itself. Some studies indeed highlight the need for a correction that varies over the period considered for initialization (e.g. Bilbao et al., 2020).

3.2 Model and Data

The CMCC-CM2-SR5 System

In this work is analyzed the retrospective forecasts (hindcasts) decadal system produced with the CMCC-CM2-SR5 model. Namely, this is the CMCC Climate Model version 2 in Standard Resolution, with the atmosphere component in the CAM5.3 configuration. This experiment contributes to the Decadal Climate Prediction Project - Component A (DPCC-A).

CMCC-CM2 (Cherchi et al., 2019) is a coupled atmosphere–ocean general circulation model, based on the Community Earth System Model (CESM) project of the National Centre for Atmospheric Research (NCAR) in the United States. The difference with the NCAR coupled model is the oceanic component, since the CESM ocean component is replaced with the NEMO physical core. The model in version 2 is an evolution of the CMCC-CM model (Scoccimarro et al., 2011) used within CMIP5, which in turn has been an update of the SINTEX-G (Scale Interaction Experiment) model used for the CMIP3 effort. For CMCC-CM2, upgraded physics

of ocean, sea ice, atmosphere and land components and an advanced coupling environment have been developed compared to the previous version. The model is composed by the Community Atmosphere Model (CAM v5), the NEMO ocean engine (v3.6), the Community Land Model (CLM v4.5) and the Community sea-Ice CodE (CICE v4).

The physical core of NEMO resolves the classic primitive equations of ocean circulation subject to the hydrostatic, incompressibility and Boussinesq approximations. The model prognostic variables are the three velocity components, salinity, the potential seawater temperature and the sea surface height. The horizontal space is discretized on an Arakawa C type staggered grid, while in the vertical can be used terrain-following or geopotential coordinates, or a hybrid combination of them. NEMO offers a wide range of numerical schemes to solve critical physical processes, such as diffusion, advection, vertical dynamics, seawater equation of state, air-sea interactions and lateral and bottom boundaries. The adopted linear free-surface formulation (Roullet and Madec, 2000) allows computing fluxes of tracers, volume, and momentum. In this model configuration is used the United Nations Educational, Scientific and Cultural Organization nonlinear equation of state (Fofonoff and Millard Jr, 1983). At the ocean lateral boundaries are applied free-slip boundary conditions, while at the ocean floor are applied an intensified tidally driven mixing, a diffusive boundary layer scheme and a non-linear friction.

In CMCC-CM2-SR5, the ocean component uses the tripolar ORCA grid based on Mercator projection with a horizontal resolution of primarily 1°, with a meridional refinement down to 1/3° in the tropics. In the vertical are considered 50 geopotential levels, ranging from 1 to 400 m. For all the other components of the system is used a grid spacing of 1°, except for the River Transport Model that has a resolution of 0.5°. More information about model settings can be found in "Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model", Cherchi et al. (2019).

For this experiment a six-members ensemble has been generated. Hindcasts are initialised every year in November, from November 1960 to November 2018. Thus, every hindcast has 122 monthly values, from the November of initialization to the December of the 10th year.

In the initialization phase, a full-field technique is adopted. The ocean-sea ice initial state is selected from an ensemble of ocean synthesis differing for assimilated data, assimilation methodologies, and atmospheric forcing fluxes (Yang et al., 2016; Storto and Masina, 2016). In particular we have used the CMCC Historical Ocean Reanalysis system (CHOR) to initialize the start dates from 1960 to 2010, and CMCC Global Ocean Reanalysis system (C-GLORS) from 2011 to 2018. The different ocean states are combined with two different syntheses of the land, providing an estimate for the land surface state uncertainty. The atmosphere is initialized with the ERA40/Era-Interim reanalyses. The required initial perturbation of the three-dimensional ocean/sea ice/land state, aimed at generating the ensemble member

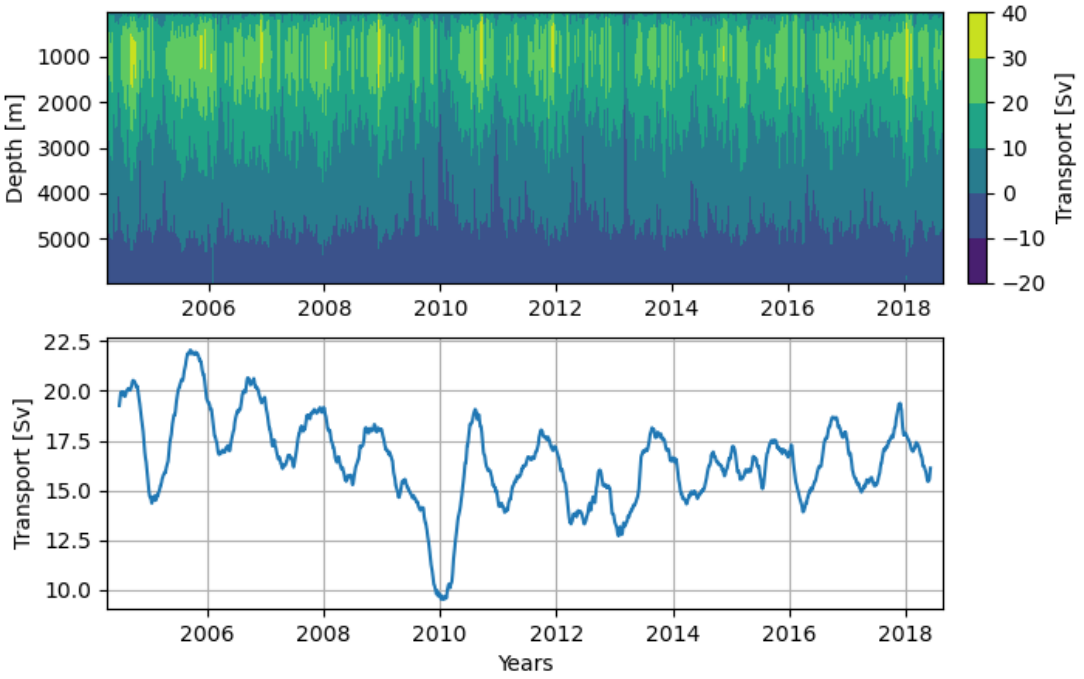

Figure 3.1: Transport measured by RAPID from 2004 to 2018 (data taken from RAPID website). **Up:** vertical profile at 26.5°N. **Bottom:** index taken as the average of transport at depths of 800-1300 m (VI26N index, see section 3.3). The index is smoothed with a six-months moving average to highlight seasonal variability.

spread, is thus provided by using the alternative ocean-sea ice and land reanalyses. Historical radiative forcing conditions are used, including green-house gases, aerosols and solar irradiance variability, followed by SSP2-4.5 scenario settings for the future.

Data for verification

Direct observations of the AMOC state began in 2004 with the RAPID-array experimental measurements. Due to the lack of measures in the greater part of the period covered by hindcasts, we have to rely on ocean syntheses. We compare our set of hindcasts with two types of reanalysis: CHOR and C-GLORS.

CHOR is the CMCC Historical Ocean Reanalysis system. Its ocean general-circulation model is NEMO3.4. The configuration is an ORCA tripolar grid and has a horizontal resolution of 0.5° (zonal) × 0.5° (meridional) and 75 vertical depth levels. Four members of the CHOR reanalyses are available at the moment. Differences among them consist in the atmospheric forcing and in the way SST data are assimilated: CHOR_AS and CHORE_AS directly ingest in-situ SST data from International Comprehensive Ocean–Atmosphere Data Set (ICOADS) in the variational

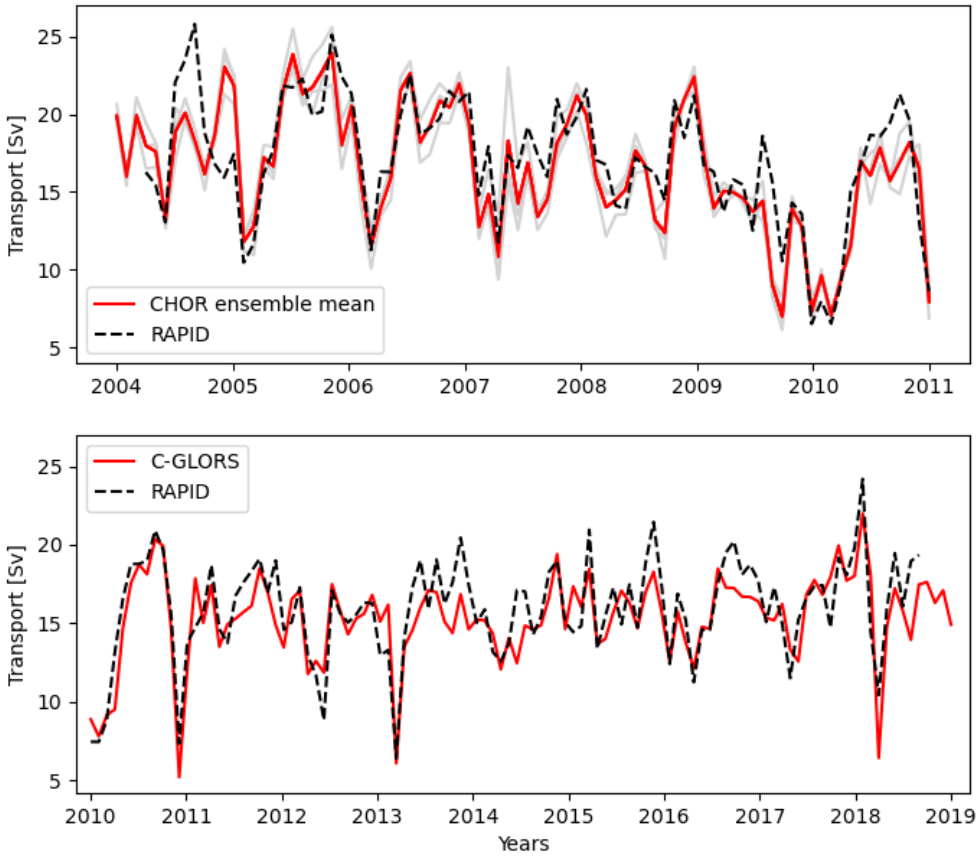

Figure 3.2: Monthly time series for RAPID (dashed line) and reanalyses, index taken as the average of transport between 800-1300m at 26.5°N (VI26N, sse section 3.3). **Up:** RAPID data from 2004 to 2010 are compared with CHOR single members (grey lines) and ensemble mean (red line). The ensemble mean has a correlation of 0.88 with RAPID. **Bottom:** C-GLORS reanalysis (red line) compared with RAPID data from 2010 to 2018, with a 0.85 correlation value between them.

data assimilation scheme; CHOR_RL and CHORE_RL implement a SST relaxation scheme towards Hadley Centre's sea ice and sea-surface temperature dataset (HadISST) analyses. CHOR_ reanalyses are forced by the ensemble mean of the Twentieth-Century Reanalysis v2 (20CRv2) from NOAA/CIRES, while CHORE_ from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-20C historical reanalysis. All of them assimilate hydrographic profiles from UK Met Office Hadley Centre series of hydrographic profile collections v4 (EN4). Relative merits of the two strategies for the SST data assimilation are discussed in detail by Yang et al., 2016.

C-GLORS, the CMCC Global Ocean Physical Reanalysis System v5, is based on NEMO3.2 configured at about 1/4° resolution on an ORCA tripolar grid, with 50 vertical depth levels (Storto and Masina, 2016). The system includes (i) a three-dimensional variational data assimilation scheme that assimilates hydrographic profiles from the UK Met Office Hadley Centre EN3 dataset (ENACT/ENSEMBLES v3) until 2012 and EN4 dataset (Good et al., 2013) afterwards, and observations provided by AVISO Satellite Altimetry Data; (ii) a nudging scheme that assimilates space-bourne SST observations supplied by NOAA (Reynolds et al., 2007) and sea-ice data; and (iii) a large-scale bias correction scheme that corrects the model tendencies to limit the large-scale biases induced by systematic errors of the NEMO model parameterizations and the atmospheric forcing. C-GLORS is forced by the ECMWF ERA-Interim atmospheric reanalysis (Dee et al., 2011).

Hindcasts start dates from 1960 to 2010 have been initialized with the three realizations of CHOR used for verification, which are CHORE_AS, CHOR_RL and CHORE_RL. We have only one realization of C-GLORS from 2010 to 2018, furthermore, it is not the one used to initialize the hindcasts, but it is consistent with the model setup and the applied forcings of the C-GLORS reanalysis used to initialize the model. For this reason, we do not expect that after 2010 hindcasts initial state coincides exactly with the reanalysis; nevertheless, we expect that this dataset is reliable enough to serve as a reference for a skill assessment. In figure 3.2 are compared the reanalyses and RAPID transports at 26°N. Reanalyses reproduce very well the seasonal variability, at least in the period covered by RAPID measures. CHOR ensemble mean has a correlation with RAPID of 0.88, while the single C-GLORS member has a correlation equal to 0.85, confirming their reliability since 2004.

Hindcasts, of course, have also been compared with RAPID-array data. Transport time series are available from April 2004 to October 2018, with data taken twice a day. The vertical profile at 26.5°N is shown in figure 3.1, upper panel. Data from the RAPID AMOC monitoring project is funded by the Natural Environment Research Council and are freely available from the website www.rapid.ac.uk/rapidmoc.

A non-initialized historical simulation, with the same model components and external forcings, has been used in order to evaluate the benefit of initialization on the forecast skill. This has been initialized in 1850 to have pre-industrial starting conditions and arrives to 2014.

3.3 Physical indicators of AMOC variability

To measure the Atlantic Meridional Overturning Circulation (AMOC) strength we use the Meridional Stream Function, which is obtained by integrating the meridional component of velocities as in eq. 2.17. Transport of water is conventionally measured in Sverdrups (1 Sv = 10^6 m^3/s).

AMOC temporal variability is monitored taking as indicator the average of transport over the area where it reaches its maximum. Looking at hindcasts climatology in figure 4.1 we can see that the area of maximum at equilibrium (i.e., in the last years of the forecast) is at 5-35°N and 800-1600 m depths. We call this variability index **VI5-35N**. This region is wide enough to represent the large-scale variability without the interference of localized effects.

When comparing data with the transport measured by the RAPID array, we take as indicator the average over a small region centered on 26.5°N, at 800-1300 m depths. Indeed, this is the region where RAPID measures the maximum transport (see fig. 3.1, top panel). We call this indicator **VI26N**.

To study the intensity of convection we evaluate the mixed layer depth (MLD). We want to focus on the Labrador Sea since it is the main region of deep water formation, so we take as indicator the MLD average over the area 55-65°N 45-60°W. We consider an average over months from February to April: this indeed is when convection is maximum. This index is called **MLD_LS**. In this experiment, the MLD is evaluated by means of the σ_T criterion.

Almost always in the diagnostics are used annual values of the forecast. When computing the average value for lead-year 1 we exclude the first two months of the forecast, i.e. the first November and December. For coherence with annual averages, the first two months are not considered also in decadal climatologies.

3.4 Verification methodology

Before going into details of the methodology we must fix our mathematical notation.

Suppose we have m forecast start dates and m annual observations in the corresponding interval of years, with $m \geq 10$. Let be $f_{(i,l),j}$ a variable of the forecast for a start date denoted by $i = 1, ..., m$, a lead-year $l = 1, ..., 10$ and an ensemble member j. This variable corresponds to a year $y_{i,l} = y_{i,l}(f_{(i,l),j})$. The same variable for observations is denoted by $o_{k,j}$, with $k = 1, ..., m$, and this corresponds to a year $y_k = y_k(o_{k,j})$. We say that the forecast value denoted by the couple (i, l) corresponds to an observation $o_{k,j}$ if

$$y_k = y_{i,l} = y_{i,1} + l - 1. \tag{3.1}$$

Clearly, $y_{i,1}$ is the year of initialization of the i-th forecast. We can write $o_{k,j}$ also as $o_{(i,l),j}$ and this means that we are considering the value of observation in the year $y_{i,l} = y_k$.

The average over the index x is denoted as $\langle . \rangle_x$. The annual ensemble mean of the forecast variable and the climatology are thus

$$F_{i,l} = \left\langle f_{(i,l),j} \right\rangle_j , \qquad C = \left\langle f_{(i,l),j} \right\rangle_{i,l,j} , \tag{3.2}$$

and for observations are respectively

$$O_k = \left\langle o_{k,j} \right\rangle_j , \qquad C_O = \left\langle o_{k,j} \right\rangle_{k,j} . \tag{3.3}$$

For the forecast we can also define a lead-year dependent climatology:

$$C_l = \left\langle f_{(i,l),j} \right\rangle_{i,j} . \tag{3.4}$$

Forecast verification metrics and statistical methods

In order to remove eventual drifts, the procedure recommended by CMIP protocols is to work in the anomaly space (CLIVAR, 2011). We assume that raw data can be decomposed into two independent signals: the mean drift and the fluctuations around the mean drift. Thus, by observing forecast anomalies with respect to the climatology, we perform a removal of the drift.

$$F'_{i,l} = F_{i,l} - C_l \tag{3.5a}$$

$$O'_k = O_k - C_O \tag{3.5b}$$

are respectively the forecast and observation annual anomalies of ensemble averages with respect to their climatologies.

Bearing that in mind, we evaluate the predictive skill of the hindcasts system with the *anomaly correlation coefficient* (ACC). This is the Pearson correlation between the anomalies of forecasts and observations and is designed to detect similarities in the patterns of departures from the climatological mean field (see Wilks, 2006, "Statistical Methods for Atmospheric Science", chapter 7). We compute a lead-year dependent ACC, thus we compare each variable in year $y_{i,l}$ with the corresponding observation in the same year. Knowing that the observation in year $y_k = y_{i,l}$ is $o_{k,j} = o_{(i,l),j}$, the lead-year dependent ACC is:

$$\mathrm{ACC}_l = \frac{\sum_{i=1}^{m} (F'_{i,l})(O'_{i,l})}{\left[\sum_{i=1}^{m} (F'_{i,l})^2 \sum_{i=1}^{m} (O'_{i,l})^2 \right]^{1/2}} . \tag{3.6}$$

The ACC is bounded between -1 and 1: indeed, it is a covariance normalized dividing it by the product of standard deviations. Eq. 3.6 is the definition for the *uncentered* ACC, i.e. in which the average is not subtracted from anomalies. Because anomalies are obtained subtracting from the initial variable its temporal

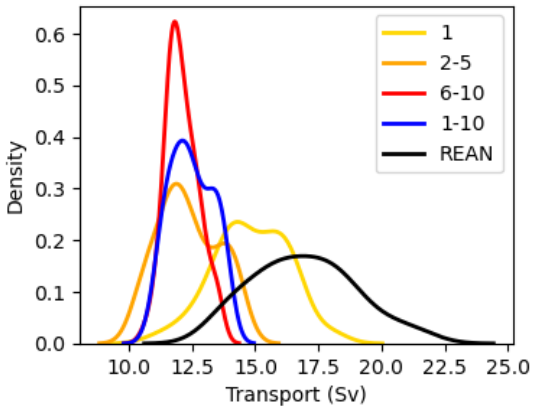

Figure 3.3: Probability Density Function of hindcasts (colored) and reanalyses (black) VI5-35N as function of some groups of lead-years.

average, anomalies mean value over time is zero. Thus, using a centered or uncentered correlation is irrelevant. This clearly would not be true if we were looking at spatial instead of time anomalies. However, as noticed by Wilks, conditional or unconditional biases are not penalized by the anomaly correlation. It is reasonable to say that the anomaly correlation reflects the potential skill that might be achieved in the absence of biases, but we cannot regard the anomaly correlation -or any correlation- as measuring actual skill.

The statistical significance of ACC is verified with a *bootstrap* procedure (Wilks, 2006). Forecast and observation time series are resampled many times to generate random correlations. The non-resampled time series have a statistically significant correlation if this falls outside the acceptance region, given the hypothesis that the obtained distribution has a mean correlation equal to zero. Following the convention, in this work we have chosen a significance level of 5%.

Let us now talk about the metrics related to an assessment of the potential predictability. For simplicity, from now on we will omit the lead-year index in metric names, although they are all defined in a lead-year dependent way. In order to choose the most suitable metrics, we have computed the probability density function (PDF) of hindcasts (at different lead-years) and reanalyses transport (figure 3.3). Based on the PDF, we have chosen the statistical metrics assuming that the error follows a normal distribution.

We define the *spread* of the forecast in year i as

$$S_i = \sqrt{\left\langle (f_{(i,l),j} - F_{i,l})^2 \right\rangle_j}. \tag{3.7}$$

The *total variance*, *signal variance* (i.e. interannual variance of the ensemble mean) and *noise variance* (i.e. mean variance around the ensemble mean or square of the

mean spread) are respectively

$$\sigma_t{}^2 = \left\langle (f_{(i,l),j} - C_l)^2 \right\rangle_{i,j}, \tag{3.8a}$$

$$\sigma_s{}^2 = \left\langle (F_{i,l} - C_l)^2 \right\rangle_{i}, \tag{3.8b}$$

$$\sigma_n{}^2 = \left\langle S_i^2 \right\rangle_{i}. \tag{3.8c}$$

It can be demonstrated that $\sigma_t{}^2 = \sigma_n{}^2 + \sigma_s{}^2$ (Boer, 2004).

If a signal exists and it raises above noise with appreciable magnitude, it is assumed to comes from physical processes which are, at least potentially, predictable. Then, *potential predictability* is defined by

$$\mathrm{PP} = \frac{\sigma_s{}^2}{\sigma_t{}^2} \tag{3.9}$$

and it represents the predictable fraction of model variance: it is 1 when climate is completely determined by predictable factors and 0 when variability is dominated by unpredictable noise (Boer, 2004). Potential predictability is not a measure of forecast skill: it just tells if prediction on certain timescales may be possible. While the ACC represents the ability of the model to predict the real world, the PP can be seen as the ability of the model to "predict itself": in a set of m forecasts, the PP is the probability of $m-1$ members to predict the m-th forecast. As such, it represents an upper bound to the skill that might be attained for prediction of this timescale component: if the model is "perfect", the PP is equal to the ACC (Boer, 2011).

Following the definition of the signal variance, the *observed interannual variance* is

$$\sigma_o{}^2 = \left\langle (O_k - C_O)^2 \right\rangle_{k}. \tag{3.10}$$

Finally, the *root-mean-square-error* is defined as

$$\mathrm{RMSE} = \sqrt{\left\langle \left[(F_{i,l} - C_l) - (O_{i,l} - C_O) \right]^2 \right\rangle_{i}}. \tag{3.11}$$

Forecast systems are said to be underdispersive/overconfident when $\sigma_n < \mathrm{RMSE}$ or overdispersive/underconfident when $\sigma_n > \mathrm{RMSE}$. These two limits are both problematic. An overdispersive system has on average an annual spread that exceeds the error between predicted and observational signals. The system ensemble mean thus may be incorrect in representing the signal of the real world: even if there is the possibility to correctly represent it, it would be a meaningless coincidence. An underdispersive system for its part has a spread that is not high enough to represent the mean distance between forecast and observation ensemble averages.

Someone has suggested that the RMSE is not a good indicator of model performance because it might be a misleading evaluation of average error, and that the mean absolute error (MAE) would be a better metric for that purpose (Willmott and Matsuura, 2005). However, Chai and Draxler (2014) have shown that the

RMSE is not ambiguous and, on the contrary, it is more appropriate than the MAE when model errors follow a normal distribution. The main disadvantage of RMSE is its sensitivity to outliers. In light of this, the choice of this metric in our case is supported by the PDF trend in figure 3.3.

Experimental procedure

First, the system's lead-year dependent forecast skill is evaluated comparing the VI5-35N index of hindcasts and reanalyses, and VI26N of hindcasts and RAPID data. When estimating the ACC between hindcasts and reanalyses, to get the anomalies we use the most extended period available for every lead-year: we consider the values from 1961 to 2018 for lead-year 1, and values from 1970 to 2018 for lead-year 10. For comparisons with RAPID, instead, we can also rely on hindcasts initialized before 2004, which is the year in which measures started. Hence, we can compute the ACC with the same number of values in every lead-year: i.e. 13, considering only annual values from 2005 to 2017 (these are the only complete years of measures).

We also give an assessment of the historical simulation forecast skill, comparing its VI5-35N index with those of the reanalyses. In this way we can understand how much the initialization affects the forecasting ability of the model. In this case the historical simulation arrives until 2014, so the climatology for lead-year 1 is taken between 1961 and 2014, while for lead-year 10 we use the period 1970-2014. We do not perform a comparison between the historical and RAPID because the period in which they overlap is too short (only 10 years).

Then, we estimate the potential predictability, variances and RMSE in a lead-year dependent way. As for the ACC, the longest available period in which are both present hindcasts and observations is considered. This is done to give an evaluation of variances that could be coherent with the RMSE, which involves both forecast and observation values. In this case we use as observations only the reanalyses, which form the longest observational dataset.

After the general assessment of the system skill and signal predictability, we experiment new procedures that take into account the biases start-date dependence, to test whether it is possible to obtain a higher predictive ability. First, we study the dependence of the forecast skill on the period selected to compute the climatology. Observing drifts and anomalies we can seek for groups of hindcasts with behaviour that deviates significantly from the average. Thus, we can evaluate the impact of these groups on the forecast skill of the system. This study is performed for both VI5-35N and VI26N indexes. Reanalyses and observation anomalies are computed with respect to their climatology over the same period selected for hindcasts. This strategy selects groups of hindcasts that have similar biases and drifts, to adjust each group separately. However, this strategy can only be used to study the impact of the bias start-date dependence. To use it as an effective procedure to improve the forecast, we should state and follow a specific method and this does not seem

an easy task. How would we choose groups in a rigorous manner? Instead, the following strategy, which uses of a "moving climatology", has the advantage of not relying on an arbitrary choice.

In the case of the moving climatology, the ACC is calculated using anomalies computed with a moving window of selected width. Then, hindcast anomalies are computed by subtracting the climatology of the symmetric period centered on the hindcast itself. Only a comparison between hindcasts and reanalyses is made, because of the shortness of the RAPID dataset. Reanalyses anomalies are computed using a moving climatology too. The moving windows have been chosen with a width of 10 and 20 years. Longer periods would not be efficient in capturing the time dependence of the drift, while shorter ones would not provide robust estimates of climatology.

3.4.1 Computational methods

ACC between forecasts and reanalysis

When computing the ACC, we have to compare each forecast value with the observation in the same year. Therefore we consider only those values that are comparable with observations, i.e. that do not temporally exceed the year y_m (please refer to the beginning of section 3.4 for the notation). The forecast variable denoted by the couple (i, l) has a corresponding in observations if

$$\begin{cases} l = 1, ..., 10 & if\ i = 1, ..., m - 10 \\ l = 1, ..., (1 + m - i) & if\ i = m - 9, ..., m \end{cases} \tag{3.12}$$

Our original arrays for forecasts and observations are:

$$\begin{array}{ll} F_{i,l} & i = 1, ..., m \quad l = 1, ..., 10 \\ O_k & k = 1, ..., m \end{array}$$

After these considerations, we can choose between two methods for the calculus of the ACC.

1. We compute the anomalies $F'_{i,l}$ and O'_k as in eq. 3.5a and 3.5b. This means that before we calculate the lead-year dependent climatology C_l over all the start dates from y_0 to y_m and then we subtract it from every 10-years forecast. Then, generate from the observation anomalies array O'_k a set of "strings", one each 10-years long, so that they perfectly match with the forecasts. That is, every string is initialized in the corresponding start-date of the forecast and observations array now has same dimension of the forecasts array. The total number of O'_k observation anomalies is m and thus when we build the new array we must exclude the last 9 start dates, because in those cases we would not have enough values to build a string of 10 years. In this way from the

original observations array O'_k of dimension m is created a new array $(O^{new})'_{i,l}$ with dimension $(m-9) \times 10$:

$$(O^{new})'_{i,l} = O'_{i+l-1} \qquad l = 1, ...10 \quad i = 1, ..., m-9. \tag{3.13}$$

The ACC is computed for each lead-year with

$$\mathrm{ACC}_l = \frac{\sum_{i=1}^{m-9} (F'_{i,l})((O^{new})'_{i,l})}{\left[\sum_{i=1}^{m-9} (F'_{i,l})^2 \sum_{i=1}^{m-9} (O'_{i,l})^2\right]^{1/2}}. \tag{3.14}$$

2. Select a lead-year $l = 1, ..., 10$ and build the longest possible array with the forecast values, i.e. considering all the available years between y_1 and y_m for that l. In this way, from the original array $F_{i,l}$ we generate a new array:

$$(F^l)_k \qquad k = 1, ..., m-l+1 \tag{3.15}$$

We have put the apex l to remember that it is built selecting a specific lead-year. $(F^l)_k$ has dimension 1 and size $m-l+1$ because, while we always have a value in the last year y_m whatever l is, for $l = 2$ we will have the first value of $F_{i,2}$ in year y_2, for $l = 3$ the first value of $F_{i,3}$ in y_3, and so on. Thus, the relation between the years of $(F^l)_k$ and the years of the original array elements is the same as in equation 3.1. We compute the anomalies as

$$(F^l)'_k = (F^l)_k - \left\langle (F^l)_k \right\rangle_k . \tag{3.16}$$

We then compute ACC_l comparing the forecast array built for that lead-year with the observation anomalies array, but excluding from the latter the first $l-1$ values in order to consider only the observations in the same interval of years. Thus, O'_k are computed as in eq. 3.5b but taking O_k with $k = l, ..., m$. For consistency of notation with forecasts, let us call this anomalies array

$$(O^l)'_k \qquad k = 1, ..., m-l+1 \tag{3.17}$$

The ACC then is

$$\mathrm{ACC}_l = \frac{\sum_{k=1}^{m-l} ((F^l)'_k)((O^l)'_k)}{\left[\sum_{k=1}^{m-l} ((F^l)'_k)^2 \sum_{k=1}^{m-l} ((O^l)'_k)^2\right]^{1/2}}. \tag{3.18}$$

The first method has the advantage of giving a most robust estimate of climatology in the last lead-years of the forecast since, in that case, also the values that exceed

y_m are considered. However, the correct method to compute climatology, from a theoretical point of view, is the second one. Furthermore, the second method gives a more robust estimate of the ACC for the first lead-years because it does not exclude the last start-dates: for $l = 1$ it considers 10 values more than method 1, 9 more for $l = 2$, etc.

A test has been made to compare the two methods. In a first moment we computed the ACC with method 2, but obtaining the forecast climatologies with the two different explained ways. In this case the ACC did not seem to be affected significantly by the climatology method. In fact, those forecast values that exceed y_m are very few with respect to the total period of years and consequently they do not have a significant impact on the climatology when we exclude or consider them. Then, we directly compared method 1 and method 2. Since we had verified that the climatology method did not affect the ACC, eventual differences could derive only from the method by which we computed the ACC. We found that the first method overestimates the ACC with respect to the second one in a non-negligible way. For these reasons we decided to follow the second method, which seems the most correct from a theoretical point of view and the most reliable.

ACC between forecasts and RAPID

All we said above is valid for the case in which the first stat-date of the forecasts coincides with the first year of observations, as when we compare hindcasts and reanalyses. However, we also have observational data from RAPID, and in this case the first year of measures, which we will call y^*, is much greater than the first year of initialization $y_{1,1}$. That is, our starting arrays are

$$\begin{array}{ll} F_{i,l} & i = 1, ..., m \quad l = 1, ..., 10 \\ O_k & k = 1, ..., n \end{array} \tag{3.19}$$

with $n \leq m - 10$. Further,

$$y^* = y(O_1) > y(F_{1,l}), \quad \forall l. \tag{3.20}$$

So, for $l > 1$ we can use the forecasts initialized before y^*. We can select our $(F^l)_k$ choosing the start-dates from $y^* - l + 1$ to $y^* + n - l + 1$. This means that we will always have the same number of values at each lead-year. Then we can find the anomalies with

$$\begin{aligned} (F^l)'_k &= (F^l)_k - \left\langle (F^l)_k \right\rangle_k \\ O'_k &= O_k - \langle O_k \rangle_k \end{aligned}$$

and directly compute the correlation between the two 1-dimensional arrays $(F^l)'_k$ and O'_k in the standard way.

ACC between historical and observations

When we estimate the correlation between the historical simulation and the reanalysis, it would seem the easiest thing to get a single value from the direct comparison of the two arrays. Nevertheless, the aim is to use the ACC between the historical simulation and observations as a reference for the ACC between forecasts and observations, hence we must provide a lead-year dependent value. To do so, for each lead-year l we exclude from both the historical and the observations the first $l-1$ values: in this way we use only those years that we have considered for the ACC between forecasts and observations. Therefore we get two arrays with dimension $m-l+1$ that we can directly compare to get the ACC.

Anomalies computed with a moving climatology

The ACC is also computed for the case of forecasts adjusted using a moving climatology. Let's call w the width of the moving window. Starting from $F_{i,l}$, the ACC is obtained from the following steps:

(i) Select a lead-year l. As in the base case we get a 1-dimensional array $(F^l)_k$, with $k = 1, ..., m-l+1$ because we exclude the last $l-1$ years.

(ii) Compute the climatology over the period w centered on the year corresponding to each $(F^l)_k$. For the forecast values in the first $w/2$ years, thus from y_1 to $y_{w/2}$, we cannot compute the climatology over a symmetric period of w years. This equally applies to the last $w/2$ years, from $y_{m-w/2}$ to y_m. Hence, for all those values at the edge of the period we consider the climatology over the first (or the last) w years:

$$\begin{cases} (C^l_{mw})_j = \left\langle (F^l)_k \right\rangle_{k=j-w/2,...,j+w/2} & if \quad 1+w/2 \leq j \leq m-w/2 \\ (C^l_{mw})_j = \left\langle (F^l)_k \right\rangle_{k=1,...,1+w} & if \quad j \leq 1+w/2 \\ (C^l_{mw})_j = \left\langle (F^l)_k \right\rangle_{k=m-w,...,m} & if \quad j \geq m-w/2 \end{cases} \tag{3.21}$$

where $(C^l_{mw})_j$, $j = 1, ..., m-l+1$, is the climatology obtained with the moving window method at the specified lead-year.

(iii) Each anomaly is computed as

$$(F^l)'_k = (F^l)_k - (C^l_{mw})_k. \tag{3.22}$$

(iv) Observation anomalies are equally computed with a moving climatology. At borders the procedure is the same as for the forecasts. After the anomalies computation, the first $l-1$ values are excluded, so to have an array comparable with the forecast array. The anomalies obtained are $(O^l)'_k$.

(v) The ACC is computed as usual as in equation 3.18.

Probabilistic metrics

Also in the case of probabilistic metrics we are constrained by the length of observations array, because we make a comparison between forecasts and observations through the RMSE. Based on that, we have to choose a method which is consistent among all the probabilistic metrics (PP, variances and RMSE), in order to be able to make meaningful comparisons.

Let us start then from the RMSE. We must take forecasts and observations over the same period of years, for each lead-year. To do so, we take the original forecast array

$$F_{i,l} \qquad i = 1, ..., m \quad l = 1, ..., 10$$

and, based on the considerations already made for the ACC, we select a lead-year l and extrapolate an array excluding the last $l - 1$ values:

$$(F^l)_k \qquad k = 1, ..., m - l + 1.$$

This is a 1-dimensional array, with the apex l to signal the lead-year dependence. To obtain an observation array in the same interval of years, we consider the values of the original array starting from $i = l$, obtaining:

$$(O^l)_k \qquad k = 1, ..., m - l + 1 \tag{3.23}$$

So we define a lead-year dependent climatology also for the observations:

$$C_O^l = \left\langle (O^l)_k \right\rangle_k \tag{3.24}$$

Then, the RMSE for each lead-year is

$$\mathrm{RMSE}_l = \sqrt{\left\langle [((F^l)_k - C_l) - ((O^l)_k - C_O)]^2 \right\rangle_k}. \tag{3.25}$$

The arrays $(F^l)_i$ are used also for the PP, σ_t, σ_n and σ_s as in equations 3.9, 3.8a-c. Thus, we underline that in these metrics are considered m values for $l = 1$, $m - 1$ for $l = 2$, and so on until $l = 10$ for which are used $m - 10$ values. For consistency among the used metrics also the observed variance σ_o is computed in a lead-year dependent way, hence using the arrays $(O^l)_i$.

Chapter 4

Results

The main findings are presented. In section 4.1 we provide a first picture of the Atlantic Meridional Overturning Circulation (AMOC) in the CMCC-CM2-SR5 decadal predictions (DPs), describing its climatology and time variability. The deterministic and probabilistic assessment of predictability is given in section 4.2, also by means of reanalyses (REAN), RAPID-array data and the historical simulation (HIST). In section 4.3 we study Labrador Sea convection and its potential influence on the AMOC simulated in the model. Finally, section 4.4 is devoted to the time dependence of drift and to the results of the "moving climatology" drift adjustment.

4.1 AMOC climatology and variability

Before going into details of the system's ability to forecast AMOC, we must start from a more general study: that is, how the AMOC climatology is made and how the strength of circulation varies in the forecasts.

The model climatology over the entire decade is shown in figure 4.1a, as function of latitude and depth. We can recognize the main features of the AMOC: the circulation is clockwise in the upper layer and anti-clockwise in the abyssal layer, indicating a northward flow near the surface and a cold return in the deeper ocean. Moreover, transport reaches its maximum at 28.5°N, in accordance with observational estimates, with a value of 14.2 Sv. The maximum standard deviation from the decadal climatology is located in the equatorial region, between -5°N and 10°N: it reaches a maximum of 5.2 Sv at 7°N. At 26.5°N the maximum is 14.0 Sv at about 1000 m, almost 3 Sv less than the RAPID climatological maximum of 16.8 Sv, but at the same depth.

The maximum region is subject to a sudden change in the first two forecast years. In fact, looking at the lead-year dependent climatologies in figures 4.1b-d, we see that the maximum undergoes an attenuation and, we could say, a progressive spreading: in lead-year 1 it is 17.1 Sv at 13°N, reaching at lead-year 2 a value that remains almost stationary for the entire decade, i.e. 13.3 Sv in a region between 5°N

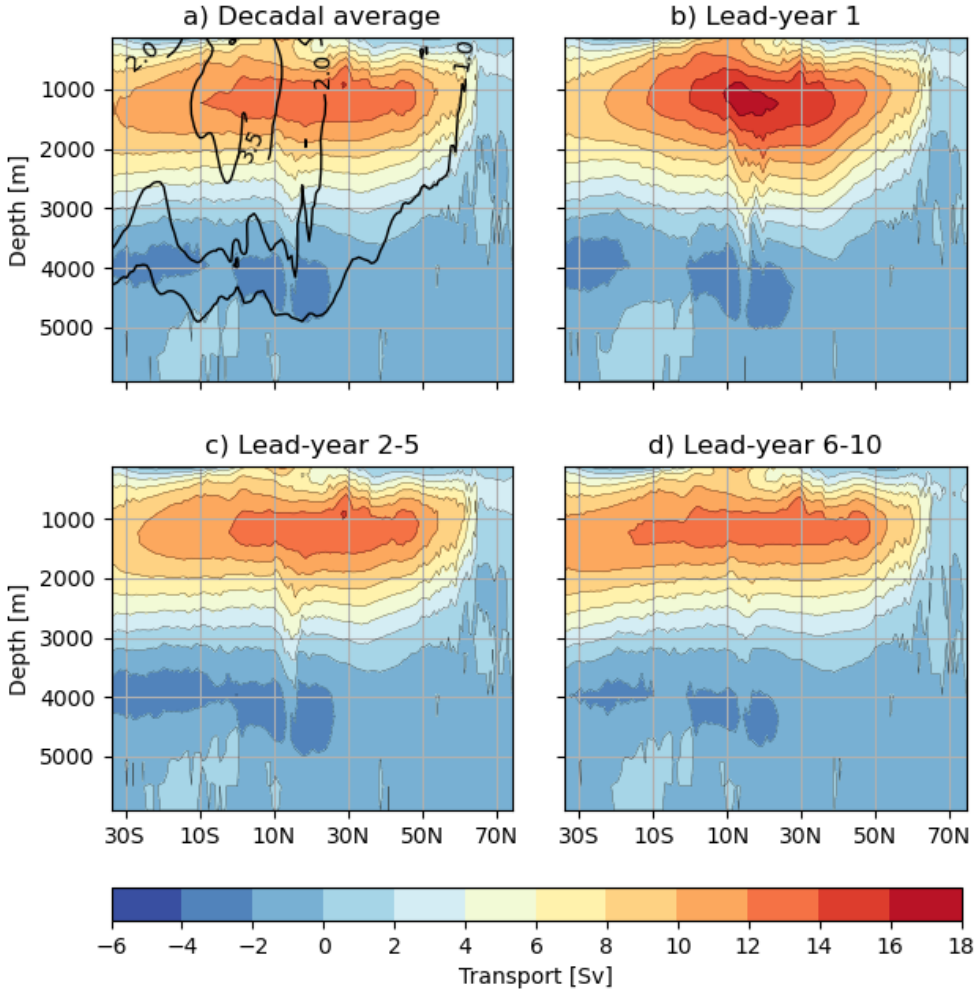

Figure 4.1: Meridional Streamfunction climatologies of DPs initialized between 1960 and 2018 (transport measured in Sverdrups). **a)** Climatology for the decadal average. Black lines represent levels of standard deviation. **b-d)** Climatologies in different lead-years, or averaged over groups of lead-years. The first two months of initialization are excluded from both lead-year 1 and the decadal average.

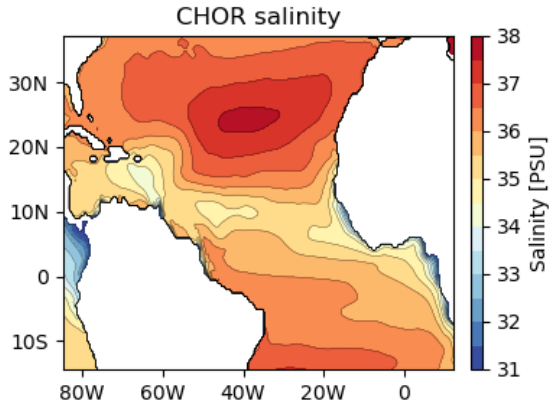

Figure 4.2: Climatological salinity field at surface for the CHOR ensemble in the equatorial and north-tropical region.

and 35°N. If the model starts on average from a configuration in which the maximum value is about 3 Sv below the RAPID's observational one, it underestimates real transport by about 7 Sv in the last years of the forecast.

At all lead-years is present a narrow band of high transport at 10-15°N, which starts from the maximum region and arrives at depth. It is reasonable to attribute this band to a strong density gradient created by a massive release of freshwater by the Amazon Delta (0-2°N) and the Orinoco Delta (8-9°N). This freshwater is carried northward along the eastern coast of South America by the North Brazil Current and then eastward towards African coasts by the Equatorial Counter Current. In figure 4.2 is shown the climatological salinity field at surface in the north-tropical region for the CHOR reanalysis ensemble: it is evident a low salinity band cutting the Atlantic ocean between 5°N and 15°N. This feature fades as lead-years increase, suggesting that the release of freshwater is a spurious behaviour due to the reanalysis or to a problem following initialization.

Let us look now at AMOC time variability in single hindcasts. Annual values of VI5-35N for DPs, REAN and HIST are shown in figure 4.3. Given that the CHOR reanalyses used for verification are used to initialize the model, we expect the initial state of DPs start dates before 2010 to coincide with the REAN state. However, DPs first annual average is computed excluding the first two months, so that the first annual value can be different from the REAN transport. To make a check we have computed the difference between DPs and REAN in the month of initialization (November), and on average it is 6.7 Sv. In addition to a numerical error, that we do not expect high enough to justify such a big difference, we can suppose that a strong initial shock is present. It is a problem reported in literature that in a coupled prediction system initialized by assimilating uncoupled atmosphere and ocean reanalyses can suffer from an imbalanced initial state, which triggers the rise of initial shocks in the forecasts (Mulholland et al., 2015; Kröger et al., 2018). This

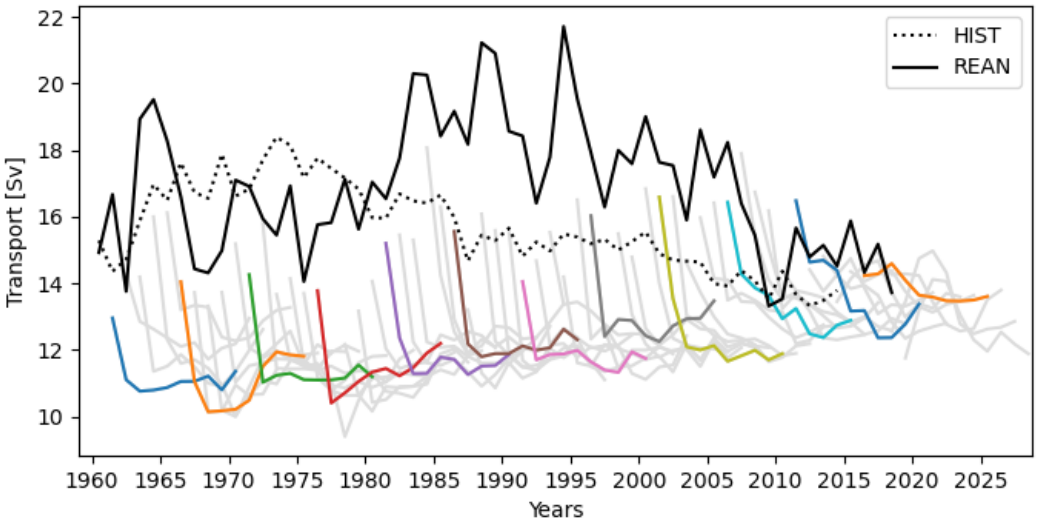

Figure 4.3: VI5-35N annual time series for the DPs initialized between 1960 and 2018 (grey lines, colored one each five), REAN (black solid line) and HIST (dotted line).

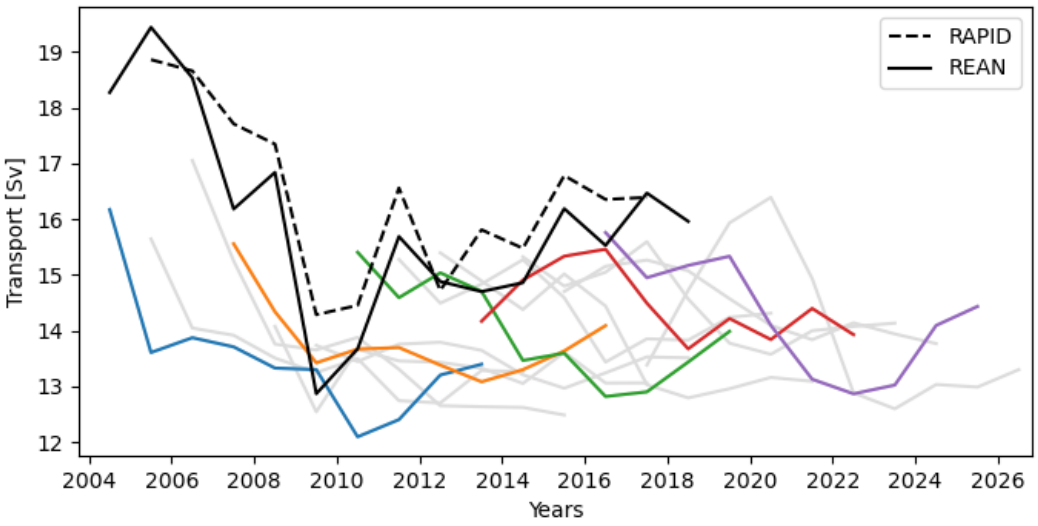

Figure 4.4: VI26N annual time series for the DPs initialized between 2004 and 2018 (grey lines, colored one each three), REAN (black solid line) and RAPID (dashed line).

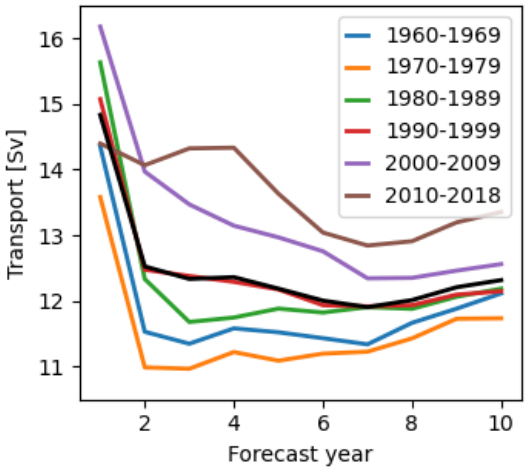

Figure 4.5: Lead-year dependent climatologies of DPs VI5-35N in single decades. The black line is the climatological average over the whole period 1961-2018.

is not in contradiction with the fact that we use the reanalyses to initialize the model. In fact, the output is given in monthly average values, but the model runs at shorter time scales: thus, there is space for the development of an initial error in the first month of the forecast.

The most prominent feature of DPs VI5-35N is a huge collapse in the first one or two years: this means that the model is initialized in a state consistent with REAN and then generally tends to a state in which transport is suppressed. This behaviour is expected in a full-field initialized model such as CMCC-CM2-SR5, and documented in literature (a detailed study can be found, e.g., in Bilbao et al., 2020). Because simulations are initialized in a state different from the model attractor, they tend to drift towards it during the run. We can see the same behaviour in figure 4.4, where RAPID, REAN and DPs VI26N are compared. This drift is plausibly linked to the spread of the climatological maximum in figure 4.1, in the sense that also the variation of maximum is related to the tendency of the system to move towards its equilibrium state.

We could expect that, since the historical simulation has been generated with the same model and thus has the same attractor on DPs, the initialized simulations should tend to the state represented by the historical one. However, this does not happen, and it is not the first case in literature. Bilbao et al. (2020) compare the transport of an EC-Earth3.3 decadal system with an ensemble of historical simulations, and they find that the equilibrium state of hindcasts is represented only by some of the historical ensemble members. More specifically, hindcasts tend only to the historical simulations in which convection is suppressed. In our case, we have only one realization of the historical available, so we cannot draw more

conclusions beyond the fact that the DPs tend to a lower state with respect to HIST.

Then, we have to notice an important aspect: that the drift has a start-date dependence, and this is related to the AMOC variability in the reanalyses. REAN VI5-35N has an increasing trend until mid '90s and a decreasing trend later, with significant multi-annual variations. DPs are initialized in a state near to REAN state but they globally tend to the model's equilibrium state; thus, the drift entity strongly depends on REAN variability. In particular, after the AMOC collapse of 2010, REAN VI5-35N is much lower than in the preceding years and DPs are initialized in a state nearer to the equilibrium state. This makes drifts after 2010 much smoother than before. This start-date dependence is well represented in figure 4.5, where the lead-year dependent climatologies are shown. Before 2010, DPs have in general different initial states but, at least, the drift gradients in the first year are similar. The 2010-2018 climatology instead has an entirely different behaviour with respect to the preceding ones: after a drift of just 0.4 Sv, transport grows until year 4 and then begins to decline. A particular feature that we can detect in more or less all DPs, is a slight recovery of transport in the last forecast years.

4.2 AMOC predictability

4.2.1 Forecast skill

As explained in section 3.4, to evaluate the forecast skill we use the longest available periods of DPs, REAN and RAPID values. In figure 4.6a-c are represented their anomalies with respect to the corresponding climatologies.

The skill assessed for VI5-35N, hence using REAN as comparison (figure 4.7a), gives an ACC of 0.42 in the first year of the forecast, followed by a sudden reduction since the second year, with values near to zero or negative, such as -0.41 in year 4. A slight increase of the ACC happens as lead-years go by: this could be related to the partial recovery of the DPs after the initial drift that we have seen in figure 4.5. We can try to infer this low skill from the anomalies in figure 4.6a. Qualitatively, anomalies seem to follow at least the general trend of reanalyses, often coinciding with them at lead-year 1. Since lead-year 2 however some DPs get anomalies values of opposite sign with respect to REAN. This can be seen especially in some groups of DPs anomalies, and the ones that stand out are those after 2003: since lead-year 2 they tend to a mean transport of about 2.5 Sv, while REAN anomalies are set at about -2.5 Sv. This behaviour is related to the start-date dependence of drift that we mentioned in section 4.1. In fact, after 2010 REAN set at a lower value and consequently the drift becomes smoother (figure 4.3). Hence, subtracting the climatology over the whole period 1961-2018 from these DPs leads to an over-correction of their drift, to an over-estimate of their anomalies and in turn to a negative ACC. This happens especially since lead-year 2, because at initialization

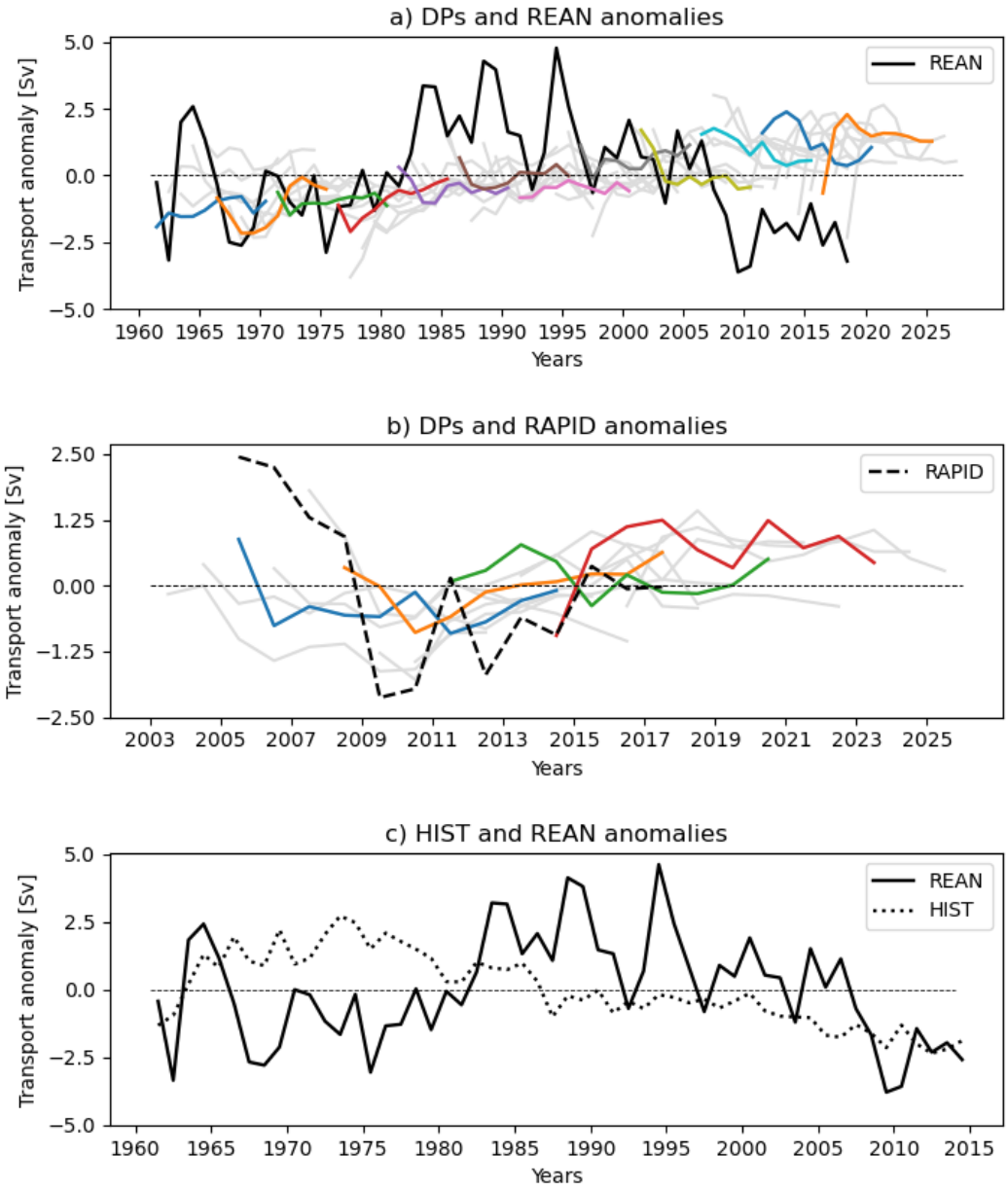

Figure 4.6: **a)** Anomalies of VI5-35N annual time series, for DPs (grey lines, colored one each five) initialized between 1961 and 2018 and REAN (black solid line). **b)** Anomalies of VI26N annual time series, for DPs (grey lines, colored one each three) and RAPID (black dashed line). **c)** Anomalies of VI5-35N annual time series, for HIST (dotted line) and REAN (solid line) between 1961 and 2014.

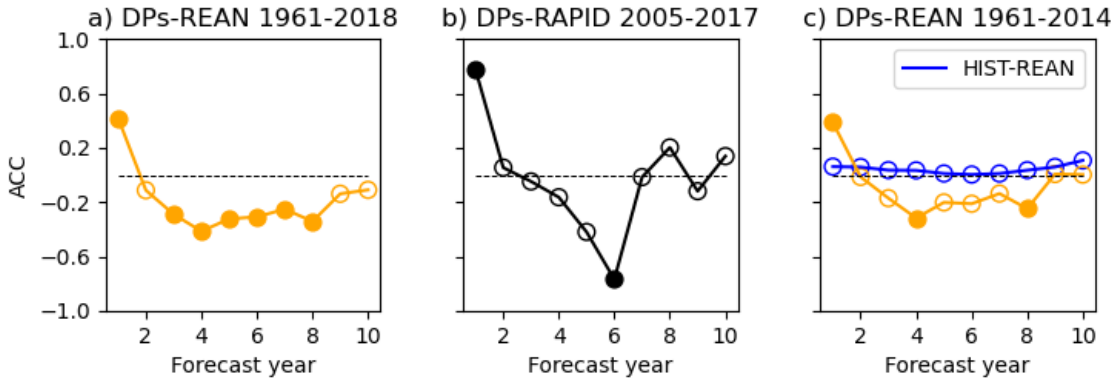

Figure 4.7: Lead-year dependent ACC for: **a)** VI5-35N anomalies, REAN compared with DPs, computed with respect to climatology between 1961 and 2018, **b)** VI26N anomalies, RAPID compared with DPs for the period 2005-2017, and **c)** VI5-35N anomalies of HIST (blue) and DPs (yellow) compared with REAN in the period 1961-2014. Solid dots represent statistically significant measures at 95% level.

REAN and DPs almost coincide.

A quite similar reasoning can be done when the skill of DPs is evaluated using RAPID data. In figure 4.7b is represented the lead-year dependent ACC for the VI26N index: at lead-year 1 it is 0.78, at lead-year 2 is almost zero and at lead-year 6 it reaches even -0.77. Looking at the corresponding DPs and RAPID anomalies in figure 4.6b indeed, there is a strong coincidence in the year of initialization but, since lead-year 2 and in particular before 2009, DPs anomalies get opposite values with respect to RAPID. In the VI26N DPs variability in figure 4.4 is evident a difference of drifts before and after 2009, due to the AMOC collapse of 2010.

Hence, it is reasonable to refer a low correlation with REAN and RAPID to the specific behaviour of some groups of DPs. This supports the idea that a start-date dependent drift adjustment may improve the forecast skill. We will better see this aspect in section 4.4, where we will investigate deeper the impact of the time dependence of drift on the system's predictive ability. Without such an investigation and a skill improvement, it would be impossible to draw useful predictions of the AMOC state after the first year of the forecast.

After the skill evaluation with the VI5-35N and VI26N indexes, we have explored also the dependence of the DPs-REAN ACC on the latitude: it is shown in figure 4.8, top panel, for different lead-years averages. The transport at each certain latitude is computed averaging the MSF on the vertical but without considering the Ekman layer. It is interesting to notice three features:

(i) at year 1 the correlation is quite high at all latitudes;

(ii) the correlation is very high at all lead-years in the South Hemisphere while, especially since year 2, there is a rapid decline between 5°N and 10°N;

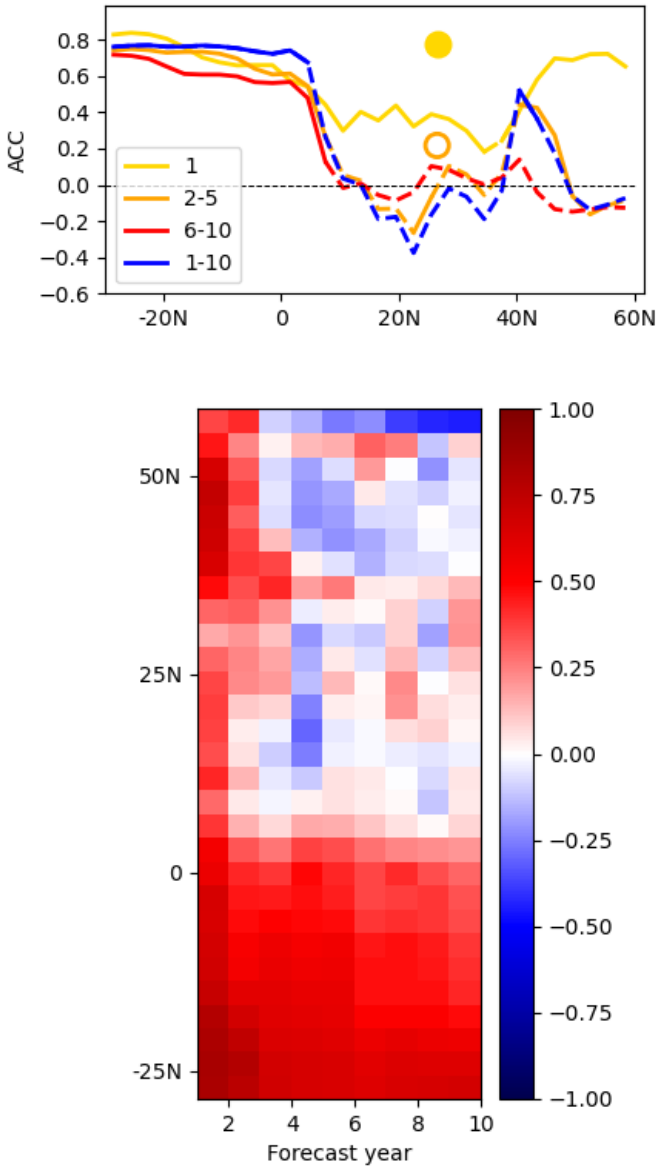

Figure 4.8: Latitude-dependent ACC between DPs and REAN transport. The transport at each latitude is computed averaging the MSF over the vertical, without considering the upper 200 m. **Top:** ACC for different groups of forecast years. Solid lines are where measures are statistically significant (95%). The correlation with RAPID is overimposed at 26.5°N (dots, full dots are statistically significant). **Bottom:** Latitude-dependent ACC as in the top panel, but in a Hovmöller plot so that we can gain resolution on the lead time.

(iii) at about 40°N there is an impressive recovery restricted to a narrow band of latitudes.

As comparison we added the correlation with RAPID at 26.5°N. The correlation at lead-year 1 is the same that we find in figure 4.7. The ACC for the average over the years 6-10 and 1-10 is not shown. Indeed for RAPID, when averaging over groups of lead-years, we cannot keep the same number of values over all the forecast time as we can do when considering only one lead-year, as we did e.g. in figure 4.7b. Hence, for averages over high lead-years we shall use very few values: specifically, when considering averages comprising year 10 we can rely only on three couples of values, making the correlation meaningless.

It seems that the region we have selected for the VI5-35N index is precisely the one with the lowest correlation. The fact that the region of maximum transport coincides with the one in which the model is worse at predicting the AMOC may not be a coincidence, because we have seen in the AMOC climatology pattern (figure 4.1) that the maximum undergoes a displacement, or better, a spread, as forecast time goes by. Nevertheless, it could be true also the opposite, i.e. that the skill is low at all latitudes and some localized effects increase the ACC only in some regions. Indeed, from the time series of transport in the Southern Hemisphere (not shown), i.e. 30-0°S, emerges a increasing global trend. This trend affects the ACC, which is remarkably high. In any case, in the bottom panel in figure 4.8 it is evident a loss of forecast skill after the first year at all latitudes, that we can always attribute to the drift problem, and a partial recovery in the last year. We can recognize only two bands of latitudes which show a much lower loss of skill, that is 40-45°N and 55-60°N. The former is significant because it is where the North Atlantic current originates from the Gulf Stream (Rossby, 1996). We can suppose that the presence of the Gulf Stream increases the predictability: indeed, the skill tends to remain higher until year 3 between 20°N and 45°N.

Impact of initialization

To understand how the initialization affects the skill, we compared the DPs-REAN ACC with the HIST-REAN ACC. The ACC between HIST and REAN is comprised between 0 and 0.1 at all lead-years, as we can see in figure 4.7c. The anomalies of HIST and REAN are indeed of opposite signs in the great part of the period 1961-2014 (figure 4.6c), confirming the randomness of AMOC variations predicted using only the historical forcings.

Initialization significantly improves the forecast skill at least at lead-year 1, when the ACC between DPs and REAN over the period 1961-2014 is 0.39. In some of the following years, the ACC obtained from initialized simulations deviates significantly from zero, although with negative values. This could be interpreted as the fact that, even though the predicted anomalies of the AMOC state are not coherent with those of observations, initialization brings to less random variations.

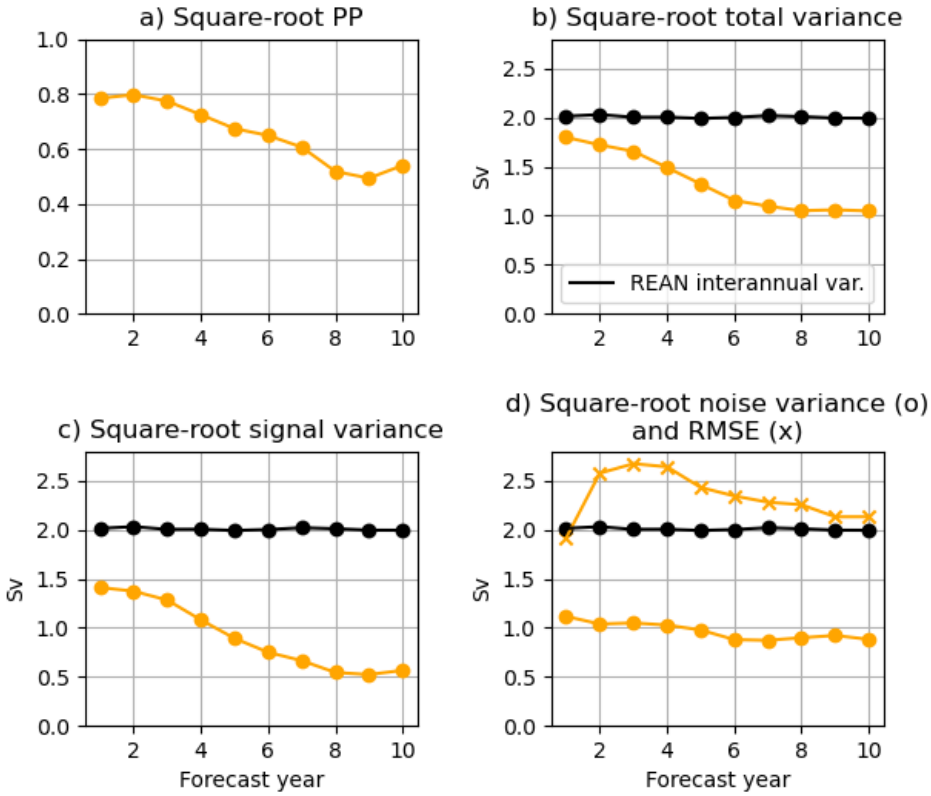

Figure 4.9: Probabilistic metrics computed with DPs VI5-35N. **a)** Lead-year dependent PP. **b-d)** Square-root of total, signal and noise variance of DPs (yellow dots). We put as reference the square-root REAN interannual variance (black dots). In **d)** is also shown the RMSE between DPs and REAN (crosses).

That is, initialization is in general beneficial for predictions, although in our case there is an issue in initialization itself that brings the model to drift towards a particular state and then to give bad predictions.

4.2.2 Potential predictability

In figure 4.9 are shown all the measures obtained for a probabilistic assessment of predictability. The system shows high potential predictability at all lead-years (figure 4.9a), in particular in the first half of the decade where it reaches a maximum of 0.8 at year 2. Subsequently, it decreases but never falling below 0.5. The PP is very high compared to the ACC computed over the same period (figure 4.7a). Because the PP is the upper limit for the model to forecast the signal of the physical process, our system shows a big limitation in that but, on the other side, has the potential for a significant improvement. The square-root total variance (figure 4.9b)

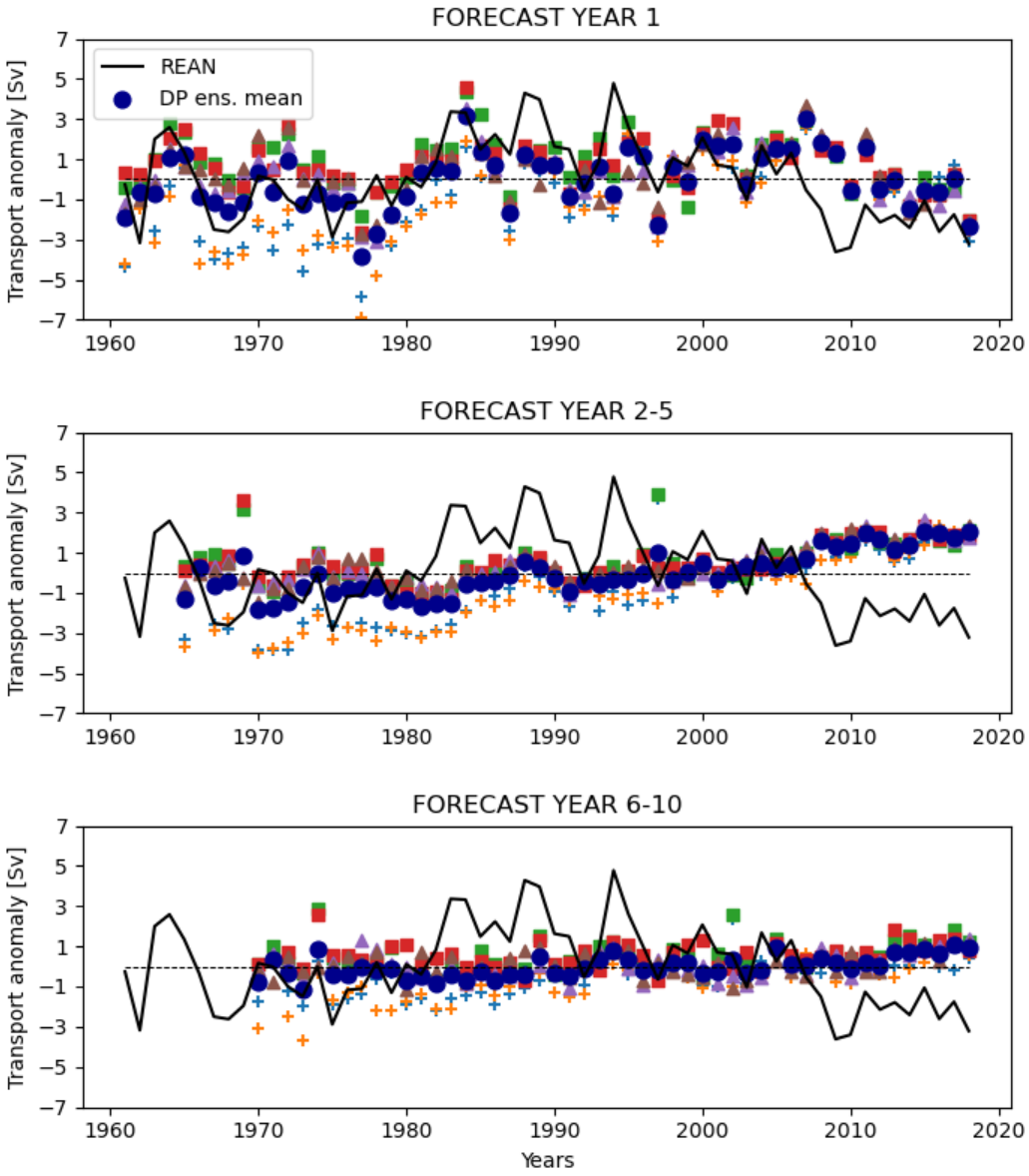

Figure 4.10: DPs and REAN Vi5-35N anomalies in different lead-years, or averaged over groups of lead-years. Blue dots represent DPs ensemble means, while the other symbols refer to single realizations. REAN anomalies are represented by the black line.

also reduces over the decade, with a total reduction of 0.8 Sv. This is due to a decrease in square-root signal variance of about 1 Sv, while the noise variance remains almost constant (figures 4.9c-d), explaining then the progressive loss of PP. Looking at figure 4.10, where the DPs anomalies at different lead-years are represented, it is evident a signal flattening, in particular in the second half of the decade. The main signal collapse is between year 1 and years 2-5. This is probably related to the ACC trend that we have seen in section 4.2.1: the decrease of skill happens primarily between year 1 and 2. Also, the reanalysis interannual variance is always greater than the total (and thus signal) variance of the system (figure 4.9b-c), revealing the already mentioned problem of the system to capture AMOC variability correctly.

The noise variance, which represents the mean interannual spread of the system, is always below the RMSE (figure 4.9d): that is, the system is overconfident and its internal variance is not high enough to explain the average distance between forecasts and observations. From the figure 4.10 we can get more information about the system dispersion. The greatest spread in general is in the decades from '60s to '90s. Indeed, ocean measures in the past were more sparse and less accurate, so we expect a biggest reanalyses dispersion in that period. Moreover, the spread has a lead-time dependence: it is greater in the first lead-years, at least for start-dates until the end of the '90s. For most recent years instead, when we have more accurate reanalyses, ensemble members are very close to their mean in lead-year 1 and then they slightly disperse. Hence in noise variance at lead-year 1, most recent decades compensate the big spread of the first ones. However, from the noise variance in figure 4.9d this does not emerge, because we are averaging over the whole period. From figure 4.10 we can conclude that all the forecast members tend to the same state, and this leads to both a reduction of signal and noise variance, even if the loss of signal is bigger than the loss of noise.

Another feature is noteworthy: since lead-year 2 there is a positive trend in DPs anomalies. After 2010 the excessive distance between DPs and REAN anomalies is attributable to the over-adjustment we have already been talking about in section 4.2.1. However, apart from this aspect, the trend persists over all decades and seems to be related to the trend we can see also in figures 4.3 and 4.6a, where it is clear that DPs equilibrium state has an increasing transport over the years.

4.3 Impact of initialization on Labrador Sea convection

As we already said, AMOC strength directly depends on convection and deep water formation (DWF). The principal place in the North Atlantic (and thus in the world) where DWF happens is the Labrador Sea (LS, see figure 4.11a). A study by Bilbao et al. (2020) has demonstrated that, in models, there may be a connection between

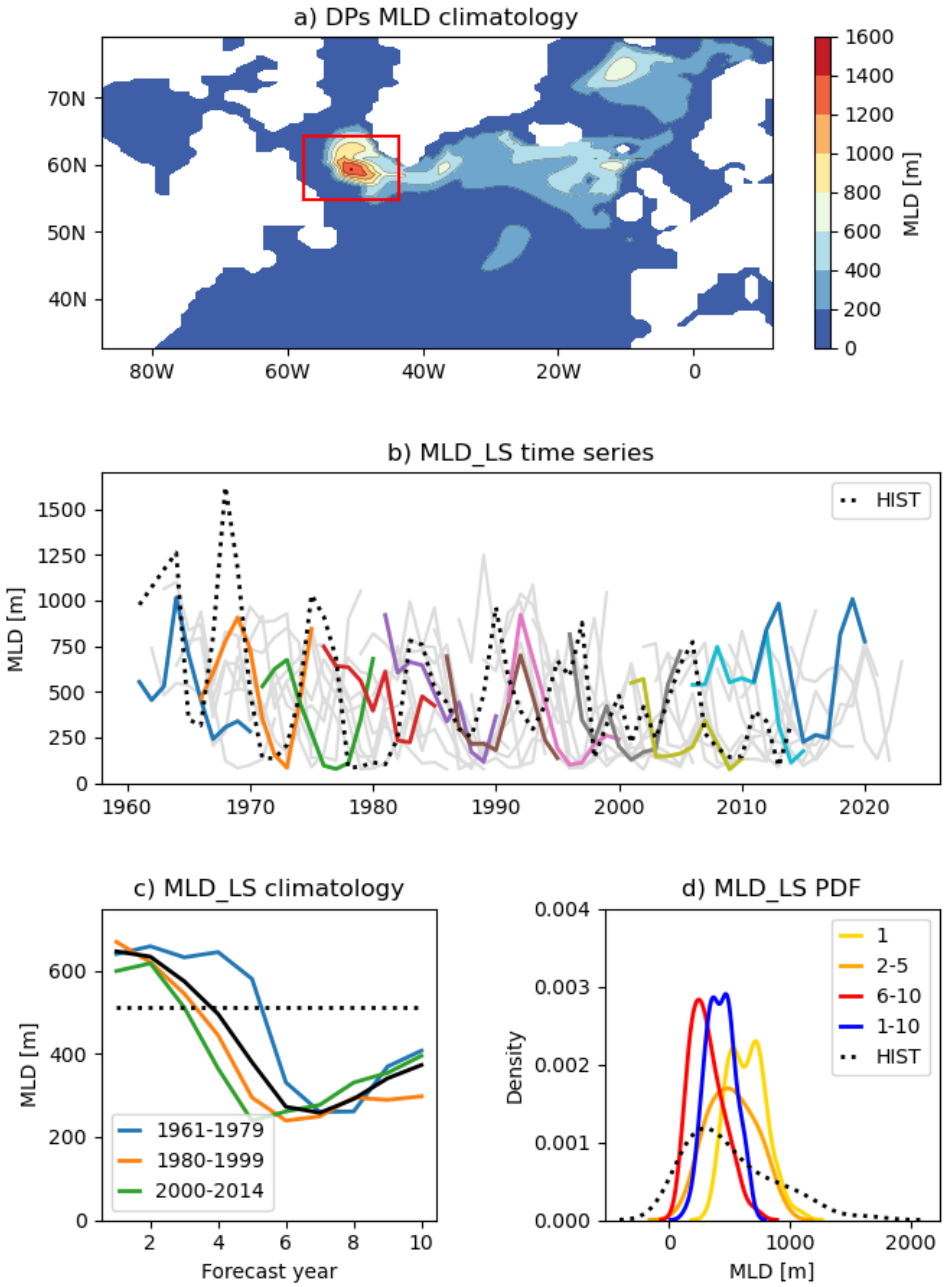

Figure 4.11: a) MLD decadal climatology pattern of the DPs over the subpolar region. The red box delimits the area taken to compute the MLD_LS index. **b)** Time series of the MLD_LS index for HIST and DPs (colored one each five start dates). **c)** MLD_LS lead-year dependent climatologies in different periods. The black line is the climatology over the entire period 1961-2014. The dashed line represents the climatology of the HIST (hence over the period 1961-2014). **d)** Probability Density Function of the DPs MLD_LS for different lead-years, and for HIST. In all panels are considered only the DPs initialized between 1960 and 2014, for consistency with HIST.

the drift-related collapse of AMOC and the stop of convection in the LS, because LS densities act as precursors for ocean circulation changes (e.g. Ortega et al., 2017; Ortega et al., 2021). We can investigate the mixed layer depth (MLD) in that region to see whether also in our DPs there is a reduction of convection, and if this can be somehow referable to the AMOC drift. We also investigate the MLD in HIST because the drift is due to the tendency of the model to return to the non-initialized state, so we could expect that it returns towards the state of the historical simulation. We use the MLD_LS index (see section 3.3). Because HIST arrives at maximum at 2014, we consider only the DPs initialized between 1961 and 2014.

Looking at the MLD_LS time series (figure 4.11b), HIST has a highly oscillating MLD between a state in which the depth is around 1000 m and a state of suppressed convection, with a period between 2 and 5 years. The oscillation is particularly high before the '90s and this is consistent with the higher uncertainty in the oceanic state for that period. It is not easy to understand whether there is an average trend in DPs MLD_LS just by looking at the time series: hence we computed the corresponding lead-year dependent climatologies in different periods, which are in figure 4.11c. Actually, there is an average reduction of convection, but not in the forecast years that we would expect from the collapse of AMOC. In fact the AMOC drift happens in the first two forecast years, while the MLD reduction happens primarily after year 2. Moreover, there is a partial recovery of convection in the second half of the forecasts, exactly as happens for AMOC (figure 4.5). Hence, it is evident that there is a stop of convection, but there is not enough evidence to link it to an AMOC trend or to understand if one trend is the cause of the other. A possibility could be that the sudden AMOC suppression in the first two years, being related to a reduction of northward heat transport, leads to ice formation in the polar region. Ice would act as insulator and inhibit convection and DWF. Certainly, the MLD collapse is a problem that must be taken in consideration for future model improvements.

The climatology plot in figure 4.11c gives us more information about the MLD of DPs, but not of HIST (dotted line): the average performed over the whole period indeed do not highlight the bi-modal behaviour found in the time series. The probability density function (PDF) in figure 4.11d instead shows this duality, even if with a preference for the lower value. At higher depths there is a knee which is distinguishable from the main distribution. Looking at the DPs PDF, they tend to start with a higher MLD in the first year and set in a state with the lower value of MLD reached by HIST in the last ones. In any case, the distribution reached at equilibrium by DPs is not the one of HIST. Thus, neither DPs MLD_LS nor AMOC indexes tend to the historical simulation's corresponding values.

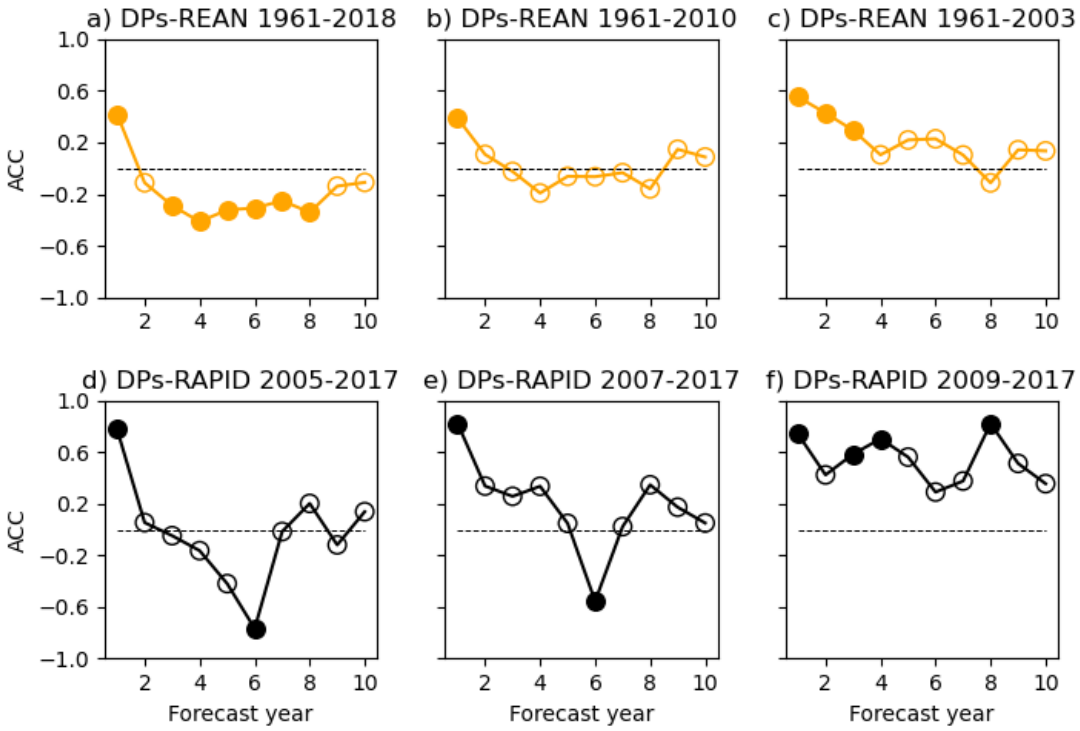

Figure 4.12: a-c) ACC between DPs and REAN VI5-35N for different periods. The longest one is in **a)**, where it is represented the skill found in section 4.2.1 for the general skill assessment over the whole period 1961-2018. **d-f)** ACC between DPs and RAPID VI26N for different periods, starting from the base case 2005-2017 in **d)**. Full dots are statistically significant measures at 95% level. Gradually excluding some years from the climatological mean the skill undergoes a substantial improvement.

4.4 A novel strategy for drift correction

In section 4.2.1 we have given an evaluation of the forecasting ability of the system and we have seen that probably the low correlation at certain lead-years is related to some groups of hindcasts, which have a drift that deviate from the average. We have qualitatively identified these groups observing the anomalies trends (figure 4.6a and 4.6b). We want to verify if our supposition is correct, i.e. if effectively these DPs we have selected have a negative impact on the ACC. If this is the case, this will support our intention to use a start-date dependent drift adjustment.

When we analyzed the forecast skill of the DPs VI5-35N through the correlation with REAN, we found a low ACC especially since lead-year 2 (figure 4.7a). We noticed that the VI5-35N DPs anomalies after 2003 had an opposite sign with respect to REAN anomalies and that this could negatively affect the correlation

(figure 4.6a). In light of this, now we can try to exclude from the assessment these last start-dates. The ACC is re-calculated: in this case, the DPs and REAN climatology used to get the anomalies is taken as the average on the selected period. Excluding gradually the years after 2003 we obtain the ACCs in figure 4.12a-c. The ACC at lead-year 1 increases from 0.42 to 0.55, but the improvement is notable in particular for all the lead-years after the first. After the selection we performed on hindcasts, we get a good forecast ability until year 3: this is a good result because we have seen that it is challenging to obtain reliable predictions at long forecast years.

Besides this, we have seen in figure 4.7b that the correlation between RAPID and the DPs is particularly low at lead-years 4, 5 and 6. Looking at the anomalies in figure 4.6b, we made the assumption that this is related to those DPs anomalies values which lie before 2009. To verify this, let us try to gradually exclude from the ACC computation the years from 2005 to 2009. For the period 2007-2017 we have 11 couples of values available at all forecast years, while for 2009-2017 we have 9 couples. The result is displayed in figure 4.12d-f. As the period shrinks, the ACC undergoes an outstanding improvement at all lead-years after the first one. The impressive feature is that the highest ACC is reached in year 8 (0.82), even higher than in year 1 (0.74). This gives hope to obtaining good predictions of the AMOC state also at long forecast years.

The two cases above show that the forecast skill strongly depends on how we compute the anomalies. Moreover, if we consider that the drift depends on the initial value (and thus is non-stationary) and we apply a specific correction for each period, we should be able to improve the predictive ability.

We can try now to perform a drift correction, computing the anomalies in a way which takes into account the start-date dependence. We apply the method of the "moving climatology", explained in section 3.4.1: we subtract from each hindcast the climatology computed over a short period centered on the hindcast itself. In figure 4.7 we can see the ACC between DPs and REAN VI5-35N for all the start-dates between 1961 and 2018, for three cases: the basic one that we have already seen (4.7a), the ACC computed with a moving-window climatology of 10 years and the ACC for a moving window of 20 years.

In both the 10 and 20-years cases there is an improvement of the forecast skill, at all lead-years. The 20-years window leads to a better improvement at lead-years 1 and 2, while the 10-years window at lead-years from 3 to 10.

The result is not very satisfying from the point of view of the prediction quality of our system: a quite good result is obtainable only until lead-year 2, while in the other years the correlation remains near or below zero. The only exception is a slight improvement in the last years. However, despite the problems related to our system, this experiment opens the possibility of studying new innovative approaches to analyse decadal simulations.

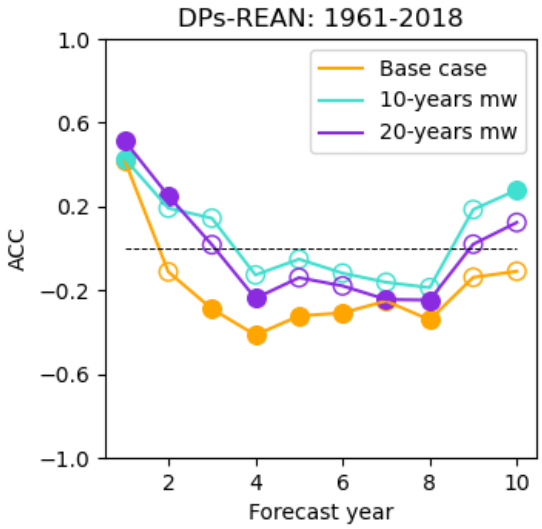

Figure 4.13: ACC for DPs and REAN VI5-35N between 1961 and 2018, for anomalies computed by means of moving climatologies. Two different widths of the moving window (mw) have been tried: 10 years (light blue) and 20 years (violet). The base case, that is the standard ACC for the period 1961-2018, is put as reference (yellow). Full dots represent statistically significant measures (95% level).

Chapter 5

Discussion

AMOC variability and predictability

The departure of the decadal predictions (DPs) from the initial state is a feature consistent among all the results in sections 4.1 and 4.2, from the lead-year dependent climatology in figure 4.1, to the time series in figures 4.3 and 4.4. DPs do not just undergo a great collapse in the first two years, they also show a slow increase immediately after the drift. This slight recovery translates in an increasing multi-decadal trend of the equilibrium state. This general variation is also reflected in the reduction of the anomaly correlation coefficient (ACC) between DPs and reanalyses since year 4 and in its slow increase after, and in the root mean square error (RMSE), which is maximum in lead-years 2-4 and successively slightly decreases. The drift is an expected feature, but the origin of the recovery is not so clear and needs further research.

From the latitude-dependent ACC (figure 4.8) it seems that the skill undergoes a reduction over the entire North Atlantic. The only exceptions are the latitudes corresponding to the beginning of the North Atlantic Current (NAC) and a band of latitudes between 55-60°N, where the southern branch of the subpolar gyre flows. Because in the Southern Hemisphere there is a trend which affects positively the ACC, we cannot say that the skill reduction affects only particular areas of the Atlantic basin. The two narrow bands of high ACC instead suggest that some oceanic phenomena may prevent the signal loss in restricted regions.

Strictly speaking about the predictability, we have seen that the system has a high potential predictability (PP) compared to the forecast skill. As underlined by Scaife and Smith (2018), we must be cautious in considering the PP as the upper limit of the achievable predictability, because the current model generation tends to underestimate the signal-to-noise ratio in the Atlantic region. This is true also for the AMOC (Yan et al., 2018). In our system there is a signal variance reduction of about 1 Sv over the decade, and this seems reasonable since all DPs tend to the same equilibrium state, but we do not know if this reduction has a true analog in the real world. Moreover, from the figure 4.10 it is evident that the noise variance

(ensemble spread) is start-date dependent, but we did not consider this feature in our analysis. In the past, the signal appears more "noisy", due to an uncertainty in the ocean state. In more recent times instead, the noise variance is so low that it makes the system overconfident. This highlights the need of studying in a start-date dependent way not just the deterministic verification metrics, such as the ACC, but also the probabilistic ones.

Having said this, we can try to understand why the skill is so low. There is more than a possible explanation, and maybe different causes act together:

- In section 4.4 we have seen that the drift start-date dependence lowers the ACC, in particular since lead-year 2, because after the correction some hindcasts result in too-high or too-low anomalies. This is a problem related to how we analyse data, rather than a problem in the model itself.

- Scaife and Smith (2018) underline that in current models the number of ensemble members should be high enough (of the order of tens) to reduce the noise component of variability and obtain an accurate estimate of the predictable signal. The uncertainty related to the model and to internal variability, which have the highest weight at decadal scale, is reducible increasing the ensemble size, but we can rely on only six members. Moreover, this is a single-model study, but it is ascertained that multi-model analyses outperform the predictions of individual models (Bellucci et al., 2015b). It is expected that an experiment with more ensemble members or more analyzed models would provide better predictions.

- An initialization shock in particular areas of the North Atlantic may lead to internal problems in the model, damaging the representation of the physical processes involved and lowering the possibility to capture the correct signal.

Initialization and Labrador Sea convection

Initialization shocks in full-field initialized models are found in different studies, but very few of them directly study their impact on the AMOC. Bilbao et al. (2020) infer that, in EC-Earth3.3, the AMOC shutdown is as a consequence of an initialization shock in LS convection, which in turn is due to surface and subsurface temperature and salinity which drift towards their preferred mean state at different paces. Thus, in our study the investigation about Labrador Sea (LS) convection has been conducted to understand if the AMOC drift is referable to a reduction of convection. The problem is that we cannot directly link the convection in the LS to the transport represented by the VI5-35N index, because there is no assurance that in the model there is a strong connection between the subpolar and the tropical/subtropical Atlantic. The AMOC between 5-35°N may not respond readily to deep convection variations. Ortega et al. (2021) indeed show that the connectivity

of LS densities with the subtropics is sensitive to different model features, including the mean density stratification in the LS, the depth at which the LS densities propagates southward, and the strength and depth of the AMOC. Therefore, we extracted another AMOC variability index, averaging the MSF over latitudes from 35°N to 55°N. The drift is always present, but with a magnitude of just about 1 Sv. It happens mainly in the first two years, followed by a recovery since year 4 which brings the transport back to the initial value at year 10. While in Bilbao et al. (2020) the AMOC drift appears as a consequence of the convection drift, in our case it is more likely that the AMOC stops before convection. If the stop of convection is a consequence of the AMOC drift, it should be better tested. A possibility could be that, with a reduced northward heat transport, in the subpolar region forms more ice on the sea surface and, acting as insulator, it turns the convection off.

It is true that, even if it seems intuitive to think at LS convection as the process that most directly affects AMOC variations, our knowledge of the problem is still at the beginning and there may be other equally contributing factors. Another area of deep water formation are the Nordic Seas, but there we found a trend in convection similar to that of the LS. The other important area is the Southern Ocean, in which the Antarctic Bottom Water (ABW) is produced. Even if the greatest part of deep waters are generated in the North Atlantic, there are other fundamental mechanisms involved in the AMOC closure: besides ABW formation, the Southern Ocean upwelling (Marshall and Speer, 2012). This is largely driven by winds, and we cannot exclude the possibility of a shock due to errors in wind stress, like that found in the tropical Pacific by Pohlmann et al. (2017).

We have also to remember that, before assimilating the ocean reanalyses, we "regridded" the reconstruction onto the hindcasts coordinates grid, and this could have altered the integrated zonal velocities. Moreover, a study by Kröger et al. (2018) poses the attention on the assimilation phase: in assimilation runs, dynamical changes in the North Atlantic are induced by full-field nudging, resulting in altered mass and heat transport. This causes an initialization shock in hindcasts and degrades ocean heat content predictions in the Subpolar gyre. During our study we have found some inconsistencies between the pattern of meridional transport in hindcasts and of the reanalyses used for assimilation. Hence, we underline the importance of studying the potential impact of these inconsistencies and of improving full-field assimilation strategies.

Drift start-date dependence

When studying the impact of the drift start-date dependence on the forecast skill (figure 4.12), it is interesting to notice that the most impressive results in terms of ACC are obtained when comparing the DPs to RAPID data. However, DPs are correlated with RAPID at a different latitude with respect to REAN, and this limit our capacity to compare the results. Probably the DPs start dates since 2004 are initialized with a most accurate state. It is also true that the reanalyses are based on

observations and thus on RAPID, but we tried to compare DPs with reanalyses from 2004 to 2018 and the result was not equally satisfying as with direct comparison with RAPID. Certainly in the future, as direct observations of the AMOC increase, it will be ever more preferable to base the evaluation of the forecast skill on RAPID data.

Concerning the "moving climatology" drift correction, it may be argued that it is not a real new post-processing method, because we have not implemented a completely diverse strategy with respect to the usual anomalies computation. Instead, we have just selected different periods to compute the climatologies. In any case, it is a method to demonstrate that near-term predictions initialized today (thus which should have practical purpose for our society) can become more reliable if they are adjusted considering only one or two decades from today. This, at least, is true for our model: in general, this strategy could be a test to see whether in other models the drift start-date dependence plays a significant role in the forecast skill.

Comparison with the historical simulation

We compared the DPs with a non-initialized simulation with boundary conditions given from historical forcings (HIST). Since the DPs are initialized near to the observations, we expect that during the run they drift towards HIST, which has been produced with the same model and thus should have the same attractor of the DPs. From the figures 4.3 and 4.11b-d however, it is clear that HIST does not represent the equilibrium state of the model. Bilbao et al. (2020) evaluate the AMOC forecast skill of another full-field initialized model and they use an ensemble of historical simulations to understand which is the preferred state of the model. They find that only some of the historical simulations have a suppressed convection, and that the hindcasts drift precisely towards them. Our HIST have an oscillating convection and from the PDF in figure 4.11d it seems that the HIST state of suppressed convection does not coincide with that of the DPs at equilibrium (years 6-10). However, in Bilbao et al. (2020) the initialization shock was produced exactly in the LS, thus it was clear that their hindcasts had to drift towards "non-convective" historical simulations. The fact that our DPs do not drift towards HIST is in line with these findings, because HIST is produced in a separate run and, as we have demonstrated by means of its ACC with the reanalyses, it simulates random transport variations at decadal scale. Hence, in an ensemble of historical simulations, we expect that they assume different states. The question that arises is: why do the DPs always tend to the same state, even if HIST can be represented by different states? The answer certainly must be searched in initialization and, specifically, in the effect of initial shocks. A limit of this study is that we can rely on only one realization of HIST: an ensemble will be essential to answer this question in future.

Chapter 6

Conclusion

This work provides an assessment of the AMOC predictability in a state-of-the-art decadal prediction system (DPS) based on the CMCC-CM2-SR5 model. We have investigated the AMOC climatology and variability, and assessed the AMOC predictability through deterministic and probabilistic metrics.

The model captures the expected inter-hemispheric circulation cell, with the maximum transport located at 28.5°N and at a depth of 1000 m, in agreement with experimental measures. Nevertheless, the model underestimates the magnitude of the mass transport by about 3 Sv compared to RAPID experimental measures at 26.5°N. Moreover, the model climatology is lead-year dependent, with a stronger drift in the first two years.

The forecast skill, evaluated by means of the Anomaly Correlation Coefficient (ACC), shows a good agreement between DPs and observations only in the first forecast year while, since the second year, the ACC is near-zero. Despite the poor ability of the model to skillfully forecast AMOC variability, the potential predictability (PP) is high, especially in the first half of the decade. This result suggests that the prediction system does not fully exploit predictable components of the AMOC variability and hints an issue in the representation of key processes that drive this variability.

To investigate the causes of the model bias and drift, we focused on Labrador Sea (LS) convection. Indeed, as outlined by Heuzé (2017), models often show limitations in representing convection in the subpolar region. A progressive reduction of convection is evident until year 6, followed by a slight recovery. However, it seems that the reduced convection follows the AMOC suppression and, therefore, we must exclude that it is a primary cause on the AMOC drift, but it could be part of a feedback process.

From the time series of AMOC transport and LS mixed layer depth, it is clear that our realization of the historical simulation does not represent the DPs equilibrium state. The fact that the model's equilibrium state is represented by only some realizations of the historical is also documented in Bilbao et al. (2020). Hence, we may conclude that the initialization forces the model to drift always towards a

particular state. This finding should be corroborated by an analysis of an ensemble of historical runs and the explanation of this behaviour must be searched in the impact of initial shocks, possibly related to potential inconsistencies in meridional transport arising after assimilation.

An important and interesting feature, found also in other studies (e.g. Bilbao et al., 2020), is that the drift has a dependence on the initialization state. This feature can emerge as a detrimental factor in the validation of the forecast, as consolidated metrics used to evaluate the skill typically assume a stationary drift. By developing a novel postprocessing method, indeed we demonstrate that accounting for the non-stationarity leads a higher skill of the forecast that is indicative of an improved prediction over the examined period.

The start-date dependence of the drift is well exploited by the novel postprocessing method that we have experimented, what we called the "moving climatology" strategy. Even though this strategy applied to CMCC-CM2-SR5 DPs does not improve the skill enough to obtain good predictions after the second year, this is a further reason to focus more efforts on the post-processing phase in order to extract as much signal as possible.

As underlined by Scaife and Smith (2018), nowadays large ensembles are required to reduce the unpredictable noise and thus to accurately estimate the predictable (ensemble mean) signal, and a multi-model analysis is essential to obtain statistical robustness and confirm the results. Nonetheless, the findings herein presented suggest the need to improve ocean initialisation in the Atlantic and the representation of physical drivers of AMOC variability in models, but also sheds light on the opportunity to provide better forecasts by accounting a posteriori for model deficiencies. Considerable advances in this direction are expected in the upcoming years.

Bibliography

A. Bellucci et al. “Advancements in decadal climate predictability: The role of nonoceanic drivers”. In: *Reviews of Geophysics* 53.2 (2015), pp. 165–202.

A. Bellucci et al. “An assessment of a multi-model ensemble of decadal climate predictions”. In: *Clim Dyn* 44 (9 2015), pp. 2787–2806.

R. Bilbao et al. “Assessment of a full-field initialised decadal climate prediction system with the CMIP6 version of EC-Earth”. In: *Earth System Dynamics Discussions* 2020 (2020), pp. 1–30.

J. Bjerknes. “Atlantic Air-Sea Interaction”. In: ed. by H.E. Landsberg and J. Van Mieghem. Vol. 10. Advances in Geophysics. Elsevier, 1964, pp. 1–82.

G. J. Boer. “Decadal potential predictability of twenty-first century climate”. In: *Clim Dyn* 36 (2011), pp. 1119–1133.

G. J. Boer. “Long time-scale potential predictability in an ensemble of coupled climate models”. In: *Clim Dyn* 23 (1 2004), pp. 29–44.

G. J. Boer et al. “The Decadal Climate Prediction Project (DCPP) contribution to CMIP6”. In: *Geoscientific Model Development* 9.10 (2016), pp. 3751–3777.

G. Branstator and H. Teng. “Potential impact of initialization on decadal predictions as assessed for CMIP5 models”. In: *Geophysical Research Letters* 39.12 (2012).

S. Brune and J. Baehr. “Preserving the coupled atmosphere–ocean feedback in initializations of decadal climate predictions”. In: *WIREs Climate Change* 11.3 (2020).

H. L. Bryden et al. “Impact of a 30% reduction in Atlantic meridional overturning during 2009/2010”. In: *Ocean Science* 10.4 (2014), pp. 683–691.

M. W. Buckley and J. Marshall. “Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review”. In: *Reviews of Geophysics* 54.1 (2016), pp. 5–63.

L. Caesar et al. “Current Atlantic Meridional Overturning Circulation weakest in last millennium”. In: *Nature Geoscience* 14.3 (2021), pp. 118–120.

C. Cassou et al. “Decadal Climate Variability and Predictability: Challenges and Opportunities”. In: *Bulletin of the American Meteorological Society* 99 (3 2018), pp. 479–490.

T. Chai and R. R. Draxler. "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature". In: *Geoscientific Model Development* 7.3 (2014), pp. 1247–1250.

A. Cherchi et al. "Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model". In: *Journal of Advances in Modeling Earth Systems* 11.1 (2019), pp. 185–209.

CLIVAR. *Drifts and bias correction for decadal climate predictions*. Tech. rep. International CLIVAR Project Office, 2011.

D. P. Dee et al. "The ERA-Interim reanalysis: configuration and performance of the data assimilation system". In: *Quarterly Journal of the Royal Meteorological Society* 137.656 (2011), pp. 553–597.

T. Delworth, S. Manabe, and R. J. Stouffer. "Interdecadal Variations of the Thermohaline Circulation in a Coupled Ocean-Atmosphere Model". In: *Journal of Climate* 6 (11 1993), pp. 1993–2011.

T. Delworth, R. Zhang, and M. E. Mann. "Decadal to Centennial Variability of the Atlantic from Observations and Models". In: 2013.

F. Doblas-Reyes et al. "Initialized near-term regional climate change prediction". In: *Nat Commun* 4 (1 2013).

H. Du et al. "Sensitivity of decadal predictions to the initial atmospheric and oceanic perturbations". In: *Climate Dynamics* 39 (7 2012).

T. Eldevik et al. "Observed sources and variability of Nordic seas overflow". In: *Nature Geoscience* 2 (6 2009).

N. Fofonoff and R. Millard Jr. "Algorithms for the computation of fundamental properties of seawater". In: UNESCO Technical Papers in Marine Sciences; 44 (1983).

E. Frajka-Williams et al. "Atlantic Meridional Overturning Circulation: Observed Transport and Variability". In: *Frontiers in Marine Science* 6 (2019).

J. García-Serrano and F. J. Doblas-Reyes. "On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast". In: *Clim Dyn* 39 (7 2012), pp. 2025–2040.

A. Germe et al. "On the robustness of near term climate predictability regarding initial state uncertainties". In: *Clim Dyn* 48 (2017), pp. 353–366.

S. Good, M. Martin, and N. Rayner. "EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates". In: *Journal of Geophysical Research: Oceans* 118.12 (2013), pp. 6704–6716.

GSOP. *Report on the 10th Session of the CLIVAR Global Synthesis and Observations Panel*. Tech. rep. 2019.

J. Guiot and W. Cramer. “Climate change: The 2015 Paris Agreement thresholds and Mediterranean basin ecosystems”. In: *Science* 354.6311 (2016), pp. 465–468.

S. Ham, S.-Y. Hong, and S. Park. “A study on air–sea interaction on the simulated seasonal climate in an ocean–atmosphere coupled model”. In: *Clim Dyn* 42 (5 2014), pp. 1175–1187.

E. Hawkings and R. Sutton. “The Potential to Narrow Uncertainty in Regional Climate Predictions”. In: *Bulletin of the American Meteorological Society* 90 (8 2009), pp. 1095–1108.

C. Heuzé. “North Atlantic deep water formation and AMOC in CMIP5 models”. In: *Ocean Sci.* 13 (4 2017), pp. 609–622.

J. Hirschi et al. “A monitoring design for the Atlantic meridional overturning circulation”. In: *Geophysical Research Letters* 30.7 (2003).

J.W. Hurrell et al. “Decadal Climate Predictions: Opportunities and Challenges”. In: *Proceedings of OceanObs’09: Sustained Ocean Observations and Information for Society*. ESA Publication, 2010.

M. Ishii and M. Kimoto. “Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections”. In: *Journal of Oceanography* 65 (3 2009), pp. 287–299.

W. E. Johns, T. Kanzow, and R. Zantopp. “Estimating ocean transports with dynamic height moorings: An application in the Atlantic Deep Western Boundary Current at 26°N”. In: *Deep Sea Research Part I: Oceanographic Research Papers* 52.8 (2005), pp. 1542–1567.

W. E. Johns et al. “Continuous, Array-Based Estimates of Atlantic Ocean Heat Transport at 26.5°N”. In: *Journal of Climate* 24.10 (2011), pp. 2429–2449.

E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge Univ. Press, 2009.

T. Kanzow et al. “Seasonal Variability of the Atlantic Meridional Overturning Circulation at 26.5°N”. In: *Journal of Climate* 23 (21 2010).

A. R. Karspeck et al. “An Ensemble Adjustment Kalman Filter for the CCSM4 Ocean Component”. In: *Journal of climate* 26 (19 2013), pp. 7392–7413.

N. S. Keenlyside et al. “Advancing decadal-scale climate prediction in the North Atlantic sector”. In: *Nature* 463 (7191 2008), pp. 957–982.

Y. Kostov, K. C. Armour, and J. Marshall. “Impact of the Atlantic meridional overturning circulation on ocean heat storage and transient climate change”. In: *Geophysical Research Letters* 41.6 (2014), pp. 2108–2116.

J. Kröger et al. “Full-field initialized decadal predictions with the MPI earth system model: an initial shock in the North Atlantic”. In: *Clim Dyn* 51 (7 2018), pp. 2593–2608.

T. Kuhlbrodt et al. “On the driving processes of the Atlantic meridional overturning circulation”. In: *Reviews of Geophysics* 45.2 (2007).

A. Kumar. “On the Interpretation and Utility of Skill Information for Seasonal Climate Predictions”. In: *Monthly Weather Review* 135 (5 2007).

Y. Kushnir et al. “Towards operational predictions of the near-term climate”. In: *Nature Climate Change* 9 (2 2019).

M. Latif et al. “A Review of Predictability Studies of Atlantic Sector Climate on Decadal Time Scales”. In: *Journal of Climate* 19.23 (2006), pp. 5971–5987.

S. Levitus. *Climatological Atlas of the World Ocean*. NOAA Professional Paper 13, 1982.

J. Marshall and F. Schott. “Open-ocean convection: Observations, theory, and models”. In: *Reviews of Geophysics* 37.1 (1999), pp. 1–64.

J. Marshall and K. Speer. “Closure of the meridional overturning circulation through Southern Ocean upwelling”. In: *Nature Geoscience* 5 (3 2012), pp. 171–180.

G. Meehl et al. “Decadal Climate Prediction: An Update from the Trenches”. In: *Bulletin of the American Meteorological Society* 95.2 (2014), pp. 243–267.

G. Meehl et al. “Decadal Prediction: Can It Be Skillful?” In: *Bulletin of the American Meteorological Society* 90.10 (2009), pp. 1467–1486.

G. A. Meehl et al. *Global climate projections. Chapter 10*. 2007.

B. I. Moat et al. “Pending recovery in the strength of the meridional overturning circulation at 26°N”. In: *Ocean Science* 16.4 (2020), pp. 863–874.

D. P. Mulholland et al. “Origin and Impact of Initialization Shocks in Coupled Atmosphere–Ocean Forecasts”. In: *Monthly Weather Review* 143 (11 2015), pp. 4631–4644.

P. Ortega et al. “Labrador Sea subsurface density as a precursor of multidecadal variability in the North Atlantic: a multi-model study”. In: *Earth System Dynamics* 12.2 (2021), pp. 419–438.

P. Ortega et al. “Mechanisms of decadal variability in the Labrador Sea and the wider North Atlantic in a high-resolution climate model”. In: *Clim Dyn* 49 (2017), pp. 2625–2547.

OSNAP. *Measuring the Meridional Overturning Circulation from the OSNAP array*. Tech. rep. 2016.

H. Pohlmann et al. “Initialization shock in decadal hindcasts due to errors in wind stress over the tropical Pacific”. In: *Clim Dyn* 49 (7 2017), pp. 2685–2693.

H. Pohlmann et al. “Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic”. In: *Journal of Climate* 22 (14 2009), pp. 3926–3938.

D. Rayner et al. “Monitoring the Atlantic meridional overturning circulation”. In: *Deep Sea Research Part II: Topical Studies in Oceanography* 58.17 (2011), pp. 1744–1753.

R. W. Reynolds et al. “Daily High-Resolution-Blended Analyses for Sea Surface Temperature”. In: *Journal of Climate* 20 (22 2007), pp. 5473–5496.

J. Robson, R. T. Sutton, and D. M. Smith. “Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid 1990s”. In: *Geophysical Research Letters* 39.19 (2012).

J. Robson et al. “Atlantic overturning in decline?” In: *Nature Geoscience* 7 (1 2014).

J. Robson et al. “Causes of the Rapid Warming of the North Atlantic Ocean in the Mid-1990s”. In: *Journal of Climate* 25.12 (2012), pp. 4116–4134.

J. Robson et al. “Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model”. In: *Clim Dyn* 50 (2018), pp. 921–937.

T. Rossby. “The North Atlantic Current and surrounding waters: At the crossroads”. In: *Reviews of Geophysics* 34.4 (1996), pp. 463–481.

G. Roullet and G. Madec. “Salt conservation, free surface, and varying levels: A new formulation for ocean general circulation models”. In: *Journal of Geophysical Research: Oceans* 105.C10 (2000), pp. 23927–23942.

E. Sanchez-Gomez et al. “Drift dynamics in a coupled model initialized for decadal forecasts”. In: *Clim Dyn* 46 (2016), pp. 1819–1840.

A. Scaife and D. Smith. “A signal-to-noise paradox in climate science”. In: *npj Clim Atmos Sci* 1 (1 2018).

Jr. Schmitz and J. William. “On the interbasin-scale thermohaline circulation”. In: *Reviews of Geophysics* 33.2 (1995), pp. 151–173.

E. Scoccimarro, S. Gualdi, and A. Bellucci. “Effects of Tropical Cyclones on Ocean Heat Transport in a High-Resolution Coupled General Circulation Model”. In: *Journal of Climate* 24 (16 2011), pp. 4368–4384.

F. Sévellec and B. Sinha. “Predictability of Decadal Atlantic Meridional Overturning Circulation Variations”. In: *Oxford Encyclopedia of Climate Science, Oxford University Press*. Oxford University Press, Sept. 2018.

F. Sévellec et al. “Optimal Surface Salinity Perturbations of the Meridional Overturning and Heat Transport in a Global Ocean General Circulation Model”. In: *Journal of Physical Oceanography* 38 (12 2008), pp. 2739–2754.

D. M. Smith, Eade R., and H. Pohlmann. “A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction”. In: *Clim Dyn* 41 (2013), pp. 3325–3338.

D. M. Smith et al. “Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model”. In: *Science* 317.5839 (2007), pp. 796–799.

D. M. Smith et al. “North Atlantic climate far more predictable than models imply”. In: *Nature* 583 (7818 2020), pp. 796–800.

D. M. Smith et al. “Predicted Chance That Global Warming Will Temporarily Exceed 1.5°C”. In: *Geophysical Research Letters* 45.21 (2018), pp. 895–903.

D. M. Smith et al. “Real-time multi-model decadal climate predictions”. In: *Clim Dyn* 41 (2012), pp. 2875–2888.

M. A. Srokosz and H. L. Bryden. “Observing the Atlantic Meridional Overturning Circulation yields a decade of inevitable surprises”. In: *Science* 348.6241 (2015).

A. Storto and S. Masina. “C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis”. In: *Earth System Science Data* 8.2 (2016), pp. 679–696.

S. H. Strogatz. *Nonlinear dynamics and chaos with applications to physics, biology, chemistry and engineering.* Ed. by Perseus book. 1994.

D. Swingedouw et al. “Bidecadal North Atlantic ocean circulation variability controlled by timing of volcanic eruptions”. In: *Nat Commun* 6 (1 2015).

K.E. Taylor, R.J. Stouffer, and G.A. Meehl. “An overview of CMIP5 and the experiment design”. In: *Bull. Amer. Meteor. Soc.* 93 (4 2012), pp. 485–498.

G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation.* Cambridge University Press, 2019.

D. S. Wilks. *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Elsevier, 2006.

C. J. Willmott and K. Matsuura. “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance”. In: *Clim Res* 30 (1 2005), pp. 79–82.

X. Yan, R. Zhang, and T. R. Knutson. “Underestimated AMOC Variability and Implications for AMV and Predictability in CMIP Models”. In: *Geophysical Research Letters* 45.9 (2018), pp. 4319–4328.

C. Yang, S. Masina, and A. Storto. “Historical ocean reanalyses (1900–2010) using different data assimilation strategies”. In: *Quarterly Journal of the Royal Meteorological Society* 143.702 (2016), pp. 479–493.

S. Yeager, W. Kim, and J. Robson. “What caused the Atlantic cold blob of 2015?” In: *US Clivar Variations Newsletters* 14.2 (2016).

S. Yeager and J. Robson. “Recent Progress in Understanding and Predicting Atlantic Decadal Climate Variability”. In: *Current Climate Change Reports* 3 (2 2017), pp. 112–127.

S. Yeager et al. “A Decadal Prediction Case Study: Late Twentieth-Century North Atlantic Ocean Heat Content”. In: *Journal of Climate* 25.15 (2012).

S. Yeager et al. “Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model”. In: *Bulletin of the American Meteorological Society* 99 (9 2018), pp. 1867–1886.

S. Yu and M. Pritchard. “A Strong Role for the AMOC in Partitioning Global Energy Transport and Shifting ITCZ Position in Response to Latitudinally Discrete Solar Forcing in CESM1.2”. In: *Journal of Climate* 32 (8 2019), pp. 2207–2226.

R. Zhang et al. “A Review of the Role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts”. In: *Reviews of Geophysics* 57.2 (2019), pp. 316–375.