https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions

A causality-based method for multi-model comparison: Application to relationships between atmospheric and marine biogeochemical variables

BÉNARD Germain[1], GEHLEN Marion[1], and VRAC Mathieu[1]

[1]Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), CEA/CNRS/UVSQ, Université Paris-Saclay, Centre d'Etudes de Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France

Correspondence: Germain Bénard (germain.benard@lsce.ipsl.fr)

Abstract. We introduce an novel approach to compare Earth System Model output using a causality-based approach. The method is based on the PCMCI+ algorithm, which identifies causal relationships between multiple variables. We aim to investigate the causal relationships between atmospheric (North Atlantic Oscillation - NAO), oceanic (gyre strength, stratification, circulation), and biogeochemical variables (nitrate, iron, silicate, net primary production) in the North Atlantic subpolar gyre, a critical region for the global climate system with a well characterised multi-year variability in physical and biogeochemical properties in response to the North Atlantic Oscillation. We test a specific multivariate conceptual scheme, involving causal links between these variables. Applying the PCMCI+ method allows us to differentiate between the influence of vertical mixing and horizontal advection on nutrient concentrations, spring bloom intensity, as well as to highlight model-specific dynamics. The analysis of the causal links suggests a dominant contribution of vertical mixing to peak spring bloom intensity compared to transport. The strength of the links is variable among models. Stratification is identified as an important factor controlling spring bloom NPP in some, but not all, models. Horizontal transport also significantly influences biogeochemistry. However, horizontal transport generally exhibits lower contributions than vertical mixing. Most of the links found are model-specific, hence likely contributing to inter-model spread. The limitations of the method are discussed and directions for future research are suggested.

1 Introduction

Earth System models (ESMs) are essential tools in climate science. They are designed to unravel the intricate workings of the planet's climate system. The simulation of historical and present dynamics have proven to be skillful in many aspects, including the representation of ocean physics and marine biogeochemistry (Flato et al., 2014; Séférian et al., 2020; Vautard et al., 2021; Tsujino et al., 2020). Despite advances in complexity, ESMs simplify the real world and this is a source of errors and uncertainties. These models create their own climate system shaped by specific spatial resolutions and parameterizations of various physical, chemical or biogeochemical processes. Multiple reasons explain the differences in architecture between models (Bonan and Doney, 2018): the constraint of computational efficiency which calls for the parameterization of small-scale processes unresolved by the large-scale variables in the model, but also the selection of which processes to include. Given the

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024


Earth System Dynamics Discussions Open Access EGU

inherent complexity of the climate system, encompassing myriad of processes currently impossible to be represented in a model, these disparities can lead to differences in future climate states and dynamics leading to projections with still scattered outcomes (e.g., Kwiatkowski et al., 2020; Tebaldi and Knutti, 2007; Flato et al., 2014; Zelinka et al., 2020).

Intercomparing model output is essential to characterize inter-model differences and to identify the uncertainty around projected values. Most model intercomparison studies still rely on the analysis of the spatial and temporal distribution of selected model outputs presented as maps of mean values, their range, along with the computation of statistics (e.g. extremes, variability) (e.g., Wang et al., 2020; Flato, 2011; Jacob et al., 2007; Chen and Knutson, 2008). For intercomparisons targeting the historical period, observations or reanalyses are commonly used as references. Model performance is evaluated against real-world data or derived products and bias maps are used to locate model differences (e.g., Séférian et al., 2020; Yool et al., 2021; Chen and Knutson, 2008; Schaller et al., 2011). Yet, none of these methods characterises the interactions between the variables or quantifies their inter-model differences. Hence, correlation-based approaches have been developed to try to overcome these limitations (e.g., Charakopoulos et al., 2018; Anagnostopoulos et al., 2010; Gleckler et al., 2008). However, if correlation allows the identification of co-variation between variables or phenomena, it fails to identify causal relationships.

To overcome the limitations of correlation-based approaches, Krich et al. (2020) and Nowack et al. (2020) employed a causality approach to analyse or evaluate model output. Here, we extend these earlier studies to the analysis and comparison of multi-model output by adopting a causality framework. This approach allows to describe model differences in a novel way, potentially revealing model-specific dynamics, quantifying the strength of interactions and its range across Earth System Models.

Our study focuses on the subpolar North Atlantic, a critical region for the global climate system with a well characterised multi-year variability in physical and biogeochemical properties in response to the North Atlantic Oscillation, the dominant mode of regional climate variability. At the seasonal time scale, deep winter mixing replenishes the sunlit surface ocean in nutrients and sustains an intense spring bloom. The subpolar North Atlantic is also a region undergoing rapid changes including freshening and cooling (Tesdal et al., 2018; Holliday et al., 2020), with potential impacts on large-scale circulation (Fox-Kemper et al., 2021; Hakkinen and Rhines, 2004). Along with these physical changes, an important variability in surface ocean nutrient levels has been documented (Johnson et al., 2013; Hátún et al., 2017) potentially foreshadowing future changes in primary productivity under climate change.

The impact of climate change on net primary production (NPP) in the North Atlantic region remains highly uncertain with a larger spread between model projections compared to other regions (Tagliabue et al., 2021; Fu et al., 2022). While vertical mixing is crucial for injecting nutrients to the productive surface ocean and fuelling primary production (D'Asaro, 2008; Williams et al., 2006), recent studies highlighted the contribution of horizontal transport to observed and projected variability of nutrients and primary production (Whitt and Jansen, 2020; Kwiatkowski et al., 2020; Hátún et al., 2017). The eastern part of the subpolar North Atlantic was proposed by Hátún et al. (2017) as a key area for the mixing of nutrient-poor subtropical waters into the subpolar North Atlantic gyre. The variability of the gyre circulation would be one mechanism controlling the advection and mixing of nutrients and ultimately NPP at regional scale. Moreover, (Pelegrí et al., 1996; Williams et al., 2011)

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024


Earth System Dynamics Discussions Open Access EGU

highlight the contribution of subsurface nutrient-rich waters of subtropical origin (the nutrient streams) as a source of nutrients to the subpolar Atlantic through vertical mixing.

This study targets the eastern subpolar North Atlantic and seeks to differentiate between the variability of vertical mixing and horizontal advection of nutrients as controls of fluctuations of NPP. It draws on an ensemble of Earth System Model simulations to explore causal links from atmospheric processes to marine biogeochemistry. We rely on the "PCMCI+" method (Runge et al., 2019) for the analysis of causal interactions represented as causal graphs. In these graphs, nodes represent variables and edges indicate potential causal links with associated strengths. Our objective is twofold: (1) to propose a novel approach for model intercomparison, (2) to understand model-specific causal links between variability in atmospheric processes, advection and vertical mixing of nutrients to the surface ocean and NPP, as well as the consequences on inter-model spread.

This article is structured as follows: The next section 2 describes the causality discovery algorithm we use as well as the conceptual scheme that guides our study. Section 3 details the data used and the pre-processing steps applied to each variable. Section 4 presents the results obtained, highlighting specific key links and their similarities and differences between models. Section 5 discusses the limitations of the method employed, providing details on its strengths and weaknesses. Results are discussed in section 6 and section 7 concludes this article as well as proposing future perspectives.

2 Methodology

2.1 Causality approach: PCMCI+

To investigate the causal links among the variables in our study, we use the PCMCI+ method (Runge et al., 2019; Runge, 2020). PC stands for “Peter & Clark”, the first step of PCMCI+ (Spirtes and Glymour, 1991), and MCI stands for “Momentary Conditional Independence”, the other step of PCMCI+. This method is based on Granger-Causality (Granger, 1969). Granger Causality examines whether past values of variable A enhances the prediction of B more than using B’s own past values. In such a case, it is said that A Granger-causes B. PCMCI+ relies on Granger causality to create a causal graph by detecting contemporaneous or lagged relationships between the (physical or biogeochemical) variables, noted X^i from now on. These variables, within the PCMCI+ framework, become the nodes in the causal graph. The edges of the graph correspond to the contemporaneous or lagged links. A causal link between two nodes has a lag (0 if contemporaneous) and a strength ranging from -1 to 1. The strength is the result of a correlation so if A causes B with a certain strength, the sign indicates if the relationship is positive or negative and the value indicates the intensity of the correlation. The graphs provided by PCMCI+ for the different climate models will be compared to the conceptual scheme discussed in Section 2.3 and intercompared to each other.

To investigate lagged interactions, we set a maximum lag value T. This means that all causal links established will have a lag within $[\![0, T]\!]$. To consider the lagged interactions, each variable X^i is considered T times. For each timestep $t \in [\![0, T]\!]$, a new set of variables X^i_t is obtained, by considering the variable X^i shifted by t timesteps. Then, for each X^i_0 a first selection of possible candidate explanatory variables is done based on correlation tests. The variables in the selection are ranked in terms of decreasing correlation order. These explanatory variables able to “cause” X^i_0 are called the parents of X^i_0.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions
Open Access
EGU

Before explaining the PCMCI+ algorithm in details, it is important to remind the concept of conditional independence of A and B given C (Dawid, 1979), defined by Equation (1). It describes the irrelevance of the event B to explain the event A when conditioned by C:

$$A \perp\!\!\!\perp B|C \iff P(A|B,C) = P(A|C) \text{ with } A,B,C \text{ being events.} \tag{1}$$

The first step of PCMCI+, PC, consists in an iterative process where the conditional independence of two variables given a third one is tested (Spirtes and Glymour, 1991). More concretely, let X, Y and Z be three variables and $t_i, t_j \in [\![0,T]\!]$. We want to test if variable $X_{\tau=0}$ is still related to $Y_{\tau=t_i}$ when the influence of the most significant variable (according to the ranking previously made) to $X_{\tau=0}$ (e.g., $Z_{\tau=t_j}$) is removed. This is achieved through Partial Correlation (noted $ParCorr$ in equation (2)), to test the conditional independence of $X_{\tau=0}$ and $Y_{\tau=t_i}$ given $Z_{\tau=t_j}$, by conducting a regression on each variable:

$$ParCorr(X_{\tau=0}, Y_{\tau=t_i}|Z_{\tau=t_j}) = Corr(\epsilon_{X_{\tau=0}|Z_{\tau=t_j}}, \epsilon_{Y_{\tau=t_i}|Z_{\tau=t_j}}) \tag{2}$$

with

$$\begin{cases} X_{\tau=0} = \alpha_{X_{\tau=0}} * Z_{\tau=t_j} + \epsilon_{X_{\tau=0}|Z_{\tau=t_j}} \\ Y_{\tau=t_i} = \alpha_{Y_{\tau=t_i}} * Z_{\tau=t_j} + \epsilon_{Y_{\tau=t_i}|Z_{\tau=t_j}} \end{cases} \tag{3}$$

and $Corr()$ being the Pearson correlation. Applied to our field of interest, let's consider an example where we have nitrate (X) related to the gyre strength (Y) and Mixed Layer depth (Z). The algorithm's purpose is to determine if nitrate and the gyre strength remain correlated when the influence of mixed layer depth is removed. The first iteration of the PC step tests, for each variable, the conditional independence with all the possible parents (explanatory variables) given the most significant parent. The subsequent iterations of PC test the conditional independence of each parent according to the k most significant variables (according to a new ranking made at the end of the previous iteration). Regressions are made with k variables. False Discovery Rates (hereafter FDR, Benjamini and Hochberg, 1995) are then used to find the equivalent of a p-value that will be the threshold to keep only the most significant links.

At the end of PC, each variable has a set of parents, corresponding to the possible explanatory variables. The concluding step MCI ("Momentary Conditional Independence") can then start, testing if the "grand-parents" influence the variable. For each variable, the conditioning variables are not taken among the parents but among the parents of the parents. Testing the grand-parents influence allows to investigate more deeply the causal relationships and to see if more distant variables have an influence.

2.2 Quantification of causal graph: Dissimilarity measure

The output of PCMCI+ is a causal graph with a set of nodes (the variables given to PCMCI+) and potential edges (i.e., causal relationships) between those nodes. We can look at the PCMCI+ results via the adjacency matrix M representing the presence and strength of links between nodes. To compare the graphs (i.e., the relationships within models), it is then possible to compute dissimilarity metrics between adjacency matrices. However, many of these measures have limitations that make

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions

them inappropriate in our case. Firstly, the methods often assume edge's weight to be positive (e.g., Wicker et al., 2013; Koutra et al., 2013), which is not suitable in our case, as our edges can be positive or negative. The second problem is that the lag associated to the link also has to be taken into account. Even if different, when the lags are close, the distance should be small. In the other hand, once we exceed a certain difference in terms of lags, we can consider the link already too different that an even bigger difference in terms of lags will not give a bigger distance. To address these limitations, we propose an alternative approach that allows for negative weights and lags.

An Euclidean distance (as dissimilarity) allows us to deal with the weights of positive and negative edges noted w_{ij} where $i, j \in [\![0, N]\!]$. To account for the lags, we extend the matrices M from size $N \times N$ to size $N \times N \times T$ by adding a third dimension, indicating the lag at which the link occurs:

$$M[i,j,t] = \begin{cases} w_{ij}, & \text{if there is a link from } i \text{ to } j \text{ with a lag } t \\ 0, & \text{otherwise.} \end{cases}$$

Next, each vector $M[i,j]$ is blurred with a Gaussian filter convolution. The kernel K is a gaussian blur 3x1 ($[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}]$). We note M' the matrix after the application of the Gaussian filter convolution. This approach helps to consider the lag difference by blurring the links in the third dimension, similar to how a blur would make two red squares on blank images share some red pixels if they are close enough. If we consider an example where $M[i,j] = [w_{ij}, 0, 0, 0]$ (i.e., a relationship from i to j only at $t = 0$), the transformed vector $M'[i,j]$ is thus given by:

$$M'[i,j] = M[i,j] * [\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}] = [w_{ij}, 0, 0, 0] * [\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}] = [\tfrac{w_{ij}}{4}, \tfrac{w_{ij}}{2}, \tfrac{w_{ij}}{4}, 0, 0, 0].$$

The Euclidean distance can then be applied to the blurred vectors/matrices, considering both the variations in lags and strengths. Our new dissimilarity measure DIS between two (blurred) adjacency matrices M_1 and M_2 can thus be written as:

$$DIS(M_1, M_2) = \sqrt{\sum_{(i,j,k)} (M_1'[i,j,k] - M_2'[i,j,k])^2}. \quad (4)$$

With this dissimilarity, we take into account how different the lags are for the same link while also taking into account the difference in intensity (strength of interaction).

2.3 Conceptual scheme

We selected the Eastern subpolar North Atlantic (Figure 1) as a test case for this study focusing on the comparison of physical-chemical drivers of spring bloom dynamics in an ensemble of ESMs through a causality approach. The region is known for its large variability in contemporary nutrient concentrations (Johnson et al., 2013; Hátún et al., 2017). The latter has been linked to the strength of the subpolar gyre circulation through its control on the inflow of nutrient poor subtropical waters (Hátún et al., 2017). Fluctuations in gyre circulation, in turn, are driven by regional atmospheric forcing, leading to a causal chain from atmospheric processes to ocean dynamics and finally biogeochemistry.

This region is also prone to great sensitivity to climate change with a high spread of projected values of primary production. However, the mechanisms behind the projected impacts of climate change on NPP are not well understood(Kwiatkowski et al.,

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions
Open Access
EGU

2019; Tagliabue et al., 2021). While vertical mixing is crucial for injecting nutrients to the productive surface layers and fueling NPP (e.g. deepening of the mixed layer during winter for nutrient supply and spring stratification for the initiation of the bloom), changes in subsurface nutrients brought by horizontal transport are also important and are likely to contribute to the between-model spread of projected NPP. (Whitt and Jansen, 2020; Kwiatkowski et al., 2020) (Johnson et al., 2013).
155 Our conceptual scheme aims to highlight model-specific representations of the chain of causality running from atmospheric forcing to NPP. It attempts to disentangle the respective contributions of variability in vertical mixing and horizontal transport to modelled fluctuations of nutrient levels and NPP in the Eastern North Atlantic basin (Fig. 1).

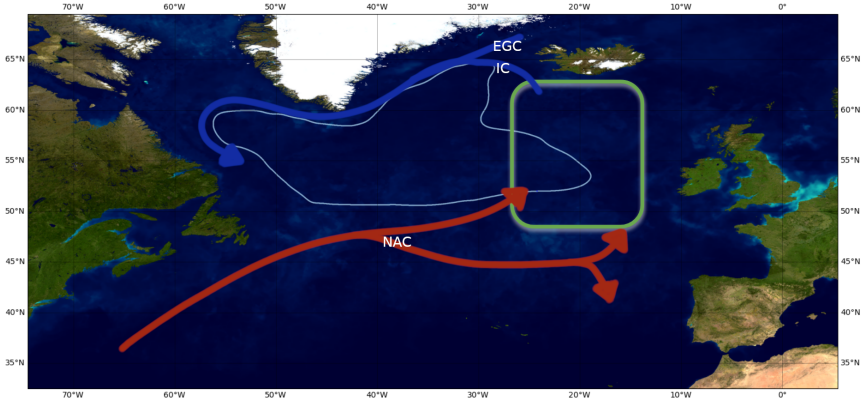

Figure 1. Study area "Eastern North Atlantic Subpolar Gyre" (green square). The main surface circulation patterns are indicated by the arrows (inspired by Daniault et al. (2016)). IC stands for Irminger Current, EGC for Eastern Greenland Current and NAC for North Atlantic Current. The gyre corresponds to the zone in light blue.

The anticipated chain of causality is represented as a conceptual scheme in Fig. 2. The target variable is the NPP of the spring bloom. Predictors of its variability include atmospheric and ocean physical processes, but also nutrient concentrations (nitrate,
160 silicate and dissolved iron). Atmospheric processes are represented by the North Atlantic Oscillation (NAO), a prominent mode of regional climate variability and an important driver of variability of ocean physical and biogeochemical dynamics (Yamamoto et al., 2020; Feucher et al., 2022; Keller et al., 2012; Oschlies, 2001; Herceg-Bulić and Kucharski, 2014; Delworth and Zeng, 2016). Physical processes included are vertical and horizontal transports. Winter-time deepening of the mixed layer injects nutrients to the surface ocean, replenishing the pre-bloom nutrient stock.(Williams et al., 2006). Stratification during
165 spring and the shoaling of the mixed layer contribute to initiate the spring bloom. Mixed Layer Depth (MLD) and stratification are thus chosen as predictors of vertical transport. For the advection of nutrients we will focus on the net transport into the area of study, as well as consider the strength of the North Atlantic subpolar gyre. Lastly, three nutrients are considered: nitrate, silicate and dissolved iron.

We anticipate numerous interconnections across the scheme. The NAO is known to exert influence over multiple processes
170 in the North Atlantic affecting physical variables such as gyre strength, water transport and MLD (Yamamoto et al., 2020; Feucher et al., 2022), but also nutrient concentrations. Diving into the oceanic variables, the gyre strength, relevant for physical

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions
Open Access
EGU

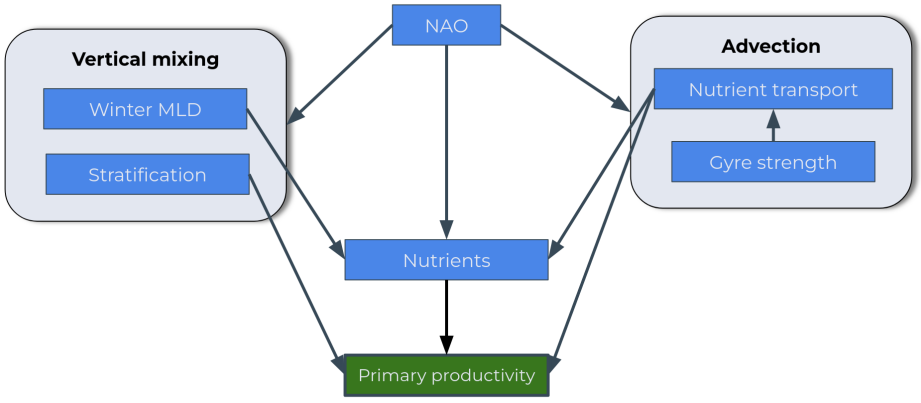

Figure 2. Suggested conceptual scheme. The target variable is in green, the explanatory variables are in blue.

variables such as MLD (Swingedouw et al., 2021), could as well be explanatory for nutrient concentrations (Hátún et al., 2017; Johnson et al., 2013) through its impact on nutrient transport. For the vertical mixing component of the scheme, a link from MLD to nutrient concentrations is anticipated, because winter MLD deepening is crucial for replenishing the nutrient stock (Williams et al., 2006). We also expect a relationship between stratification and NPP because high stratification during the productive period will inhibit the upward mixing of nutrients and thus NPP (Lozier et al., 2011; Tagliabue et al., 2021). The scheme will be explored for each nutrient (nitrate, dissolved iron and silicate) and we anticipate differences in the intensity of the relationships for each nutrient, especially for the relationship between nutrient and productivity.

3 Experimental setup

3.1 Selection of Earth System Models and simulations

We selected a set of Earth system models with biogeochemical models of similar complexity to explore the variability of causal links connecting atmospheric processes to NPP. These biogeochemical models include 4 plankton functional types (PFT) (2 phyto- and 2 zoo-plankton groups) with model-specific parameterisations of interactions between PFTs, as wall as nutrient limitation of phytoplankton growth. We made sure not to repeat the coupling of ocean and biogeochemical models twice. This led to the selection of 5 Earth System Models presented in table 1. There is nevertheless some redundancy in our selection with 4 ESMs sharing the same ocean model (3 with the same version).

We selected pre-industrial control simulations (piControl) from the 6th Coupled Model Intercomparison Project (CMIP6) (Eyring et al., 2016). These simulations were run for 500 years with constant external forcings (e.g. volcanoes, solar radiation, greenhouse gases) fixed at pre-industrial level. It allows access to an extented period for assessing model specificites through causal graphs. Due to the absence of variability and trend in external forcings, climate variability is restricted to its unforced internal component. Interactions between variables should not evolve during the 500 years of simulation.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions Open Access EGU

Table 1. Models used in the study.

ESM model	Oceanic model	Oceanic grid	Atmospheric model	Atmospheric grid	Biogeochemical model	ESM reference
IPSL-CM6A-LR	NEMO-OPA 3.6	1° x 1° with 75 levels	LMDZ	1.3° x 2.5° with 79 levels	PISCES-v2	Boucher et al. (2020)
CESM2	POP2	1°x1.125° with 60 levels	CAM6	0.9°x 1.25° with 32 levels	MARBL	Danabasoglu et al. (2020)
CMCC-ESM2	NEMO-OPA 3.6	1° x 1° with 50 levels	CAM5.3	0.9° × 1.25° with 30 levels	BFM5.1	Lovato et al. (2022)
CanESM5-CanOE	CanNEMO 3.4.1	1° x 1° with 45 levels	CanAM5	isotropic triangular 2.8° grid with 49 level	CanOE	Swart et al. (2019); Christian et al. (2021)
UKESM1-0-LL	NEMO-OPA 3.6	1° x 1° with 75 levels	MetUM-HadGEM3-GA7.1	1.25° x 1.88° gridpoints with 85 levels	MEDUSA2	Sellar et al. (2019); Yool et al. (2021)

3.2 Definition and preprocessing of variables for PCMCI+

We compute time series of annual means from monthly output fields for each model variable. From these time series variables for PCMCI+ are computed as described below. Each processed variable is normalized prior to running PCMCI+. Variables selected for PCMCI+ and their definition are presented in table 2. The following paragraphs add details to the preprocessing. With the exception of transport of nutrients and the subpolar gyre index which are computed on the original model grid, all other variables are interpolated on a regular 1°x1° grid prior to preprocessing.

3.2.1 Bloom and productivity

We focus on the variability of maximum of NPP reached during the year. This corresponds to the maximum reached during the spring months, the spring bloom. For several other variables (e.g. atmospheric forcing, nutrient transport, gyre circulation) it is important to time the start of the bloom as their variability during the bloom might impact bloom dynamics. Following the threshold method in (Brody et al., 2013) we consider the bloom has started once a certain threshold has been exceeded. This threshold value corresponds to the median NPP plus 5%. The starting date for the bloom is the first month m verifying Eq. (5):

$$intpp(m) > 1.05 \times median(intpp). \quad (5)$$

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions
Open Access
EGU

Table 2. Variables and their formulation. The variables used in PCMCI+ are in bold.

Full name	Abbreviation	Formulation	Period	Grid
Net Primary Productivity	**intpp**	$Max(\int_{depth} pp)$	Peak of the bloom	Regular Grid
Nutrients	**no3, dfe, si**	$Max([nutrients]_{0m-100m})$	Peak reached before the bloom	Regular Grid
Mixed Layer Depth	**mlotstmax**	$Max(mlotst)$	Peak reached before the bloom	Regular Grid
North Atlantic Oscillation	**NAO**	$slp_\alpha - slp_\beta$	From the start of the bloom to the peak of the bloom	Regular Grid
Sea Level Pressure of the high pressure zone	slp_α	$\alpha = \{(i,j) \mid \lvert\{t \mid slp(i,j,t) > P_{90}(t)\}\rvert > 0.9 * T\}$ with (i,j) indicating a grid point, and $P_{90}(t)$ the 90^{th} percentile of slp at time t		Regular Grid
Sea Level Pressure of the low pressure zone	slp_β	$\beta = \{(i,j) \mid \lvert\{t \mid slp(i,j,t) < P_{10}(t)\}\rvert > 0.9 * T\}$ with (i,j) indicating a grid point, and $P_{10}(t)$ the 10^{th} percentile of slp at time t		Regular Grid
Strength of the gyre	**Gyre**	$Mean(\psi_{7.5}) \quad where \quad \psi_{7.5} = \{\psi > 7.5 \ Sverdrup\}$ with ψ the barotropic streamfunction	From the start of the bloom to the peak of the bloom	Original Grid
Barotropic Streamfunction	ψ	$\int_{West}^{East} \int_{-2000m}^{0m} v dx dz$ with v the meridional component of the velocity		Original Grid
Stratification	**Strati**	$\rho_{100} - \rho_0$ with ρ_z the density at depth z	From the start of the bloom to the peak of the bloom	Regular Grid
Transport	**Trsp**	$\sum_{(i,j)\in B} Max(\int_{MLD}^{0m} vel[nutrient]dz, 0)$ with B the group of bordering points of the study area and vel the velocity, meridional or zonal component according to the orientation of the border)	From the start of the bloom to the peak of the bloom	Original Grid

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024


Earth System Dynamics Discussions
Open Access
EGU

3.2.2 Nutrients

The study focuses on nitrate, silicate, and dissolved iron before the beginning of the bloom to qualify the biogeochemical context of the spring bloom. Specifically, we will consider the average concentration in the upper 100 meters of the water column. When the bloom starts, nutrient concentrations of the pre-bloom period begin to decline. The timing of the spring bloom varies across different regions and models, as does the timing of the maximum nutrient concentration. Therefore, we will define our annual signal as the peak (i.e., maximum) concentration reached during the winter-spring season. This approach will allow us to capture the yearly variability in nutrient concentrations associated with the intensity of the spring bloom.

3.2.3 North Atlantic Oscillation

The NAO (North Atlantic Oscillation) index quantifies the difference between the high and low-pressure zones in the North Atlantic region. However, the location of the high and low-pressure centers varies among different climate models, making it necessary to use a model-independent method for its computation. Historically the NAO was computed as the difference of Sea Level Pressure (slp) between two specific stations (Lisbon, Portugal and Stykkisholmur, Iceland) (Hurrell et al., 2003). Nowadays, the use of EOF over a large geographic area has become common (Hurrell et al., 2003; Hurrell and Deser, 2010). We preferred using the method consisting in computing the difference of slp on two selected zones (Hurrell, 1995; Hurrell et al., 2003) but the selection of the zone needs to be model independent.

Our proposed solution to select the zones here uses a counting approach. At each time step, the points belonging to the higher and lower 10_{th} percentiles of sea level pressure are identified. For each of these points, the counting index for the high and low pressure is incremented accordingly. As a result, each point is associated with two numbers α and β indicating how many times it belonged to the high (α) or low-pressure zone (β). To compute the NAO index, we consider 10_{th} percentile for respectively α and β as the high and low-pressure areas to compute the sea level pressure difference. In the end, we consider the zones that were most of the time high or low pressure as the high and low pressure poles. We display the poles selected for each model in the supplementary material Figure S1. The selected period for computing the index was from the start to the peak of the bloom. Choosing this period ensures coherence with the other variables considered during this period (e.g. gyre circulation, nutrient transport)

3.2.4 Gyre index

The gyre circulation is characterized by the gyre index which is computed on the months between the start and the peak of the bloom as it is linked to the NAO (Hurrell and Deser, 2010; Koul et al., 2020). The stream function is computed over 2000 m depth on the original grid and integrated from east to west. Once obtained, a threshold of 7.5 Sverdrup is applied on each grid point as in Biri and Klein (2019). From this method we extract the strength i.e the mean of the stream function at the selected point). The shape of each model gyre and its geographical variability is displayed in the supplementary material in Figure S2.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024


Earth System Dynamics Discussions Open Access EGU

3.2.5 Vertical mixing

The MLD depth in winter will determine the stock of nutrients available (Williams et al., 2006) for the spring bloom. Here we want to capture the variability of MLD intensity. Therefore, we focus on the maximum depth reached between November and April. The impact of stratification on NPP is an important aspect of our study, and we will quantify this relationship using the difference in density as an indicator (Lozier et al., 2011; Van De Poll et al., 2013). We focus on the difference of density between 100m and the surface $\rho_{100} - \rho_0$. To assess the impact of stratification on productivity, we will calculate the mean value of $\rho_{100} - \rho_0$ from the beginning to the peak of the bloom. This approach will allow us to examine the relationship between stratification and NPP during the most active period of phytoplankton growth.

3.2.6 Nutrient transport

The inflow of nutrients via advection is also an important factor to consider for the nutrient concentration at the start of the bloom and the resulting NPP. We integrate the transport of nutrients from the ocean surface to the mixed layer depth (MLD) for each model. For each depth level and each month considered (between the beggining of the bloom and the peak of the bloom), the nutrient concentration is multiplied by the transport. The transport is computed from the meridional and zonal velocity variable. For our analysis, we will focus only on the inflow of nutrients into our study region. As for the period selected for the signal we will take the transport of nutrients from the beginning of the bloom to the peak of the bloom.

4 Results

We obtain one distinct graph for each nutrient : nitrate, dissolved iron and silicate. With the exception of CanESM5-CanOE for which silicate was not available, all 3 nutrients are considered for the remaining models. Causal links differ slightly between nutrients. For example, for a given link between two variables, the nutrient may not be directly involved but indirectly influence the relationships through changes in conditioning variables, resulting in slight variations in the calculated strength. Nutrients are identified in the following by adding "_no3" (nitrate), "_dfe" (dissolved iron) or "_si" (silicate).

4.1 Similar models

The similarity between models is evaluated based on the dissimilarity introduced earlier. Focusing on sub-matrices provides insight into differences between models with respect to specific causal relationships. For instance, computing the dissimilarity on the sub-matrix corresponding to the drivers of one variable allows to explore specific dynamics shared by two models (similar impacts, similar lags). Figure 3 represents the dissimilarity between models for links with dfe. UKESM1-0-LL has the largest dissimilarity with other models. These high values of dissimilarity have two possible explanations (interactions between nutrients and lagged links) which will be discussed in section 4.2.

To further quantify the similarity or dissimilarity between models and to identify the corresponding variables, we count how many times two models are the closest to each other or the most different. The results are synthesised in Figure 4 for IPSL-

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics
Discussions
Open Access
EGU

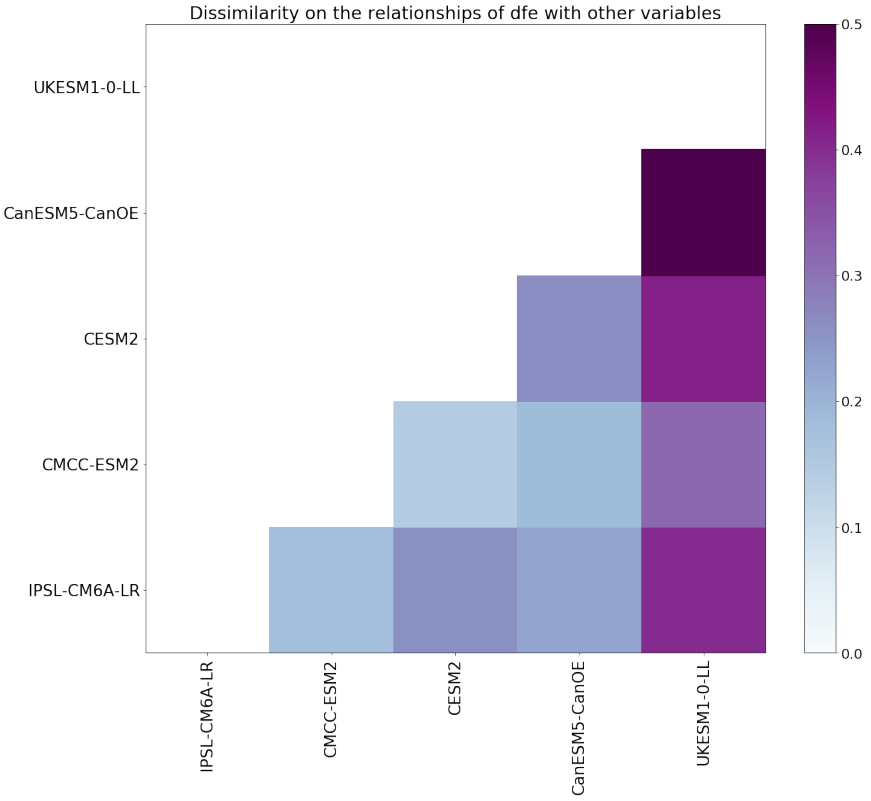

Figure 3. Dissimilarity between each model for iron links only. The darker the colour the more different the models.

CM6A-LR with the most similar links displayed on 4(a) and the most different on 4(b) The results for the other models are given as supplementary material from Figure S2 to S5. From Figure 4, it appears that CMCC-ESM2 is the closest model, particularly for impacts from NAO, gyre, and transport. It has to be noted that the converse is not true. To understand this, we can look at the most different model on the right of Figure 4. CESM2 is the most different model from IPSL with differences located mainly on Gyre and Transport but UKESM1-0-LL also appears to be rather different from IPSL-CM6A-LR. UKESM1-0-LL is also the most different model for all the other models (Figure S1-S4). However, UKESM1-0-LL obviously also has a most similar model, which is CMCC-ESM2, but this similarity is not reciprocal, as UKESM1-0-LL seems to have the most distinct behaviour. CanESM5-CanOE frequently appears as the most different model from UKESM1-0-LL concerning nutrients, gyre, NAO, and stratification. Therefore, these two models are the most dissimilar. Grouping models according to their similarity or dissimilarity raises the question of which specific links differ between the models and for which links the models agree. To answer this question, specific key links are analysed next starting with contemporaneous links (at lag 0).

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

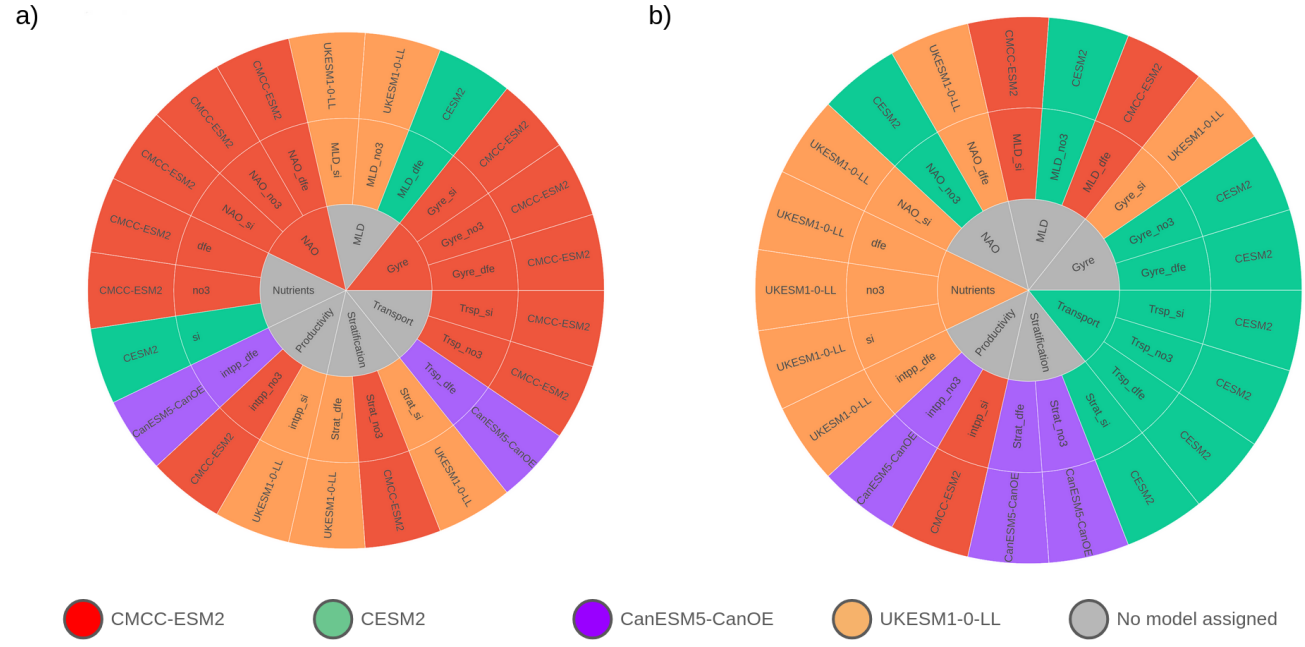

Figure 4. Most similar (a) and most different (b) models to/from IPSL-CM6A-LR according to dissimilarity introduced in 4. The interior circle is the variable studied. Each variable is presented for each nutrient in the middle circle. The exterior circle indicates which model is the closest/most different for this specific variable.

4.2 Control of nutrients

On Figure 5, we have selected a subset of links that directly control pre-bloom nutrient concentrations. The first panel shows the model agreement for the control of MLD on nitrate and silicate (highlighted in green). As expected, the maximum winter MLD has a strong impact on pre-bloom nutrient concentrations with a consensus between models (mean strength of link: 0.61 for silicate and 0.65 for nitrate). This agreement brings coherence among models, although links for nitrate (standard deviation of 0.15) are slightly more scattered compared to silicate (standard deviation of 0.08). Regarding iron concentration, UKESM1-0-LL stands out with a non-significant link from MLD to iron concentration. For the other models the strength of the link is of the same order of magnitude as for nitrate. This reflects the variability in underlying parameterisations of iron biogeochemistry across Earth System Models (Tagliabue et al., 2016). The median strength is about the same for the three nutrients.

The models also agree in the lack of direct impact of NAO on nutrient concentrations. However, it is possible that the NAO indirectly affects nutrient concentrations by impacting a physical variable which in turn influences nutrient concentrations. As discussed in (Patara et al., 2011), this intermediary variable could be the advection of nutrients, which is considered here via the transport of nutrients. The intermediary variable could also be the MLD, partially driven by the NAO (Yamamoto et al., 2020), also considered in our scheme. These indirect links will be discussed in Section 4.4. Lastly, the influence of nutrient transport

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

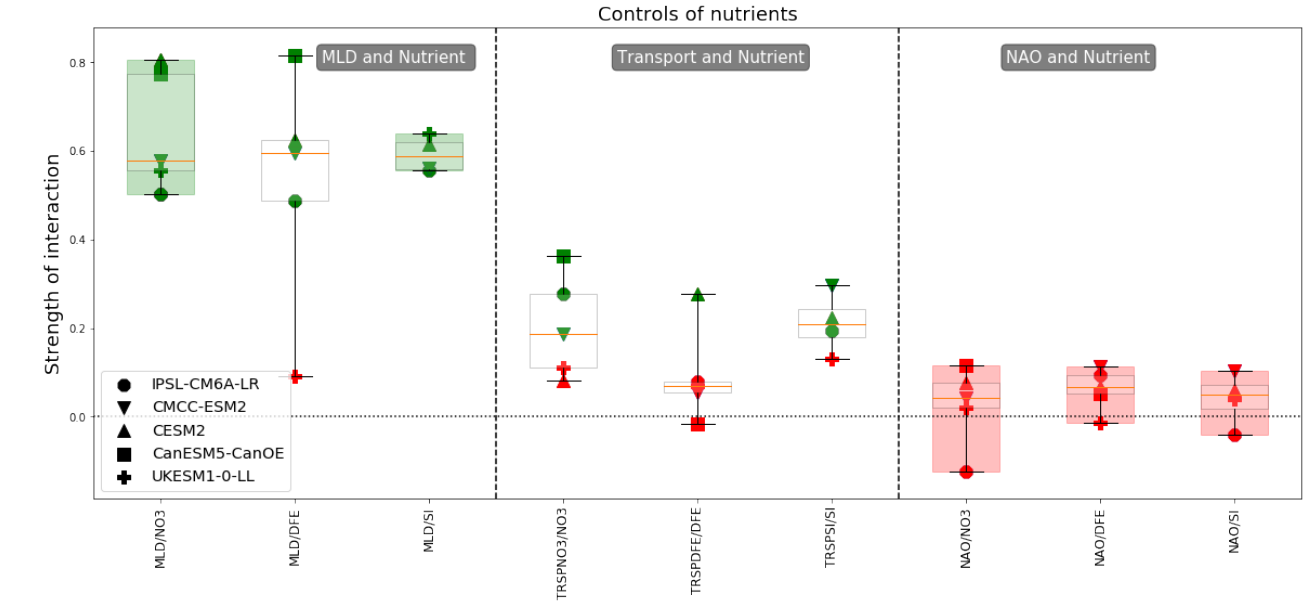

Figure 5. Controls of pre-bloom nutrients concentrations: Strength of links for each model with median and quartile values (boxplot). Each model is represented with a different marker and the color of the marker shows the significance of the link according to PCMCI+ (green for significant and red for not significant). Boxplots highlighted in green (red) indicate model agreement (disagreement).

on pre-bloom nutrient concentrations is scattered. Models CMCC-ESM2 and IPSL-CM6A-LR agree on a positive impact for nitrate and silicate, (CMCC-ESM2:0.18, IPSL-CM6A-LR:0.27 for nitrate and CMCC-ESM2:0.3, IPSL-CM6A-LR:0.19 for silicate) but neither shows a significant impact on iron. CESM2 also shows a positive impact on silicate (0.24) and is the only one to have an impact on iron (0.27). CanESM5-CanOE shares some of the behaviour of CMCC-ESM2 and IPSL-CM6A-LR, with a positive impact for nitrate (0.36, the strongest link for transport) and no impact for iron. UKESM1-0-LL is the only one to have no influence from advection for any nutrient. In CESM2, nitrate is not influenced by advection but is strongly influenced by MLD. In contrast, for iron and silicate, MLD has a less strong influence, and a positive impact from transport is observed. This suggests that for nitrate, the strong influence of MLD leaves no opportunity for advection to affect nutrient variability.

A general conclusion of those results is the major importance of vertical mixing for nutrients. While there is a consensus on the importance of MLD for nitrate and silicate concentrations, inter-model differences still exist, particularly for iron cycling and nutrient transport.

4.3 Productivity drivers

On Figure 6 we isolated a subset of links between variables known to control NPP and the peak intensity of the spring bloom. The left panel reveals two consistent links among the models: nitrate and NPP, silicate and NPP. However, while the models

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

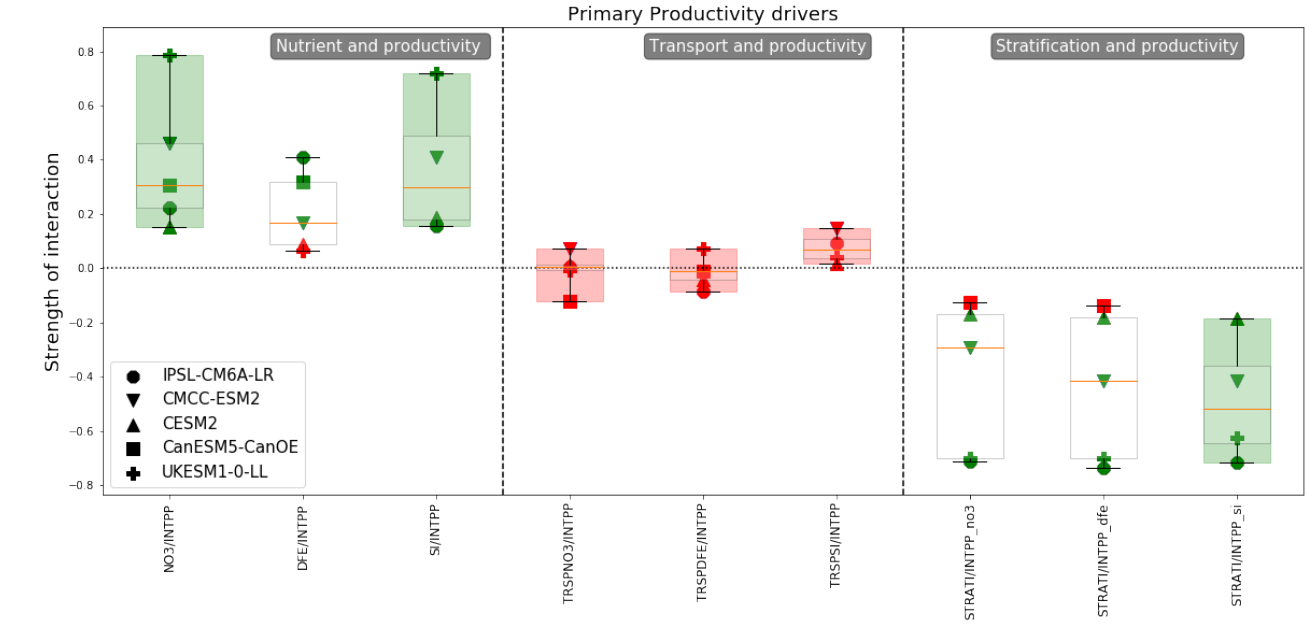

Figure 6. Same as 5 but for controls of net primary productivity.

agree on the significance of these interactions, they disagree on their strength, with a wide range of values (nitrate: mean of 0.39 and standard deviation of 0.23; silicate: mean of 0.37 and standard deviation of 0.23). The UKESM1-0-LL model exhibits the strongest links for both nitrate (0.79) and silicate (0.72), while the CESM2 and IPSL-CM6A-LR models show the weakest, with all values below 0.19 for both nutrients. The variability of nitrate and silicate concentrations before the bloom has a variable impact on NPP across the selected models. For iron the range of interaction strengths is narrower, with both the CESM2 and UKESM1-0-LL models showing non-significant results. The maximum strength value is 0.41, observed for IPSL-CM6A-LR model. Notably, the impact of iron variability on NPP in the IPSL-CM6A-LR model is twice as strong as that of silicate or nitrate. Conversely, the CMCC-ESM2 and UKESM1-0-LL models show the opposite trend. For CMCC-ESM2, the strength decreases from approximately 0.44 for both nitrate and silicate to 0.17 for iron. The difference is even more pronounced in the UKESM1-0-LL model, which exhibits the strongest links for nitrate and silicate, but a non-significant one for iron. In the CESM2 model the strength of those links remains fairly stable, with low strength values for both nitrate and silicate and an iron residual value (the last partial correlation value obtained which did not pass the significativity test) that is close to these.

The last panel of Figure 6 presents the relationship between stratification during the productive period (from the start to the peak of the bloom) and NPP. With the exception of CanESM5-CanOE, which shows a non-significant link, the impact of stratification is negative across all models, as anticipated. However, the range of interaction strengths is quite broad. The UKESM1-0-LL and IPSL-CM6A-LR models exhibit strong negative links (approximately -0.7), while CESM2 shows a weak strength of around -0.2, and CMCC-ESM2 has a moderate strength of approximately -0.3. This marked difference in dynamics among the models suggests that they may respond differently to projected climate conditions, as stratification is known to be

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions
Open Access
EGU

increasing under global warming (Li et al., 2020). Specifically, if stratification continues to increase, the intensity of the bloom in the eastern part of the subpolar gyre may decrease more significantly in the UKESM1-0-LL and IPSL-CM6A-LR models compared to the others.

Lastly, the middle panel of Figure 6 indicates that nutrient transport does not have a significant direct impact on NPP in any of the models. However, as discussed in the previous subsection, transport plays a role in determining pre-bloom nutrient concentrations for some models. These nutrient concentrations, in turn, significantly affect NPP, as shown in the first panel of Figure 6. Therefore, transport has an indirect impact on NPP via nutrient concentrations. However, this impact is relatively low, as the influence of transport on nutrient concentrations is moderate, and the effect of nutrient concentrations on NPP can also be moderate or strong depending on the model. As a result, two consecutive moderate impacts lead to a rather low overall impact. For instance, if the variability of transport explains 30% of the variability of nutrient concentrations and the latter explains 30% of NPP variability, then transport explains 9% of NPP variability.

4.4 Physical interactions

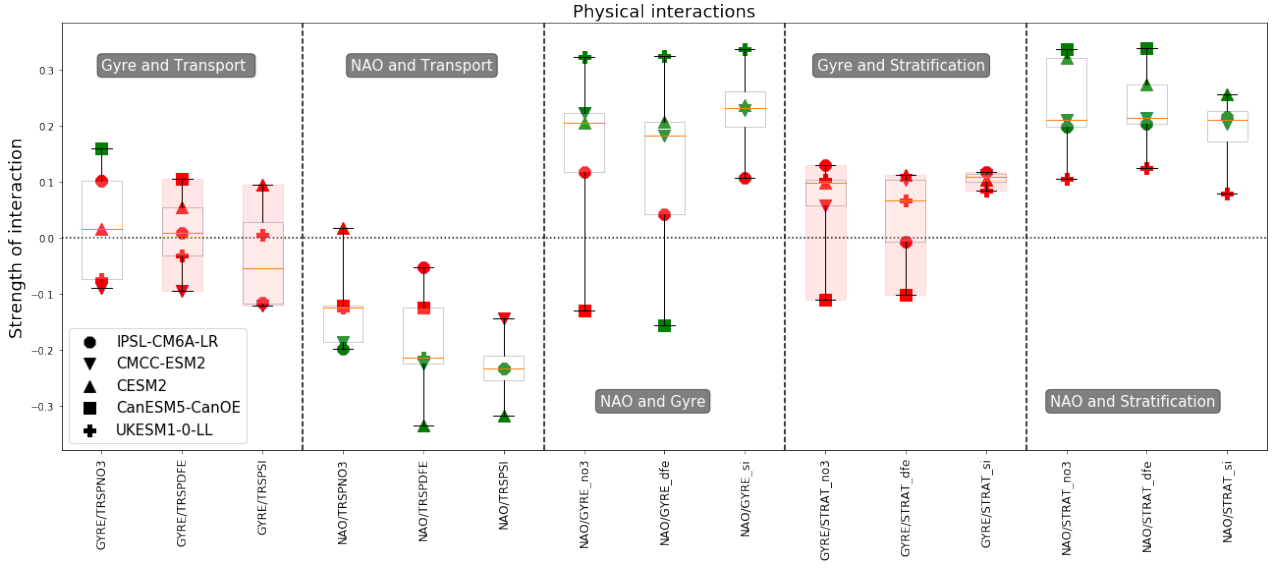

Figure 7. Same as Figure 5 but for selected physical interactions. Other physical interactions are displayed in Supplementary Figure S7

Figure 7 presents a subset of physical interactions included in our conceptual scheme. Other interactions are displayed in the supplementary Figure S7. The focus of Figure 7 is on the atmospheric impact on the ocean component, specifically the impact of the North Atlantic Oscillation (NAO) on transport, gyre strength, and stratification. For the CMCC-ESM2 model a significant impact of NAO was found for most variables, except for silicate transport, which has a residual of around 0.14. The same holds for CESM2 which only misses an impact on nitrate transport with a residual around 0.03. Comparing only these two models, the strength of relationship between NAO and gyre strength is similar (around -0.2), but the impact on transport

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

is slightly stronger for CESM2 (around 0.3 for CESM2 and around 0.2 for CMCC-ESM2). CESM2 also has a stronger effect of NAO on stratification (around -0.3 for CESM2 and around -0.2 for CMCC-ESM2), making it the model with the strongest overall impact from NAO on its oceanic variables.

For the other models many model-specific links were identified, suggesting inter-model differences in dynamics. Starting with UKESM1-0-LL, NAO has a significant impact on the gyre strength (-0.33) but none on the stratification in our zone. The lack of a significant link between NAO and stratification sets this model apart. It does not result from the approach taken to compute NAO as there are no major differences in the placement of high and low-pressure zones between UKESM1-0-LL and the other models. The UKESM1-0-LL model also shows significant links from NAO on iron and silicate transport (with a strength around 0.24 slightly bigger for silicate), but not on nitrate transport (residual around -0.12). The IPSL-CM6A-LR model is the only one with a non-significant impact on gyre strength. However, the NAO significantly impacts stratification (with a strength of about -0.2) and the transport of nitrate and silicate (around -0.22, slightly bigger for silicate).

Lastly the CanESM5-CanOE model, it is the only one with a negative impact of NAO on the gyre opposed to a positive one observed in the other models (or at least a positive residual when not significant). However, we see on the third panel Figure 7 that on the first boxplot it is not significant and on the second one it is. This relationship is on the verge of significance (with a strength of -0.14 when significant). This suggests a potential unique behaviour in this model with respect to the relationship between the NAO and gyre strength. The CanESM5-CanOE model also exhibits one of the strongest impacts on stratification (with a strength of about -0.33). Lastly, the transport of nutrients is not significantly impacted by the NAO.

In the previous analysis, we identified for all models, except UKESM1-0-LL, the transport of at least one nutrient impacting the variability of nutrient concentrations. Therefore, it is possible for NAO to indirectly influence nutrient concentrations through its impact on transport. This indirect link between NAO and nutrient concentrations is present in IPSL-CM6A-LR, CESM2 and CMCC-ESM2. In the case of IPSL-CM6A-LR nitrate and silicate transport significantly impact their respective nutrient concentrations. Likewise, for the CESM2 model, iron and silicate transport significantly affect their corresponding nutrient levels. The CMCC-ESM2 model also exhibits this indirect link through nitrate transport, but not for iron transport, as NAO does not impact it. In all these cases the transport variability is partly driven by the NAO. In summary, the indirect link between NAO and nutrient pre-bloom concentration through transport is model-dependent.

Considering the interaction between gyre strength and nutrient transport and the interaction between gyre strength and stratification, we observe that the models agree on the non-significance of these interactions. However, there is an exception for the link between gyre strength and nitrate transport, where the interaction is significant with a strength of 0.16 for CanESM5-CanOE only. In this model, nitrate transport significantly impacts pre-bloom nitrate concentration with a strength of 0.36. Thus, there is an indirect influence of gyre strength on nutrient concentration via nitrate transport.

5 PCMCI+ Limitations

Like for many other algorithms, it is crucial to carefully select variables given to PCMCI+, as the selection will significantly affect results. While it is necessary to include all variables containing relevant information, a too high number will decrease the

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics
Discussions
Open Access
EGU

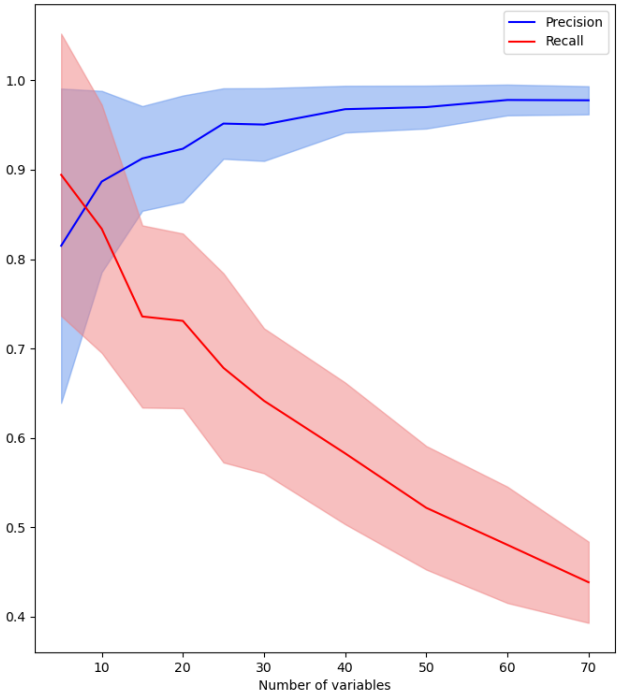

Figure 8. Precision (blue) and recall (red) according to the number of variables with a max time lag fixed at 5.

performance of the algorithm. To evaluate the decrease in performance as a function of the number of variables and the time lag, two criteria are commonly used: "precision" and "recall":

- The precision represents the ratio of True Positives among all the positives found by the algorithm (True Positives + False Positives): $precision = \#true\ pos./(\#true\ pos. + \#false\ pos.)$.
- The recall corresponds to the ratio of True Positives among all the positives that the algorithm should have found (True Positives + False Negatives): $recall = \#true\ pos./(\#true\ pos. + \#false\ neg.)$.

To quantify the recall and precision performances of PCMCI+, we applied it to artificially generated random auto-regressive vectors. Auto-regressive vectors are time series built so that they depend on past values of themselves.A set of N vectors $(X_i)_{1\leq i\leq N}$ is defined according to:

$$X_i(t) = \sum_{0\leq\tau\leq\tau_{max}} \sum_{1\leq j\leq N} \alpha_{i,j,\tau} X_j(t-\tau) \tag{6}$$

with $\alpha_{i,j,\tau}$ being randomly chosen dependence coefficient and τ_{max} the maximum lag value. Each non null random dependence between the variables has been constrained to have a value of at least 0.2. To be close to the configuration of this study, vectors are generated on 500 time steps. These auto-regressive vectors are now taken as the "ground truth" and thus allow to evaluate the precision and recall performances of the PCMCI+ algorithm.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024


Earth System Dynamics Discussions Open Access EGU

We tested the performances according to a varying number of variables and 50 times for each number of variables considered.. The result is displayed in Figure 8, where the blue curve in Figure 8 illustrates that precision is not critical, since an increasing number of variables does not lower the it. However, the red curve shows that an increasing number of variable affects the recall. With 60 variables, we retrieve less than half of the expected links (recall criterion). Thus, this experiment clearly illustrates that it is important to give a relatively low number of variables to the algorithm.

Hence, variables need to be selected with great care in order to capture every important feature of the scheme under investigation. In our study, we selected 7 variables to consider key elements for our study and to also have a good compromise with the performance by not having too many variables. If certain key variables are omitted from the set of variables given to PCMCI, there is a risk that some of the causal links found by PCMCI will be incomplete or inaccurate. Forgetting key explanatory variables turns a causality analysis into a correlation analysis.

6 Discussion

PCMCI+ has demonstrated its efficiency in identifying model-specific dynamics and highlighting differences between models that may not be immediately apparent on traditional model output intercomparisons. We were able to separate the role of vertical mixing and advection on controling the variability of NPP. Our analysis suggests that vertical mixing plays a more pronounced role in modulating the intensity of spring bloom NPP in Eastern North Atlantic subpolar gyre than horizontal transport, in line with previous studies (Ólafsson, 2003). We also observed that stratification during the bloom serves as a significant inhibitor of bloom intensity (Sarmiento et al., 2004), but this relationship is significantly strong in only two of the five models included in this study (IPSL-CM6A-LR, UKESM1-0-LL). As climate change progresses, it is anticipated that NPP in these models will show a greater sensitivity to increasing stratification (Wilson et al., 2022). Moreover, models that currently show no significant impact of stratification on NPP may exhibit a negative effect appearing as global warming intensifies. This could unveil a novel causal relationship emerging under certain climate change scenarios. This inter-model variability of the response of spring bloom NPP to stratification may contribute to the multi-model spread on projected values of NPP.

Furthermore, the nutrient transport in this study has been computed along the borders of a large region, with each boundary point of the area (Figure 1) taken into consideration. With the exception of CanESM5-CanOE, this transport criteria did not allow to retrieve a consistent link between the gyre strength and the transport. However, this criteria allowed to highlight impacts of NAO on nutrient variability via its impact on transport for most of the models. Addressing this indirect link from NAO to nutrient concentration via the transport underscores the importance of a causality based approach and the importance of considering key explanatory variables. If we do not consider transport we may find a spurious link between NAO and nutrient concentration, obscuring a hidden variable. This illustrates results that would have been missed by a correlation-based approach only.

The above highlights the complexity of the North Atlantic subpolar gyre and the potential for indirect relationships between variables. Although the NAO does not directly impact nutrient concentrations, it can still influence nutrient availability through indirect pathways. By examining these indirect links, we obtain a more comprehensive understanding of the dynamics of the

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

North Atlantic subpolar gyre and the factors that control nutrient concentrations and NPP. This also emphasises the additional information gained from a causality-based intercomparison approach compared to a correlation-based intercomparison.

Despite the redundancy of the oceanic component (NEMO) across selected ESMs, with only CESM2 using a different oceanic model (POP2), we do not observe any unique behaviour specific to CESM2 or a commonality shared by the other four ESMs. It is noteworthy that the two models exhibiting the strongest NAO impact, namely CESM2 and CMCC-ESM2, share the same atmospheric model CAM6. Simpson et al. (2020) did not identify a distinct behaviour for this atmospheric model compared to others among CMIP models, but did not look at the impact of NAO on specific variables via causality. With our study, it is hard to establish a firm conclusion based on a single member, but applying a causality-based approach to a wider set of simulation using this atmospheric model could offer alternative insights.

The proposed causality-based comparison method also allows to quantify inter-models variations in the strength of variables interactions. While it is expected that winter mixing influences the pre-bloom nutrient stock, variations between the models still exist. All models agree on the significant role of MLD for silicate within a narrow range of strengths. However, the range is somewhat broader for nitrate and even more for iron, with CanESM5-CanOE finding MLD not significant. Tagliabue et al. (2016) found a similar result with models having a stronger disagreement for iron than for nitrate. Also, the unique behaviour we observe in the MEDUSA2 model from UKESM1-0-LL is also discussed in this article with an overestimation of surface iron possibly due to a too long residence time and a too low scavenging rate. We also found a strong auto-correlation for iron (not shown) only for this model which could have the same explanation.

7 Conclusion and perspectives

This study uses a set of Earth System Models to explore the variability of nutrients and NPP in the Eastern North Atlantic Subpolar Gyre. This region is characterized by significant mixing and inter-model uncertainty in nutrient and NPP projections. We established a conceptual scheme to determine which mechanism, vertical mixing or advection, is crucial in explaining multi-model dispersion in nutrient and NPP dynamics in this region.

The method allowed us to extract the causal links from a multivariate conceptual scheme. Our findings indicate that vertical mixing during winter is generally more important for peak spring bloom NPP in this region than transport. However, the influence of vertical mixing on peak spring bloom NPP still has some variability among the models expressed by the spread of the strength of the link. Combined with model-specific dynamics (e.g. NAO impacting transport, transport impacting nutrient concentration), such differences are likely to contribute to inter-model spread. Stratification is identified as an important factor controlling spring bloom NPP in some, but not all, models. As climate change progresses, it is expected that NPP will suffer from increasing stratification with potentially different reactions to it from one model to another.

Despite its limitations, the causality method used in this study has proven effective in identifying model-specific dynamics and highlighting differences between models that may not be immediately apparent based on usual difference metrics. The proposition of a new dissimilarity measure to be applied on causal graphs has enhanced model comparison by synthesising and quantifying between-model differences.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

455 The method has been verified and can be applied for a variety of purposes such as model evaluation. For example, the comparison of model output and observational data will allow to identify discrepancies in links and associated strengths, suggesting unrealistic model dynamics. The dissimilarity metric proposed could serve as a validation tool. However, it is essential to acknowledge as the recall of PCMCI+ decreases with increasing variables, an important part of the ground truth (True Links existing in the data) can be missed and will add a bias for the validation. Another possible application of the 460 PCMCI+ method involves comparing various experiments, such as different climatic scenarios. This approach facilitates the discovery of new dynamics or the observation of changes in the intensity of existing ones, leading to an improved understanding of models' behaviours under different environmental conditions. Applying the method to climate change scenarios will improve our understanding of the inter-model spread of projected variables (e.g. NPP).

Author contributions. All authors contributed to the development of the study, the interpretation of the results, and the writing of the 465 manuscript. The lead author processed the data, obtained the results, and proposed the visualization of the results.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. This study has been supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 862923 (project AtlantECO, Atlantic Ecosystems Assessment, Forecasting & Sustainability). This work has also been supported by the "COESION" project funded by the French National program LEFE (Les Enveloppes Fluides et l'Environnement). We ac- 470 knowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Earth System Dynamics Discussions
Open Access
EGU

References

Anagnostopoulos, G., Koutsoyiannis, D., Christofides, A., Efstratiadis, A., and Mamassis, N.: A comparison of local and aggregated climate model outputs with observed data, Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55, 1094–1110, 2010.

Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal statistical society: series B (Methodological), 57, 289–300, 1995.

Biri, S. and Klein, B.: North Atlantic sub-polar gyre climate index: a new approach, Journal of Geophysical Research: Oceans, 124, 4222–4237, 2019.

Bonan, G. and Doney, S.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, Science, 359, eaam8328, 2018.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., et al.: Presentation and evaluation of the IPSL-CM6A-LR climate model, Journal of Advances in Modeling Earth Systems, 12, e2019MS002 010, 2020.

Brody, S. R., Lozier, M. S., and Dunne, J. P.: A comparison of methods to determine phytoplankton bloom initiation, Journal of Geophysical Research: Oceans, 118, 2345–2357, 2013.

Charakopoulos, A., Katsouli, G., and Karakasidis, T.: Dynamics and causalities of atmospheric and oceanic data identified by complex networks and Granger causality analysis, Physica A: Statistical Mechanics and its Applications, 495, 436–453, 2018.

Chen, C.-T. and Knutson, T.: On the verification and comparison of extreme rainfall indices from climate models, Journal of Climate, 21, 1605–1621, 2008.

Christian, J. R., Denman, K. L., Hayashida, H., Holdsworth, A. M., Lee, W. G., Riche, O. G., Shao, A. E., Steiner, N., and Swart, N. C.: Ocean biogeochemistry in the canadian earth system model version 5.0. 3: CanESM5 and CanESM5-CanOE, Geoscientific Model Development Discussions, 2021, 1–68, 2021.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., Emmons, L., Fasullo, J., Garcia, R., Gettelman, A., et al.: The community earth system model version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, e2019MS001 916, 2020.

Daniault, N., Mercier, H., Lherminier, P., Sarafanov, A., Falina, A., Zunino, P., Pérez, F. F., Ríos, A. F., Ferron, B., Huck, T., et al.: The northern North Atlantic Ocean mean circulation in the early 21st century, Progress in Oceanography, 146, 142–158, 2016.

Dawid, A. P.: Conditional independence in statistical theory, Journal of the Royal Statistical Society Series B: Statistical Methodology, 41, 1–15, 1979.

Delworth, T. L. and Zeng, F.: The impact of the North Atlantic Oscillation on climate through its influence on the Atlantic meridional overturning circulation, Journal of Climate, 29, 941–962, 2016.

D'Asaro, E. A.: Convection and the seeding of the North Atlantic bloom, Journal of Marine Systems, 69, 233–237, 2008.

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

Feucher, C., Portela, E., Kolodziejczyk, N., and Thierry, V.: Subpolar gyre decadal variability explains the recent oxygenation in the Irminger Sea, Communications Earth & Environment, 3, 279, 2022.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation of climate models, in: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 741–866, Cambridge University Press, 2014.

Flato, G. M.: Earth system models: an overview, Wiley Interdisciplinary Reviews: Climate Change, 2, 783–800, 2011.

Fox-Kemper, B., Hewitt, H., Xiao, C., Aðalgeirsdóttir, G., Drijfhout, S., Edwards, T., Golledge, N., Hemer, M., Kopp, R., Krinner, G., Mix, A., Notz, D., Nowicki, S., Nurhati, I., Ruiz, L., Sallée, J.-B., Slangen, A., and Yu, Y.: Ocean, Cryosphere and Sea Level Change, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 1211–1362, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896.011, 2021.

Fu, W., Moore, J. K., Primeau, F., Collier, N., Ogunro, O. O., Hoffman, F. M., and Randerson, J. T.: Evaluation of ocean biogeochemistry and carbon cycling in CMIP earth system models with the international ocean model benchmarking (IOMB) software System, Journal of Geophysical Research: Oceans, 127, e2022JC018 965, 2022.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres, 113, 2008.

Granger, C. W.: Investigating causal relations by econometric models and cross-spectral methods, Econometrica: journal of the Econometric Society, pp. 424–438, 1969.

Hakkinen, S. and Rhines, P. B.: Decline of subpolar North Atlantic circulation during the 1990s, Science, 304, 555–559, 2004.

Hátún, H., Azetsu-Scott, K., Somavilla, R., Rey, F., Johnson, C., Mathis, M., Mikolajewicz, U., Coupel, P., Tremblay, J.-É., Hartman, S., et al.: The subpolar gyre regulates silicate concentrations in the North Atlantic, Scientific reports, 7, 14 576, 2017.

Herceg-Bulić, I. and Kucharski, F.: North Atlantic SSTs as a link between the wintertime NAO and the following spring climate, Journal of climate, 27, 186–201, 2014.

Holliday, N. P., Bersch, M., Berx, B., Chafik, L., Cunningham, S., Florindo-López, C., Hátún, H., Johns, W., Josey, S. A., Larsen, K. M. H., et al.: Ocean circulation causes the largest freshening event for 120 years in eastern subpolar North Atlantic, Nature communications, 11, 585, 2020.

Hurrell, J., Kushnir, Y., Ottersen, G., and Visbeck, M.: An overview of the North Atlantic Oscillation, vol. 134, pp. 1–36, 2003.

Hurrell, J. W.: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, Science, 269, 676–679, 1995.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: the role of the North Atlantic Oscillation, Journal of marine systems, 79, 231–244, 2010.

Jacob, D., Bärring, L., Christensen, O. B., Christensen, J. H., de Castro, M., Déqué, M., Giorgi, F., Hagemann, S., Hirschi, M., Jones, R., et al.: An inter-comparison of regional climate models for Europe: model performance in present-day climate, Climatic change, 81, 31–52, 2007.

Johnson, C., Inall, M., and Häkkinen, S.: Declining nutrient concentrations in the northeast Atlantic as a result of a weakening Subpolar Gyre, Deep Sea Research Part I: Oceanographic Research Papers, 82, 95–107, 2013.

Keller, K. M., Joos, F., Raible, C. C., Cocco, V., Frölicher, T. L., Dunne, J. P., Gehlen, M., Bopp, L., Orr, J. C., Tjiputra, J., et al.: Variability of the ocean carbon cycle in response to the North Atlantic Oscillation, Tellus B: Chemical and Physical Meteorology, 64, 18 738, 2012.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Koul, V., Tesdal, J.-E., Bersch, M., Hátún, H., Brune, S., Borchert, L. F., Haak, H., Schrum, C., and Baehr, J.: Unraveling the choice of the north Atlantic subpolar gyre index, Scientific Reports, 10, 1005, https://doi.org/10.1038/s41598-020-57790-5, 2020.

Koutra, D., Vogelstein, J. T., and Faloutsos, C.: Deltacon: A principled massive-graph similarity function, in: Proceedings of the 2013 SIAM international conference on data mining, pp. 162–170, SIAM, 2013.

Krich, C., Runge, J., Miralles, D. G., Migliavacca, M., Perez-Priego, O., El-Madany, T., Carrara, A., and Mahecha, M. D.: Estimating causal networks in biosphere–atmosphere interaction with the PCMCI approach, Biogeosciences, 17, 1033–1061, 2020.

Kwiatkowski, L., Naar, J., Bopp, L., Aumont, O., Defrance, D., and Couespel, D.: Decline in Atlantic primary production accelerated by Greenland ice sheet melt, Geophysical Research Letters, 46, 11 347–11 357, 2019.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., et al.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, Biogeosciences, 17, 3439–3470, 2020.

Li, G., Cheng, L., Zhu, J., Trenberth, K. E., Mann, M. E., and Abraham, J. P.: Increasing ocean stratification over the past half-century, Nature Climate Change, 10, 1116–1123, 2020.

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., Fogli, P., Cherchi, A., Bellucci, A., Gualdi, S., et al.: CMIP6 simulations with the CMCC Earth system model (CMCC-ESM2), Journal of Advances in Modeling Earth Systems, 14, e2021MS002 814, 2022.

Lozier, M. S., Dave, A. C., Palter, J. B., Gerber, L. M., and Barber, R. T.: On the relationship between stratification and primary productivity in the North Atlantic, Geophysical Research Letters, 38, 2011.

Nowack, P., Runge, J., Eyring, V., and Haigh, J.: Causal networks for climate model evaluation and constrained projections, Nat. Commun., 11, 1415, 2020.

Ólafsson, J.: Winter mixed layer nutrients in the Irminger and Iceland Seas, 1990-2000, 2003.

Oschlies, A.: NAO-induced long-term changes in nutrient supply to the surface waters of the North Atlantic, Geophysical Research Letters, 28, 1751–1754, 2001.

Patara, L., Visbeck, M., Masina, S., Krahmann, G., and Vichi, M.: Marine biogeochemical responses to the North Atlantic Oscillation in a coupled climate model, Journal of Geophysical Research: Oceans, 116, 2011.

Pelegrí, J., Csanady, G., and Martins, A.: The north Atlantic nutrient stream, Journal of Oceanography, 52, 275–299, 1996.

Runge, J.: Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets, in: Conference on Uncertainty in Artificial Intelligence, pp. 1388–1397, PMLR, 2020.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets, Science advances, 5, eaau4996, 2019.

Sarmiento, J. L., Slater, R., Barber, R., Bopp, L., Doney, S., Hirst, A., Kleypas, J., Matear, R., Mikolajewicz, U., Monfray, P., et al.: Response of ocean ecosystems to climate warming, Global Biogeochemical Cycles, 18, 2004.

Schaller, N., Mahlstein, I., Cermak, J., and Knutti, R.: Analyzing precipitation projections: A comparison of different approaches to climate model evaluation, Journal of Geophysical Research: Atmospheres, 116, 2011.

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., et al.: Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6, Current Climate Change Reports, 6, 95–119, 2020.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'connor, F. M., Stringer, M., Hill, R., Palmieri, J., et al.: UKESM1: Description and evaluation of the UK Earth System Model, Journal of Advances in Modeling Earth Systems, 11, 4513–4558, 2019.

Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., Lauritzen, P. H., Marsh, D. R., Mills, M. J., Medeiros, B., et al.: An evaluation of the large-scale atmospheric circulation and its variability in CESM2 and other CMIP models, Journal of Geophysical Research: Atmospheres, 125, e2020JD032 835, 2020.

Spirtes, P. and Glymour, C.: An algorithm for fast recovery of sparse causal graphs, Social science computer review, 9, 62–72, 1991.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., et al.: The Canadian earth system model version 5 (CanESM5. 0.3), Geoscientific Model Development, 12, 4823–4873, 2019.

Swingedouw, D., Bily, A., Esquerdo, C., Borchert, L. F., Sgubin, G., Mignot, J., and Menary, M.: On the risk of abrupt changes in the North Atlantic subpolar gyre in CMIP6 models, Annals of the New York Academy of Sciences, 1504, 187–201, 2021.

Tagliabue, A., Aumont, O., DeAth, R., Dunne, J. P., Dutkiewicz, S., Galbraith, E., Misumi, K., Moore, J. K., Ridgwell, A., Sherman, E., et al.: How well do global ocean biogeochemistry models simulate dissolved iron distributions?, Global Biogeochemical Cycles, 30, 149–174, 2016.

Tagliabue, A., Kwiatkowski, L., Bopp, L., Butenschön, M., Cheung, W., Lengaigne, M., and Vialard, J.: Persistent uncertainties in ocean net primary production climate change projections at regional scales raise challenges for assessing impacts on ecosystem services, Frontiers in Climate, 3, 2021.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philosophical transactions of the royal society A: mathematical, physical and engineering sciences, 365, 2053–2075, 2007.

Tesdal, J.-E., Abernathey, R. P., Goes, J. I., Gordon, A. L., and Haine, T. W.: Salinity trends within the upper layers of the subpolar North Atlantic, Journal of Climate, 31, 2675–2698, 2018.

Tsujino, H., Urakawa, L. S., Griffies, S. M., Danabasoglu, G., Adcroft, A. J., Amaral, A. E., Arsouze, T., Bentsen, M., Bernardello, R., Böning, C. W., et al.: Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), Geoscientific Model Development, 13, 3643–3708, 2020.

Van De Poll, W., Kulk, G., Timmermans, K., Brussaard, C., Van Der Woerd, H., Kehoe, M., Mojica, K., Visser, R., Rozema, P., and Buma, A.: Phytoplankton chlorophyll a biomass, composition, and productivity along a temperature and stratification gradient in the northeast Atlantic Ocean, Biogeosciences, 10, 4227–4240, 2013.

Vautard, R., Kadygrov, N., Iles, C., Boberg, F., Buonomo, E., Bülow, K., Coppola, E., Corre, L., van Meijgaard, E., Nogherotto, R., et al.: Evaluation of the large EURO-CORDEX regional climate model ensemble, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 344, 2021.

Wang, B., Jin, C., and Liu, J.: Understanding future change of global monsoons projected by CMIP6 models, Journal of Climate, 33, 6471–6489, 2020.

Whitt, D. B. and Jansen, M. F.: Slower nutrient stream suppresses Subarctic Atlantic Ocean biological productivity in global warming, Proceedings of the National Academy of Sciences, 117, 15 504–15 510, 2020.

Wicker, N., Nguyen, C. H., and Mamitsuka, H.: A new dissimilarity measure for comparing labeled graphs, Linear Algebra and its Applications, 438, 2331–2338, 2013.

Williams, R. G., Roussenov, V., and Follows, M. J.: Nutrient streams and their induction into the mixed layer, Global Biogeochemical Cycles, 20, 2006.

https://doi.org/10.5194/esd-2024-31
Preprint. Discussion started: 14 October 2024

Williams, R. G., McDonagh, E., Roussenov, V. M., Torres-Valdes, S., King, B., Sanders, R., and Hansell, D. A.: Nutrient streams in the North Atlantic: Advective pathways of inorganic and dissolved organic nutrients, Global Biogeochemical Cycles, 25, 2011.

Wilson, J. D., Andrews, O., Katavouta, A., de Melo Viríssimo, F., Death, R. M., Adloff, M., Baker, C. A., Blackledge, B., Goldsworth, F. W., Kennedy-Asser, A. T., et al.: The biological carbon pump in CMIP6 models: 21st century trends and uncertainties, Proceedings of the National Academy of Sciences, 119, e2204369 119, 2022.

Yamamoto, A., Tatebe, H., and Nonaka, M.: On the emergence of the Atlantic multidecadal SST signal: A key role of the mixed layer depth variability driven by North Atlantic oscillation, Journal of Climate, 33, 3511–3531, 2020.

Yool, A., Palmiéri, J., Jones, C., de Mora, L., Kuhlbrodt, T., Popova, E., Nurser, A., Hirschi, J., Blaker, A., Coward, A., et al.: Evaluating the physical and biogeochemical state of the global ocean component of UKESM1 in CMIP6 historical simulations, Geosci. Model Dev., 14, 3437–3472, 2021.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of higher climate sensitivity in CMIP6 models, Geophysical Research Letters, 47, e2019GL085 782, 2020.