ARTICLE   OPEN

Check for updates

Early-onset of Atlantic Meridional Overturning Circulation weakening in response to atmospheric CO_2 concentration

Mihai Dima [1,2✉], Denis R. Nichita[1,3], Gerrit Lohmann[2], Monica Ionita [2] and Mirela Voiculescu[4]

The Atlantic Meridional Overturning Circulation (AMOC), a tipping component of the climate system, is projected to slowdown during the 21st century in response to increased atmospheric CO_2 concentration. The rate and start of the weakening are associated with relatively large uncertainties. Observed sea surface temperature-based reconstructions indicate that AMOC has been weakening since the mid-20th century, but its forcing factors are not fully understood. Here we provide dynamical observational evidence that the increasing atmospheric CO_2 concentration affects the North Atlantic heat fluxes and precipitation rate, and weakens AMOC, consistent with numerical simulations. The inferred weakening, starting in the late 19th century, earlier than previously suggested, is estimated at 3.7 ± 1.0 Sv over the 1854–2016 period, which is larger than it is shown in numerical simulations (1.4 ± 1.4 Sv).

INTRODUCTION

In the context of global warming, a major concern is related to climatic components which can suffer rapid transitions between two distinct states[1] (e.g., Atlantic deep water formation, Arctic sea ice, Greenland ice-sheet, among others). Such a tipping element, AMOC, has a quasi-global impact and played a central role in past abrupt climate changes[2,3]. Its fate during the twenty first century is a topic of major scientific and socio-economic interest.

Most climate projections indicate that AMOC will suffer a centennial scale slowdown during the twenty-first century, mainly in response to intensified North Atlantic freshwater and heat fluxes, induced by increased atmospheric CO_2 concentration[4]. However, there are significant quantitative differences between the results of various model simulations, which imply large uncertainties regarding the future evolution of AMOC.

Numerical integrations over the historical period simulate a modest AMOC decrease of 1.4 ± 1.4 Sverdrup (Sv) between preindustrial (1850–1900) and present day (2006–2015), which is most pronounced during the last decades, indicating that anthropogenic warming may have already weakened it[5]. Direct observations indicate that the overturning weakened by 30% during the 1957–2004 period[6] and that it was in a relatively weak state between 2008 and 2017[7]. Alternatively, indirect measures of ocean circulation changes, based on historical sea surface temperature (SST) fields, suggest that AMOC has been weakening since the mid-twenty century[8]. When calibrated with an ensemble of model simulations from the CMIP5 project, the weakening over 1870–2006 period was estimated to be of 3 ± 1 Sverdrup (Sv) (15%) and it was most pronounced since the mid-twenty century[9]. Proxy and reconstructed data suggest that the reduced AMOC intensity during the 1975–1995 period is at the lowest level in the last millennium[10]. The attribution of this reconstructed trend to external or internal factors remains an open problem of fundamental importance in climate research.

Here we investigate a potential contribution of atmospheric CO_2 concentration to AMOC slowdown, based on observational and reanalysis data. First, we separate the SST-based reconstructed

long-term AMOC weakening trend and emphasize the routes through which this greenhouse gas could affect AMOC. Then we probe the associated causal chains using the Convergent Cross Mapping (CCM) technique, a method based on the theory of dynamical systems used to identify causation in weakly coupled systems based on two timeseries[11].

In practical terms, CCM causation is tested using the technique of "cross mapping": a time delay embedding is constructed from the time series of Y and the ability to estimate the values of X from this embedding quantifies how much information about the later has been encoded into the former variable. The accuracy of the prediction is measured using Pearson's correlation coefficient (ρ), between observed and predicted values. One notes the counter-intuitive fact that the cross map estimate runs in the reverse direction of causality: if Y predicts X, then X causes Y. However, one key property that distinguishes causality from mere correlation is the *convergence* of the cross estimation. When constructing the embedding, only a given portion of the time series is used. Increasing this library length should improve the accuracy of the prediction, since the additional points fill in the trajectories in the attractor, resulting in closer nearest neighbors. At some point, the information contained in the affected variable has been exhaustively harnessed and the cross map saturates to a plateau. The asymptotic increase of the cross-map skill with library length is called convergence. The strength of the causal interaction may be linked with the rate of growth toward the convergence level with library size, but also with the level of the cross map skill.

RESULTS

Isolating the AMOC trend

In order to investigate the AMOC response to this greenhouse gas, in a preliminary analysis, we aim to increase the signal-to-noise ratio by separating the centennial from multidecadal overturning variations emphasized in previous numerical and observational studies[8,12,13], with the former having the same characteristic time scale as the increasing trend of the atmospheric CO_2

[1]University of Bucharest, Faculty of Physics, Măgurele, Romania. [2]Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany. [3]"Horia Hulubei" National Institute of Physics and Nuclear Engineering, Bucharest-Măgurele, Romania. [4]University "Dunărea de Jos", Galati, Romania. ✉email: mihai.dima@unibuc.ro
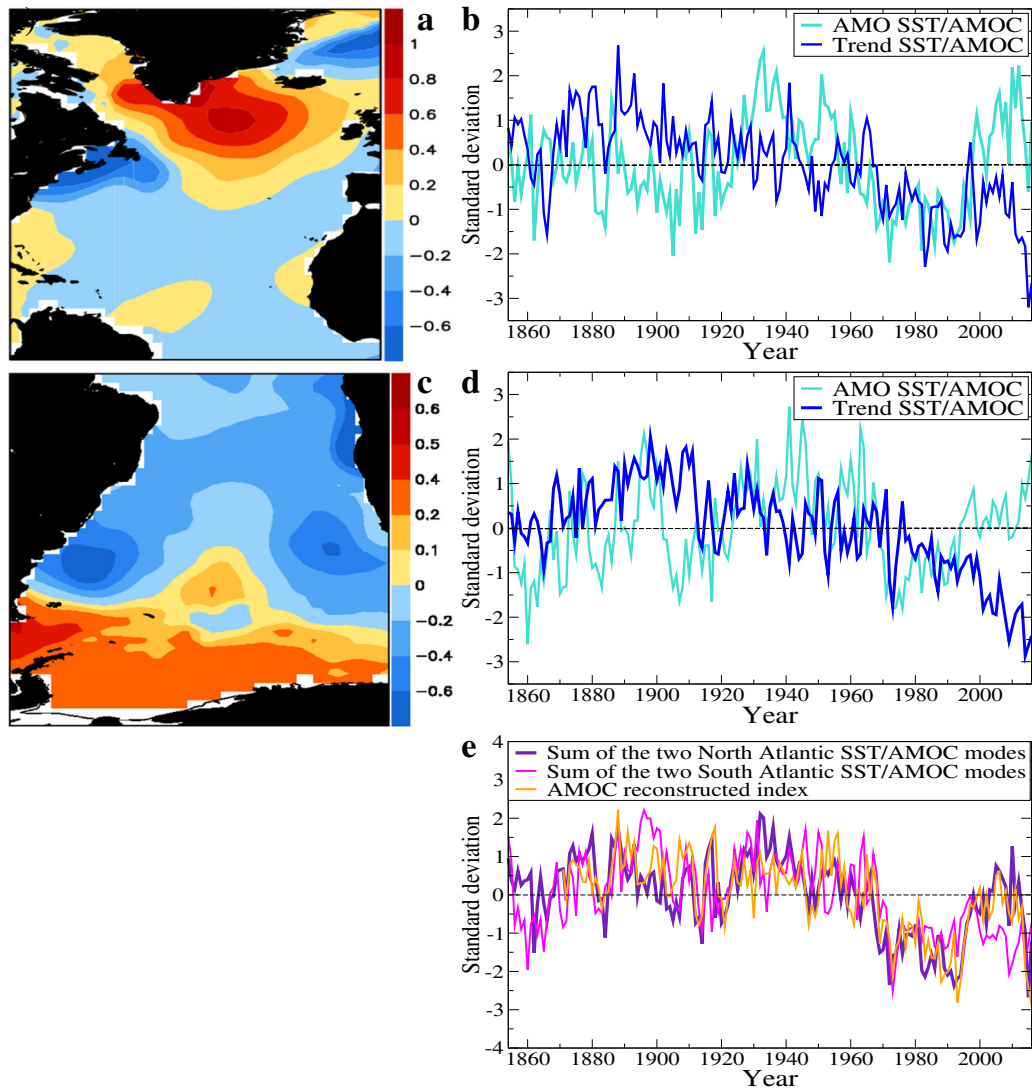
Fig. 1 North Atlantic and South Atlantic SST modes linked to AMOC changes. The pattern of the SST Trend Mode (°C) (**a**) and the associated time series (dark blue line), together with the AMO time series (turquoise line) (**b**), derived through EOF analysis of North Atlantic sector (80°W-0°E, 0°N-80°N) annual SST fields. The corresponding pattern and time series, for the South Atlantic sector, are shown in panels **c** and **d**. The sums of the two timeseries in panel **b** and that of the pair of time components in panel **d** are shown in panel **e** (indigo and magenta lines, respectively), together with an AMOC reconstruction[9] (orange line). Details on both EOF analyzes are presented in the supplement.

concentration[4]. The separation is performed through two EOF analyzes performed on North Atlantic and South Atlantic annual SST anomalies from the ERSST.v5 dataset, extending over the 1854–2016 period[14] (Supplementary Figs. 1–4). Similar results are obtained if the EOF method is applied to the Hadley SST fields[15].

Whereas the time component of one North Atlantic mode is marked by a decreasing centennial-scale trend (hereafter Trend Mode—TM), which starts around 1890, that of the other one is dominated by multidecadal fluctuations, typical for the Atlantic Multidecadal Oscillation (AMO) (Fig. 1b). While the TM pattern is dominated by three centers of alternating signs, disposed from SW to NE of Greenland (Fig. 1a), a structure which is linked to AMOC changes[9,10,16,17], the other mode has a typical monopolar AMO structure (Supplementary Fig. 2a), reflecting multidecadal AMOC variations[18,19]. The superposition of the time components of the two North Atlantic modes is strongly correlated (0.77) with a SST-based reconstruction[9] (Fig. 1e). The time series of the corresponding modes identified based on the South Atlantic SST fields show similar centennial-scale and multidecadal variations

(Fig. 1d). The South Atlantic TM pattern has a north-south oriented dipolar pattern (Fig. 1c). The other AMOC mode is marked by multidecadal fluctuations (Fig.1d) and by monopolar structures in both hemispheres of the Atlantic basin (Supplementary Figs. 2a and 4c), which are typical for AMO[8,20]. The TM(AMO) derived from North Atlantic SSTs has a 0.65 (0.49) maximum correlation with the corresponding mode obtained from South Atlantic fields, when the former leads the later by 19 (9) years.

The superposition of the time components of the two South Atlantic modes is significantly correlated (0.60) with the SST-based reconstruction[9] (Fig. 1e). Therefore, these two SST modes are characterized by distinct spatial and temporal features and are responsible for the largest part of decadal and longer AMOC variability.

A qualitatively identical and quantitatively similar bimodal decomposition of decadal and longer-term AMOC variability over the instrumental period was emphasized in an ensemble of historical simulations[21]. When these were combined with control simulations with constant radiative forcing, AMO appears

associated with internal AMOC variations, whereas the centennial scale trend was interpreted as an externally forced mode. One notes that in some of the historical simulations the AMOC centennial-scale decreasing trend starts at the end of the nineteenth century[21] (Fig. 11 in their study), as does the time component of TM, derived from observed SSTs (Fig. 1b). Furthermore, the Atlantic TM and AMO SST patterns were linked to forced and internal variations, respectively, based on model simulations and observations[22]. Consequently, hereafter we consider TM as an indicator of externally forced AMOC centennial-scale variations, separated here from the multidecadal internal fluctuations, reflected by AMO.

It was shown that the average SST anomalies over the subpolar gyre can be translated to AMOC changes, by using a calibration factor of 3.8 ± 0.5 Sv K^{-1}, derived from CMIP5 simulations[9]. The fact that the dominant center of TM coincides with the SST anomalies in this area (Fig. 1a), makes possible an estimation of the AMOC change associated with this mode. Consequently, the linear decreasing trend of TM (Fig. 1b) translates in an AMOC weakening of 3.7 ± 1.0 Sv over the 1854–2016 period.

The North Atlantic SST dipole pattern located south of Greenland in the TM structure (Fig. 1a) was associated with the AMOC slowdown in climate model simulations with increasing atmospheric CO_2 concentration[9,23–25]. This suggests that the TM's weakening trend, which starts in the late nineteenth century, is induced by this greenhouse gas (Fig. 1b).

Mechanisms of CO_2 influence on the AMOC trend

Model integrations indicate that an increase of atmospheric CO_2 concentration can weaken AMOC through increases of North Atlantic surface heat and freshwater fluxes[4,26–28]. Here we investigate these simulated connections, based on observational (SST—ERSSTv5 dataset) and reanalysis (Sea Level Pressure (SLP), heat fluxes and precipitation rate—NOAA/CIRES/DOE 20th Century Reanalysis V3 datasets) fields[14,29] (sea Methods).

The regression of spring North Atlantic Ocean heat fluxes on the time series of atmospheric CO_2 concentration[30] (Fig. 2a) is dominated by positive values (Fig. 2c), consistent with a direct thermodynamic influence of this greenhouse gas on ocean surface temperature. A quasi-identical pattern is obtained if a composite map is constructed as difference between average values over the 1935–2015 and 1854–1934 periods (Supplementary Fig. 5). The regression of the North Atlantic spring precipitation rate on CO_2 record is dominated by negative anomalies (not shown).

Model projections show that increasing atmospheric CO_2 concentration results in a positive trend of the dominant atmospheric mode in this sector, the North Atlantic Oscillation[31]. The pattern and time series of this mode are derived through EOF analysis (not shown). The regression map of the North Atlantic precipitation rate field on the NAO index is dominated by a prominent center of positive values located in the low-pressure center of this mode, consistent with an influence from the later to the former (Fig. 2d).

A similar regression map of heat fluxes on NAO time series reveals a dipolar structure, which has a small projection on the average value of the North Atlantic sector (not shown). Therefore, these analyzes suggest influences of increased atmospheric CO_2 concentration on spring North Atlantic surface heat fluxes (directly, thermodynamically) and on precipitation rate (indirectly, dynamically, through NAO), which were shown to weaken AMOC in model simulations[4,26–28].

The dominant growing character of these two fluxes, associated with increasing atmospheric CO_2 concentration (Fig. 2c, d), is reflected also in the upward trends of the integrated North Atlantic (70°W-20°E, 40°N-80°N) heat flux and precipitation rate time series (Fig. 2e, f). In order to test these observed connections linking increasing CO_2 with weakening AMOC, anticipated by

model simulations, we apply CCM on their associated pairs of variables.

CO_2-to-AMOC causal links

TDCCM is first applied to two possible connected time series in order to estimate the lag for which the cross-map estimate is maximum and to infer the correct sense of the causal links, when an unambiguous interpretation is available[11]. With the identified lag, CCM is used (Supplementary Fig. 9) to check for convergence and its statistical significance, based on which a potential causal relationship can be inferred (see Methods).

The cross map skill from North Atlantic spring surface heat flux to atmospheric CO_2 concentration increases with the library length and reaches a plateau around $\rho \cong 0.8$, well above the 95% significance level ($p < 0.03$), indicating a causal relationship from the later to the former (Fig. 3a). The convergent ($\rho \cong 0.7$) and significant ($p < 0.05$) cross map of the TM to heat flux (Fig. 3c) suggests a causal link from the latter to the former, thus completing the first causal chain from CO_2 to the AMOC weakening trend, via the heat flux. The second channel of causality starts again with CO_2 as a cause, but this time its dynamic signature is found in the spring SLP attractor. This causal character is inferred from the convergent ($\rho \cong 0.65$) and significant ($p < 0.05$) cross map (Fig. 3b). SLP further influences spring precipitation rate, this link having a convergence level of $\rho \cong 0.5$ and p value < 0.05 (Fig. 3d). Finally, the cross map from TM to spring precipitation rate shows clear convergence ($\rho \cong 0.6$) and statistical significance ($p < 0.02$), therefore indicating a causal connection from the later to the former (Fig. 3f). The CCMs in the opposite directions for all the above pairs are generally non-significant, with the exception of the TM-precipitation rate pair of time series, which are thus part of a feedback (Supplementary Figs. 6 and 7). Other potential causal channels linking CO_2 with AMOC are also tested, but are not significant (Supplementary Fig. 8).

Interhemispheric connections between the North Atlantic and South Atlantic components of TM are also explored (Fig. 3e). The CCM analysis reveals a robust causal influence from the northern component to the southern one ($\rho \cong 0.7$) significant above the 95% level ($p < 0.03$). No reversed significant convergence is detected (Supplementary Fig. 7e). Finally, the indirect link from the primary causal factor (CO_2) to the final recipient (North Atlantic TM) has a clear causal nature as it is revealed by the convergent ($\rho \cong 0.9$) and significant ($p < 0.02$) cross map skill from the latter to the former (Fig. 3g).

The relatively high convergence level may reflect synchronicity, but this is dismissed by the inverse cross map, which is not statistically significant ($p > 0.05$, Supplementary Fig. 7g). The thermodynamical and dynamical causal links discussed above are synthesized in Fig. 4. Similar possible causal channels are investigated for all other seasons. The only significant one is found for winter: $CO_2 \rightarrow$ Heat flux $\rightarrow$ North Atlantic TM (Supplementary Figs. 10 and 11).

An *increasing* $CO_2 \rightarrow$ *AMOC weakening* causal connection inferred here based on observed and reanalysis data, is consistent with the anticorrelated millennial record levels of high atmospheric CO_2 concentrations[32] and the reconstructed record low level of the AMOC strength over the last decades[10].

DISCUSSION

An annual SST-based AMOC reconstruction shows a pronounced long-term slowdown since 1950s and no significant trend before[9]. Based on observed Atlantic SST fingerprints we separate associated centennial and multidecadal AMOC variations. The centennial-scale component indicates that an AMOC weakening trend starts earlier, in the late nineteenth century, several decades after the onset of the sustained industrial-era warming[33].
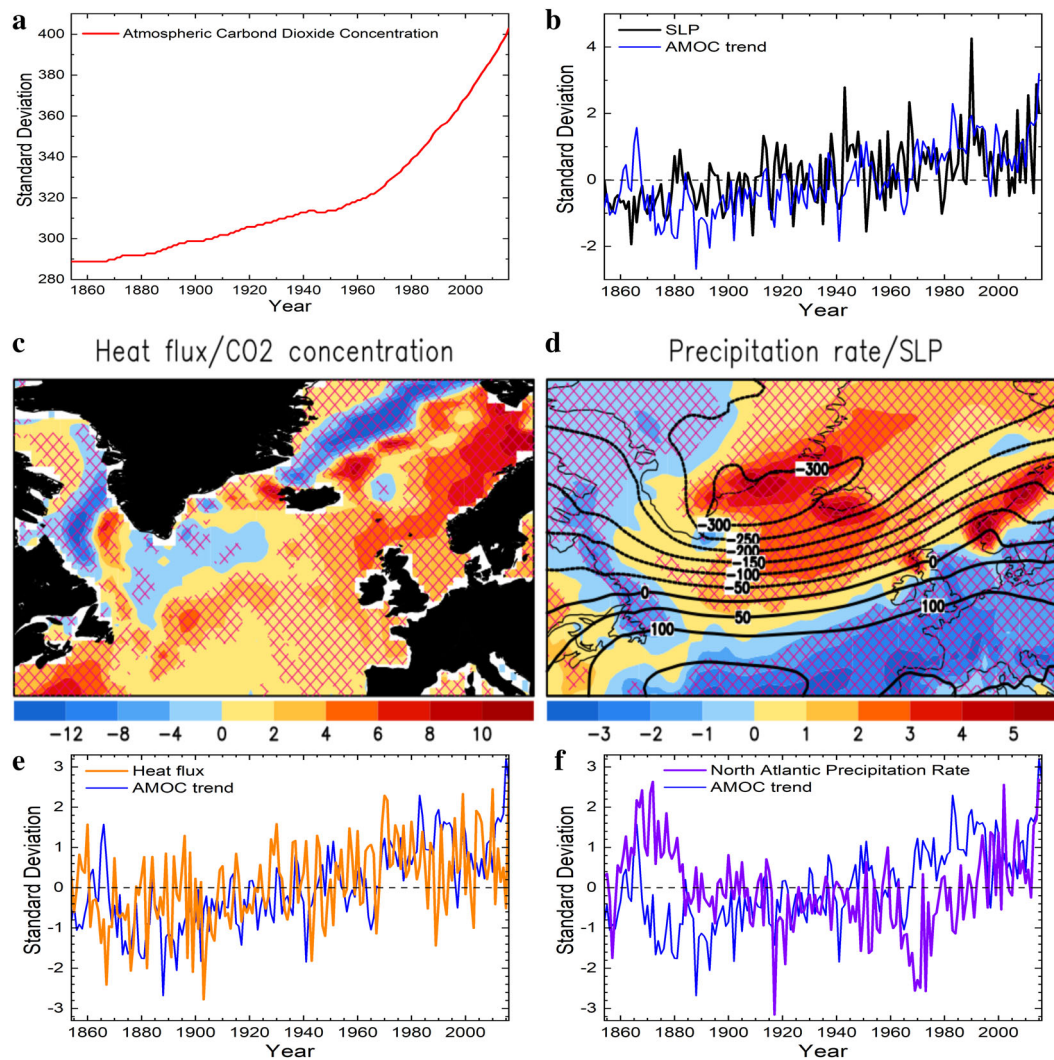
Fig. 2 CO₂ footprints on surface fluxes. a Reconstruction of annual atmospheric CO_2 concentration (ppm); **b** Time series of the dominant North Atlantic spring SLP mode (black line) and of the SST/AMOC Trend Mode (blue line) (**c**) regression map of spring surface heat fluxes on the annual atmospheric CO_2 concentration (W/m²/ppm); for positive values the flux is directed from the atmosphere into the ocean (**d**) regression map of spring precipitation rate on the SLP time series shown in panel **b** ([kg/(m² × s)/Pa] × 10⁶); the dominant North Atlantic SLP mode is also shown (solid black contour lines) (Pa); **e** time series of average North Atlantic ocean spring heat fluxes (orange line) and time series of the SST/AMOC Trend Mode (blue line). **f** Time series of average North Atlantic spring precipitation rate (violet line) and the time series of the SST Trend Mode (blue line). In panels **c** and **d** the regions for which the regression is significant above the 95% level are marked by red grids.

In model integrations, centennial-scale increasing atmospheric CO_2 concentration affects North Atlantic spring heat fluxes and precipitation rate, which results in an AMOC weakening trend[4,28]. We construct regression maps based on observed and reanalysis data, which support these simulated mechanisms. Furthermore, by applying the CCM method on pairs of observed and reanalysis time series, we identify the causal connections linking increasing atmospheric CO_2 concentration with AMOC weakening (Fig. 4).

Our analyzes of observational and reanalysis data suggest that the AMOC linear slowdown over the 1854–2016 period, estimated at 3.7 ± 1.0 Sv, is larger than the of 1.4 ± 1.4 Sv weakening from the preindustrial period (1850–1900) to present days (2006–2015), exhibited in climate projections[5].

METHODS

Sea surface temperature

Due to their relatively long-time span, observed SSTs were used to infer ocean circulation variations from surface measurements[9]. Here we use

fields from the ERSSTv5 dataset, distributed on a 2° × 2° grid and extending over the 1854–2016 period[14]. Similar results with that presented here, were obtained based on the SST fields from the HADISST1 dataset, distributed over 1° × 1° grids[15].

Atmospheric CO₂ concentration

Reconstructed values of annual means of atmospheric CO_2 concentration[30] were obtained from: https://climexp.knmi.nl/start.cgi.

Reanalysis SLP, heat fluxes and precipitation rate

SLP, heat fluxes and precipitation rate are from the monthly NOAA/CIRES/ DOE 20th Century Reanalysis V3 data set[29], were obtained from https:// climexp.knmi.nl/start.cgi. They are distributed over a 1° × 1° grid and extend over the 1836–2015 period.

Preprocessing

Before all analyzes a pre-filtering procedure is applied to the SST fields in order to remove the uniform global warming trend. The yearly global average is subtracted from each grid point. The procedure
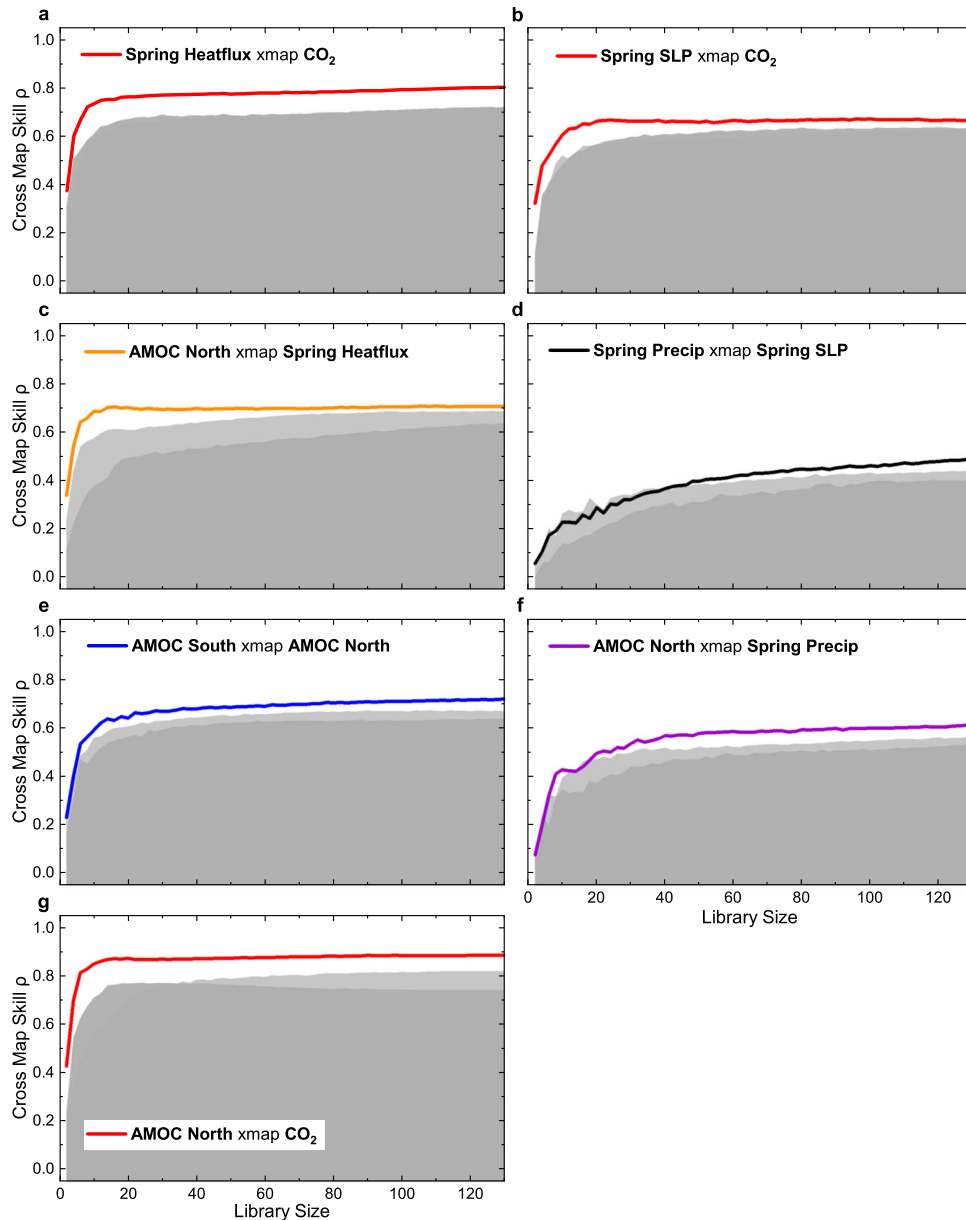
Fig. 3 CCM analyzes. Cross map skill given by the correlation between predicted and observed values, as a function of library length for: **a** spring heat flux and CO_2; **b** spring SLP and CO_2; **c** AMOC North and heat flux; **d** spring precipitation rate and spring SLP; **e** AMOC South and AMOC North; **f** AMOC North and spring precipitation rate; **g** AMOC North and CO_2. The light (dark) gray shaded areas correspond to the 95th cross map between each affected variable and the surrogate of the cause generated by a swap (Ebisuzaki) model. All variables have the embedding dimension, $E = 8$ and embedding lag, $l = 1$. The saturation of a cross map between X and Y variables, at a plateau above the significance levels, indicates a causal connection from Y to X. All cross maps exhibit statistically significant levels of convergence. In order to damp interannual fluctuations and focus on decadal and longer variations, a 3-year running mean filter was applied to the time series (excepting the heat flux record, for which the length of the filter was 5 years), before the CCM analyzes. The CCM results for the opposite causal relationships are shown in Supplementary Fig. 7.

has the advantage that it removes the spatially quasi-uniform nonlinear trend determined from the data, without need to a priori choose a linear or nonlinear shape to be removed. The global mean time series is quasi-identical with that of the average SST anomalies over a smaller domain (e.g., 0–360°E, 70°S–70°N) and therefore the subtracting method is not sensitive to scarcity of data in high latitudes. Because the globally uniform warming trend explains a large amount of variance in the initial SST fields, but the focus of this study is on spatially heterogeneous patterns, this preliminary operation increases significantly the signal-to-noise ratio in the SST data.

EOF analysis

The North and South Atlantic modes of SST variability, which are linked to AMOC changes, are identified through Empirical Orthogonal Functions (EOF) analyzes (Supplementary Figs. S1–S3).

This method is also used to identify the dominant mode of North Atlantic spring SLP variability and its associated time series. The first EOF, explaining 42% of variance, is the North Atlantic Oscillation.

Convergent cross mapping

The identification of causal relationships based on empirical data represents a critical problem across a wide range of scientific fields, which
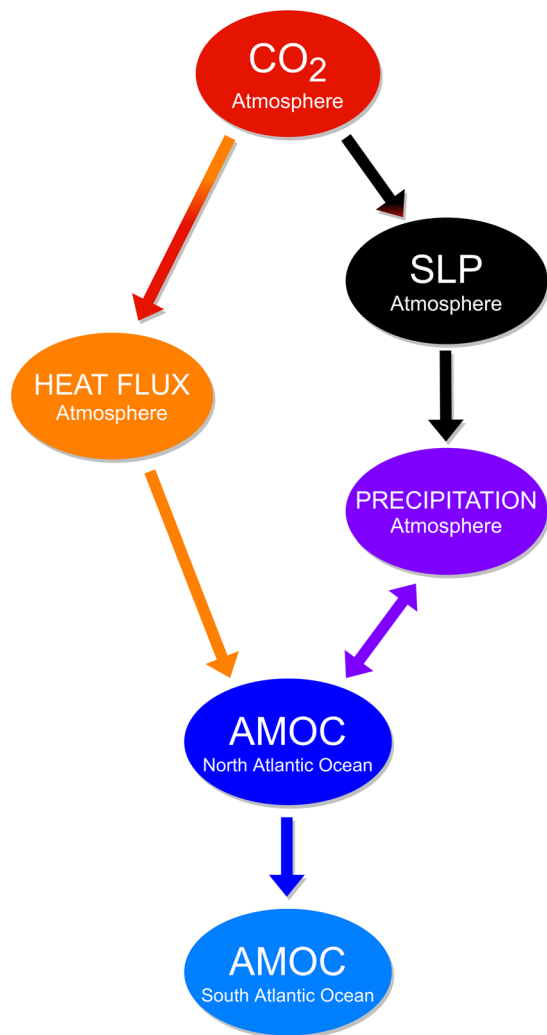
Fig. 4 Causal chains of the atmospheric CO_2 concentration influence on AMOC. It includes a direct, thermodynamic branch (left route), through heat fluxes and an indirect, dynamical one (right route), through NAO-driven precipitation rates. Also, the southward interhemispheric connection is depicted in the bottom link. The diagram is the conceptual analog of the CCM results shown in Fig. 3.

cannot be satisfactorily solved using correlations, which is a poor indicator of causality. A significant correlation between two variables does not imply a direct causal link between them. For example, a third variable could drive both of them. Similarly, a weak correlation between two variables does not imply a lack of causal relationship as it is often the case for systems governed by nonlinear dynamics. A recently proposed method, Convergent Cross Mapping (CCM), relying on time embedded state space reconstruction based on data, provides significant progress on this problem[11].

The dynamics of the system is represented by coherent trajectories in state space which organize into an (usually lower dimensional) attractor manifold. Time is implicit and is represented by the direction along the state space trajectory. For deterministic dynamical systems, the variables are not independent and the system must be understood as a whole, rather than the sum of its parts. This non-separability translates into the fact that information about past states is carried forward through time and any variable in the system contains information about the states of the other. The historical values of a variable contain information about both, its past behavior and the instantaneous interdependence between the variables of the system. Thus Takens' theorem applies, stating that we may reconstruct the underlying attractor manifold of a system by a time embedding of only one variable in the system, say Y. If two variables, X and Y, are bidirectionally causally linked then they contain information about each other and share a common attractor manifold. Their time embedded

attractor reconstructions will be topologically equivalent (diffeomorphic) and nearby points in the attractor of Y will correspond to nearby points in the attractor of X. On the other hand, if only X causes Y, then Y will contain information about X, but the time evolution of X is independent of Y and the former variable does not contain information about the later. The CCM method extracts causal signatures using prediction as a criterion: from the E-dimensional time embedded attractor manifold of variable Y, find nearest neighbors to a given point at time t and construct weights from the identified neighbors. An estimate of X at time t is generated using these weights. This procedure is repeated for the values of Y at all times and a correlation measure between the predicted and observed time series is computed. Here we use Pearson's correlation coefficient, ρ, as a measure of correlation. Most importantly, to truly distinguish between causality and correlation, one should check for the property of convergence of the cross estimation with the library size, that is the increase in estimation precision when considering more points for the prediction. The accuracy and convergence of the cross prediction may be limited by noise, observational error, time series length, but also by the complexity of the real-world systems, which could exhibit transitory and non-stationary causal behavior. As mentioned before, the method applies to nonlinear deterministic systems, even to stochastic ones as long as they are not completely random[11]. Thus, CCM becomes a necessary condition for causation.

CCM depends on several parameters. First of all, as is the case with any time embedded state space reconstruction method, we have a dependency on the embedding dimension, E and the embedding lag, l. Here we choose $E = 8$ and $l = 1$ everywhere. For our analyzes, Simplex Projection is essentially constant for any embedding dimension (Fig. S9a). On the other hand, $E = 8$ seems to be the right number of dimensions for cross prediction between the variables of our physical climatic system (Fig. S9b). Each cross-map value in the CCMs is computed as the average over 100 random samples (without replacement) at each library size.

Time delayed convergent cross mapping

The question still remains if CCM could indicate a false positive causal link. In cases of strong unidirectional forcing (X causes Y) the affected variable (Y) becomes a dynamical slave of its cause (X) so that they vary synchronously. In such a situation, the CCM could indicate a virtual bidirectional causal relationship, even if there is no information transfer from the effect to the cause. However, in such cases causality can still be inferred through surrogate analysis or Time Delayed CCM (TDCCM)[34]. Here, we preliminarily use TDCCM to find the lag for maximum predictability and use surrogate analysis to single out the unidirectional causal connections.

TDCCM consists of calculating the cross map skill for different lags in order to reveal one for which the prediction skill is optimal (i.e., maximum)[34]. This lag is negative since we make the prediction backwards in time, from the effect $Y(t)$ to the cause $X(t - \tau)$ and cause must precede effect ($X(t - \tau)$ causes $Y(t)$). Each cross-map value is computed for the maximum library size available for each time delay. The identified lag is then used in CCM. As the exact value of the delay is highly unstable with respect to embedding dimension and embedding lag[35], we don't rely on its physical relevance, but rather use it as an optimization tool. TDCCM also provides qualitative insight into the correct causal directions through the sign of the maximum cross map skill lag (negative for the true causal direction). Nevertheless, the correct causal directions are primarily identified through surrogate analysis: we apply CCM between the presumed effect and the surrogate of the cause generated under two surrogate models.

This technique, together with CCM, was used, for example, to investigate potential causal relationships between atmospheric CO_2 concentration and global temperature[36].

Statistical significance of CCM

We estimate statistical significance under the null hypothesis that the (possible) effect does not contain information about the (possible) cause. To test it, we use surrogate randomization models for the cause. Under the null hypothesis, cross maps from the effect to the surrogate cause might be generated by information which was not destroyed by the randomization procedure, or by spurious correlation in the time series. The rejection of the null hypothesis means that we can find a cross map estimate above the 95% percentile of the estimates for the surrogates (or a p value of $p < 0.05$). As mentioned before, we can use statistical significance to single out the true direction of causality in the cases of synchrony or ambiguous

TDCCM[37]. Here, we employ two surrogate tests[36]: (a) Ebisuzaki phase shift: one keeps the same frequency spectrum as the original time series and randomize the phases, generating 100 surrogates of the cause and try to estimate them from the effect. In Fig. 3 we have represented 95 of the cross maps obtained, eliminating the top 5; (b) Swap model: one chooses a random point in the time series and swap the two segments; this procedure randomizes the phases, while preserving nearly all short-term deterministic dependencies; 100 surrogates of the effect variable are created and 95 of them are represented in Fig. 3, eliminating the top 5.

DATA AVAILABILITY

All data sources are mentioned in the Methods section.

CODE AVAILABILITY

Code for CCM analysis was obtained from[11]. Codes to produce the figures are available from the corresponding author.

REFERENCES

1. Lenton, T. M. et al. Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
2. Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature* **419**, 207–214 (2002).
3. Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic Thermohaline Circulation. *Clim. Change* **54**, 251–267 (2002).
4. Collins, M., et al. Long-term climate change: Projections, commitments and irreversibility, 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 (Cambridge Univ. Press, 2013).
5. Collins M. et al. Extremes, abrupt changes and managing risk. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate (2019). Available online: https://www.ipcc.ch/srocc/.
6. Bryden, H. L., Longworth, H. R. & Cunningham, S. A. Slowing of the Atlantic meridional overturning circulation at 25°N. *Nature* **438**, 655–657 (2005).
7. Smeed, D. A. et al. Observed decline of the Atlantic meridional overturning circulation 2004–2012. *Ocean Sci.* **10**, 29–38 (2014).
8. Dima, M. & Lohmann, G. Evidence for two distinct modes of large-scale ocean circulation changes over the last century. *J. Clim.* **23**, 5–16 (2010).
9. Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G. & Saba, V. Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature* **556**, 191–196 (2018).
10. Rahmstorf, S. et al. Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nat. Clim. Change* **5**, 475–480 (2015).
11. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
12. Park, W., & Latif, M. Multidecadal and multicentenial variability of the meridional overturning circulation. *Geophys. Res. Lett*. **35**, https://doi.org/10.1029/2008GL035779 (2008).
13. Cheng, W., Chiang, J. C. H. & Zhang, D. Atlantic Meridional Overturning Circulation (AMOC) in CMIP5 models: RCP and historical simulations. *J. Clim.* **26**, 7187–7197 (2013).
14. Huang, B. P. et al. Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5), upgrades, validations, and intercomparisons. *J. Clim*. https://doi.org/10.1175/JCLI-D-16-0826.1 (2017).
15. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407, https://doi.org/10.1029/2002JD002670 (2003).
16. Born, A., Levermann, A. & Mignot, J. Sensitivity of the Atlantic Ocean circulation to a hydraulic overflow parametrisation in a coarse resolution model: response of the subpolar gyre. *Oc. Model.* **27**, 130–142 (2009).
17. Zhang, R. et al. Sensitivity of the North Atlantic Ocean Circulation to an abrupt change in the Nordic Sea overflow in a high resolution global climate model, *J. Geophys. Res*. **116**, https://doi.org/10.1029/2011JC007240 (2011).
18. Latif, M. et al. Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with sea surface temperature. *J. Clim.* **17**, 1605–1614 (2004).
19. Knight, J. R., Allan, R. J., Folland, C. K., Vellinga, M. & Mann, M. E. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32**, L20708, https://doi.org/10.1029/2005GL024233 (2005).
20. Schlesinger, M. E. & Ramankutty, N. An oscillation in the global climate system of period 65–70 years. *Nature* **367**, 723–726 (1994).
21. Schmith, T., Yang, S., Gleeson, E. & Semmler, T. How much have variations in the meridional overturning circulation contributed to sea surface temperature trends since 1850? A study with the EC-Earth global climate model. *J. Clim.* **27**, 6343–6357 (2014).
22. DelSole, T., Tippett, M. K., Shukla, J. A significant component of unforced multi-decadal variability in the recent acceleration of global warming, *J. Clim.* **24**, 909–926 (2011).
23. Saba, V. S. et al. Enhanced warming of the Northwest Atlantic Ocean under climate change. *J. Geophys. Res. Oceans* **120**, https://doi.org/10.1002/2015JC011346 (2015).
24. Sgubin, G., Swingedouw, D., Drijfhout, S., Marry, Y. & Bennabi, A. Abrupt cooling over the North Atlantic in modern climate models. *Nat. Commun*. **8**, https://doi.org/10.1038/ncomms14375 (2017).
25. Menary, B. M. & Wood, R. A. An anatomy of the projected North Atlantic warming hole in CMIP5 models. *Clim. Dyn.* **50**, 3063–3080 (2018).
26. Dixon, K., Delworth, T., Spelman, M. & Stouffer, R. The influence of transient surface fluxes on North Atlantic overturning in a coupled GCM climate change experiment. *Geophys. Res. Let.* **26**, 2749–2752 (1999).
27. Mikolajewicz, U. & Voss, R. The role of the individual air-sea flux components in CO_2-induced changes of the ocean's circulation and climate. *Clim. Dyn.* **16**, 627–642 (2000).
28. Gregory, J. M. et al. A model intercomparasion of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 concentration. *Geophys. Res. Let.* **32**, https://doi.org/10.1029/2005GL023209 (2005).
29. Slivinski, L. C. et al. Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis System. *Q. J. R. Met. Soc.* **145**, 2876–2908 (2019).
30. Ballantyne, A. P., Alden, C. B., Miller, J. B., Tans, P. P. & White, J. W. C. Increase in observed net carbon dioxide uptake by land and oceans during the last 50 years. *Nature* **488**, 70–72 (2012).
31. Kuzmina, S. I., Bengtsson, L., Johannessen, O. M. & Drange, H. The North Atlantic Oscillation and greenhouse-gas forcing. *Geophys. Res. Let.* **32**, https://doi.org/10.1029/2004GL021064 (2005).
32. Meure, C. M. et al. LowDome CO_2, CH4 and N2O ice core records extended to 2000 years BP. *Geophys. Ress. Lett.* **33**, L14810, https://doi.org/10.1029/2006GL026152 (2006).
33. Abram, J. N. et al. Early onset of industrial-era warming across the oceans and continents. *Nature* **536**, 411–418 (2016).
34. Ye, H., Deyle, E. R., Gilarranz, L. J. & Sugihara, G. Distinguishing time-delay causal interactions using convergent cross mapping. *Sci. Rep.* **5**, 1–9 (2015).
35. Ma, H. et al. Detection of time delays and directional interactions based on time series from complex dynamical systems. *Phys. Rev. E* https://doi.org/10.1103/PhysRevE.96.012221 (2015).
36. van Ness, E. H. et al. Causal feedback in climate change. *Nat. Clim. Change* https://doi.org/10.1038/nclimate2568 (2015).
37. Sugihara, G., Deyle, E. R. & Ye, H. Misconceptions about causality with synchrony. *PNAS* **114**, E2272–E2274, https://doi.org/10.1073/pnas.1700998114 (2017).

AUTHOR CONTRIBUTIONS

M.D. initiated and led this research. M.D. and D.N. performed the analyses. All authors participated in discussions during this study and contributed to the writing and revising the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-021-00182-x.

Correspondence and requests for materials should be addressed to M.D.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published in partnership with CECCR at King Abdulaziz University

Supplementary Information

Early-onset of Atlantic Meridional Overturning Circulation weakening in response to atmospheric CO$_2$ concentration

Mihai Dima[1,2]*, Denis R. Nichita[2,3], Gerrit Lohmann[1], Monica Ionita[1], Mirela Voiculescu[4]

[1]Alfred Wegener Institute for Polar and Marine Research
[2]University of Bucharest, Faculty of Physics
[3]"Horia Hulubei" National Institute of Physics and Nuclear Engineering, Bucharest-Măgurele, Romania
[4]University "Dunărea de Jos", Galati

Content

Supplementary Methods

As numerical simulations show a centennial-scale weakening trend of AMOC in response to CO_2 increase[1], we assume that a similar one is contained in the observational data and we aim to identify and separate it. An AMOC weakening trend extending over the last decades was linked with center of SST anomalies located south of Greenland[2,3]. Similarly, several previous studies associated AMOC changes with AMO, a mode characterized by a monopolar North Atlantic SST pattern and by a time component marked my multidecadal fluctuations[4,5]. We aim to separate in observational North Atlantic SSTs these two distinct modes (patterns and corresponding time series), one associated with a secular trend and another one with multidecadal fluctuations. If we identify an AMOC mode associated with a time component characterized by centennial scale variability, then we can use the CCM method in order to test if it is caused by increasing atmospheric CO_2 concentrations. Consequently, we performed EOF analysis on annual mean North Atlantic (80^oW-0,0-80^oN) anomalous SSTs from the ERSSTv5 dataset, which are distributed on a 2° x 2° grid and extend over the 1854-2016 period.

In order to increase the signal-to-noise ratio, the quasi-uniform global warming trend was removed before the EOF analysis, by subtracting the annual global mean from all points in each grid. This preprocessing stage is motivated by several considerations. The increasing atmospheric CO_2 concentration has a complex influence on climate. One direct impact is the global warming, which can be considered a first order effect. However, indirectly, the global warming could induce further changes on the climate system, like, for example, altering the North Atlantic freshwater and heat fluxes, which can subsequently on AMOC, as suggested by model simulations[6,7]. Such an indirect impact

51 could be considered a second order effect of increasing atmospheric CO_2 concentration.

52 One notes that changes in AMOC are associated with interhemispheric dipolar SST

53 patterns[8], which have small projection on the global average (i.e. when including in the

54 global average, the southern hemisphere part of the SST dipole induced by AMOC is

55 largely canceled by the northern hemisphere part). Therefore, by subtracting the uniform

56 global mean SST, one largely removes the first order effect of the increasing atmospheric

57 CO_2 concentration on the climate system. However, second order effects (e.g. weakening

58 AMOC and the associated interhemispheric dipolar structure), with insignificant projection

59 on the global mean, are not eliminated. This is of specific interest in our study. One notes

60 also that the global warming trend explains the largest part of variance in the SST field,

61 although it is not of interest in our study, which is focused on the second order effects of

62 the increasing atmospheric CO_2 concentration on the Atlantic ocean circulation.

63 Consequently, by removing the global warming trend, the signal-to-noise ratio is

64 significantly increased in our study. If the EOF analysis is performed without previously

65 removing the global warming trend, the obtained dominant mode looks like a

66 superposition of EOF1 obtained in the analysis without the global warming trend and the

67 warming trend. A regression of the global SST field on the increasing atmospheric CO_2

68 concentration reveals a similar pattern.

69 The first four EOFs are explaining 25%, 14%, 11% and 9% of variance of the annual SST

70 field, respectively. Among these modes, only EOF3 and EOF4 are not well separated

71 according to the North's criterion[9]. This implies that they represent random mixtures of

72 the true eigenvectors. However, our goal is to select and combine all modes which show

73 multidecadal fluctuation and centennial scale trends, regardless if they represent a

74 degenerated multiplet or true eigenvectors. In other words, as long as all temporal

75 fluctuations of these types are captured, it is of no importance if they are provided by a

76 degenerated multiplet or by a true eigenvector. In particular, in this case only EOF4/PC4

77 includes such variations, unlike EOF3/PC3. Therefore, we conclude that the degeneracy

78 of EOF3 and EOF4 has no impact on our results.

79 In order to express the EOFs' structures in degrees, they are multiplied with the standard

80 deviations of the corresponding PCs. The last are normalized by their standard

81 deviations.

82 From the resulting EOFs only the time components of the first four show multidecadal

83 and/or longer time scale variations (Supplementary Fig. 1). However, among these, PC3

84 includes a centennial fluctuation (with minima around 1870 and 1970), but does not how

85 a centennial trend or multidecadal fluctuations (Supplementary Fig. 1f). Consequently,

86 PC3 is not considered further. One observes also that none of the remaining three PCs

87 (PC1, PC2 and PC4) is characterized only by a centennial trend or only by multidecadal

88 fluctuations, but they show a mixture of these two types of temporal variations. In

89 particular, PC1 includes multidecadal fluctuations but also a decreasing trend over the

90 second part of the analyzed period (Supplementary Fig. 1b). PC2 presents multidecadal

91 fluctuations which appear to be superimposed on a centennial decreasing trend starting

92 at the end of the 19th century (Supplementary Fig. 1d). In PC4 the fluctuations have a

93 smaller amplitude and a secular trend is more prominent (Supplementary Fig. 1h).

94 Therefore, it appears that these three PCs are containing mixed information related to

95 multidecadal fluctuations and to secular trends.

4

Given that 1) we aim to disentangle two modes which are linked to AMOC changes, which are mutually linked through various complex linear and nonlinear processes and 2) the separation is performed through the EOF decomposition, which is linear, one could expect that essential characteristic information of each of the two AMOC modes is distributed over several EOFs/PCs. This is consistent with the observation that the PC1, PC2 and PC4 are containing mixed information about the two timescales of interest here (centennial trend and multidecadal fluctuations). Consequently, we attempt to separate the two AMOC modes by screening all linear combinations of normalized PC1, PC2 and PC4, and of their corresponding EOFs expressed in degrees. All four possible combinations of these PCs and EOFs are shown in Supplementary Fig. 2.

PC1+PC2+PC4 shows clear multidecadal fluctuations but no secular trend (Supplementary Fig. 2b) and the corresponding combination of EOFs is marked by a monopolar structure (Supplementary Fig. 2a). These features are typical for AMO (e.g. Latif et al. 2004), indicating that this combination provides an effective isolation of AMO.

PC1-PC2+PC4 appears to be characterized by interdecadal fluctuations (40-50 years), but no centennial trend (Supplementary Fig. 2d). The corresponding pattern resembles the NAO fingerprint on SST[10]. As this resulting time component does not include a centennial scale trend or multidecadal fluctuations, it is not of interest here.

PC1+PC2-PC4 (Supplementary Fig. 2f) shows no multidecadal fluctuations or centennial trend, but only a relatively short rapid decrease to a lower mean level around 1960. Consequently, this combination is not of interest here.

PC1-PC2-PC4 shows a clear long-term trend which starts in 1890s and extends to 2016 (Supplementary Fig. 2h). The corresponding SST pattern has a tripolar structure,

119   dominated by a center of positive values located south of Greenland and with two centers

120   of negative values located southwestward and northeastward from it (Supplementary Fig.

121   2g). Such a structure, including its positive center, was associated with AMOC

122   changes[2,3,11]. Furthermore, the positive center and the negative one located in the Gulf

123   Stream were associated to AMOC response to increasing atmospheric CO_2 in numerical

124   simulations[12].

125   We consider the separation of the two modes performed through the procedure described

126   above as optimal, based on three criteria:

127   1) They have clear distinct characteristic, without containing mixed information, as do the

128   initial EOFs/PCs. Whereas the Trend Mode is associated with a centennial trend and

129   a tripolar structure, AMO is characterized by multidecadal fluctuations and a

130   monopolar spatial structure, which were associated with AMOC variations;

131   2) A very similar bimodal decomposition of AMOC long term variability over the

132   instrumental period was emphasized based on control and historical model

133   integrations[13];

134   3) The sum of the time components of these two distinct reconstructed modes is

135   significantly correlated (0.77) with an SST-based AMOC index (Fig. 1e), suggesting

136   that together they provide a quasi-complete bimodal decomposition of AMOC decadal

137   and longer-term variability. This applies to both, North Atlantic and South Atlantic

138   analyzes and decompositions.

139

140

141   A similar EOF analysis is performed also on South Atlantic SST fields (70°W-20°E, 80°S-

142   0), in order to reconstruct the two SST-AMOC modes.

143   The first 5 PCs show multidecadal and centennial fluctuations and therefore are further

144   considered, together with the corresponding EOFs (Supplementary Fig. 3), which explain

145   25%, 22%, 10%, 6% and 5%. The first four have centennial scale trends, but only PC1,

146   PC2 and PC5 show multidecadal fluctuations. As for the North Atlantic case, we screened

147   all linear combinations of these two groups of PCs (not shown) and found out that two

148   combinations, PC1+PC2+PC3+PC4 and PC1-PC2-PC5, provide the most accurate

149   separation between two AMOC modes. The first shows a centennial decreasing trend

150   starting in 1900s (Fig. 1d) and resembles the AMOC trend derived based on North Atlantic

151   SSTs (Fig. 1b). The maximum correlation coefficient between these two time series is

152   maximum (0.65) when the later (North) leads the former (South) with 19 years. The

153   corresponding SST pattern has a dipolar structure (Fig. 1c). For the second combination

154   (PC1-PC2-PC5, Fig. 1d), the maximum correlation with the corresponding time

155   component derived based on North Atlantic SSTs (Fig. 1b) (0.49), is obtained when the

156   latter leads the former by 9 years. The associated spatial pattern is dominated by a center

157   of negative anomalies (Supplementary Fig. 4b). The superposition of the two

158   reconstructed time components is significantly correlated (0.60) with an SST based

159   AMOC reconstruction[3] (Fig. 1e).

160   In order to test the robustness of the regression maps of the heat flux field on the time

161   series of the atmospheric CO_2 concentration, we infer the projection of this time series on

162   the same field through another method. Given that over the 1854-2015 time interval the

163   atmospheric CO_2 concentration shows a monotonic increase (Fig. 2a), the composite

7

164 maps constructed based on periods corresponding to its maximum (1935-2015) and

165 minimum phases (1854-1934) (average over the former period minus mean of the last

166 time interval) reflects the projection of this record on the heat flux field. The composite

167 map (Supplementary Fig. 5) has a very similar structure with the regression map (Fig.

168 2c), indicating that they robustly reflect the CO_2 connection with the heat flux field.

169

170 In Time Delay CCM (TDCCM) representations, a maximum of the cross-map estimation

171 for a negative lag indicates a physical causal connection (i.e. the affected variable best

172 predicts past values of the driver variable). Conversely, a maximum for a positive lag (i.e.

173 the causal variable best predicts future values of the effect) indicates an unphysical

174 causal connection and it can only be the result of unidirectional forcing, even in the cases

175 of synchrony. We stress the fact that TDCCM is not a sufficient indicator of causality. It

176 optimizes the cross-map skill and singles out the physical directions of causality (when a

177 clear interpretation is available). Only in conjunction with statistically significant

178 convergence of cross map one can infer a causal connection.

179 TDCCMs involving CO_2 show very small fluctuations with respect to the variation of the

180 time-delay (Supplementary Fig. 6a,b). We choose the instantaneous CCM. Surrogate

181 analysis in Fig. 3a,b and in Supplementary Fig. 7a,b determine the unidirectional nature

182 of causality from the greenhouse gas forcing.

183 Supplementary Fig. 6c, showing the AMOC North vs Spring Heat flux, indicates

184 synchronicity reflected in symmetric maxima of $\rho \cong 0.8$ for negative and positive lags, the

185 true direction of causality being Spring Heat flux $\rightarrow$ AMOC North (red, negative lag).

186 Supplementary Fig. 6d reflects the link between atmospheric variables, Spring

187 Precipitation vs Spring SLP, and displays oscillatory behavior in cross estimate.

188 Consequently, its interpretation is not straightforward. The chosen maxima are still global,

189 but surrogate analysis ultimately elects the valid direction of causality (Fig. 3d and

190 Supplementary Fig. 7d), that is Spring SLP $\rightarrow$ Spring Precipitation Rate. In

191 Supplementary Fig. 6e clear maxima for each cross-map are observed, indicating

192 unidirectional southward influence from TM North to TM South. Supplementary Fig. 6f

193 exhibits bidirectional causality between AMOC North and Spring Precipitation as two

194 global maxima for negative lags are observed in both directions. Surrogate analysis is

195 consistent with bidirectionality (Fig. 3f and Supplementary Fig. 7f).

196 In Supplementary Fig. 6g the link $CO_2 \rightarrow$ TM North is depicted, showing an almost

197 constant correlation for negative lags in the TM North xmap CO_2 cross-map followed by

198 a sudden decrease for positive lags. This is consistent with a delayed, indirect interaction.

199 The inverse cross-estimate is manifestly constant, with a slight increase for positive lags,

200 indicating either dynamic independence of CO_2 or synchronicity (taking into account that

201 $\rho \cong 0.9$).

202 All cross-mappings are computed with embedding dimension E=8 and embedding lag

203 l=1.

204 Alternative, complementary and possibly competing causal chains with the ones identified

205 through regression and CCM in the main text could exist. We analyze two possibilities,

206 through the results presented in Supplementary Fig 8. The first column presents evidence

207 that CO_2 could influence AMOC through SLP-driven heat fluxes during spring, as follows:

208 panel a) contains the maximum of TDCCM favoring the SLP $\rightarrow$ Heat flux causal link

9

209 (Supplementary Fig. 8a), which is supported by the CCMs with the significant

210 convergence having SLP as a forcing factor (Supplementary Fig. 8c,e). Second column

211 presents evidence of direct thermodynamic effect of CO_2 on precipitation rate. An

212 invariant cross-map with respect to lag (Supplementary Fig. 8b) is associated with an

213 unidirectional causal influence from CO_2 to spring precipitation rate, as revealed by the

214 significant cross map in Supplementary Fig. 8d and non-significant one in panel

215 Supplementary Fig. 8f. The lag used is 0. Taking into account the fact that the regression

216 of precipitation rate on CO_2 is dominated by negative anomalies (not shown) and that the

217 structure of the regression map of NAO on heat fluxes is heterogeneous (not shown),

218 these two causal channels are thought to be less physically significant, but competing

219 components to those described in Fig. 3 and Fig. 4.

220 In order to construct the time-embedding necessary for CCM estimation we have to

221 choose the Embedding dimension in advance. Usually, this is done using Simplex

222 projection method. This is a nearest-neighbor method for forecasting time-series, even

223 chaotic ones involving "*tracking the forward evolution of nearby points in an*

224 *embedding*"[14]. By using a past library set of a time series to construct an embedding,

225 simplex projection predicts the values of its out-of-sample complement. Of course, points

226 are nearby depending on the chosen embedding dimension and the prediction will vary

227 accordingly. We can thus find an optimal embedding dimension for which the prediction

228 is maximally accurate. A correlation measure (here, Pearson's correlation coefficient)

229 between predicted and observed values is used to assess the accuracy.

230 We employ simplex projection for each variable in our analyses (variables in Fig 3) to

231 determine its best embedding dimension. The correlation as a function of embedding

232 dimension is plotted in Supplementary Fig. 8a, for each variable. All forecasts are

233 approximatively constant having a slight increase for larger values of E. The ambiguity of

234 simplex projection prompted us to use a more empirical measure for the embedding

235 dimension. For all pairs of variables, the instantaneous CCM skill is computed, using the

236 full library length available and it is plotted as a function of Embedding Dimension

237 (Supplementary Fig. 8b)). The cross estimate increases until it settles to an approximately

238 constant level, meaning that from a certain point on, the further enlargement of the

239 dimensionality of the state space is quasi-redundant with respect to the accuracy of the

240 prediction. A conservative choice of embedding dimension so as to ensure maximum

241 predictability before the asymptotic constant level is reached, is $E = 8$. The embedding

242 lag is left at its default value of $l = 1$.

243 The main CCM analyses of the present manuscript, presented in Fig 3 involve the time

244 series of spring observables. The analogous analyses on autumn and summer are not

245 statistically significant and do not reveal consistent causal channels among these

246 variables (not shown). Nevertheless, on winter, we were able to find significant casual

247 links for winter. The corresponding TDCCMs and CCMs are presented in Supplementary

248 Figs. 10 and 11. These results are less robust than that for spring.

249 In the TDCCM in Supplementary Fig. 10a, the winter heat flux $\rightarrow$ CO_2 cross map (red)

250 exhibits both positive and negative optimal lags. For CCM we take the negative one which

251 makes physical sense. The reversed cross-map is essentially constant, signaling

252 dynamical independence of CO_2 with respect to winter heat flux. As before, we take zero

253 lag for this cross map. In Supplementary Fig. 10b CO_2 exhibits maximum predictability for

254 a positive lag, when predicting from the winter SLP attractor (red), which might be

255   indicative of a lack of causal interaction. Winter SLP is dynamically independent from CO_2

256   (blue), ρ being constant. The winter heat flux $\rightarrow$ AMOC North causal direction

257   (Supplementary Fig. 8c) is a physical direction of causality, since the cross-mapping of

258   the former has an attenuated oscillatory character towards positive values, the optimal

259   lag being negative. The opposite causal direction, may be instantaneous. The cross map

260   from winter heat flux to winter SLP shows predominant oscillatory behavior

261   (Supplementary Fig. 8d), with a global maximum for a negative lag. The same ambiguous

262   oscillatory behavior is observed in the opposite sense, with a 0-centered maximum.

263   With the lags, we compute the CCM for the causal chains on winter. Supplementary Fig.

264   11a,b suggests a non-significant causal connection between CO_2 and Winter SLP.

265   Supplementary Fig 11c,d indicate a causal feedback between winter heat flux and winter

266   SLP. Supplementary Fig 11e,f suggests unidirectional forcing from CO_2 to winter heat

267   flux. Supplementary Fig. 11g,h indicates unidirectional influence of winter heat flux on

268   AMOC North. Surrogate analysis is consistent with the TDCCMs. The light (dark) gray

269   shaded areas correspond to surrogate cross-maps generated by a swap (Ebisuzaki)

270   model. All variables have the embedding dimension, E=8 and embedding lag, l=1. In

271   synthesis, the analyzes for the winter season indicate a marginally significant influence

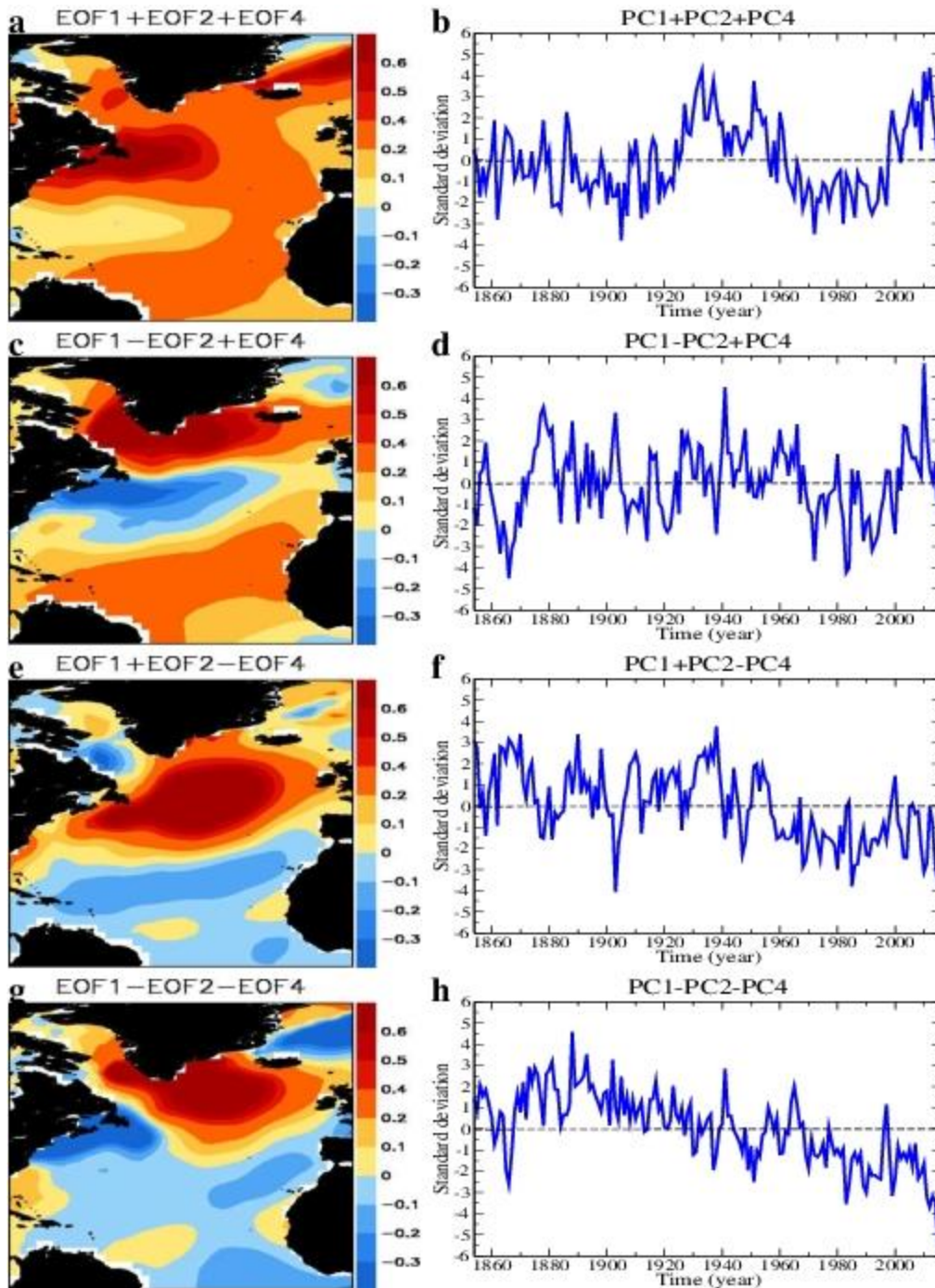272   of CO_2 on AMOC, through heat fluxes.

273

274

275

276

Supplementary Figures

278
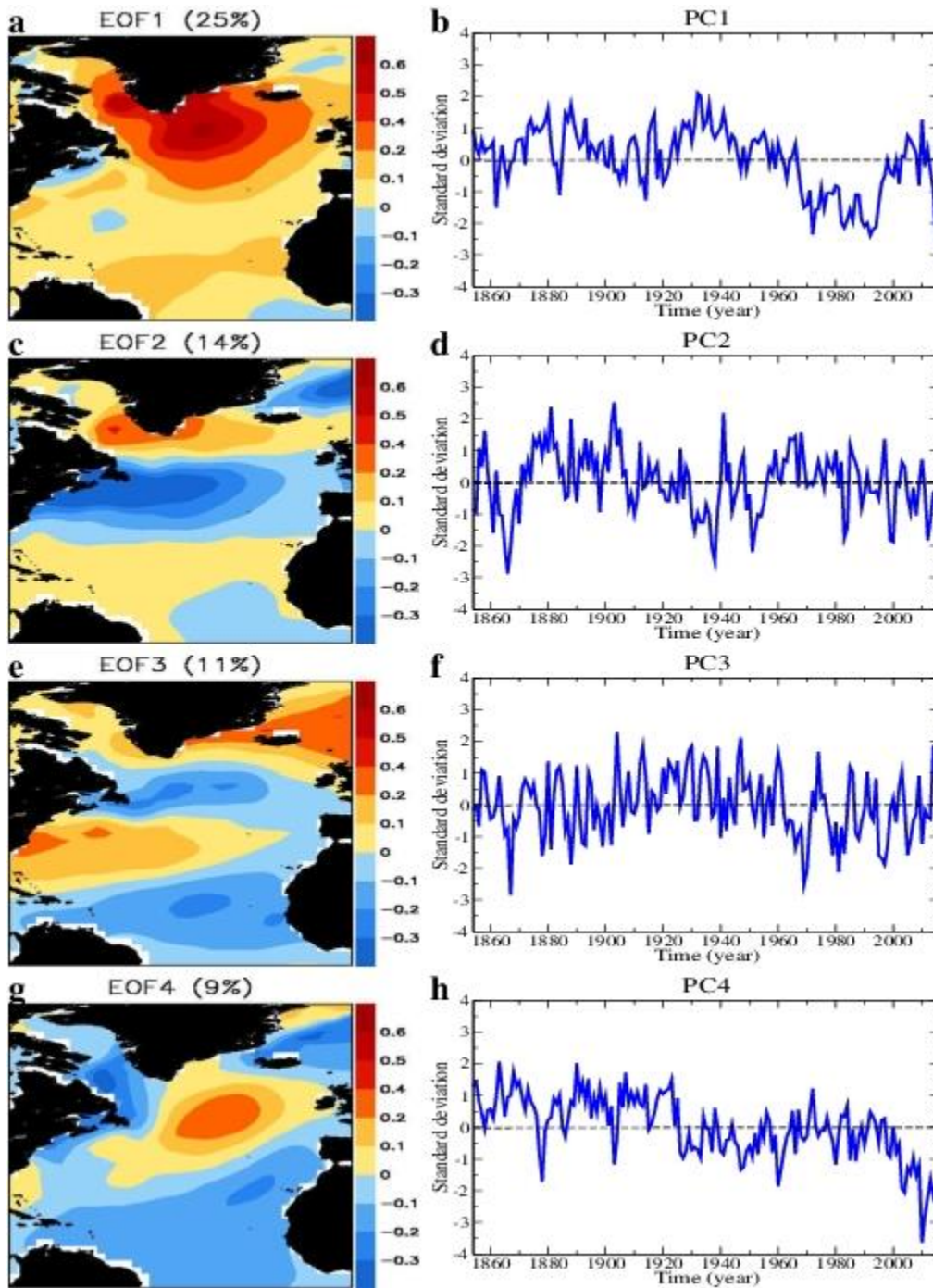
Supplementary Figure 1: EOF analysis of North Atlantic SSTs. First four EOFs (^{o}C) of annual
North Atlantic SST field (a), c), e), g)) and their associated time components (b), d), f), h)). They
explain 25%, 14%, 11% and 9% of variance, respectively. The time components were normalized
with their standard deviations. The EOFs are multiplied with the standard deviations of the
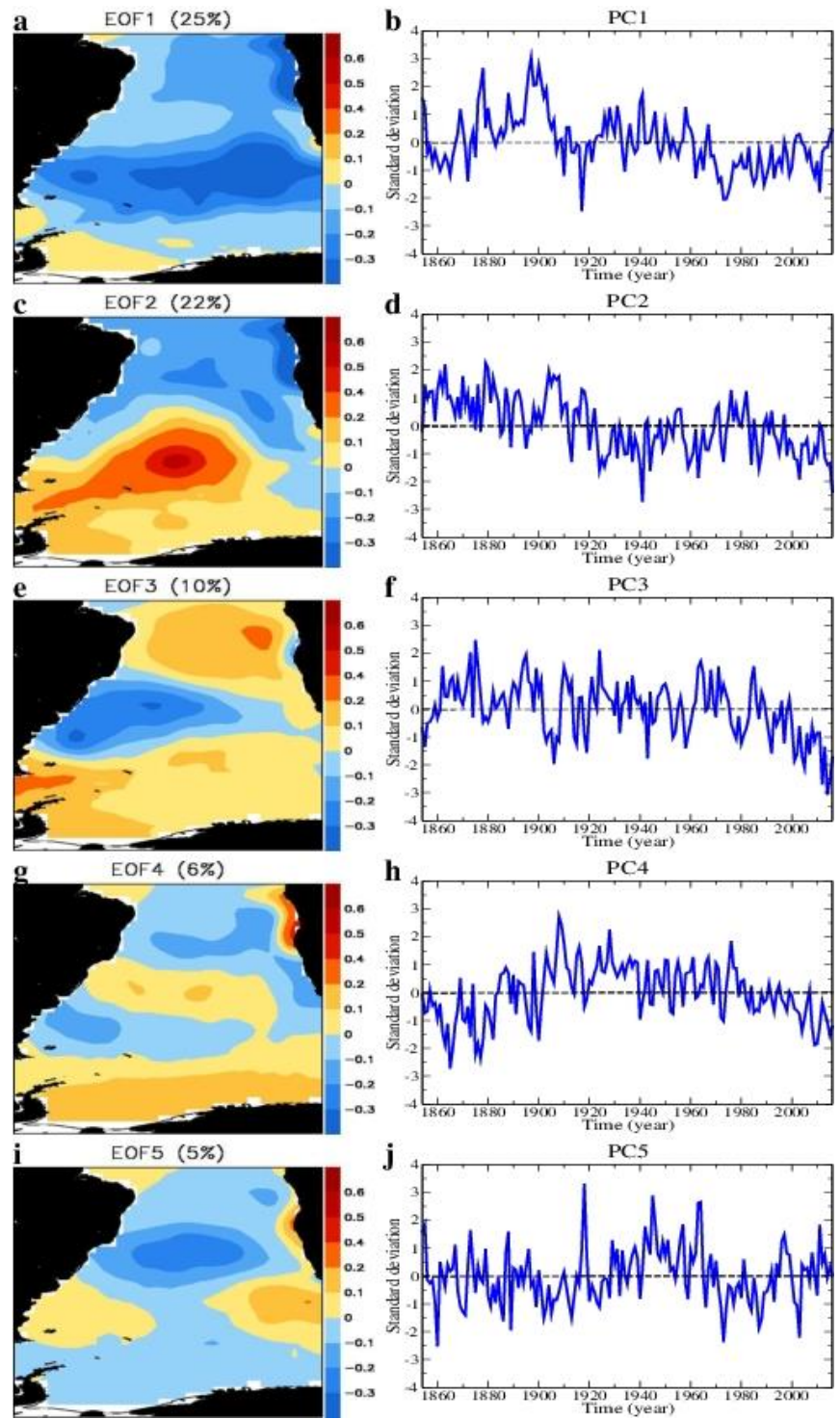corresponding time components.

13

284

Supplementary Figure 2: EOF analysis of North Atlantic SSTs. Linear combinations of the first four EOFs (°C) of annual North Atlantic SST field (a), c), e), g)) and of their associated time components (b), d), f), h)), shown in Supplementary Fig. 1.

288

289

290

Supplementary Figure 3: EOF analysis of South Atlantic SSTs. First five EOFs (ºC) of annual South
Atlantic SST field (a), c), e), g), i)) and their associated time components (b), d), f), h), j)). They explain
25%, 22%, 10%, 6% and 5%, respectively. The time components were normalized with their standard
deviations. The EOFs are multiplied with the standard deviations of the corresponding time
components.

296
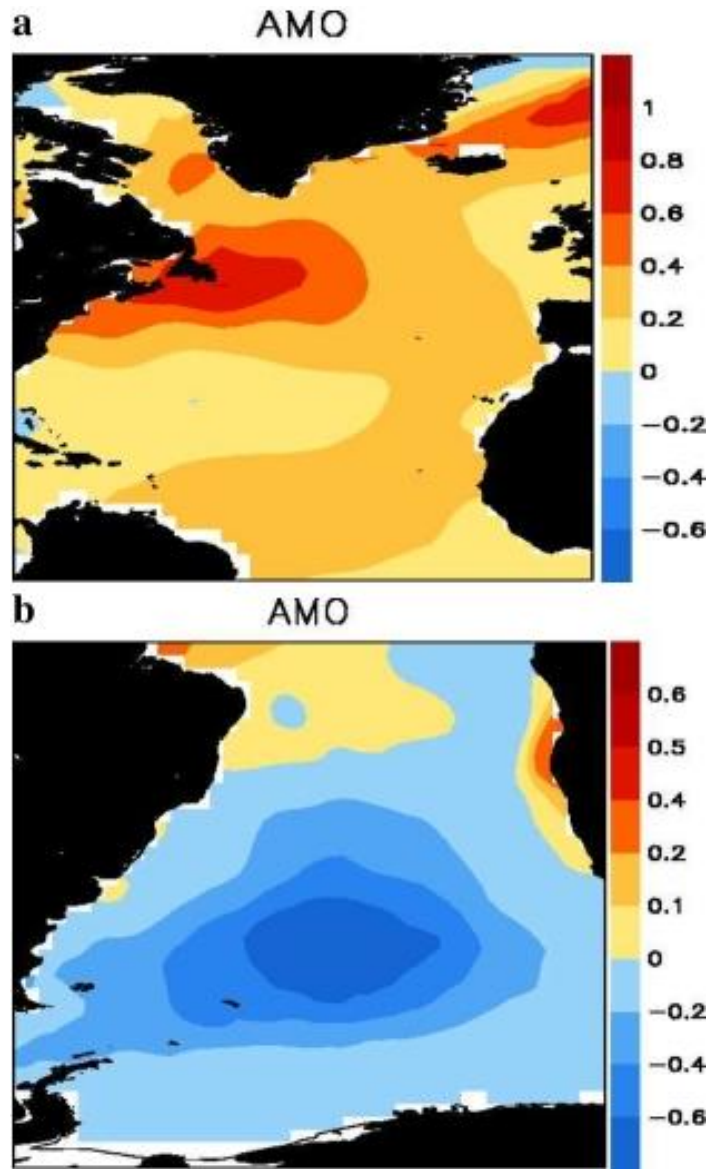
Supplementary Figure 4: EOF analysis of South Atlantic SSTs. Linear combinations of the first four EOFs (°C) of annual South Atlantic SST field (a), c), e)) and of their associated time components (b), d), f)), shown in Supplementary Fig. 3.

300

301

302

Heat flux composite map
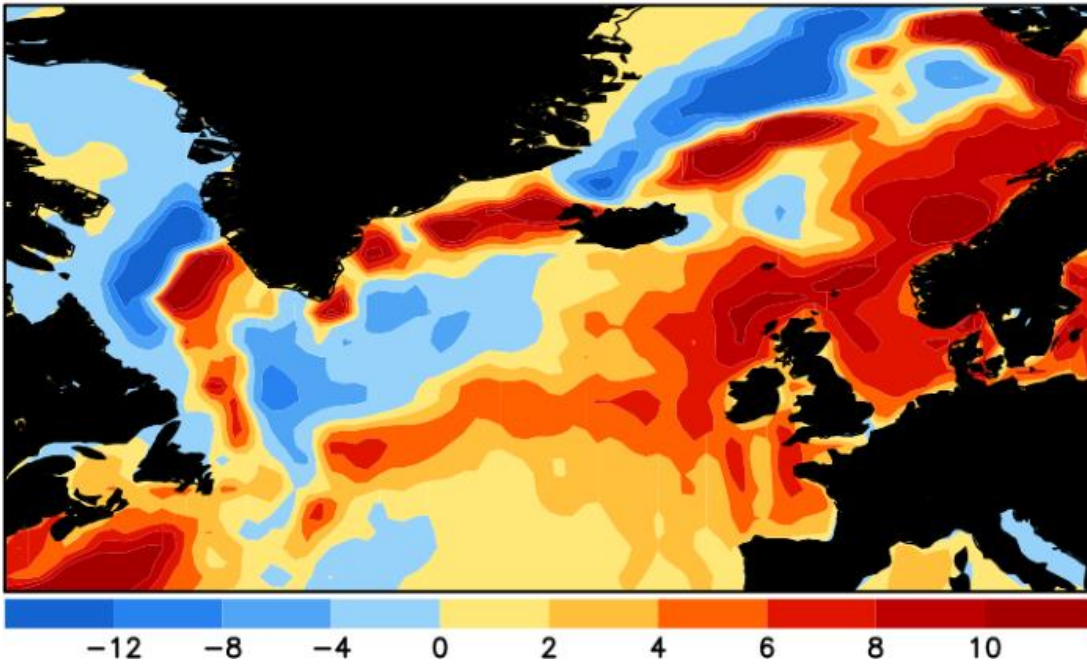
Supplementary Figure 5: Heat flux (W/m^2) composite map, constructed as a difference between the average fields over the 1935-2015 and the mean of the 1854-1934 time interval. Positive values of the fluxes are directed from atmosphere into ocean.
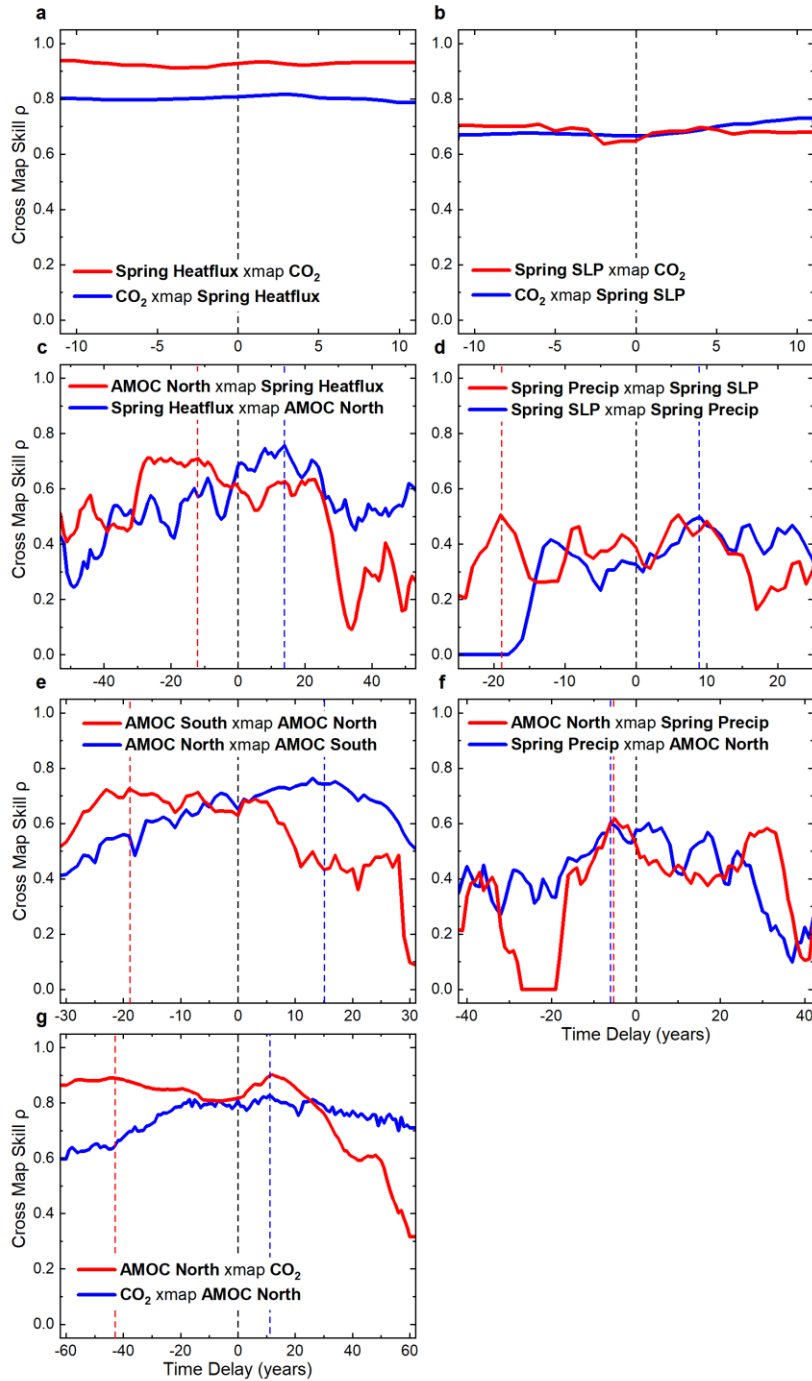
Supplementary Figure 6: Time Delay CCM analyzes for spring. Cross map skill (ρ) given by the correlation between predicted and observed values as a function of time lag in each direction. a) Spring heat flux vs CO_2; b) Spring SLP vs CO_2 are invariant with respect to time lag; c) AMOC North vs Spring heat flux; d) Spring Precipitation vs Spring SLP; e) AMOC South vs AMOC North display unidirectional causality; f) AMOC North vs Spring Precipitation indicates bidirectional interaction; g) AMOC North vs CO_2 suggests unidirectional causality with multiple, distributed time delays. The maxima for cross-mapping directions are marked by red (blue) vertical dashed lines, corresponding to the red (blue) cross-estimates. The marked lags are the ones used for CCM.

326
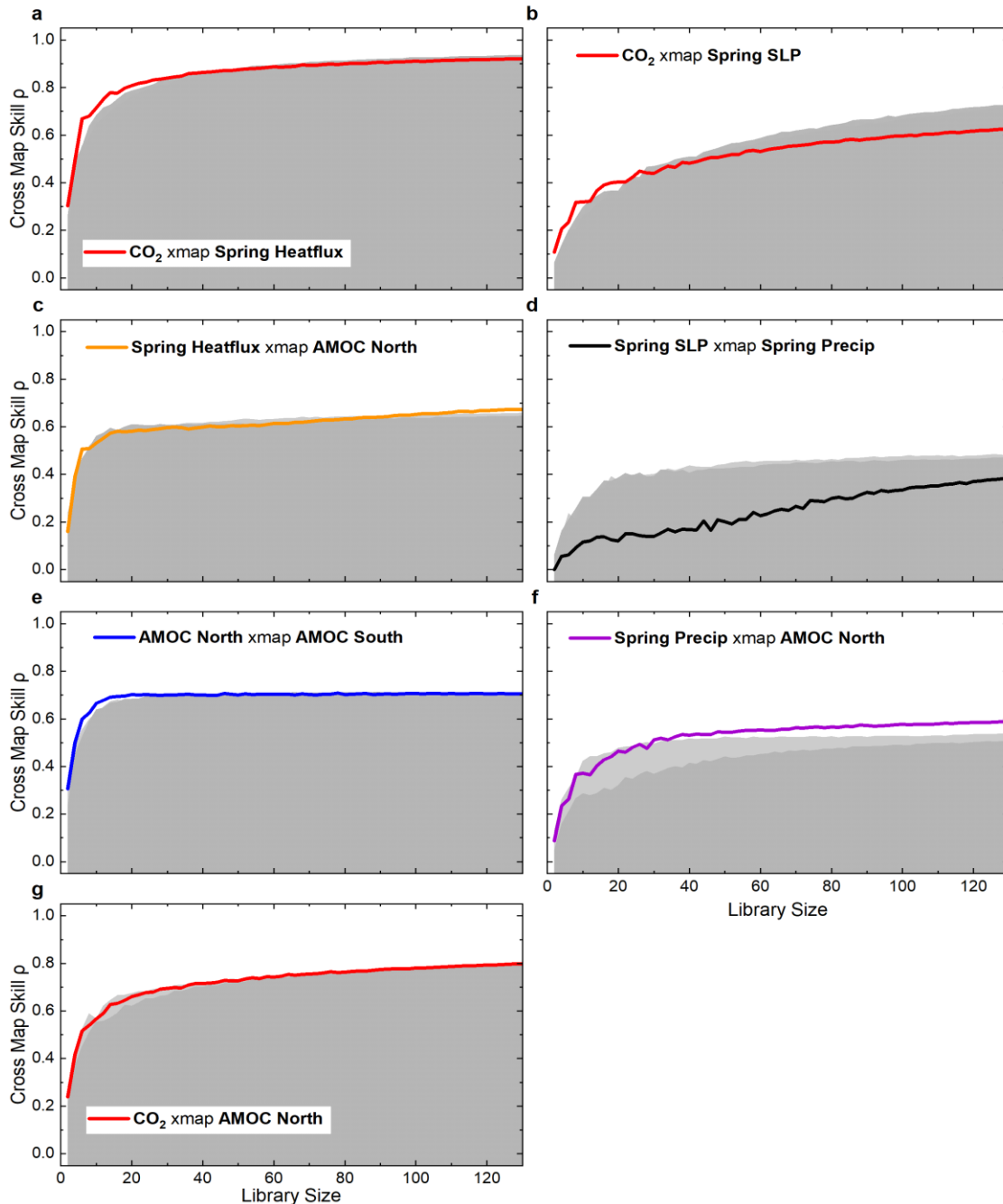
Supplementary Figure 7: Conjugate CCM analyzes, for spring. CCM representations in the inverse direction with respect to Fig 3: a) CO_2 xmap Spring heat flux; b) CO_2 xmap Spring SLP; c) Spring heat flux xmap AMOC North; d) Spring SLP xmap Spring precipitation rate; e) AMOC North xmap AMOC South; f) Spring precipitation rate xmap AMOC North; g) CO_2 xmap AMOC North. The shaded areas represent the 95th percentile of significance under a swap model (light gray) and an Ebisuzaki model (dark gray). All cross maps are implemented with the embedding dimension, E=8 and embedding lag, l=1. Panel f) is the only one exhibiting statistical significance of the cross estimate (p<0.02) which in conjunction with Fig. 3f) reveals a causal feedback between the Ocean and precipitation rates.

336
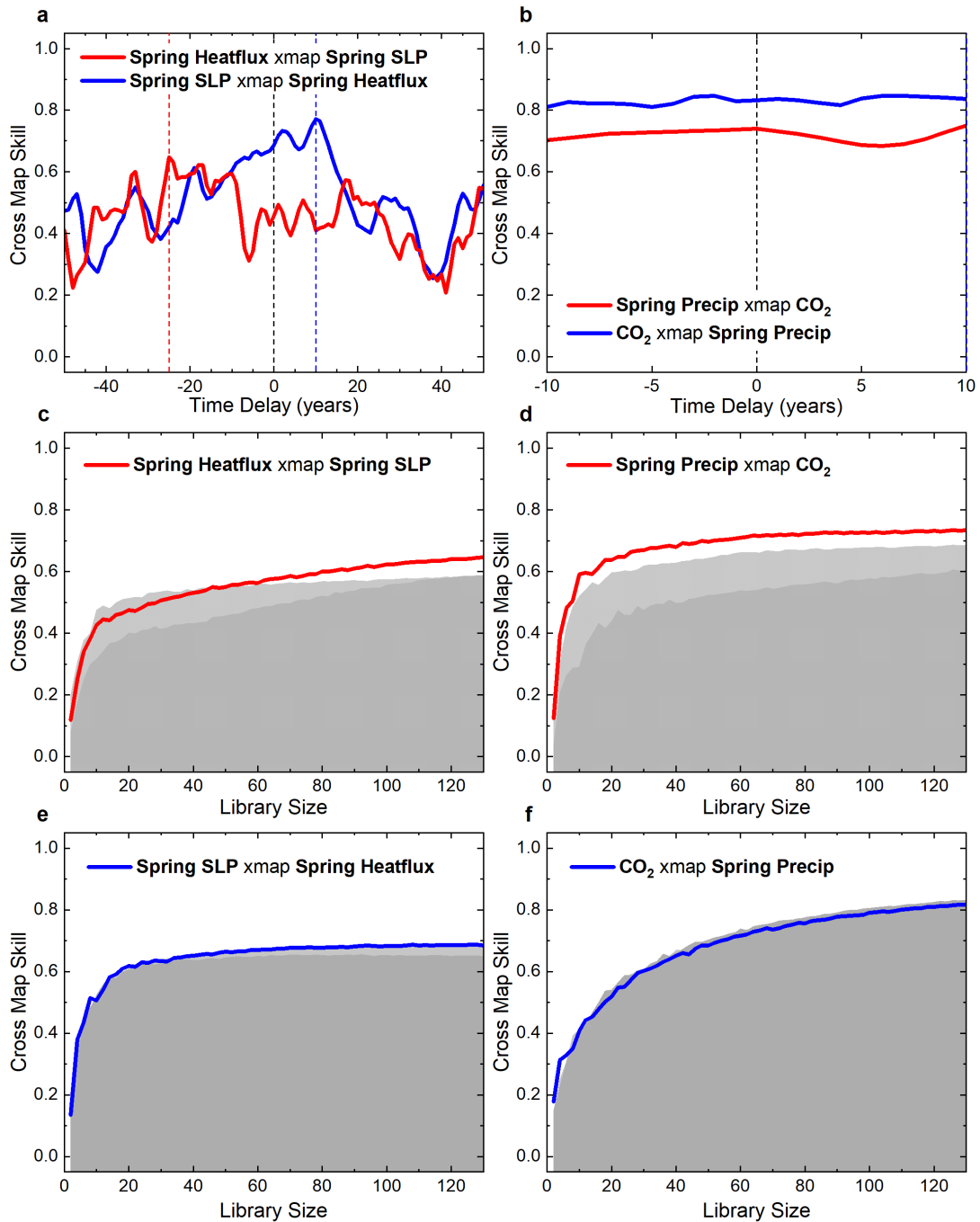
Supplementary Figure 8: CCM and TDCCM results linked with potential CO_2 influence on North AMOC weakening, for spring. a) bidirectional TDCMM for the potential spring heat flux $\rightarrow$ SLP connection; b) bidirectional TDCMM for the potential $CO_2 \rightarrow$ precipitation rate connection; CCM representations for potential causal connections: c) spring SLP $\rightarrow$ spring heat flux; e) spring heat flux $\rightarrow$ spring SLP; d) $CO_2 \rightarrow$ spring precipitation; f) spring precipitation $\rightarrow CO_2$. We take the lag 0. Surrogate analysis in panels d) and f) show unidirectional forcing of CO_2 on spring precipitation rates.
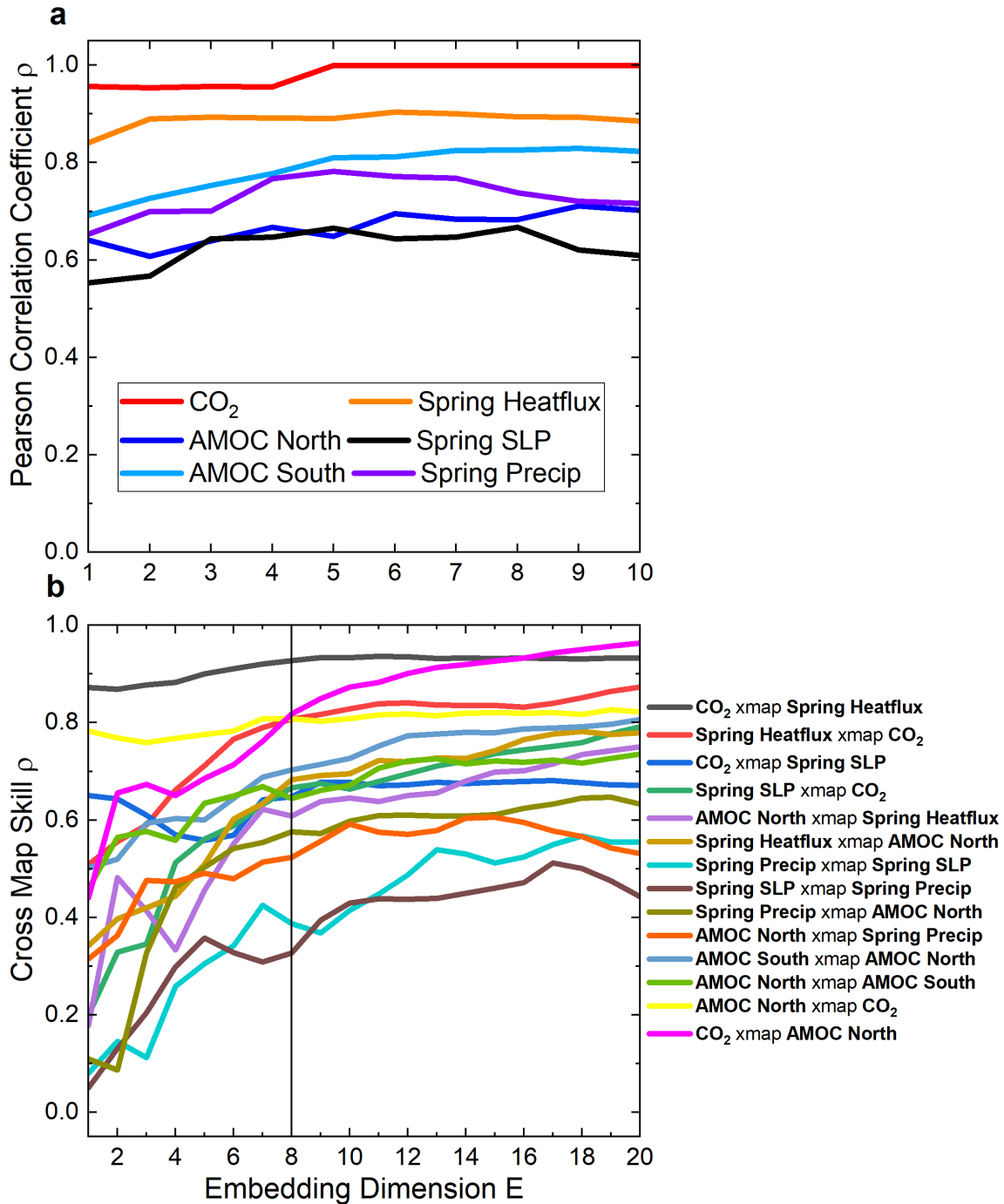
Supplementary Figure 9: Embedding Dimension analysis. a) for each variable Pearson's correlation coefficient (ρ) is ploted as a function of embedding dimension (E) is approximately constant; b) Instantaneous cross-map skill between all pairs of variables used in Fig. 3 for the full library length, for each embedding dimension. A conservative choice which ensures maximum predictability is E=8.

351
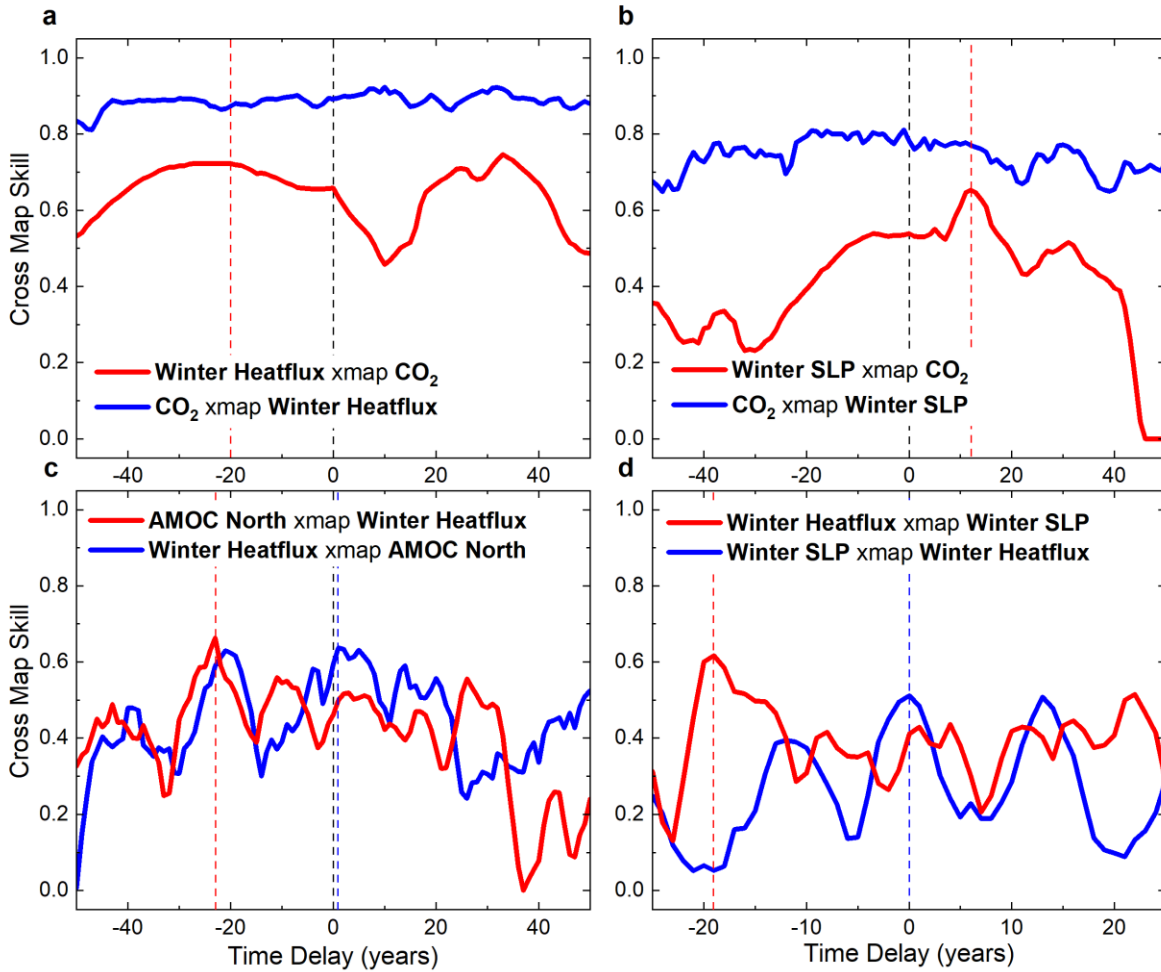
Supplementary Figure 10: TDCCM analyzes for winter. a) Cross map as a function of time delay between CO_2 and winter heat flux with an ambiguous prediction of CO_2 from winter heat flux (red) and a constant CO_2 xmap winter heat flux estimate (blue). b) Positive maximum for winter SLP xmap CO_2 (red) suggesting that CO_2 is not a forcing factor of winter SLP and an almost constant estimation of winter SLP from CO_2 showing that winter SLP is dynamically independent from CO_2. c) Negative optimal lag for the influence of winter heat flux on AMOC and quasi-instantaneous lag for the reversed causal direction. Oscillatory behavior of the cross-map clutters the clarity of the interpretation. d) The same oscillatory behavior for the TDCCM between winter heat flux and SLP, with a negative global maximum for the winter SLP $\rightarrow$ winter heat flux causal link. The opposite sense may be instantaneous.

362

363

364

365

366
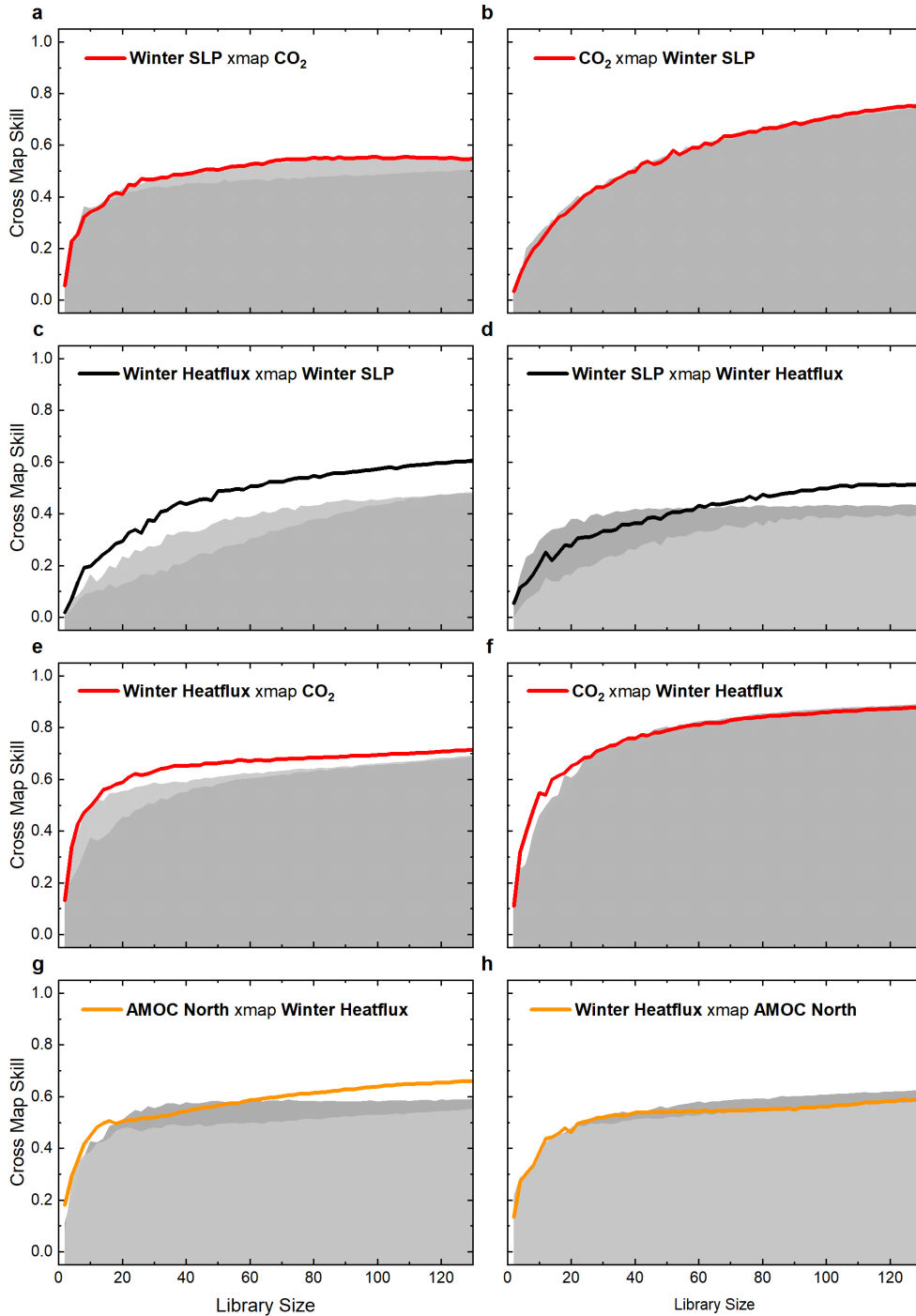
Supplementary Figure 11: CCM analyzes for winter. CCM for causal chains on winter. Panels
a) winter SLP xmap CO_2; b) CO_2 xmap winter SLP; c) winter heat flux xmap winter SLP; d) winter
SLP xmap winter heat flux; e) winter heat flux xmap CO_2; f) CO_2 xmap winter heat flux; g) AMOC
North xmap winter heat flux; h) winter heat flux xmap AMOC North. The shaded areas represent
the 95th percentile of significance under a swap model (light gray) and an Ebisuzaki model (dark
gray). All cross maps are implemented with the embedding dimension, E=8 and embedding lag,
l=1.

Supplementary References

1. Collins, M., et al. Long-term climate change: Projections, commitments and irreversibility, 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 (Cambridge Univ. Press., 2013).

2. Dima, M., Lohmann, G. Evidence for Two Distinct Modes of Large-Scale Ocean Circulation Changes over the Last Century. *J. Clim.* **23**, 5-16 (2010).

3. Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., Saba, V. Observed fingerprint of a weakening Atlantic Ocean overturning circulation, *Nature* **556**, 191-196 (2018).

4. Latif, M., et al. Reconstructing, Monitoring, and Predicting Multidecadal-Scale Changes in the North Atlantic Thermohaline Circulation with Sea Surface Temperature. *J. Clim.* **17**, 1605-1614 (2004).

5. Knight, J. R., Allan, R. J., Folland, C. K., Vellinga, M. & Mann, M. E. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32**, L20708, doi:10.1029/2005GL024233 (2005).

6. Mikolajewicz, U. & Voss, R. The role of the individual air-sea flux components in CO_2-induced changes of the ocean's circulation and climate. *Clim. Dyn.* **16**, 627–642 (2000).

7. Gregory, J. M. et al. A model intercomparasion of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 concentration. *Geophys. Res. Let.* **32(12)**, doi:10.1029/2005GL023209 (2005).

8. Crowley, T. J. North Atlantic deep water cools the Southern Hemisphere. *Paleoc.* **7(4)**, 7489–97 (1992).

9. North, G. R., Bell, T. L., Cahalan, R. F., Moeng, F. J., Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Mon. Weath. Rev.* **110**, 699-706 (1982).

10. Peng, S., Robinson, W. A., Li, S. North Atlantic SST Forcing of the NAO and Relationships with Intrinsic Hemispheric Variability. *Geophys. Res. Lett.* **29(8)**, 1276 (2002).

11. Park, W., & Latif, M. Multidecadal and multicentenial variability of the meridional overturning circulation. *Geophys. Res. Lett.* **35**, doi:10.1029/2008GL035779 (2008).

12. Saba, V. S. et al. Enhanced warming of the Northwest Atlantic Ocean under climate change. *J. Geophys. Res. Oceans* **120**, doi:10.1002/2015JC011346 (2015).

13. Schmith,, T., Yang, S., Gleeson, E., Semmler, T. How Much Have Variations in the Meridional Overturning Circulation Contributed to Sea Surface Temperature Trends since 1850? A Study with the EC-Earth Global Climate Model, *J. Clim.* **27**, 6343-6357 (2014).

14. Hsieh, C. *et al.* Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* **435,** 336–340 (2005).