

1 Deep Learning based reconstructions of the Atlantic Meridional Overturning Circulation  
2 confirm twenty-first century decline  
3 Simon L. L. Michel<sup>1,2\*</sup>, Henk A. Dijkstra<sup>1</sup>, Francesco Guardamagna<sup>1</sup>, Valérian Jacques-  
4 Dumas<sup>1</sup>, René M. van Westen<sup>1</sup>, Anna S. von der Heydt<sup>1</sup>

5 <sup>1</sup>: Institute for Marine and Atmospheric research Utrecht (IMAU), Utrecht University, Utrecht,  
6 Netherlands

7 <sup>2</sup>: Atmospheric Oceanic and Planetary Physics (AOPP), University of Oxford, Oxford, United Kingdom

8 \*simon.michel@physics.ox.ac.uk

## 9 **Abstract**

10 Gaining knowledge of the past and present variations of the Atlantic Meridional  
11 Overturning Circulation (AMOC) is crucial for the development of accurate future climate  
12 projections. The short range covered by direct AMOC observations, inconsistent  
13 paleoclimate records, and scattered hydrographic data are insufficient to realistically  
14 reconstruct the AMOC strength since 1900. An AMOC proxy index based on sea surface  
15 temperatures suggests that the AMOC has declined by 15% since the late 19<sup>th</sup> century but  
16 this index received extensive scientific criticism. Here, we use a deep learning algorithm and  
17 climate model simulations to accurately reconstruct the AMOC strength between 20°N and  
18 60°N since 1900. In contrast with the existing indices, our reconstructions are well in  
19 agreement with AMOC strength variations simulated by climate models and direct  
20 observations at 26.5°N. Our novel set of AMOC reconstructions contribute to a larger  
21 confidence in 21<sup>st</sup> century AMOC decline projections from climate models.

22

23 Modern Earth System Models (ESMs) from the Couple Model Intercomparison Project phase  
24 6 (CMIP6) (1) consistently project a decline in Atlantic Meridional Overturning Circulation  
25 (AMOC) strength throughout the twenty-first century (2) due to anthropogenic climate change  
26 (3). However, there is substantial spread in projected AMOC decline among models (2,4), with

27 a strong meridional dependency (5-7), leading to uncertainties in regional and global  
28 temperature and precipitation trends (4). Given the implications for policymaking, robust  
29 validation of ESMs using historical observations is essential, particularly regarding AMOC  
30 strength (4).

31

32 The AMOC has been monitored at 26.5°N since 2004 through the RAPID array (8-12), providing  
33 a valuable nearly 20-year observational record (10). Although this dataset reveals significant  
34 variations in AMOC strength (11,12), it is limited by measuring the large-scale circulation at  
35 only one latitude. To address this limitation, subsequent research projects such as SAMBA (13)  
36 and OSNAP (14) have begun measuring AMOC at other latitudes (34°S and 60°N, respectively),  
37 albeit with shorter time series of less than ten years. Nevertheless, the RAPID observations,  
38 referred to as AMOC<sub>RAPID</sub>, indicate an overall decline in AMOC strength since 2004 (11,12),  
39 though there has been a modest but consistent increase in the last twelve years of available  
40 data (10-11) (i.e., from 2010 to 2021). However, due to their expense and complexity (8-12),  
41 the RAPID measurements cover too short a period to establish or dismiss any consistent long-  
42 term trend in AMOC intensity in response to anthropogenic climate change (6,8-12,15).  
43 Recent studies, incorporating direct deep ocean observations from various Atlantic regions  
44 (6,14,16), hydrographic data (15), and ESMs (7), have highlighted the latitudinal dependency  
45 of AMOC variability, thereby casting doubt on the robustness of trends observed in RAPID data  
46 (10-12).

47

48 Subsurface density measurements in coastal North Atlantic areas have been utilized to  
49 contextualize direct AMOC observations (17,18). However, due to their limited spatial and  
50 temporal coverage, these data are too noisy to effectively reconstruct monthly to annual

51 AMOC strength using current aggregation methods (17). Given the constraints of using  
52 AMOC<sub>RAPID</sub> and hydrographic observations for ESM validation, AMOC fingerprints based on  
53 North Atlantic SST and SSS fields have been developed to estimate 20th-century AMOC  
54 variations (19-22). Notably, the emergence of the "warming hole" pattern in the North Atlantic  
55 subpolar gyre (SPG) surface has been identified as an indicator of declining AMOC strength  
56 since 1900 (19,23). Using the SST average over the SPG ( $SST_{SPG}$ ) subtracted from the global  
57 mean SST signal as an AMOC proxy ( $AMOC_{SPG}$ ), it has been demonstrated that the AMOC  
58 decreased by 15% overall since the mid-19th century (19). However, salinity also significantly  
59 contributes to surface density fields and is thus critical for AMOC variations (20). In this regard,  
60 Sea Surface Salinity (SSS) indices instead reveal a century-long upward trend in AMOC strength  
61 (Fig. S1), raising doubts about the relevance of sea surface-based AMOC indices.

62

63 Although an abnormally cold (resp. warm)  $SST_{SPG}$  can indicate a decreasing (resp. increasing)  
64 AMOC response to anthropogenic greenhouse gas (resp. aerosol) emissions, the respective  
65 timescales of  $SST_{SPG}$  and AMOC responses to overlapping and opposite radiative effects from  
66 different forcings differ considerably and are non-stationary (24). The  $AMOC_{SPG}$  index also  
67 assumes that the AMOC fingerprint on  $SST_{SPG}$  over the strongly forced recent decades (19)  
68 was the same as during the mainly naturally driven AMOC over the 20th century (25). This  
69 constitutes a strong assumption of stationarity in SPG and AMOC dynamics under constantly  
70 evolving radiative forcing (6,24). North Atlantic SST and  $SST_{SPG}$  are also driven by diverse  
71 factors and their interactions, including the AMOC (20-22), but also stratospheric (26,27) and  
72 tropospheric (26) aerosol concentrations, or atmospheric (28,29) and oceanic (6,16,28,30)  
73 processes. Given the variety of processes affecting North Atlantic SST patterns, the  $AMOC_{SPG}$

74 may be too simplistic as its calculation assumes an unrealistic in-phase and stationary  
75 relationship between  $\text{SST}_{\text{SPG}}$ , the AMOC, and the global warming signal (6) (Fig. S2).

76

77 Consequently, the Special Report on the Ocean and Cryosphere in a Changing Climate (2019)  
78 from the Intergovernmental Panel on Climate Change (31) evaluated the claim that the AMOC  
79 strength has significantly decreased as a response to climate change (19) with only *medium*  
80 *confidence*. With respect to the AMOC validation of ESMs, the  $\text{AMOC}_{\text{SPG}}$  indicates that the  
81 AMOC has declined over the second half of the 20th century (19), while the mean of the  
82 CMIP6 ESMs indicates that it would have increased (32,33). Are the CMIP6 ESMs, with their  
83 known biases, wrong here or is the  $\text{AMOC}_{\text{SPG}}$ , often considered as extended AMOC  
84 observations (19,34,35), inadequate? Clearly, there is an urgency to obtain better  
85 reconstructions of the 20th-century AMOC strength (6,17,25) (Fig. S1), using available  
86 observations.

87

### 88 **A deep learning approach to AMOC reconstruction**

89 Here, we provide a novel reconstruction of the AMOC strength using a deep learning algorithm  
90 trained to produce AMOC timeseries at different latitudes based on SST observations. The use  
91 of deep learning for this task is innovative, as is the production of several AMOC indices for  
92 different latitudes of the North Atlantic Ocean (Fig. 1a). This aligns with recent results  
93 indicating a strong spatial dependence of AMOC variability from weekly to decadal timescales  
94 (6,7). These findings are confirmed for our North Atlantic study area ( $0^{\circ}\text{N}$ - $80^{\circ}\text{N}$ ,  $80^{\circ}\text{W}$ - $0^{\circ}$ , Fig.  
95 1a) by a set of 51 historical climate simulations from 17 ESMs (3 simulations each, Tab. S1).  
96 This extensive model intercomparison analysis indicates that AMOC trends (Fig. 1b) and  
97 covariances (Fig. 1c) in the North Atlantic meridional streamfunction time series vary greatly

98 across latitudes (6,7) and depth (Fig. 1b,c). This analysis further confirms that AMOC variations  
99 cannot be characterized by a single index, suggesting that former attempts to reconstruct the  
100 AMOC strength from surface data can potentially lead to misinterpretations regarding  
101 historical AMOC variability.

102

103 The deep learning method we use here is a 3-layer convolutional neural network (CNN, see  
104 materials and methods), a widely used approach in image analysis (36) that has proven itself  
105 in many climate research applications (37-39). To train CNN models, we selected the best ESM  
106 from the 17 listed in Tab. S1 for representing the two main components of surface density  
107 fields: the SST and SSS fields (Tab. S2). The model-observation comparison uses SST and SSS  
108 observations from the EN4 dataset (40). To comprehensively study ESMs' ability to represent  
109 SST and SSS, we computed mean biases in simulated mean state, standard deviations, minima,  
110 and maxima over the overlapping period of historical CMIP6 simulations and the EN4 dataset  
111 (1900-2014, Fig. S3). These statistics were averaged over three historical simulations for each  
112 of the 17 ESMs (Tab. S2).

113

114 We focus on model biases calculated from annually averaged data (Fig. S3) and also provide  
115 similar analyses for monthly and seasonal averages for both SST (Fig. S4) and SSS (Fig. S5).  
116 Results indicate that ESMs can accurately represent SST only, SSS only, or both (Fig. S3, Tab.  
117 S2). We are interested in the ESM that best reproduces statistics for both SST and SSS, as they  
118 together determine surface density fields which is crucial for simulating the interplay between  
119 the ocean surface and the AMOC (20-22,28,30). Accurately simulating SST and SSS statistical  
120 properties is fundamental for adequately training CNN models to reconstruct the AMOC from

121 ocean surface properties within climate simulations, which we will then apply to observational  
122 data.

123

124 ***Novel AMOC reconstructions across latitudes***

125 The best ESM determined from the bias analysis (Figs. S3-S5, Tab. S2, materials and methods)  
126 is HadGEM3-GC31-MM (hereafter HadGEM3). To train the CNNs, we use 4 historical HadGEM3  
127 simulations, each covering 115 years (1900-2014). Additionally, we use one future simulation  
128 with low CO<sub>2</sub> emissions (SSP1-2.6, 86 years) and one piControl simulation (500 years) with  
129 constant pre-industrial CO<sub>2</sub> levels (280 ppm). Using SSP1-2.6 and piControl simulations has  
130 two main advantages: i) it provides more data for training than the relatively short historical  
131 simulations, with greenhouse gas concentrations similar to late (SSP1-2.6) and early  
132 (piControl) historical levels; ii) it prevents CNNs from overfitting to historical climate  
133 simulations' outputs, generated under similar radiative forcing (*i.e.*, observed anthropogenic  
134 and natural forcings since 1900). In addition to validations for HadGEM3, we applied the same  
135 CNN approach to the best 6 ESMs from Figs. S3-S5 and Tab. S2 to ensure our results are not  
136 model-specific. A detailed description of climate simulations used for CNN validations and  
137 training is given in Tab. S3 (HadGEM3 only) and Tab. S4 (best 6 ESMs).

138

139 We quantify the efficiency of CNNs in reconstructing AMOC variations in HadGEM3  
140 simulations across latitudes from 20°N to 60°N (Fig. 1a). We exclude one historical simulation  
141 from the training set at a time and train CNNs with the remaining simulations (historical,  
142 piControl, and SSP1-2.6). The CNN's ability to reconstruct historical AMOC variations is tested  
143 by reconstructing the AMOC time series of the excluded historical member and comparing it

144 with the original AMOC time series from this simulation (see Tab. S3). Similarly, the ESM  
145 simulations used for validation with the best 6 ESMs are listed in Tab. S4.

146

147 To ensure high performance for each CNN model trained for each excluded historical  
148 simulation and each latitude studied, we used 5-fold cross-validations (see materials and  
149 methods) to tune the CNNs' hyperparameters. We trained and tuned 109 CNNs without  
150 optimizing the number of epochs. Instead, we used a large number of epochs (i.e, 1000) with  
151 an early stopping criterion to ensure convergence of CNNs' training losses. Therefore, we only  
152 optimized the batch sizes and initial learning rates. Each CNN was tested for three batch sizes  
153 and four initial learning rates (see supplementary materials), resulting in 12 hyperparameter  
154 combinations for each excluded historical member/latitude pair. The optimal batch size and  
155 initial learning rate for each case are provided in Tab. S5 for HadGEM3 analyses and in Tab. S6  
156 for analyses based on the best 6 ESMs.

157

158 Using the validation setup described above and employing single-input CNNs based on SST  
159 only, SSS only, and two-input CNNs combining both SST and SSS, we find that CNN models  
160 perform better in reconstructing historical ESM-based AMOC time series using SST only (Fig.  
161 S6). This outcome might stem from significant biases in ESM representations of North Atlantic  
162 SSS (41). Moreover, utilizing SST for CNN training may yield more accurate reconstructions  
163 when applied to real data for practical reasons. Indeed, SST has been directly measured with  
164 better spatio-temporal coverage since 1900 (40) compared to SSS, making it less susceptible  
165 to measurement uncertainties and statistical approximations (40,42).

166

167 To further assess the performance of CNN models in reconstructing ESM-based AMOC  
168 variations across latitudes, we compared our outputs with existing AMOC indices (19,20,34)  
169 (Fig. S1) computed from the same historical ESM simulations. These indices include  $\text{AMOC}_{\text{SPG}}$   
170 based on SST (19) and four SSS-based AMOC indices computed as averages over different  
171 Atlantic areas (20,34), denoted here as  $\text{AMOC}_{\text{SSS}1}$  through  $\text{AMOC}_{\text{SSS}4}$  (see materials and  
172 methods). In Fig. 2, we demonstrate that AMOC time series generated from CNN models  
173 largely outperform other AMOC indices across all latitudes studied (Fig. 2) and for the four  
174 excluded members from training (Fig. 2a-d). While  $\text{AMOC}_{\text{SSS}3}$  and  $\text{AMOC}_{\text{SSS}4}$  exhibit low  
175 performance in estimating AMOC strength at any latitude for all four historical members (Fig.  
176 3a-d), SPG-based SSS ( $\text{AMOC}_{\text{SSS}1}$  and  $\text{AMOC}_{\text{SSS}2}$ ) and SST ( $\text{AMOC}_{\text{SPG}}$ ) indices show relatively  
177 good abilities in estimating variations of southernmost AMOC time series for two members  
178 (members 2 and 4, Fig. 2b,d). However, they generally yield poor estimations of Atlantic  
179 overturning in northernmost areas for all members (Fig. 2).  $\text{AMOC}_{\text{SSS}1}$  and  $\text{AMOC}_{\text{SSS}2}$  always  
180 have poor reconstruction skill for a range of latitudes in all members tested (Fig. 2),  
181 highlighting that basin-scale AMOC variations cannot be adequately characterized with a  
182 single index (Fig. 2). Producing a set of past century AMOC strength variations across several  
183 latitudes thus provides a more coherent and robust view of historical AMOC variability.  
184 Individual reconstructions and their comparison with other indices in the literature and the  
185 AMOC time series from HadGEM3 simulations are provided in Figs. S7-S10.

186

187 We conducted a similar analysis using the six best ESMs instead of relying solely on HadGEM3,  
188 excluding all simulations from each of these ESMs (including piControl and SSP1-2.6) from  
189 training, one after the other (Figure S11). Our conclusions mirror those drawn from the  
190 HadGEM3-only analyses (Figure S11). From Figure 2 and Figure S11, it is evident that ESMs'

191 AMOC time series generated from CNN models are significantly more accurate than previously  
192 suggested AMOC indices (19,20), across all latitudes considered, with improvements  
193 consistently significant at the 95% confidence level (Figure 2, Figure S11). Furthermore, our  
194 analysis indicates that the best CNN-based AMOC reconstructions are achieved when only the  
195 HadGEM3 data are used for training.

196

197 ***Comparison of AMOC reconstructions in historical ESM simulations***

198 To compare the performance of CNN and earlier suggested AMOC indices in historical Earth  
199 System Model (ESM) simulations, we retrained nine new CNN models (one for each latitude  
200 of interest, see Fig. 1a) using all simulations from HadGEM3 (i.e., the best ESM). We applied  
201 the same cross-validation scheme as for Fig. 2 to tune these CNN models, and the optimal sets  
202 of batch sizes and initial learning rates were determined for each latitude (Tab. S7).  
203 Comparisons were made between the CNN-based AMOC reconstructions and the AMOC  
204 results of 51 historical simulations from the 17 ESMs presented in Tab. S1. For reference, we  
205 included SST ( $\text{AMOC}_{\text{SPG}}$ ) and SSS ( $\text{AMOC}_{\text{SSS1}}$ ) indices, with the latter chosen as it performed  
206 best among the studied SSS indices when computed within historical simulations (see Fig. 2,  
207 Fig. S11). Fig. 3 illustrates that, for all AMOC timeseries, the obtained CNN-based ( $\text{AMOC}_{\text{CNN}}$ )  
208 reconstructions fall within the range of historical simulations. To quantify this, we calculated  
209 the number of years between 1900 and 2014 that fall outside of the range described by ESM  
210 simulations (Tab. S1). We performed the same calculation for  $\text{AMOC}_{\text{SPG}}$  and  $\text{AMOC}_{\text{SSS1}}$ , as well  
211 as for each of the 51 simulations (Tab. S1) where the ESM range for one model simulation is  
212 based on the 50 remaining ones (Fig. 3).

213

214 Interestingly, lower latitude AMOC<sub>CNN</sub> reconstructions (20°N to 30°N) closely align with  
215 AMOC<sub>SPG</sub>, with the latter lagging the former by about 5-10 years. Both time series depict a  
216 downward trend from the 1950s (19), but CNN-based AMOC reconstructions suggest a slight  
217 recovery of the AMOC in recent years, consistent with direct RAPID observations (10,11). This  
218 partial recovery is thus expected to continue over the coming years for AMOC<sub>SPG</sub> (Fig. 3).

219  
220 For middle latitudes (35°N and 40°N), the century-long downward trend appears to be less  
221 pronounced for 40°N or non-discriminable for 35°N. Conversely, at higher latitudes (from 45°N),  
222 a downward trend marked by recent record lows is observed, with no visible recovery yet in  
223 these regions (Fig. 3). Consistent with ESMs, the largest decrease at the centennial scale is  
224 generally observed for the highest latitudes (Fig. 1b, Fig. 3). The increase in subpolar salinity  
225 concentration (captured by AMOC<sub>SSS1</sub>) from 1950 to 1970, associated with enhanced ocean  
226 convection, is recorded in all our reconstructions, albeit with a slight delay at the highest  
227 latitudes (Fig. 3). The second salinity increase through the late 20<sup>th</sup> century is also associated  
228 with a local increase in both AMOC<sub>CNN</sub> and AMOC<sub>SPG</sub> time series.

229  
230 The AMOC<sub>SPG</sub> and AMOC<sub>SSS1</sub> indices often exhibit values outside the range of ESM simulations  
231 for most latitudes, likely due to their delayed response and inability to capture AMOC  
232 variability across the entire North Atlantic (Fig. 2, Fig. S11). In contrast, the AMOC<sub>CNN</sub>  
233 reconstructions consistently align better with the range of ESM results (Fig. 3). While  
234 reconstructions at the lowest latitudes may occasionally fall outside the ESM range, they still  
235 demonstrate improved agreement compared to previous indices (Fig. 3). This reconciliation  
236 addresses earlier discrepancies between existing estimations of historical AMOC strength and  
237 AMOC simulations from ESMs, marking a significant advancement.

238

239 ***Comparison with direct AMOC observations***

240 We do not proceed to comparisons of our reconstructions with too short records (less than  
241 ten years) from the SAMBA (13) and OSNAP (14) arrays. In addition, SAMBA measurements  
242 (at 34°S) fall outside our trained CNN domain (0°N-80°N, 80°W-0°, Fig. 1a). Thus, our analysis  
243 focuses on the RAPID time series, compared (Fig. 4a,b) with  $\text{AMOC}_{\text{CNN}}$  reconstructions at  
244 adjacent latitudes (i.e., 25°N and 30°N). Remarkably, our reconstructions closely match RAPID  
245 measurements, capturing the 2009-2010 drop and subsequent partial recovery at the  
246 interannual scale (Fig. 4c,d). Additionally, our reconstructions indicate a continuous partial  
247 recovery of AMOC strength since 2010, consistent with RAPID observations (10-12) (Fig. 4c,d).

248

249 The correlation coefficients between AMOC timeseries at 25°N and 30°N and SST fields, using  
250 historical observations from HadISST (43) and HadGEM3 simulation data, are depicted in Fig  
251 4c-f. Interestingly, the (linear) relationship between our reconstructed time series and model  
252 time series exhibits strong similarities (Fig. 4c-f). Consistent with earlier studies (19,23,25),  
253 the correlation between low latitude (25°N and 30°N) AMOC time series and North Atlantic  
254 SST under climate change is generally negative at the interannual timescale, except for a  
255 portion of the subpolar North Atlantic where correlations are positive (25). However, in  
256 agreement with simulations from HadGEM3, the region where the AMOC correlates positively  
257 with SST is shifted southward compared to the area proposed for the calculation of  $\text{AMOC}_{\text{SPG}}$   
258 (19). Specifically, the areas within the Labrador and Irminger seas, which are included in  
259  $\text{AMOC}_{\text{SPG}}$  calculations, are found to be anticorrelated with our AMOC reconstructions at 25°N  
260 and 30°N and in the HadGEM3 simulations (Fig. 4c-f).

261

262 ***Conclusions and discussions***

263 Recent findings have raised concerns that the AMOC may be nearing a tipping point, primarily  
264 based on surface AMOC indices (34,35,44). However, some of these indices have been found  
265 to produce false alarms in AMOC collapse simulations (45). Nevertheless, a more robust early  
266 warning indicator based on South Atlantic salinity transport reinforces the notion that the  
267 AMOC may be approaching a tipping point (45). The reconstructions provided in the present  
268 study offer a more detailed framework to assess the likelihood of such an event.

269

270 While climate change has already impacted many observables worldwide (3), the  
271 development of historical AMOC time series has long been very controversial (6).  
272 Disagreements between AMOC reconstructions based on surface indices, ESM simulations,  
273 and the direct, but very short, observations from the deep ocean (6-15), indicate that ESMs  
274 were not able to correctly simulate the AMOC, hence casting legitimate doubts on their  
275 projections of a future AMOC decline over the twenty-first century. The present study employs  
276 advanced deep learning techniques to produce new AMOC indices that align with historical  
277 ESM simulations across latitudes from 20°N to 60°N. These results reconcile the longstanding  
278 mismatch between AMOC fingerprint proxies (19,20,34) and ESM simulations (15), reinforcing  
279 confidence in future projections of 21<sup>st</sup> century AMOC decline (2).

280

281

282 References:

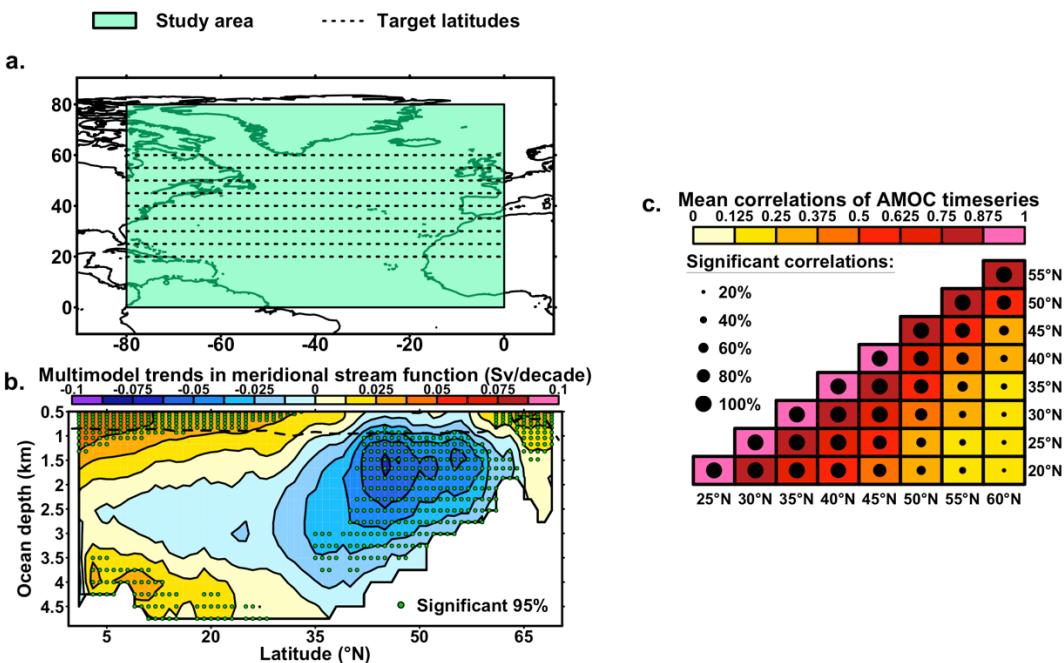
- 283 1. Eyring, V., Bony, S., Meehl, G. A. et al. Overview of the Coupled Model Intercomparison  
284 Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Mod. Dev.* **9**,  
285 1937-1958 (2016).
- 286 2. Weijer, W., Cheng, W., Garuba, O. A., et al. CMIP6 models predict significant 21st  
287 century decline of the Atlantic Meridional Overturning Circulation. *Geophys. Res. Lett.*  
288 **47(12)**, GL086075 (2020).
- 289 3. Eyring, V., Gillett, N. P., Achuta, R. et al. 2021: Human influence on the climate system.  
290 In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I  
291 to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.  
292 Eds: Masson-Delmotte, V., Zhai, P., Pirani, A. et al. *Cambridge University Press*. 432-  
293 552 (2021).
- 294 4. Bellomo, K., Angeloni, M., Corti, S. et al. Future climate change shaped by inter-model  
295 differences in Atlantic Meridional Overturning Circulation response. *Nat. Commun.* **12**,  
296 3659 (2021).
- 297 5. Årthun, M., Asbjørnsen, H., Chafik, L., et al. Future strengthening of the Nordic Seas  
298 overturning circulation. *Nat. Commun.* **14**, 2065 (2023).
- 299 6. Frajka-Williams, E., Ansorge, I. J., Baehr, J. et al. Atlantic Meridional Overturning  
300 Circulation: observed transport and variability. *Front. Mar. Sci.* **6** (2019).
- 301 7. Zou, S., Lozier, S. & Xu, X. Latitudinal structure of the Meridional Overturning  
302 Circulation variability on interannual to decadal time scales in the North Atlantic  
303 Ocean. *J. Clim.* **33(9)**, 3845-3862 (2020).
- 304 8. Cunningham, S. A., Kanzow, T., Rayner, D. et al. Temporal variability of the Atlantic  
305 Meridional Overturning Circulation at 26.5°N. *Science*, **317(5840)**, 935-938 (2007).

- 306 9. McCarthy, G. D., Smeed, D. A., Johns, W. E. et al. Measuring the Atlantic Meridional  
307 Overturning Circulation at 26°N. *Progr. Oceanogr.* **130**, 91-111 (2015).
- 308 10. Moat, B. I., Smeed, D. A., Rayner, D. et al. (2023). Atlantic meridional overturning  
309 circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning  
310 Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004  
311 to 2022 (v2022.1), *British Oceanographic Data Centre - Natural Environment Research*  
312 *Council, UK.* doi: 10.5285/04c79ece-3186-349a-e063-6c86abc0158c (2023).
- 313 11. Smeed, D. A., Josey, S. A., Beaulieu, W. E. et al. The North Atlantic Ocean Is in a state of  
314 reduced overturning. *Geophys. Res. Lett.* **45(3)**, 1527-1533 (2018).
- 315 12. Smeed, D. A., McCarthy, G. D., Cunningham, S. A. et al. Observed decline of the Atlantic  
316 meridional overturning circulation 2004–2012. *Oc. Sci.* **10**, 29-38 (2014).
- 317 13. Meinen, C. S., Speich, S., Perez, R. C., et al. Temporal variability of the meridional  
318 overturning circulation at 34.5°S: Results from two pilot boundary arrays in the South  
319 Atlantic. *J. Geophys. Res. Oc.* **118**, 6461-6479 (2013).
- 320 14. Lozier, M. S., Li, F., Bacon, S. et al. A sea change in our view of overturning in the  
321 subpolar North Atlantic. *Science* **363(6426)** 516-521 (2019).
- 322 15. Jackson, L. C., Biastoch, A., Buckley, M. W. et al. The evolution of the North Atlantic  
323 Meridional Overturning Circulation since 1980. *Nat. Rev. Earth & Environ.* **3**, 241-254  
324 (2022).
- 325 16. Chafik, L. Holliday, N. P., Bacon, S. et al. Irminger Sea is the center of action for subpolar  
326 AMOC variability. *Geophys. Res. Lett.* **49(22)**, GL099133 (2022).
- 327 17. R Rossby, R., Palter, J. & Donohue, K. What can hydrography between the New England  
328 slope, Bermuda and Africa tell us about the strength of the AMOC over the last 90  
329 years? *Geophys. Res. Lett.* **49(23)**, GL099173 (2022).

- 330        18. Worthington, E. L., Moat, B. I., Smeed, D. A. et al. A 30-year reconstruction of the  
331              Atlantic meridional overturning circulation shows no decline. *Oc. Sci.* **17(1)**, 285-299  
332              (2021).
- 333        19. Caesar, L., Rahmstorf, S., Robinson, A. et al. Observed fingerprint of a weakening  
334              Atlantic Ocean overturning circulation. *Nature* **556**, 191–196 (2018).
- 335        20. Jackson, L. C. & Wood, R. A. Fingerprints for early detection of changes in the AMOC.  
336              *J. Clim.* **33(16)**, 7027-7044 (2020).
- 337        21. Knight, J. R., Allan, R. J., Folland, C. K. et al. A signature of persistent natural  
338              thermohaline circulation cycles in observed climate. *Geophys. Res. Lett.* **32(20)**, L20708  
339              (2005).
- 340        22. Zhang, R., Sutton, R., Danabasoglu, G. et al. A review of the Role of the Atlantic  
341              Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated  
342              Climate Impacts. *Rev. Geophys.* **57(2)**, 316-375 (2019).
- 343        23. Rahmstorf, S., Box, J., Feulner, G. et al. Exceptional twentieth-century slowdown in  
344              Atlantic Ocean overturning circulation. *Nature. Clim. Change.* **5**, 475–480 (2015).
- 345        24. Little, C. M., Zhao, M. & Buckley, M. W. Do surface temperature indices reflect  
346              centennial-timescale trends in Atlantic Meridional Overturning Circulation strength?  
347              *Geophys. Res. Lett.* **47(22)**, GL090888 (2020).
- 348        25. Latif, M., Sun, J., Visbeck, M. et al. Natural variability has dominated Atlantic  
349              Meridional Overturning Circulation since 1900. *Nat. Clim. Chang.* **12**, 455-460 (2022).
- 350        26. Booth, B. B., Dunstone, N. J., Halloran, P. R. et al. Aerosols implicated as a prime driver  
351              of twentieth-century North Atlantic climate variability. *Nature* **484**, 228-232 (2012).

- 352        27. Mann, M. E., Steinman, B. A., Brouillette, D. J. et al. Multidecadal climate oscillations  
353                  during the past millennium driven by volcanic eruptions. *Science* **371**, 1014-1019  
354                  (2021).
- 355        28. Buckley, M. W. & Marshall, J. Observations, inferences, and mechanisms of the Atlantic  
356                  Meridional Overturning Circulation: A review. *Rev. Geophys.* **54(1)**, 5-63 (2016).
- 357        29. Fan, Y., Liu, W., Zhang, P. et al. North Atlantic Oscillation contributes to the subpolar  
358                  North Atlantic cooling in the past century. *Clim. Dyn.* (2023).
- 359        30. Swingedouw, D., Houssais, M.-N., Herbault, C. AMOC Recent and future trends: a  
360                  crucial role for oceanic resolution and Greenland melting? *Fron. Clim.* **4**, 838310  
361                  (2022).
- 362        31. Collins, M., Sutherland, M., Bouwer, L. et al. Extremes, abrupt changes and managing  
363                  risk. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate. Eds:  
364                  Pörtner, H.-O., Roberts, D. C., Masson-Delmotte, V. et al. *Cambridge University Press*.  
365                  585-655 (2019).
- 366        32. Menary, M. B., Robson, J., Allan, R. P. et al. Aerosol-forced AMOC changes in CMIP6  
367                  historical simulations. *Geophys. Res. Lett.* **47(14)**, GL088166 (2020).
- 368        33. Robson, J., Menary, M. B., Sutton, R. T. et al. The role of anthropogenic aerosol forcing  
369                  in the 1850–1985 strengthening of the AMOC in CMIP6 historical simulations *J. Clim.*  
370                  **35(20)**, 6843-6863 (2022).
- 371        34. Boers, N. Observation-based early-warning signals for a collapse of the Atlantic  
372                  Meridional Overturning Circulation. *Nat. Clim. Chang.* **11**, 680–688 (2021).
- 373        35. Ditlevsen, P. & Ditlevsen, S. Warning of a forthcoming collapse of the Atlantic  
374                  meridional overturning circulation. *Nat. Commun.* **14**, 4254 (2023).

- 375        36. Goodfellow, I., Bengio, Y. & Courville, A. Convolutional Networks. In: Deep learning.  
376                  *MIT Press* (2016).
- 377        37. Bone, C., Gastineau, G., Thiria, S., et al. Accepted in *JAMES*.
- 378        38. Ham, YG., Kim, JH. & Luo, JJ. Deep learning for multi-year ENSO forecasts. *Nature* **573**,  
379                  568–572 (2019).
- 380        39. Jacques-Dumas, V., Ragone, F., Borgnat, P. et al. Deep learning-based extreme  
381                  heatwave forecast. *Front. Clim.* **4** (2022).
- 382        40. Good, S. A., Martin, M. J. & Rayner, N. A. EN4: quality controlled ocean temperature  
383                  and salinity profiles and monthly objective analyses with uncertainty estimates, *Journ.*  
384                  *Geophys. Res. Oc.* **118**, 6704-6716 (2013).
- 385        41. Liu, W., Xie, S.-P., Liu, Z., et al. Overlooked possibility of a collapsed Atlantic Meridional  
386                  Overturning Circulation in warming climate. *Sci. Adv.* **3(1)**, e1601666 (2017).
- 387        42. Ben-Yami, M., Skiba, V., Bathiany, S. et al. Uncertainties in critical slowing down  
388                  indicators of observation-based fingerprints of the Atlantic Overturning Circulation.  
389                  *Nat. Commun.* **14**, 8344 (2023).
- 390        43. Rayner, N. A., Parker, D. E., Horton, E. B. et al. Global analyses of sea surface  
391                  temperature, sea ice, and night marine air temperature since the late nineteenth  
392                  century. *Journ. Geophys. Res. Atm.* **108(D14)**, 4407 (2003).
- 393        44. Michel, S.L.L., Swingedouw, D., Ortega, P. et al. Early warning signal for a tipping point  
394                  suggested by a millennial Atlantic Multidecadal Variability reconstruction. *Nat.*  
395                  *Commun.* **13**, 5176 (2022).
- 396        45. Van Westen, R. M., Kliphuis, M. & Dijkstra, H. A. Physics-based early warning signal  
397                  shows that AMOC is on tipping course. *Sci. Adv.* **10(6)**, eadk1189 (2024).
- 398



400

401 **Figure 1: Area of study and model diversity of Atlantic Meridional Overturning**402 **Circulation (AMOC) time series across latitudes in Earth System Models (ESMs)**403 **from CMIP6 (1).** **a.** Study area (light green), where Sea Surface Temperature and (SST)

404 and Sea Surface Salinity (SSS) fields are used to train and validate Convolutional

405 Neural Networks (CNNs) to reconstruct AMOC time series for the 9 latitudes

406 indicated by dashed lines (i.e., from 20 $^{\circ}$ N to 60 $^{\circ}$ N with an increment of 5 $^{\circ}$ ). **b.** Mean

407 trends in North Atlantic meridional overturning streamfunction time series below 500

408 meters of depth from an ensemble of 51 historical simulations from 17 CMIP6 ESMs

409 (3 simulations each, Tab. S1) expressed in Sverdrup (Sv) per decade, for the period

410 1900-2014. Green dots indicate where the ensemble mean from ESM simulations is

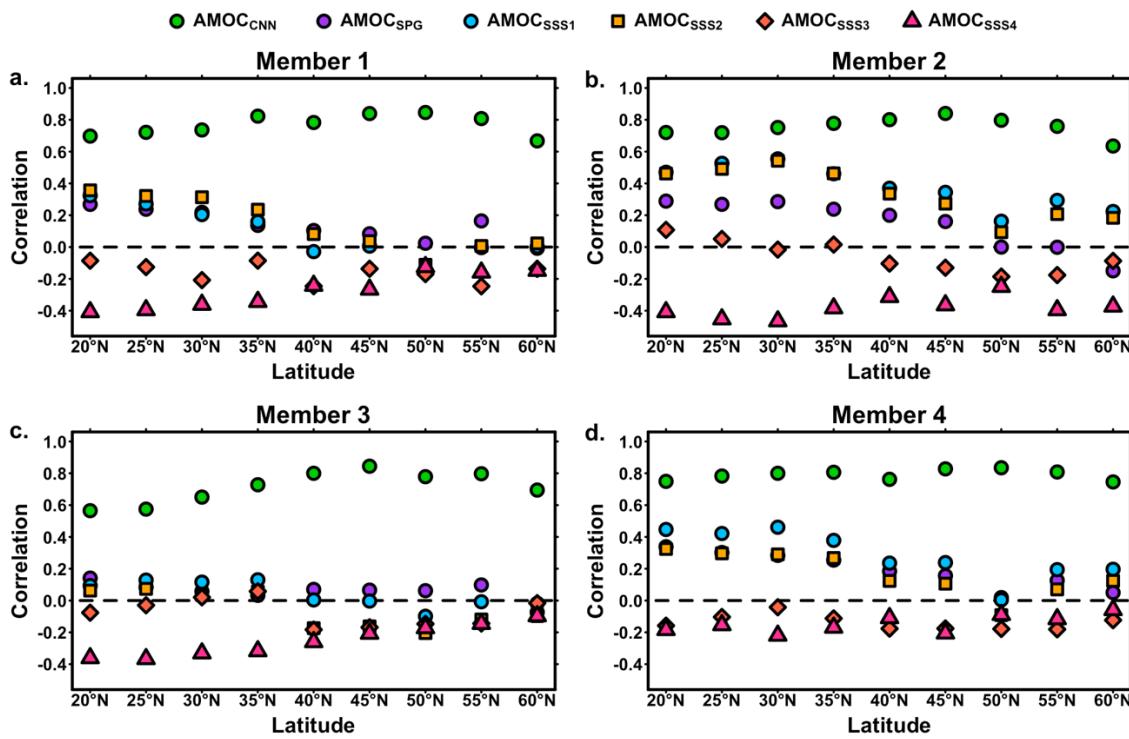
411 significantly different from 0 at the 95% confidence level. The black dashed lines

412 indicate where the maximum overturning is reached on average for the 51 ESM

413 simulations. **c.** Matrix of mean correlations (colors) of AMOC time series (each

414 calculated within the same simulation) for the 9 latitudes studied here and indicated

415       in **a**, for the same 51 historical simulations from **b**. The fraction of significant  
416       correlations is indicated by the size of dots for each couple of latitudes among the 51  
417       simulations.



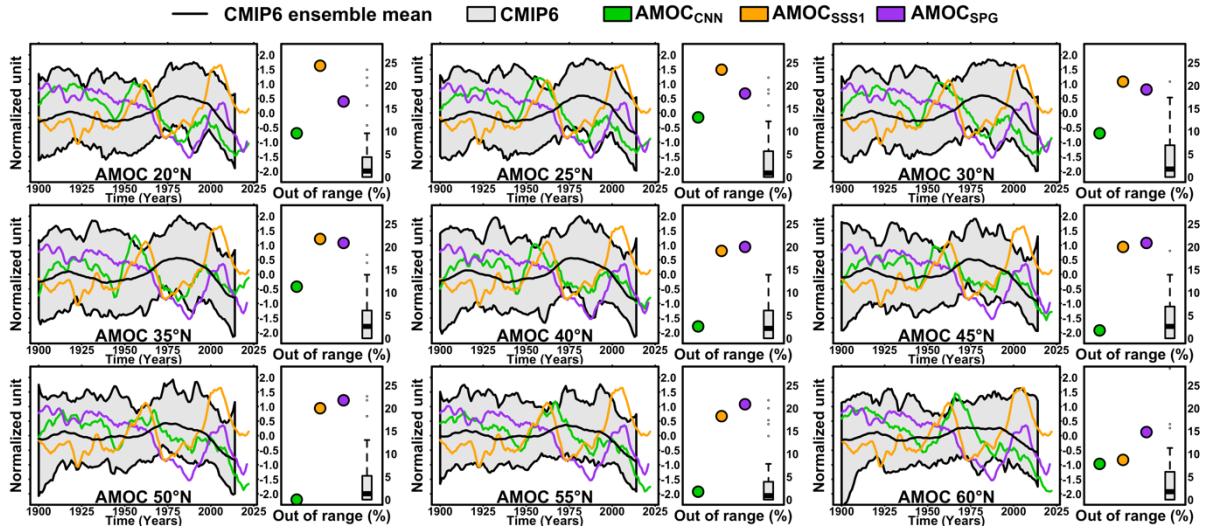
418

419 **Figure 2: Performance of Convolutional Neural Network (CNN) to reconstruct**  
 420 **Atlantic Meridional Overturning Circulation (AMOC) time series in HadGEM3**  
 421 **using Sea Surface Temperature (SST) fields, and comparison with other AMOC**  
 422 **indices.** Reconstructions were performed for 9 different latitudes (from 20°N to 60°N  
 423 with an increment of 5°) with HadGEM3, the best ESM in terms of simulating observed  
 424 SST and SSS statistics (Figs. S3-S5, Tab. S2). The training set used here is composed  
 425 of 4 historical runs, and one SSP1-2.6 and piControl runs (Tab. S3). For panels **a.** to  
 426 **d.**, green circles give the correlation of reconstructed AMOC time series for historical  
 427 simulation members 1 to 4 and actual AMOC time series from this same simulation,  
 428 when it was excluded from the training (materials and methods, Tab. S3), one after the  
 429 other, for each latitude. Blue circles, orange squares, red diamonds, pink triangles,  
 430 and purple circles show the correlations between the AMOC time series from the  
 431 given ESM simulation member with other existing surface-based indices of the AMOC

432 proposed by other studies, i.e., AMOC<sub>SSS1</sub> through AMOC<sub>SSS4</sub> (20,34) and AMOC<sub>SPG</sub> (19)

433 (materials and methods), respectively.

434

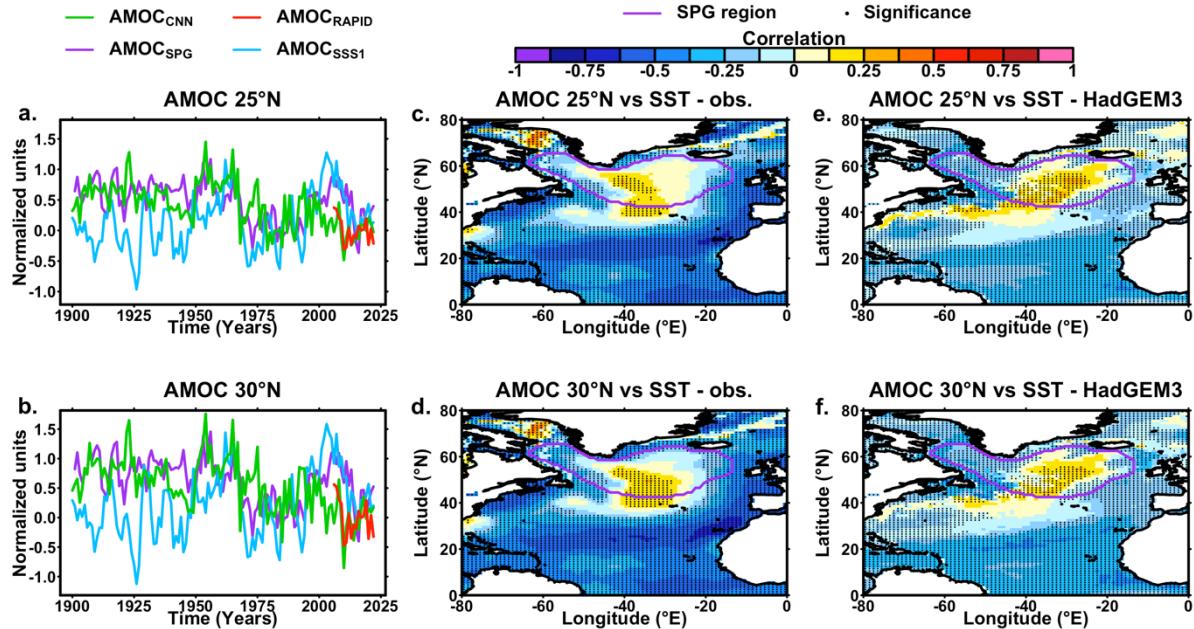


435

436 **Figure 3: Atlantic Meridional Overturning Circulation reconstructions with**  
 437 **Convolutional Neural Networks (CNNs) and comparison with other existing surface-**  
 438 **based indices and Earth System Model (ESM) simulations.** Reconstructed AMOC time  
 439 series are provided for 9 different latitudes (from 20°N to 60°N with an increment of 5°),  
 440 each obtained with a CNN trained and tuned (materials and methods, Tab. S7) using  
 441 simulations from HadGEM3 (i.e., the less biased ESM from Tab. S2) and applied to  
 442 observed Sea Surface Temperature (SST) from the HadISST dataset (43) For each latitude,  
 443 the left panel gives the normalized and 10-year smoothed reconstructed AMOC time  
 444 series from the CNN ( $\text{AMOC}_{\text{CNN}}$ , green), and one SST (19) ( $\text{AMOC}_{\text{SPG}}$ , purple lines) and one  
 445 SSS (20,34) ( $\text{AMOC}_{\text{SSS1}}$ , orange lines) indices from previous studies (materials and  
 446 methods), here computed from EN4. Grey shaded areas give the spread of normalized  
 447 and 10-year smoothed AMOC time series for respective latitudes, computed for 51  
 448 historical simulations from 17 CMIP6 ESMs (3 simulation members each, Tab. S1). The  
 449 ensemble mean of the ESMs' AMOC time series are given by black lines. Right panels give  
 450 the number of years where  $\text{AMOC}_{\text{CNN}}$  (green dots),  $\text{AMOC}_{\text{SPG}}$  (purple dots), and  $\text{AMOC}_{\text{SPG}}$   
 451 (orange dots) fall out of the range drawn by the 51 historical simulations from ESMs (Tab.

452 S1). Grey boxplots give the same out-of-range statistics, when each simulation member  
453 from the 51 ensemble is excluded, and where the ESM spread is recomputed from the 50  
454 other simulations.

455



456  
457 **Figure 4: Comparison of Atlantic Meridional Overturning Circulation (AMOC)**

458 **timeseries with direct observations from RAPID (8) and HadISST (43).** **a,b.** Green lines  
459 give the AMOC reconstructed from Convolutional Neural Network (AMOC<sub>CNN</sub>) for  
460 latitudes 25°N (**a.**) and 30°N (**b.**). Purple line: AMOC reconstruction based on Sea Surface  
461 Temperature (SST) in the Subpolar Gyre (AMOC<sub>SPG</sub>). Blue: AMOC reconstruction based on  
462 Sea Surface Salinity averaged in the subpolar North Atlantic (AMOC<sub>SSS1</sub>). Red: Direct  
463 AMOC measurements from the RAPID data array measuring overturning at 26.5°N in the  
464 North Atlantic (8). **c,d.** Correlation maps of SST from HadISST (43) against AMOC  
465 timeseries reconstructed at 25°N and 30°N from CNN. Stippled grid points indicate  
466 where the correlation is significant at the 90% confidence level using a Student t-test for  
467 correlation with corrected degrees of freedom using time series autocorrelations (44).  
468 **e,f.** Ensemble mean correlation maps of SST against AMOC timeseries at 25°N and 30°N  
469 from the four historical simulations from HadGEM3 studied here. Stippled grid points  
470 indicate where correlations are of same sign in all four members. The purple surrounded  
471 area (panels **c.** through **f.**) depicts the region used for the AMOC<sub>SPG</sub> reconstruction (19).

472 Acknowledgments:

473 The analyses of all the model output were conducted on the Dutch National  
474 Supercomputer (Snellius) within NWO-SURF project~17239. R.M.v.W. and H.A.D. are  
475 funded by the European Research Council through the ERC-AdG project TAOC (PI:  
476 Dijkstra, project~101055096). V.J.D. is funded by the European Union's Horizon 2020  
477 research and innovation program CriticalEarth under the Marie Skłodowska-Curie grant  
478 agreement (project~956170). The work of F.G. was supported by the Netherlands  
479 Organization for Scientific Research (NWO) under grant OCENW.M20.277. S.L.L.M.,  
480 H.A.D. and A.S.vdH. acknowledge funding from the European Union's Horizon 2020  
481 research and innovation program under grant agreement No 820970 (TiPES contribution  
482 no. 255). The work of A.S.vdH. was also funded by the Dutch Research Council (NWO)  
483 through the NWO-Vici project 'Interacting climate tipping elements: When does tipping  
484 cause tipping?' (project VI.C.202.081). Data for CMIP6 ESM outputs are available on the  
485 different Earth System Grid Federation nodes, such as: <https://esgf-data.dkrz.de/projects/esgf-dkrz/>, <https://esgf-index1.ceda.ac.uk/projects/esgf-ceda/>, <https://esgf-node.ipsl.upmc.fr/projects/esgf-ipsl/>. Data from the RAPID AMOC  
488 monitoring project is funded by the Natural Environment Research Council and are freely  
489 available from [www.rapid.ac.uk/rapidmoc](http://www.rapid.ac.uk/rapidmoc).

490

491 Authors contributions:

492 S.L.L.M. has led the study and was its main designer. S.L.L.M and H.A.D. have co-written  
493 the manuscript, managed its production, and determined its scope and purpose. F.G.  
494 and V.J.-D. have carried out the deep learning analyses of the study, with systematic  
495 contributions and assessments from S.L.L.M. and H.A.D. F.G. and V.J.-D. also contributed

496 to the manuscript writing. R.M.v.W. and A.S.v.d.H. contributed to results' assessments,  
497 interpretations, the design of the study, and significantly contributed to the manuscript  
498 writing.

499

500 Code availability:

501 All codes required to reproduce the study are available online on the following Zenodo  
502 link: *to\_be communicated*.

503

504 Data availability: All data required to reproduce the study are available online on the  
505 following Zenodo link: *to\_be communicated*. Output data from the study are available  
506 online on the same Zenodo link.

## Material and methods:

### **AMOC indices**

- The AMOC<sub>SPG</sub> index (19) is calculated for the period 1900-2021 using the HadISST dataset (43). The SPG<sub>SST</sub> is computed as the area-weighted and spatially averaged SST over the SPG region (19) from November to May. The global mean SST (GMST) is computed as the area-weighted and spatially averaged yearly SST over the globe, also from month November to May. The AMOC<sub>SPG</sub> index is then obtained by subtracting GMST from SPG<sub>SST</sub> and covers the period 1900-2021.
- AMOC<sub>sss1</sub> to AMOC<sub>sss4</sub> indices (20) are the same as those investigated by studies analyzing early warning signals for an AMOC tipping point (34,42). These indices all are area-averaged SSS anomalies for the following regions (80°W-0°; 45°N-65°N), (62°W-26°W, 54°N-62°N) (80°W-0°, 10°N-40°N), and (80°-0°, 34°S-10°S), for AMOC<sub>sss1</sub> to AMOC<sub>sss4</sub>, respectively. These indices were extract from the EN4 dataset<sup>45</sup>.
- AMOC time series computed from ESM simulations are obtained as the maximum of the North Atlantic meridional overturning streamfunction below 500m of depth<sup>4</sup>, for the respective latitudes studied here (i.e. from 20°N to 60°N, with an increment of 5°).
- Although CNN returns AMOC strength values in Sverdrup, there is no long enough set of observations to rescale them. Therefore, reconstruction data provided along with the present study are given in absolute Sverdrup values, Sverdrup anomalies, and normalized units. In the manuscript, we only considered reconstructed time series in normalized units relative to the historical period

studied, as it was done for the former AMOC indices (i.e.,  $\text{AMOC}_{\text{SPG}}$  is initially expressed in  $^{\circ}\text{C}$ , and  $\text{AMOC}_{\text{SSS}1}$  to  $\text{AMOC}_{\text{SSS}4}$  are in psu, see above).

## Bias analysis

The bias analysis makes use of an ensemble of 51 simulations from 17 CMIP6 ESMs, with 3 historical realizations per ESM. Biases in simulated SST and SSS fields are calculated based on a comparison with observations from the EN4 dataset (40), where all ESM simulations are regridded on  $1^{\circ}\text{x}1^{\circ}$  regular grids. The area of interest extends between latitudes  $0^{\circ}\text{N}$  and  $70^{\circ}\text{N}$  and longitudes  $80^{\circ}\text{W}$  and  $0^{\circ}$ .

Four statistics are computed for the period 1900-2014, made of  $n = 115$  annual time steps, for every grid point of the study area, in observations and all ESM simulations:

- Mean:  $m(\text{SST}) = \frac{1}{n} \sum_{t=1}^n \text{SST}(t)$
- Standard deviation:  $v(\text{SST}) = \sqrt{\left( \frac{1}{n} \sum_{t=1}^n (\text{SST}(t))^2 \right) - \left( \frac{1}{n} \sum_{t=1}^n (\text{SST}(t)) \right)^2}$
- Minimum:  $\min(\text{SST}) = \min_{t=1,\dots,n} \text{SST}(t)$
- Maximum:  $\max(\text{SST}) = \max_{t=1,\dots,n} \text{SST}(t)$

For each grid point and each statistic, the bias is calculated as the difference between the observed statistic and the simulated statistic. For each ESM and each statistic, biases are averaged over the three historical simulations of the ESM. Mean bias for a given statistic, ocean surface variable, and ESM, is computed as the spatially averaged ensemble-mean absolute biases.

Finally, for each ESM (using 3 historical members), mean biases for SST and SSS are normalized for the four statistics, so that none of them has an arbitrary larger weight (Fig. S3). These normalized biases are summed-up for each ESM to give a global bias score.

Therefore, models with globally smallest biases are those with smallest global bias score and are ranked-up according to these scores (Fig. S3, Tab. S2).

## **Convolutional Neural Network**

The Convolutional Neural Network (CNN) (36) is a deep learning method suited for the analysis of maps because it is able to capture the neighboring spatial relationships present in the data. The time varying two-dimensional SST and/or SSS field for the area [0°N-70°N, 80°W-0°] regridded on 2.5°x2.5° regular grids for the CNN is denoted  $X$  and the target AMOC index for a given latitude from ESM simulations is denoted  $Y$ . Given the resolution and area we use,  $X$  is hence composed of time-varying matrices of 33 rows and 33 columns.

For each CNN computed in the study, the entire training dataset is first divided in batches containing 32, 64, or 128 samples and the learnings were performed for 1000 epochs (see CNN optimization section in materials and methods and Tabs. S5-S7). The initial learning rate (see CNN optimization section in materials and methods and Tabs. S5-S7), but we also use a scheduler: the learning rate is divided by 10 whenever the score function has stopped decreasing for 10 consecutive epochs. In addition, we implemented an early-stopping policy to prevent overfitting, *i.e.*, if the loss does not decrease for more than 50 epochs, the training procedure stops. The score function used here is a mean squared errors and the optimizer is the classic Adam algorithm.

Our architecture is organized in three convolutional layers. Each layer uses 3x3 filters with stride 1 (36). The first, second, and third layer present 16, 32 and 64 filters, respectively. In order to conserve the dimensions of  $X$  throughout the CNN, we use zero-

padding of size 1 in each convolutional layer (in other words,  $X$  is surrounded by 0, which has no effect on the convolution).

Each layer is followed by a LeakyReLU activation function, defined by:

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ -\alpha x & \text{if } x < 0 \end{cases}, \text{with } \alpha = 10^{-2}$$

After the LeakyReLU, we apply batch normalization to every pixel of the activation map in output from the corresponding convolutional layer.  $x$  is the specific activation map, it is transformed into  $\tilde{x}$  by the formula:

$$\tilde{x} = \frac{x - \mu_m}{\sqrt{\sigma_m^2 + \varepsilon}} \gamma + \beta$$

Where  $\mu_m$  and  $\sigma_m^2$  are the mean and variance of  $x$  across the batch, respectively;  $\varepsilon = 10^{-5}$  is meant to improve numerical stability; and  $\gamma$  and  $\beta$  are two parameters to be learnt during the training using the so-called backward propagation (36). This operation greatly accelerates and improves the robustness of the learning, by first normalizing the value of each pixel, then shifting its mean and rescaling it. The new map of values  $\tilde{x}$  is then passed to the next layer.

After the three convolutional layers, the resulting activation map is flattened in a single vector and passed through a multi-layer perceptron. The hidden layer of the multi-layer perceptron has 512 neurons and a single neuron is given in output. Both the hidden layer and the output neuron use LeakyReLU as activation function.

## CNN tuning

For each CNN model used to reconstruct AMOC time series across the 9 different latitudes in ESMs (Fig. 2, Fig. S11), and with real data (45 CNN in total for the experiment with HadGEM3, and 64 CNN models in total for the 6-model experiment, 109 in total), we optimized two control parameters, by comparing different values for each: the batch size (32, 64, and 128), the initial learning rate (0.005, 0.001, 0.0005, and 0.0001). When shuffling all possible combinations of values for these parameters, we thus test 12 configurations for each of the 109 CNN models of the study. For each CNN model. For each CNN, we evaluated these 12 configurations using five-fold cross validations, which consists in separating the learning data into five sets and consecutively excluding each one as test set and using the four others to train the CNN models with all 12 possible combinations of control parameters. Tabs. S5 and S6 present the optimal set of parameters found for the CNN models tuned to reconstruct AMOC time series from ESM simulations and Tab. S7 gives them for real data. For the sake of computation time and energy use, we decided not to optimize the number of epochs for which CNNs were trained. For each CNN, the number of epochs was thus set to 1000, which is a rather large number that ensure convergence of CNNs' training losses.

### **CNNs' stability**

The CNN is not a deterministic learning model due to the random weights' initialization. Therefore, it is common to run a CNN several times to ensure that the learning is robust, *i.e.*, it does not overfit the data or underperform for certain initializations. Once we found the optimal control parameters for a given CNN (Tabs. S5-S7), we applied the CNN with the best set of parameters on the real data for 100 random initializations. The final AMOC reconstructions for each latitude presented in the study, both for ESMs (Fig. 2), and

observational data (Figs. 3,4), are the medians of these 100 reconstructions. In the reconstructions for real data provided along with the present study, we also provided the 5%-95% and 10%-90% spreads drawn by their respective set of 100 single-reconstructions.

### **Evaluation metrics for tuning and validating reconstructions**

- For each ESM run, the evaluation of reconstructed AMOC time series from CNN models ( $\hat{Y}$ ) in reproducing AMOC time series from ESMs' historical simulations ( $Y$ ), over  $N$  common time steps, is made using correlations (Fig. 2), given by:

$$cor(Y, \hat{Y}) = \frac{cov(\hat{Y}, Y)}{var(\hat{Y})var(Y)},$$

where the covariance is given by:

$$cov(\hat{Y}, Y) = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})$$

Variances are given by:

$$var(\hat{Y}) = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - \bar{\hat{Y}})^2, \text{ and, } var(Y) = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

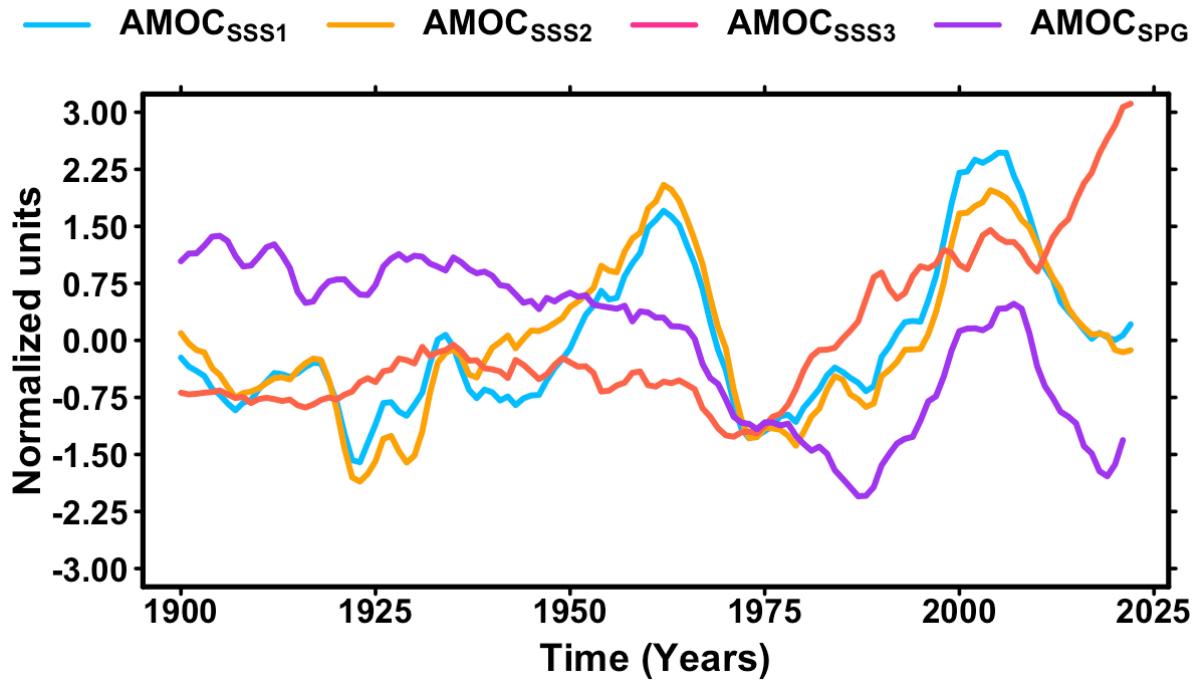
And averages are given by:

$$\bar{\hat{Y}} = \frac{1}{N} \sum_{i=1}^N \hat{Y}_i, \text{ and, } \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

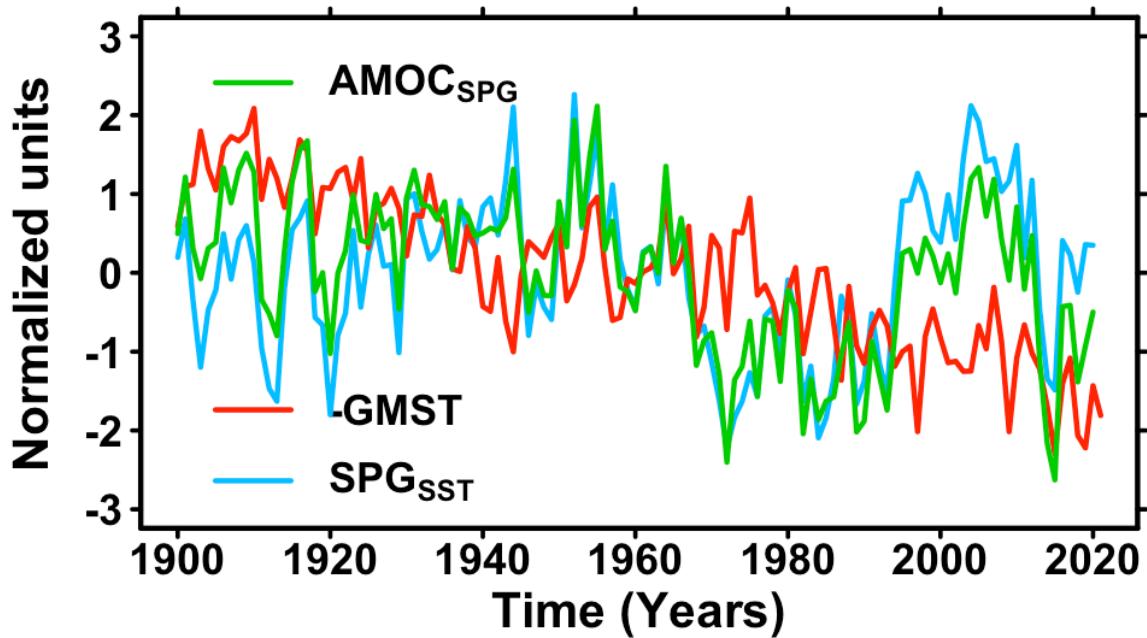
- The tuning of CNN models, is made using the Root Mean Squared Error, which is given by is given by:

$$RMSE\big(Y,\hat{Y}\big)=\sqrt{\frac{1}{N}\sum_{i=1}^N(Y_i-\hat{Y}_i)^2}$$

Supplementary figures:

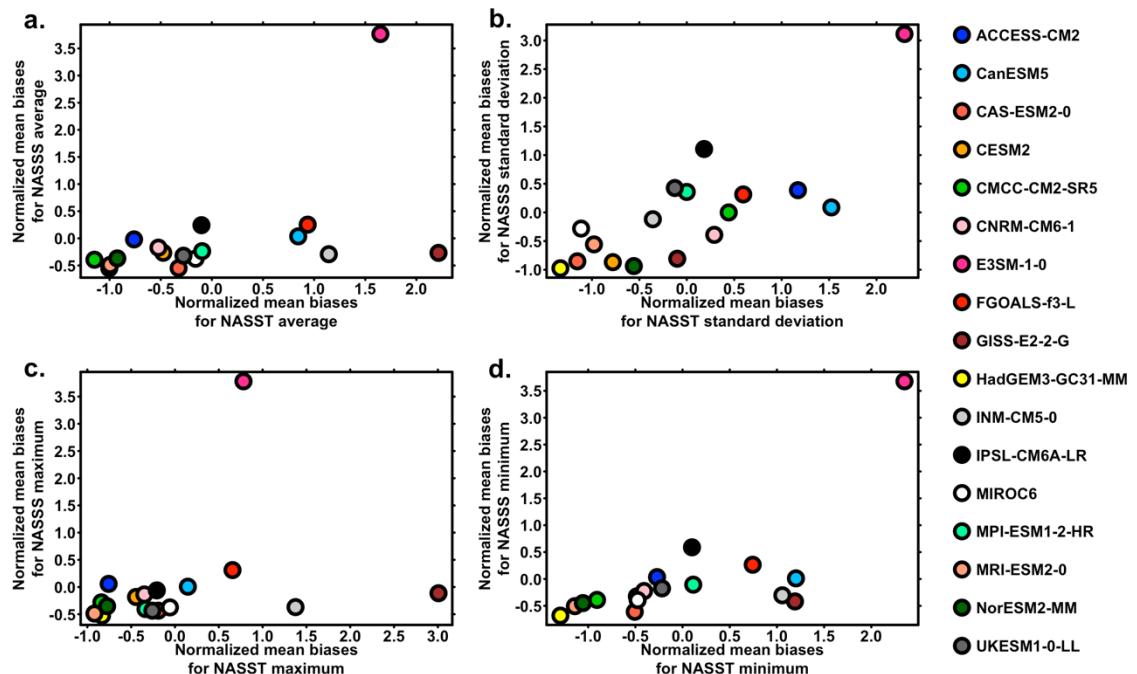


**Figure S1: State-of-the-art Atlantic Meridional Overturning Circulation (AMOC) indices based on sea surface temperature (SST) and salinity (SSS).** Subpolar gyre SST-based AMOC index ( $\text{AMOC}_{\text{SPG}}$ ) (19), in purple, and three SSS-based AMOC indices ( $\text{AMOC}_{\text{SSS}1}$ ,  $\text{AMOC}_{\text{SSS}2}$ ,  $\text{AMOC}_{\text{SSS}3}$ ) (20,34), in blue, orange, and pink, respectively (see materials and methods). All the time series are 10-year filtered and normalized for the sake of graphical representation.

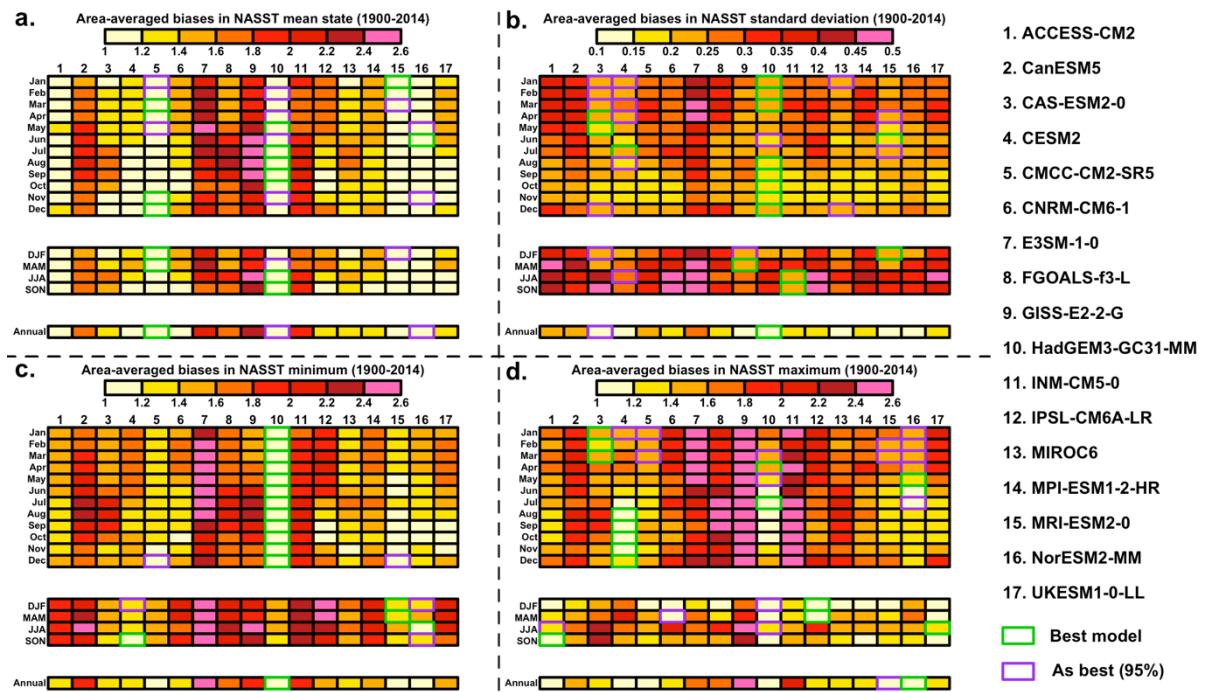


**Figure S2: State-of-the-art Atlantic Meridional Overturning Circulation (AMOC) index**

**calculation.** Subpolar gyre AMOC index (19) ( $\text{AMOC}_{\text{SPG}}$ , green) calculated as Global Mean SST anomalies (GMST, red, multiplied by -1) subtracted from area-averaged subpolar gyre SST ( $\text{SPG}_{\text{ss}}$ , blue), see materials and methods. All the time series were normalized for the sake of graphical representation.

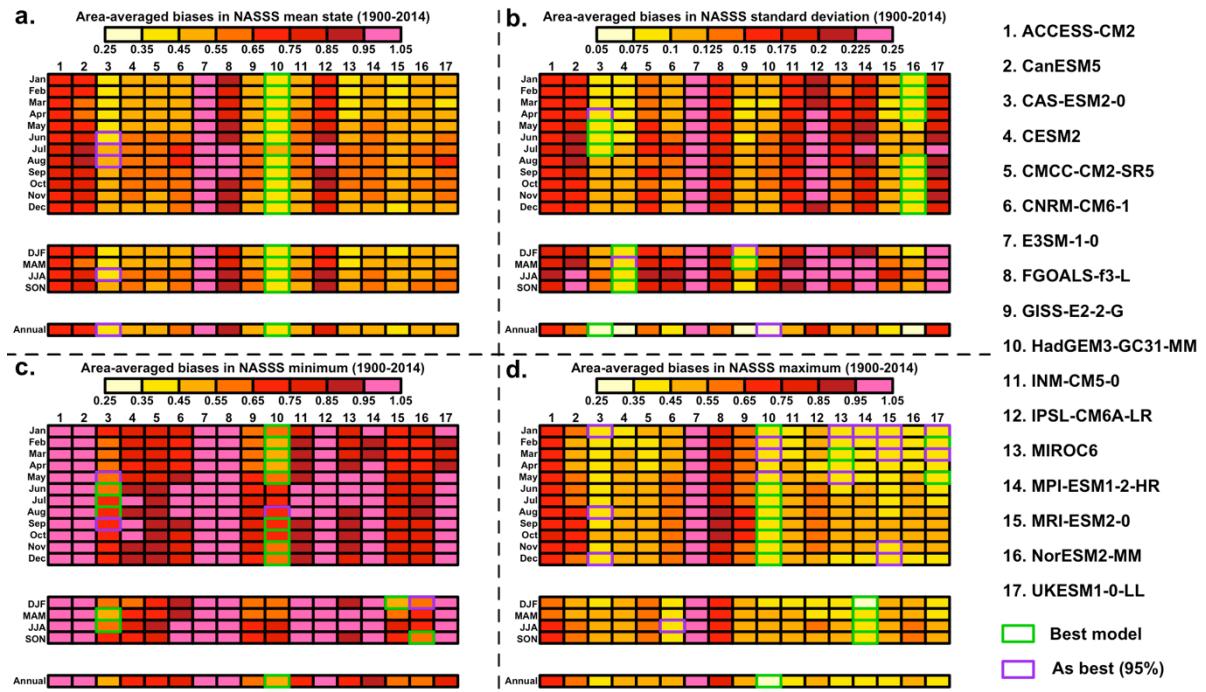


**Figure S3: Bias analysis of 17 Earth System Models (ESMs) in their ability to resolve observed North Atlantic Sea Surface Temperature (NASST) and Salinity (NASSS) statistics from the EN4 dataset (40) over the period 1900-2014.** **a.** Normalized ESM biases in resolving mean NASST (x-axis) and NASSS (y-axis) averages. **b.** Normalized ESM biases in resolving mean NASST (x-axis) and NASSS (y-axis) standard deviations. **c.** Normalized ESM biases in resolving mean NASST (x-axis) and NASSS (y-axis) maxima. **d.** Normalized ESM biases in resolving mean NASST (x-axis) and NASSS (y-axis) minima. For **a-d**, 3 historical simulation members for each ESM were used and statistical biases were averaged. For each couple of climate variable/statistic, a normalization was applied to obtain equal importance for each in the ESM selection. This means that lower (resp. higher) values for both axes indicate smaller (resp. larger) bias for the ESM and the statistic in questions. The ESM (Tab. S1) rank in terms global biases is given in Tab. S2.



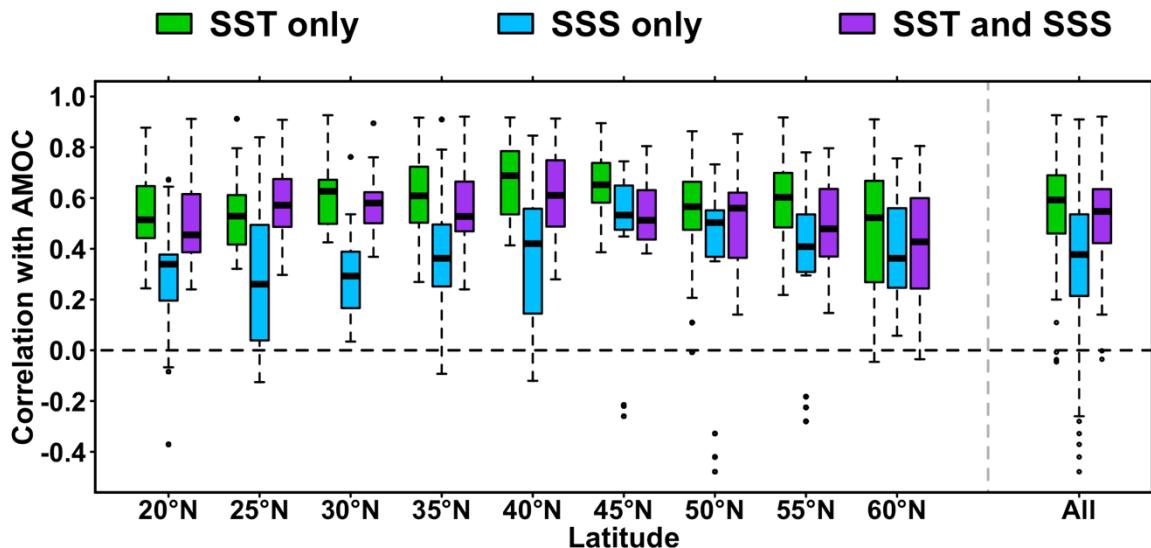
**Figure S4: Bias analysis for 17 Earth System Models (ESMs) in representing North**

**Atlantic Sea Surface Temperature (NASST).** **a-d)** Three member spatially averaged absolute biases for NASST mean, standard deviation, minimum, and maximum statistics, respectively. The study area is 20°N-70°N/80°W-0° and the period of interest is 1900-2014. Outputs from ESMs (Tab. S1) are compared to EN4 observations (40). For each panel, heat maps give monthly, seasonal, and annual biases, from top to bottom. Column numbers refer to ESMs listed on the right (Tab. S1). Green surrounding for each line of each heat map gives the ESM with lowest biases on average. Purple surroundings indicate ESMs whose biases are not significantly different from the best one on the same line (green surrounded one) at the 95% confidence level using a two-tailed Student *t*-test.

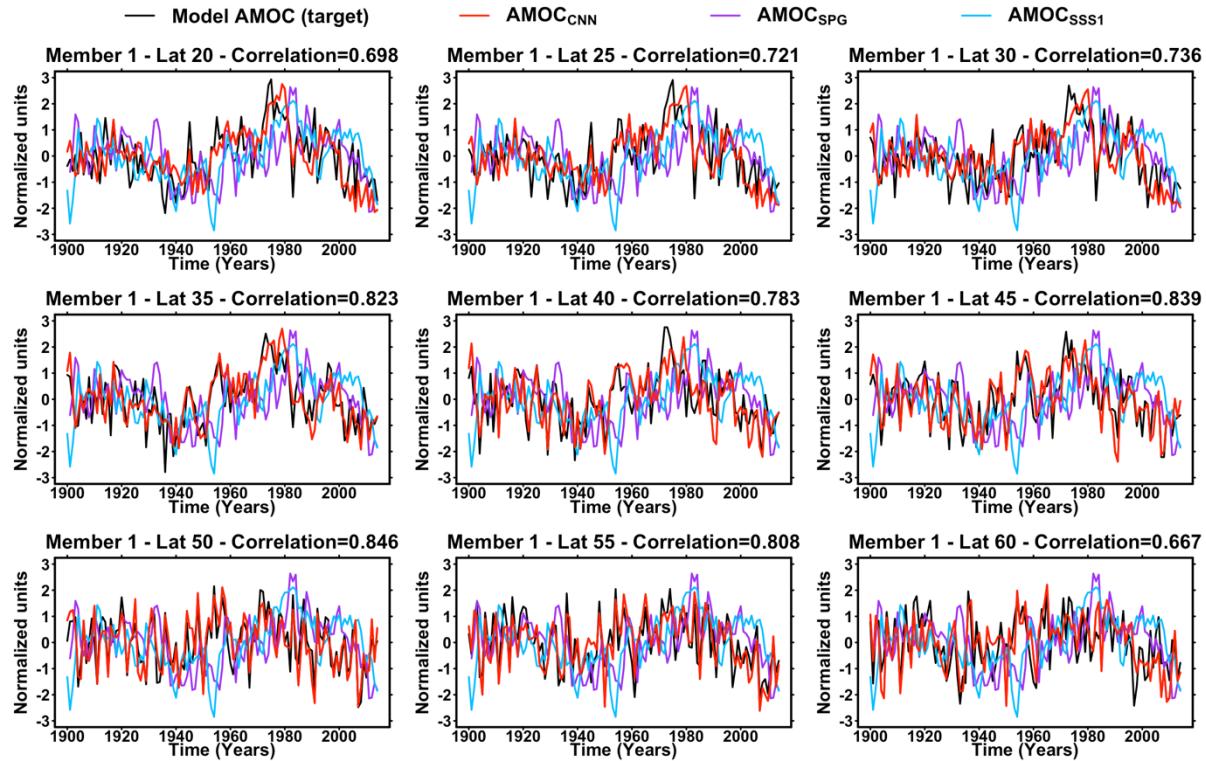


**Figure S5: Bias analysis for 17 Earth System Models (ESMs) in representing North**

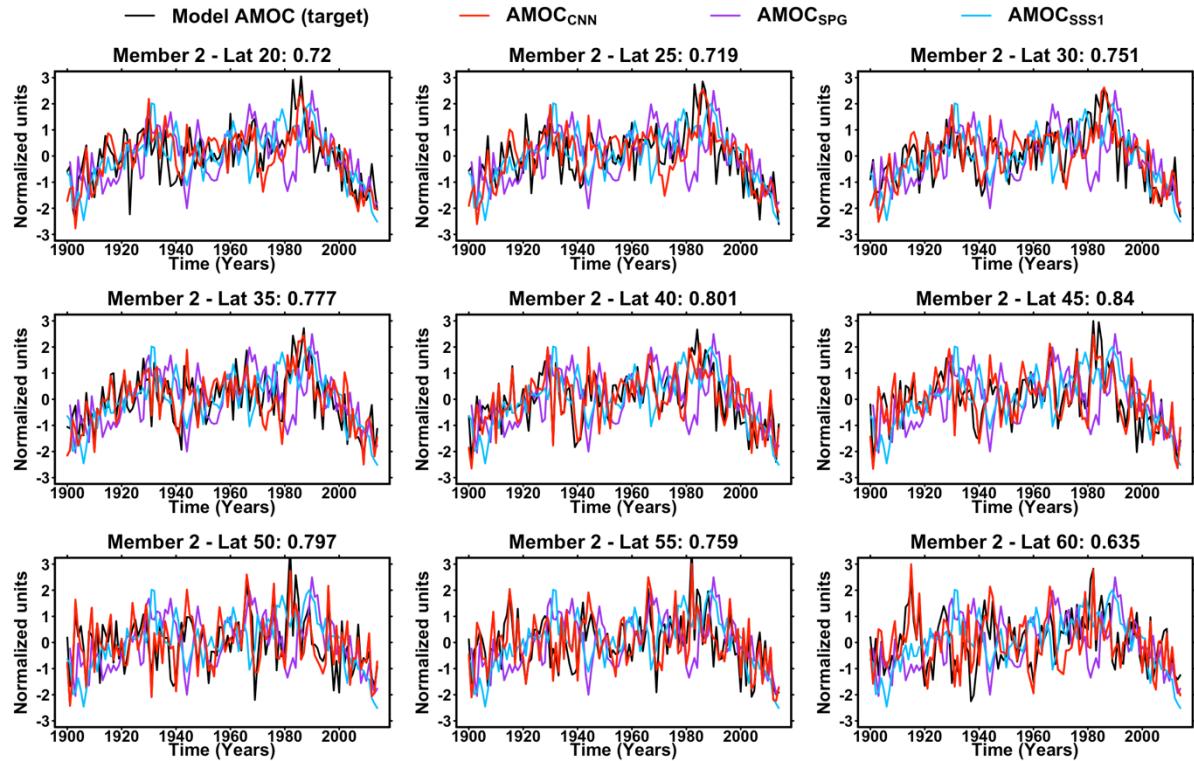
**Atlantic Sea Surface Salinity (NASSS). a-d)** Three member spatially averaged absolute biases for NASSS mean, standard deviation, minimum, and maximum statistics, respectively. The study area is 20°N-70°N/80°W-0° and the period of interest is 1900-2014. Outputs from ESMs (Tab. S1) are compared to EN4 observations (40). For each panel, heat maps give monthly, seasonal, and annual biases, from top to bottom. Column numbers refer to ESMs listed on the right (Tab. S1). Green surrounding for each line of each heat map gives the ESM with lowest biases on average. Purple surroundings indicate ESMs whose biases are not significantly different from the best one on the same line (green surrounded one) at the 95% confidence level using a two-tailed Student *t*-test.



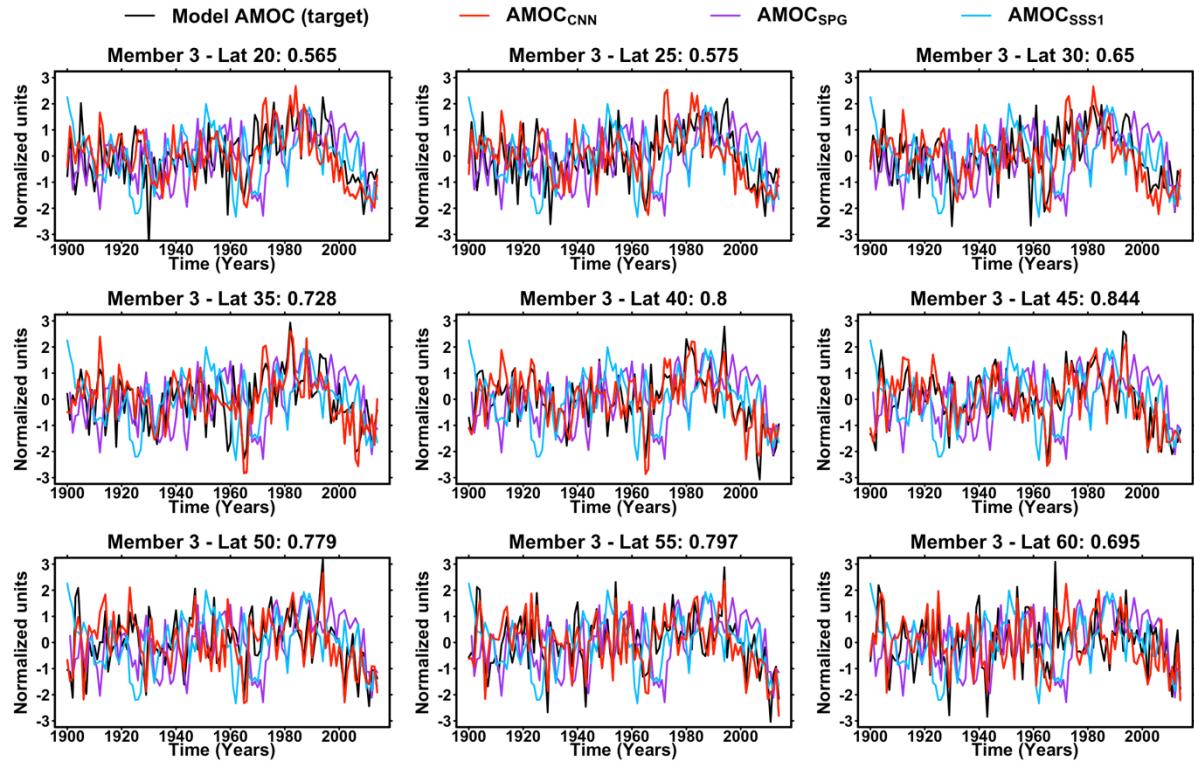
**Figure S6: Performance of Convolutional Neural Networks (CNNs) for reconstructing Atlantic Meridional Overturning Circulation (AMOC) time series from different sets of ocean surface variables using the best 6 ESMs.** Boxplots give correlations obtained by using the 6 less biased Earth System Models (ESMs) for resolving observed Sea Surface Temperature (SST) and Salinity (SSS) statistics (Figs. S3-S5, Tab. S2, materials and methods). Correlations are given when training tuned CNNs using SST only (green), SSS only (blue), SST and SSS together (purple). Correlation statistics were obtained by excluding all simulations from a single ESM, one after the other, and reconstructing the AMOC time series from this excluded ESM for its three historical runs (Tab. S4,S6). Reconstructed AMOC time series range for latitudes from 20°N to 60°N, with an increment of 5°. Correlations averaged over all latitudes are given for boxplot with the “All” x-axis label.



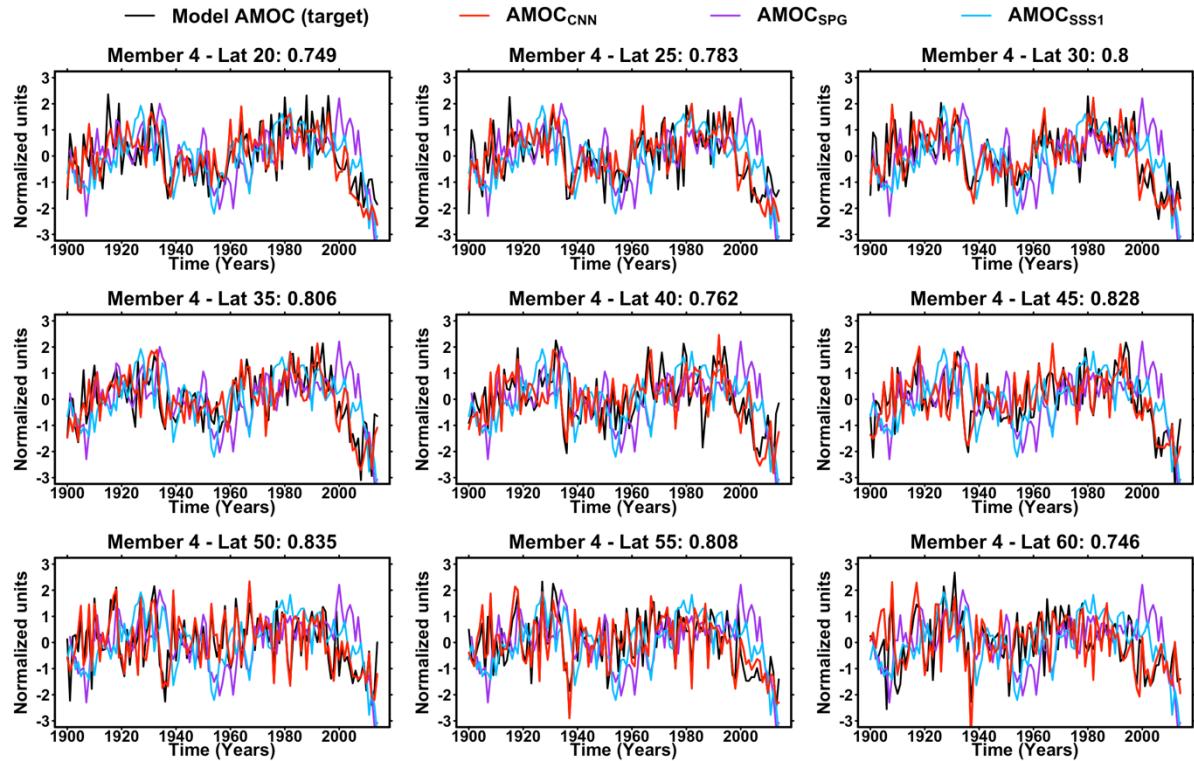
**Fig. S7: Individual reconstruction of Atlantic Meridional Overturning Circulation (AMOC) obtained when HadGEM3 historical simulation member 1 was excluded from training.** Panels a.-i. give the obtained reconstructions for latitudes ranging from 20°N to 60°N with a step of 5°, respectively. Red line: reconstructed AMOC time series from Convolutional Neural Network (CNN). Black line: real AMOC timeseries from the HadGEM3 simulation. Purple line: AMOC reconstruction based on Subpolar Gyre (SPG) Sea Surface Temperature (AMOC<sub>SPG</sub>) (19). Blue line: AMOC reconstruction based on SPG Sea Surface Salinity (AMOC<sub>SSS1</sub>) (20,34). Correlation between model simulation AMOC (black) and the AMOC reconstructed from CNN (red) are given in each panel's title.



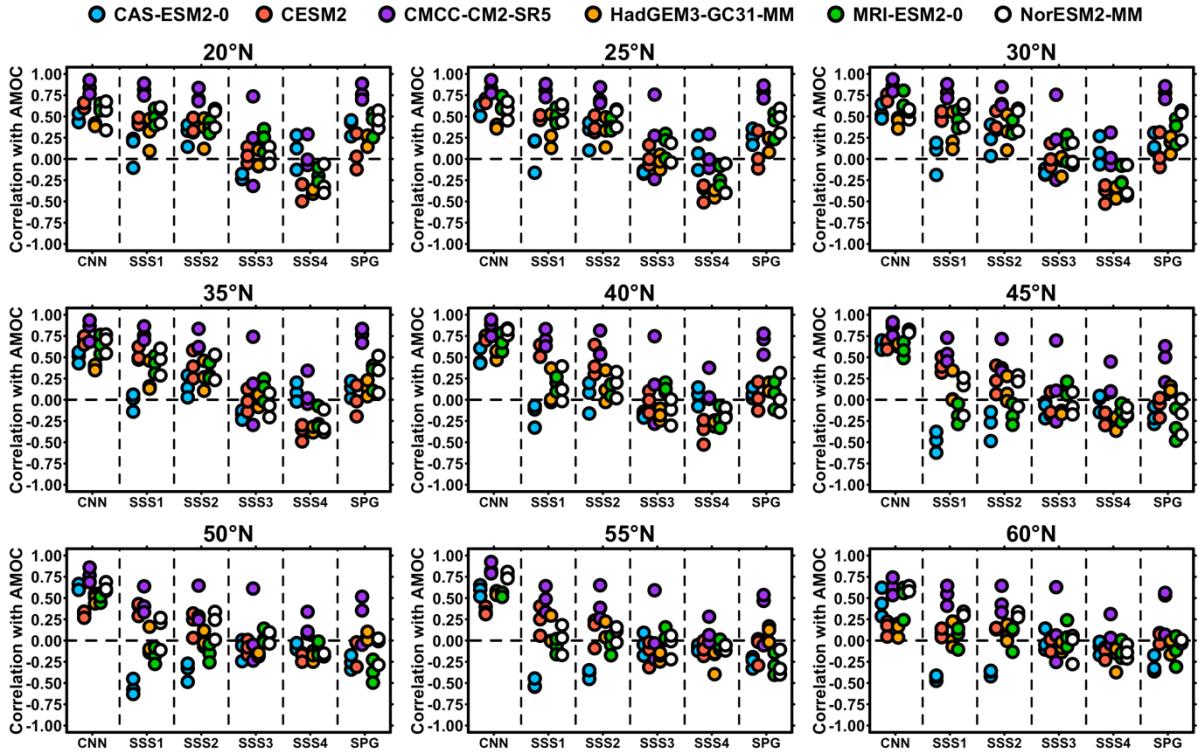
**Fig. S8: Individual reconstruction of Atlantic Meridional Overturning Circulation (AMOC) obtained when HadGEM3 historical simulation member 2 was excluded from training.** Panels (from left to right then top to bottom) give the obtained reconstructions for latitudes ranging from 20°N to 60°N with a step of 5°, respectively. Red line: reconstructed AMOC time series from Convolutional Neural Network (CNN). Black line: real AMOC timeseries from the HadGEM3 simulation. Purple line: AMOC reconstruction based on Subpolar Gyre (SPG) Sea Surface Temperature (AMOC<sub>SPG</sub>) (19). Blue line: AMOC reconstruction based on SPG Sea Surface Salinity (AMOC<sub>SSS1</sub>) (20-34). Correlation between model simulation AMOC (black) and the AMOC reconstructed from CNN (red) are given in each panel's title.



**Fig. S9: Individual reconstruction of Atlantic Meridional Overturning Circulation (AMOC) obtained when HadGEM3 historical simulation member 3 was excluded from training.** Panels a.-i. give the obtained reconstructions for latitudes ranging from 20°N to 60°N with a step of 5°, respectively. Red line: reconstructed AMOC time series from Convolutional Neural Network (CNN). Black line: real AMOC timeseries from the HadGEM3 simulation. Purple line: AMOC reconstruction based on Subpolar Gyre (SPG) Sea Surface Temperature (AMOC<sub>SPG</sub>) (19). Blue line: AMOC reconstruction based on SPG Sea Surface Salinity (AMOC<sub>SSS1</sub>) (20,34). Correlation between model simulation AMOC (black) and the AMOC reconstructed from CNN (red) are given in each panel's title.



**Fig. S10: Individual reconstruction of Atlantic Meridional Overturning Circulation (AMOC) obtained when HadGEM3 historical simulation member 4 was excluded from training.** Panels a.-i. give the obtained reconstructions for latitudes ranging from 20°N to 60°N with a step of 5°, respectively. Red line: reconstructed AMOC time series from Convolutional Neural Network (CNN). Black line: real AMOC timeseries from the HadGEM3 simulation. Purple line: AMOC reconstruction based on Subpolar Gyre (SPG) Sea Surface Temperature (AMOC<sub>SPG</sub>) (19). Blue line: AMOC reconstruction based on SPG Sea Surface Salinity (AMOC<sub>SSS1</sub>) (20,34). Correlation between model simulation AMOC (black) and the AMOC reconstructed from CNN (red) are given in each panel's title.



**Figure S11: Performance of Convolutional Neural Networks (CNNs) for reconstructing Atlantic Meridional Overturning Circulation (AMOC) time series for each excluded Earth System Models (ESMs), considering the best 6 ESMs.** Dots give correlations obtained by using the 6 less biased ESMs for resolving observed Sea Surface Temperature (SST) and Salinity (SSS) statistics (Figs. S3-S5, Tab. S2,S4, materials and methods). Correlations are given when training tuned CNNs using SST only. Correlation statistics were obtained by excluding all simulations from a single ESM (Tab. S4), one after the other, and reconstructing the AMOC time series from this excluded ESM for its three historical runs (Tab. S4). Reconstructed AMOC time series range for latitudes from 20°N to 60°N, with an increment of 5°. For each latitude (panels), each of the 3 dots per ESM indicate the obtained correlations when AMOC was reconstructed using the tuned

CNN for the three historical members of the ESM in question, respectively, and whose simulations (Tab. S4) were excluded for CNN training and tuning.

## Supplementary tables:

Model name	Modelling centre (country)	Experiment (period)	Members
ACCESS-CM2	CSIRO-ARCCSS (Australia)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
CanESM5	CCCma (Canada)	historical (1900-2014)	r1i1p2f1, r2i1p2f1, r3i1p2f1
CAS-ESM2-0	CAS (China)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
CESM2	NCAR (USA)	historical (1900-2014)	r4i1p1f1, r10i1p1f1, r11i1p1f1
CMCC-CM2-SR5	CMCC (Italy)	historical (1900-2014)	r1i1p1f1, r2i1p2f1, r3i1p2f1
CNRM-CM6-1	CNRM-CERFACS (France)	historical (1900-2014)	r1i1p1f2, r2i1p1f2, r3i1p1f2
E3SM-1-0	DOE (USA)	historical (1900-2014)	r2i1p1f1, r3i1p1f1, r4i1p1f1
FGOALS-f3-L	CAS (China)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
GISS-E2-2-G	NASA GISS (USA)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
HadGEM3-GC31-MM	MOHC (UK)	historical (1900-2014)	r1i1p1f3, r2i1p1f3, r3i1p1f3
INM-CM5-0	INM (Russia)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
IPSL-CM6A-LR	IPSL (France)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
MIROC6	JAMSTEC (Japan)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
MPI-ESM1-2-HR	MPI (Germany)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
MRI-ESM2-0	MRI (Japan)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
NorESM2-MM	NCC (Norway)	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
UKESM1-0-LL	MOHC (UK)	historical (1900-2014)	r1i1p1f2, r2i1p1f2, r3i1p1f2

**Table S1: List of 51 historical simulations from 17 CMIP6 (1) Earth System Models (3 simulations each) for the AMOC analysis presented in Fig. 1 and the bias analysis presented in Fig. 2.**

Model name	SST bias ranking	SSS bias ranking	SST+SSS bias ranking
ACCESS-CM2	10	14	11
CanESM5	15	13	14
CAS-ESM2-0	5	2	4
CESM2	4	6	5
CMCC-CM2-SR5	6	10	6
CNRM-CM6-1	9	8	8
E3SM-1-0	17	17	17
FGOALS-f3-L	13	16	15
GISS-E2-2-G	16	5	16
HadGEM3-GC31-MM	1	1	1
INM-CM5-0	14	9	13
IPSL-CM6A-LR	12	15	12
MIROC6	7	7	7
MPI-ESM1-2-HR	11	11	10
MRI-ESM2-0	2	3	2
NorESM2-MM	3	4	3
UKESM1-0-LL	8	12	9

**Table S2: Model ranking for resolving observed SST and SSS statistics.** Mean biases

for the period 1900-2014 were computed for four statistics: mean state, standard deviation, minimum and maximum, for three simulation members for each model. The ranking is determined by the minimum of the sum of normalized biases for each of the four statistics (see Figs. S3-S5).

Model name	Experiment (period)	Member(s)
HadGEM3-GC31-MM	historical (1900-2014)	r1i1p1f3, r2i1p1f3, r3i1p1f3, r4i1p1f3
HadGEM3-GC31-MM	piControl (1850-2349)	r1i1p1f1
HadGEM3-GC31-MM	ssp126 (2015-2100)	r1i1p1f3

**Table S3: List of 5 simulations from less biased Earth System Models in resolving SST**

**and SSS used for training and validating CNNs in the main text.**

Model name	Experiment (period)	Member(s)
CAS-ESM2-0	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
CAS-ESM2-0	piControl (100-499)	r1i1p1f1
CAS-ESM2-0	ssp126 (2015-2100)	r1i1p1f1
CESM2	historical (1900-2014)	r4i1p1f1, r10i1p1f1, r11i1p1f1
CESM2	piControl (701-1200)	r1i1p1f1
CESM2	ssp126 (2015-2100)	r4i1p1f1
CMCC-CM2-SR5	historical (1900-2014)	r1i1p1f1, r2i1p2f1, r3i1p2f1
CMCC-CM2-SR5	piControl (1850-2349)	r1i1p1f1
CMCC-CM2-SR5	ssp126 (2015-2100)	r1i1p1f1
HadGEM3-GC31-MM	historical (1900-2014)	r1i1p1f3, r2i1p1f3, r3i1p1f3
HadGEM3-GC31-MM	piControl (1850-2349)	r1i1p1f1
HadGEM3-GC31-MM	ssp126 (2015-2100)	r1i1p1f3
MRI-ESM2-0	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
MRI-ESM2-0	piControl (2051-2550)	r1i1p1f1
MRI-ESM2-0	ssp126 (2015-2100)	r1i1p1f1
NorESM2-MM	historical (1900-2014)	r1i1p1f1, r2i1p1f1, r3i1p1f1
NorESM2-MM	piControl (1200-1699)	r1i1p1f1
NorESM2-MM	ssp126 (2015-2100)	r1i1p1f1

**Table S4: List of 30 simulations from less biased Earth System Models in resolving SST and SSS used for training and validating CNNs for the analyses based on the best 6 ESMs.**

Excluded member	Latitude	Batch size	Initial learning rate
r1i1p1f3	20	64	0.001
r1i1p1f3	25	64	0.001
r1i1p1f3	30	64	0.001
r1i1p1f3	35	64	0.001
r1i1p1f3	40	64	0.001
r1i1p1f3	45	64	0.001
r1i1p1f3	50	32	0.001
r1i1p1f3	55	128	0.001
r1i1p1f3	60	128	0.001
r2i1p1f3	20	32	0.001
r2i1p1f3	25	32	0.001
r2i1p1f3	30	32	0.001
r2i1p1f3	35	64	0.001
r2i1p1f3	40	32	0.001
r2i1p1f3	45	64	0.001
r2i1p1f3	50	128	0.001
r2i1p1f3	55	128	0.001
r2i1p1f3	60	128	0.001
r3i1p1f3	20	64	0.001
r3i1p1f3	25	64	0.001
r3i1p1f3	30	32	0.001
r3i1p1f3	35	64	0.001
r3i1p1f3	40	64	0.001
r3i1p1f3	45	32	0.001
r3i1p1f3	50	128	0.001
r3i1p1f3	55	128	0.001
r3i1p1f3	60	128	0.001
r4i1p1f3	20	64	0.001
r4i1p1f3	25	64	0.001
r4i1p1f3	30	32	0.001
r4i1p1f3	35	64	0.001
r4i1p1f3	40	64	0.001
r4i1p1f3	45	64	0.001
r4i1p1f3	50	64	0.001
r4i1p1f3	55	128	0.001
r4i1p1f3	60	128	0.001

**Table S5: Hyperparameters for reconstructing Atlantic Meridional Overturning**

**Circulation in HadGEM3 historical members for each of the latitudes studied.** For each pair of historical simulation/latitude, the historical member in question was excluded and the tuning based on 5-fold cross-validation was performed all other simulations from Tab. S3. Each Convolutional Neural Network (CNN) (36) was run for 1000 epochs (see materials and methods) and tested for 12 sets of control parameters: for batch sizes of 32, 64, and 128 (see materials and methods) and initial learning rates of 0.005, 0.001, 0.0005, 0.0001 (see materials and methods). The 5 validation splits for

cross-validation partitions were performed using a sliding window representing 20% of the total size of the data (see materials and methods). The root mean squared error (materials and methods) across training/validation cross-validation partitions was used to determine optimal batch sizes and initial learning rates.

Model name	Latitude	Batch size	Initial learning rate
CAS-ESM2-0	20	128	0.0001
CAS-ESM2-0	25	64	0.0005
CAS-ESM2-0	30	64	0.0005
CAS-ESM2-0	35	64	0.0001
CAS-ESM2-0	40	128	0.0005
CAS-ESM2-0	45	64	0.0005
CAS-ESM2-0	50	128	0.005
CAS-ESM2-0	55	64	0.0005
CAS-ESM2-0	60	64	0.0005
CESM2	20	128	0.0005
CESM2	25	64	0.0005
CESM2	30	64	0.0005
CESM2	35	128	0.001
CESM2	40	64	0.0001
CESM2	45	64	0.0005
CESM2	50	64	0.0005
CESM2	55	64	0.0005
CESM2	60	64	0.0005
CMCC-CM2-SR5	20	64	0.0005
CMCC-CM2-SR5	25	64	0.0001
CMCC-CM2-SR5	30	128	0.0005
CMCC-CM2-SR5	35	64	0.0005
CMCC-CM2-SR5	40	64	0.0001
CMCC-CM2-SR5	45	64	0.001
CMCC-CM2-SR5	50	64	0.0005
CMCC-CM2-SR5	55	64	0.001
CMCC-CM2-SR5	60	64	0.005
HadGEM3-GC31-MM	20	64	0.0005
HadGEM3-GC31-MM	25	64	0.001
HadGEM3-GC31-MM	30	64	0.0005
HadGEM3-GC31-MM	35	128	0.0001
HadGEM3-GC31-MM	40	128	0.001
HadGEM3-GC31-MM	45	64	0.001
HadGEM3-GC31-MM	50	64	0.0005
HadGEM3-GC31-MM	55	64	0.0005
HadGEM3-GC31-MM	60	128	0.001
MRI-ESM2-0	20	64	0.0005
MRI-ESM2-0	25	128	0.0005
MRI-ESM2-0	30	64	0.001
MRI-ESM2-0	35	64	0.0001
MRI-ESM2-0	40	128	0.001
MRI-ESM2-0	45	64	0.0005
MRI-ESM2-0	50	64	0.0005
MRI-ESM2-0	55	64	0.001
MRI-ESM2-0	60	64	0.0005
NorESM2-MM	20	64	0.001
NorESM2-MM	25	64	0.0005
NorESM2-MM	30	64	0.0005
NorESM2-MM	35	64	0.0001
NorESM2-MM	40	64	0.0001
NorESM2-MM	45	64	0.0005
NorESM2-MM	50	64	0.0005
NorESM2-MM	55	64	0.001
NorESM2-MM	60	64	0.0005

**Table S6: Hyperparameters for Earth System Models (ESMs) for each of the latitudes studied for the 6-ESMs experiments.** For each pair of ESM/latitude, the ESM in question was excluded and the tuning based on 5-fold cross-validation was performed on simulations from the 4 other ESMs. Each Convolutional Neural Network (CNN) (36) was run for 1000 epochs (see materials and methods) and tested for 12 sets of control parameters: for batch sizes of 32, 64, and 128 (see materials and methods) and initial learning rates of 0.005, 0.001, 0.0005, 0.0001 (see materials and methods). The 5 validation splits for cross-validation partitions were performed using a sliding window representing 20% of the total size of the data (see Methods). The root mean squared error (materials and methods) across training/validation cross-validation partitions was used to determine optimal batch sizes and initial learning rates.

Latitude	Batch size	Initial learning rate
20	64	0.001
25	64	0.001
30	64	0.001
35	64	0.001
40	128	0.001
45	64	0.001
50	128	0.001
55	32	0.001
60	64	0.001

**Table S7: Hyperparameters for historical AMOC reconstructions for each of the latitudes studied.** For each latitude, the tuning based on 5-fold cross-validation was performed on all simulations from HadGEM3 (Tab. S3). Each Convolutional Neural Network (CNN) (36) was run for 1000 epochs (see materials and methods) and tested for 12 sets of control parameters: for batch sizes of 32, 64, and 128 (see materials and methods), and initial learning rates of 0.005, 0.001, 0.0005, 0.0001 (see materials and methods). The 5 validation splits for cross-validation partitions were performed using a sliding window representing 20% of the total size of the data (see materials and methods). The root mean squared error (materials and methods) across training/validation cross-validation partitions was used to determine optimal batch sizes and initial learning rates.