



No critical slowing down in the Atlantic Overturning Circulation in historical CMIP6 simulations

Maya Ben-Yami^{1,2}, Lana Blaschke^{1,2}, Sebastian Bathiany^{1,2}, and Niklas Boers^{1,2,3}

¹Earth System Modelling, School of Engineering and Design, Technical University of Munich, Munich, Germany

²Potsdam Institute for Climate Impact Research, Potsdam, Germany

³Department of Mathematics and Global Systems Institute, University of Exeter, Exeter, UK

Correspondence: Maya Ben-Yami (maya.ben-yami@tum.de)

1 **Abstract.** The Atlantic Meridional Overturning Circulation (AMOC) is a key component of the Earth's climate system, and is
2 theorized to have multiple stable states. Critical slowing down (CSD) can detect stability changes in Earth system components,
3 and has been found in sea-surface temperature (SST) based fingerprints of the AMOC. We look for CSD in simulations from
4 27 models from the sixth Climate Model Intercomparison Project (CMIP6). We calculate three different CSD indicators for
5 the AMOC streamfunction strengths at 26.5°N and 35°N, as well as for a previously suggested AMOC fingerprint based on
6 averaging SSTs in the subpolar gyre region. We find mixed results. Most models do not have a statistically significant sign
7 of CSD, and no model has a conclusive sign of CSD in all ensemble members. However, some models exhibit a number of
8 significant increases in the CSD indicators of the streamfunction strength that is highly unlikely to occur by chance ($p < 0.05$). In
9 addition, the number of significant increases in the AMOC SST fingerprint are as would be expected by random chance. Since
10 we do not know whether or not the AMOC in these models is approaching a critical transition, we cannot deny or confirm the
11 validity of CSD for detecting an upcoming AMOC collapse. However, we can confirm that the AMOC SST fingerprint is not
12 prone to false positives.

13 1 Introduction

14 The Atlantic Meridional Overturning Circulation (AMOC) is a vital part of the Earth's climate system. It consists of a system of
15 currents that transport large amounts of heat and salt northward near the surface of the Atlantic Ocean. The AMOC is predicted
16 to weaken under anthropogenic global warming (Lee et al., 2021), with substantial effects on the climate (Bellomo et al.,
17 2021; Liu et al., 2020). There is some observational and paleoclimate evidence that the AMOC has already weakened (Caesar
18 et al., 2021; Zhu et al., 2023), although the decline is controversial and likely still consistent with natural variability (Kilbourne
19 et al., 2022; Latif et al., 2022). The AMOC has also been identified as a potential climate tipping element (McKay et al.,
20 2022; Boers et al., 2022), with support for bistability coming from paleoclimate proxy records (Henry et al., 2016), theoretical
21 studies (Stommel, 1961; Cessi, 1994), as well as from experiments with some general circulation models (Rahmstorf, 2002;
22 Hawkins et al., 2011; Liu et al., 2017; Jackson and Wood, 2018; Westen et al., 2024). However, the AMOC appears to be
23 monostable in other models, and the question of alternative stable AMOC states in comprehensive climate models remains
24 debated (Jackson et al., 2023a). Roughly, the AMOC's bistability is theorised to arise from a key positive feedback in the



25 system, the salt advection feedback. A weaker AMOC leads to less northward salt transport and thus reduced surface density
26 in the regions of north Atlantic deep water (NADW) formation. The reduction in NADW formation in turn leads to an even
27 weaker AMOC, continuing until an alternative weak AMOC state is reached. A question of intense debate is whether or not
28 the AMOC could undergo such an abrupt collapse under future anthropogenic climate change (Mckay et al., 2022; Boers
29 et al., 2022). The 6th assessment report (AR6) of the International Panel on Climate Change (IPCC) concludes that there is
30 medium confidence that such a collapse will not happen before 2100 (Arias et al., 2021). This conclusion is in large part based
31 on results from complex climate models, and in particular on the experiments of the sixth Climate Model Intercomparison
32 Projects (CMIP6). Yet the complexity and extent of the AMOC makes it challenging to simulate, and complex models can fail
33 to capture key processes which could lead to a warming-induced collapse (Jackson et al., 2023b; Liu et al., 2017). When the
34 small number of experiments that modelled a global-warming induced AMOC collapses are used to estimate the temperatures
35 at which a collapse might happen, these range from 1.4 to 8 °C, in practice encompassing the full scope of future warming
36 (Mckay et al., 2022).

37 Another approach to understanding the stability of the AMOC is to use its characteristic statistical properties from the
38 perspective of nonlinear dynamical systems. In particular, when a system approaches a bifurcation-induced transition to a
39 different stable state it undergoes a process called Critical Slowing Down (CSD), in which the rate of the system's return
40 to equilibrium after perturbations weakens and eventually approaches zero. Statistical indicators of CSD, such as increasing
41 variance or autocorrelation, are a measure of the stability of the system. In extreme cases they can signal an approaching
42 transition, and have thus sometimes been called early-warning signals (Scheffer et al., 2009). Boers (2021) (hereafter B21)
43 studied sea surface temperature (SST)- and salinity-based observational proxies of the historical AMOC strength, and found
44 a significant signal of CSD in these proxies. B21 also applied this analysis to time series of AMOC streamfunction strengths
45 and SSTs from the historical runs of the fifth Climate Model Intercomparison Project (CMIP5), but found no consistent sign
46 of CSD in those runs. In this study we repeat this CSD analysis on a range of more recent state-of-the-art models from CMIP6,
47 expand it to other CSD indicators and AMOC fingerprints, and look for patterns in the results.

48 There are two main reasons why such an analysis is of interest. The first is that to trust the projections of climate models we
49 have to understand their accuracy and biases, and one of the best ways to do that is to compare their results to observations of
50 the climate. Signs of CSD are consistently present in observational proxies that have been suggested to represent the AMOC,
51 and so we can evaluate models by the presence or lack of CSD in the corresponding model proxy time series from the historical
52 runs.

53 The second reason is to validate the usefulness of these SST- and salinity-based proxies for detection of CSD for the AMOC.
54 In an ideal scenario, one would calculate CSD indicators directly from the time series that characterises the system: the AMOC
55 streamfunction strength. However, CSD detection requires long timeseries, and direct observations of AMOC strength only start
56 in 2004 (Frajka-Williams et al., 2021). Studies have thus identified multiple AMOC fingerprints that use long-term observation-
57 based time series that in models have been shown to be correlated – at different strength and different lags – with the AMOC
58 strength (Jackson and Wood, 2020).



59 One of the most commonly used AMOC fingerprints is based on the SSTs in the subpolar gyre (SPG), and in this work
60 will be called the AMOC SST index (Caesar et al., 2018; Rahmstorf et al., 2015). Although there are some concerns about
61 this fingerprint (Little et al., 2020; Zhu et al., 2023), studies have found correlations between this index and the AMOC
62 streamfunction strength that are sufficient for understanding the historical AMOC trend (Caesar et al., 2018; Jackson and
63 Wood, 2020). Thus the observed decrease in the SST index likely indicates a decrease in the strength of the AMOC (Caesar
64 et al., 2018). However, so far no study has examined if this or other fingerprints can be used to detect CSD in the AMOC
65 under anthropogenic forcing. The best evidence for using the SST index to detect CSD would be to consistently see CSD in
66 the fingerprint in models where the AMOC then subsequently undergoes a warming-induced collapse. Unfortunately, there are
67 very few models in which the AMOC collapses due to global warming, and there are no studies that test the AMOC SST index
68 in this manner.

69 In this work, we calculate CSD indicators for a large range of CMIP6 models. We examine our results to see if they can be
70 used to evaluate the climate models, and whether they validate the use of the AMOC SST Index for CSD detection.

71 **2 Results**

72 **2.1 AMOC timeseries slopes**

73 In this work, the AMOC streamfunction for the different CMIP6 models is computed in the same way as in Menary et al. (2020).
74 To get the streamfunction for a larger number of models, the oceanic meridional velocities are used to calculate the overturning
75 stream function directly for each model. For the streamfunction strength we take the maximum strength of overturning at
76 latitudes 35°N and 26.5°N. These time series will in the following sometimes be abbreviated as S35 and S26, respectively. The
77 AMOC SST Index is calculated following Caesar et al. (2018) but with the SPG area simplified as in Menary et al. (2020). The
78 index is the average SSTs in the area between 41° and 60° N and 20° and 55° E, minus the average global SSTs. The AMOC
79 SST Index will sometimes be shortened to ASSTI. The list of CMIP6 models and their ensemble members can be found in
80 Table C1.

81 As expected, the slope of the AMOC streamfunction strength at the two different latitudes is highly correlated (Fig. 1).
82 Throughout this work, slope stands for the slope of a fitted linear trend. The correlation of the streamfunction slope with the
83 AMOC SST Index slope is present in CMIP6 (Fig. 2), as seen in other studies in CMIP5 (Caesar et al., 2018; Menary et al.,
84 2020), and has an R value of 0.7 and 0.67 for 26.5°N and 35°N, respectively. As discussed in many previous works, a large part
85 of the AMOC time series in CMIP6 have a positive slope, i.e. an increasing AMOC strength in the historical period (Menary
86 et al., 2020; Weijer et al., 2020). For the SST Index this is in opposition to the slope in the observed SSTs, which is around -0.5
87 C°/Century. This increase in AMOC strength in CMIP6 is mostly present in the time span of 1850-1985, and has in part been
88 attributed to too-strong aerosol forcing (see Menary et al. (2020); Robson et al. (2022); Weijer et al. (2020) for discussion of
89 this forcing and for full AMOC time series). In addition, other work has shown that the large scatter in slopes is also caused
90 by internal AMOC variability (Bonnet et al., 2021). Indeed, although for each model the ensemble members have the same
91 forcing, in some cases there is a large spread in ensemble member slopes (Fig. C1).

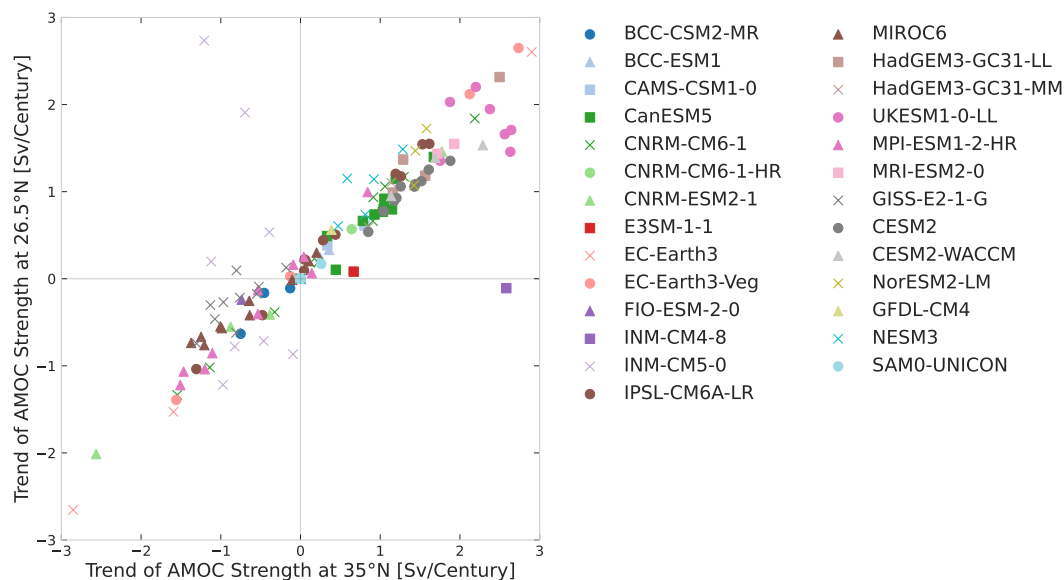


Figure 1. slope of AMOC streamfunction strength at 35°N vs 26.5°N for the individual ensemble members of 27 CMIP6 models. For each model all ensemble member slopes are shown using the same coloured symbol.

92 2.2 AMOC CSD indicators

93 Following B21 the main CSD indicator we will use is the restoring rate λ . Whilst it is common to use the variance and
 94 autocorrelation of a timeseries as CSD indicators, they can result in a false positives or negatives when the statistics of the
 95 external conditions change. The restoring rate λ as introduced by B21 is estimated under the assumption of non-stationary
 96 correlated noise driving the system, and thus avoids biases induced by non-stationary noise (see Methods for details). When a
 97 system is in a stable equilibrium state, λ is negative, and it increases toward 0 from below as a critical transition is approached.
 98 In addition to λ , we also calculate the variance and autocorrelation of the time series.

99 Since we are dealing with a large number of time series, statistically λ is likely to increase in some cases even without
 100 any underlying CSD; this is the so-called multiple comparisons issue. We approach this problem from two different angles.
 101 The first is that in the case of true AMOC CSD, it is likely that both streamfunction strengths at the two latitude bands, and
 102 possibly the AMOC SST Index, will have an increase in λ , as well as in the other two classical CSD indicators (variance and
 103 autocorrelation). This will be discussed later in this section. The second is statistical significance, which is tested individually
 104 for each time series: in particular, for an AMOC time series of given variance and autocorrelation function, how likely is it
 105 to have the given increase in λ ? We generate 1000 Fourier surrogates for each AMOC time series, and use those to calculate
 106 a p-value for the slope of λ in each time series. Whilst we define $p \leq 0.05$ as a significant signal, we also use the p-values
 107 themselves to compare the models. Note that we calculate the surrogates from the original AMOC time series and not the CSD
 108 indicator time series, which is a more conservative approach than followed in B21 (see Ben-Yami et al. (2023) for details).

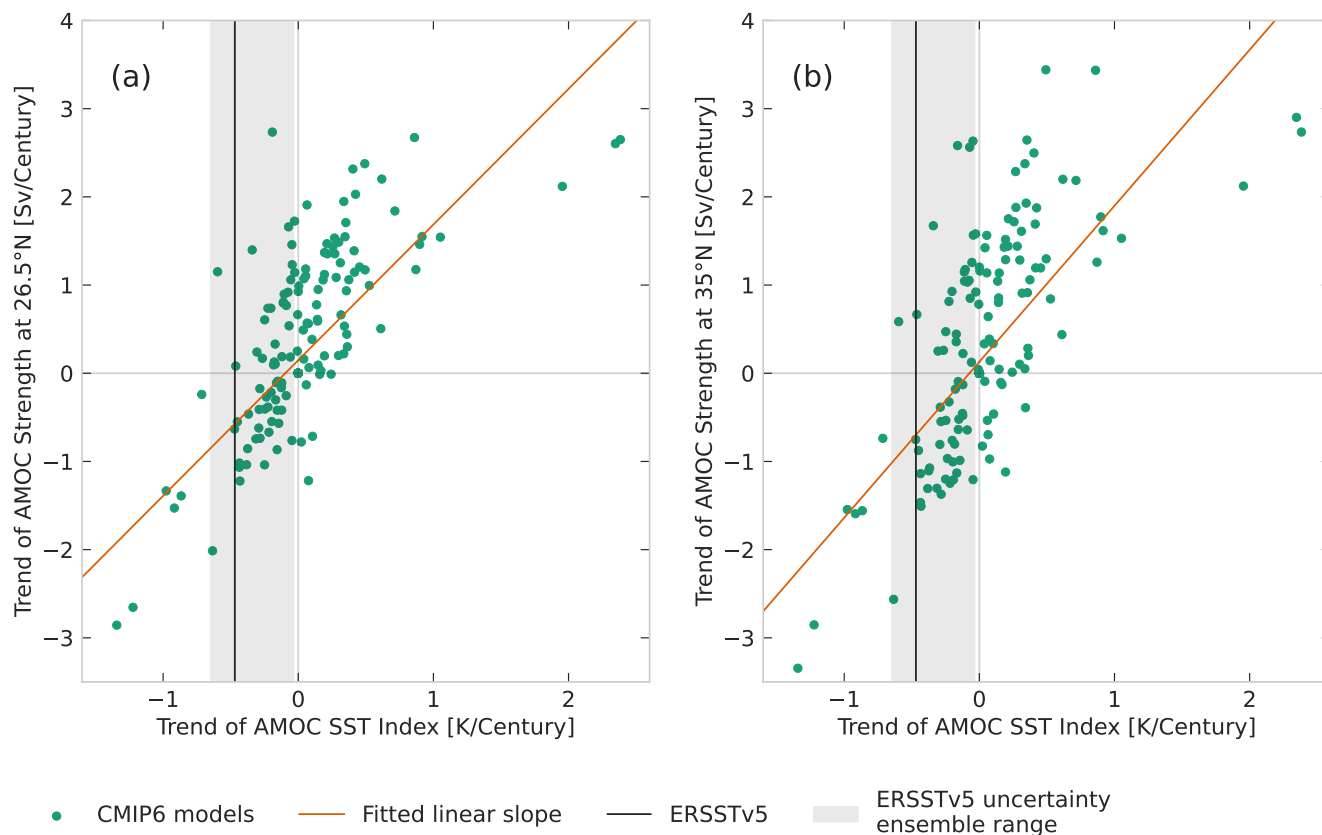


Figure 2. slope of AMOC streamfunction strength at 26.5°N (a) and 35°N (b) vs. the slope of the AMOC SST Index. All individual ensemble member slopes of 27 CMIP6 models are shown using the same turquoise dot to emphasize the correlation. The slopes of the linear regressions (orange lines) are 1.54 and 1.77 SvK⁻¹ for 26.5°N and 35°N, respectively. The black line is the slope in the ERSSTv5-derived observational AMOC SST Index, with the uncertainty range from the dataset’s 1000 member uncertainty ensemble.

109 Out of the 399 time series for both streamfunction strengths and SST Index for the 27 models, 36 have an increasing slope
 110 in the restoring rate λ with p-value below the 0.05 significance level (Fig. 4). However, only 4 ensemble members have a
 111 significant λ increase in both streamfunction strengths (CanESM5 r7, MIROC6 r8 and CESM2 r3 and r8, see Fig. 4). There
 112 is only one case in which both the SST Index and a streamfunction strength have coinciding significant positive λ slopes
 113 (CanESM5 r10).

114 In addition, if the significant λ increases were due to AMOC CSD, we might expect to see them consistently across all
 115 ensemble members in a model. The result here is again ambiguous: the significant increases are definitely concentrated amongst
 116 some models, with 11 models (96 ensemble members) having none at all. For example, at one end, IPSL-CM6A-LR has 10
 117 ensemble members (30 time series) but no significant λ increase, whilst at the other end, CanESM5 has 10 ensemble members
 118 and 8 significant λ increases (about a fourth of its 30 time series). The fraction of significant λ increases for each model are

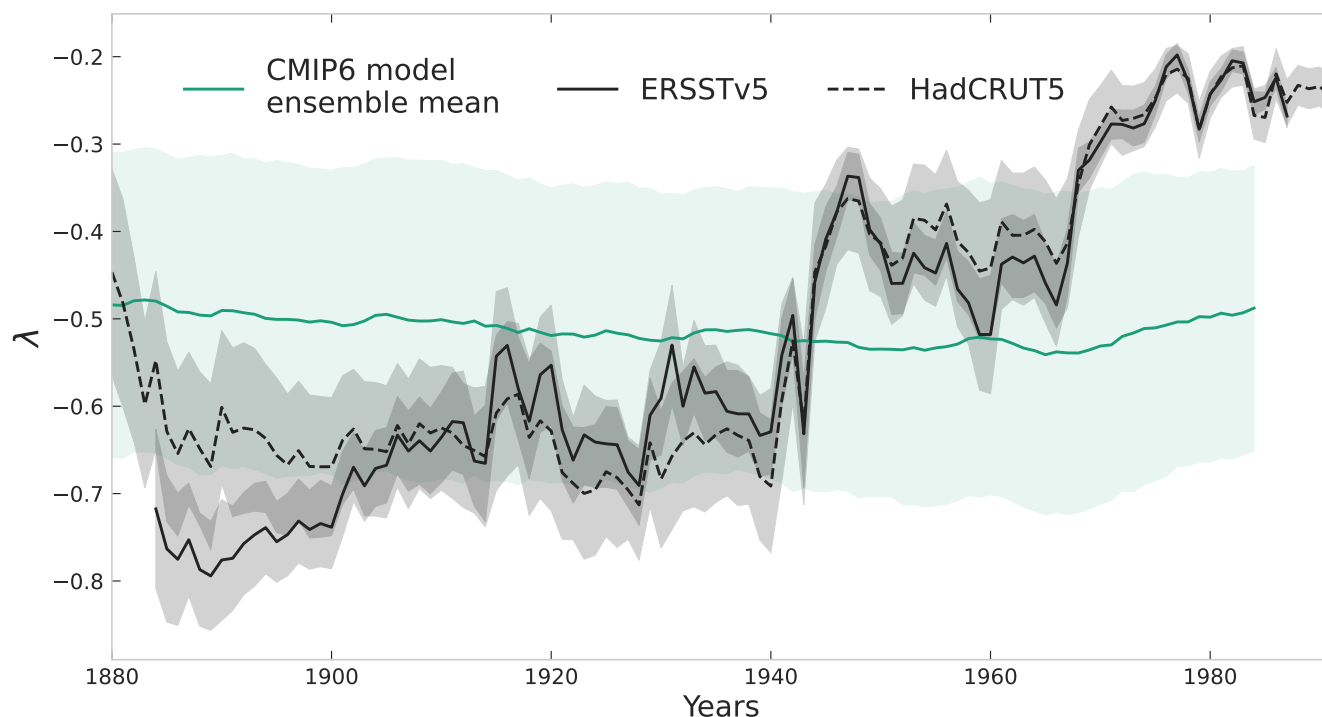


Figure 3. Recovery rate λ in the observed and modelled SST Index. The mean of the λ time series from CMIP6 models (turquoise line) is calculated based on the “1 ensemble member, 1 vote” principle. The one standard deviation of the ensemble is also shown (light turquoise shading). The λ timeseries of the SST Index calculated from the SST observational datasets ERSSTv5 (solid black) and HadCRUT5 (dashed black) are also shown with their respective uncertainty standard deviations from Ben-Yami et al. (2023).

119 shown in Figure C2 in the SI. Similarly, when we look beyond the 0.05 hard limit and look at the value of p , we can see that
 120 some models have much smaller p -values than others (Fig. C3).

121 The next step in testing the validity of a CSD signal is to check the agreement between different CSD indicators. We also
 122 calculate the variance and autocorrelation of the time series, and check for their significant increases following the same
 123 methods. Out of the 399 time series, there are 73 time series in total with any sort of significant CSD indication (Fig. 5). Of
 124 these, 47 have only one significant indicator, 20 have two, and 6 have all three (Table B1). These last six all occur in different
 125 models and all except one occur in the streamfunction at 35°N . Again, there is an uneven distribution of these indicators among
 126 the models (see Fig. C4 for the fraction of significance per model).

127 Because our definition of statistical significance is a p -value below 0.05, even if all time series are completely random
 128 we would still expect about 5% that fit our criteria for statistical significance. Thus, we need to take that into account and
 129 calculate probabilities if we want to know whether our results are random or not (see Methods for details). The different CSD
 130 indicators and the AMOC strengths and index are interdependent (Figs. C5, 6d, C6d and C7d), and thus we consider each
 131 indicator-strength combination separately when calculating probabilities.



132 When the results for all models are considered together, the number of significant increases in all three CSD indicators is
133 significant for the two streamfunction strengths. However, the results for the AMOC SST Index are not significant, and in fact
134 the probability is only 10-15% because the number of significant ASSTI time series is less than the expected number, not more.
135 When we consider the probability for each model individually, there are only five model indicator-strength combinations that
136 are below the 0.05 probability level: CanESM5 λ S35, CESM2 λ S35, CESM2 AR1 S26, HadGEM3-GC31-LL AR1 S35 and
137 MIROC6 AR1 S35 (Table B2). Note however that these probabilities are highly conservative (see Methods).

138 However, even the models with the statistically unlikely result still have ensemble members with an inconclusive or nonex-
139 istent signal, so it is uncertain if the fact that the result is unlikely to happen by chance is a definite sign of CSD. If the results
140 were more consistent within models, we could try to formulate necessary criteria for CSD. But to match the observations, we
141 would need to see a significant increase in all three of the CSD indicators in the AMOC SST Index, something which is not
142 present in any of the models' ensemble members. If we can interpret the behavior of the AMOC in historical CMIP6 simu-
143 lations as a conservative null model (since no model has an AMOC collapse before 2100), our results suggest that the CSD
144 detected in observation-based AMOC fingerprints (Boers, 2021; Ben-Yami et al., 2023) would be highly unlikely to occur by
145 chance.

146 Whilst the list of time series with a significant λ slope changes slightly when the CSD analysis parameters are changed (see
147 Figures C8, C9, C10, C11 and Tables C2 and C3), the statistics and inconclusiveness of the results remain the same.

148 2.3 Correlation of CSD indication with other parameters

149 Finally, we will try to make some sense of the chaotic list of ensemble members and indicators by looking for correlations
150 between the p-values of the different CSD indicator slopes and other properties and observables. This could, for example,
151 explain why only a few ensemble members in each model show CSD, since we have seen the AMOC slope also varies between
152 ensemble members. We compare the p-values with other p-values, with the slope of the historical AMOC, and with the mean
153 AMOC strength.

154 Unfortunately, there is almost no correlation between p-values of λ -slopes and these properties (Fig. 6). The same is also
155 true for the AR1 and variance (Figs. C6 and C7). The only exception is the correlation between the p-values of the two
156 streamfunction strengths, which is strongest for the variance but also present for the AR1 and λ slopes (Figs. 6d, C6d and
157 C7d). This correlation is also present for the ensemble means (Fig. C12). This is not very surprising, as the streamfunction
158 strengths at different latitudes are correlated to some extent. For the ensemble means we also find a negative correlation between
159 each streamfunction strength p-value and the corresponding mean strength (Fig. C12g-i). However, this correlation disappears
160 when two outliers with very low mean streamfunction strength are removed.

161 3 Discussion and Conclusions

162 We have analysed CSD indicators in time series representing AMOC strength at different latitudes as well as an SST-based
163 AMOC fingerprint in the historical runs of CMIP6. We have chosen to focus on the historical time period so that we can

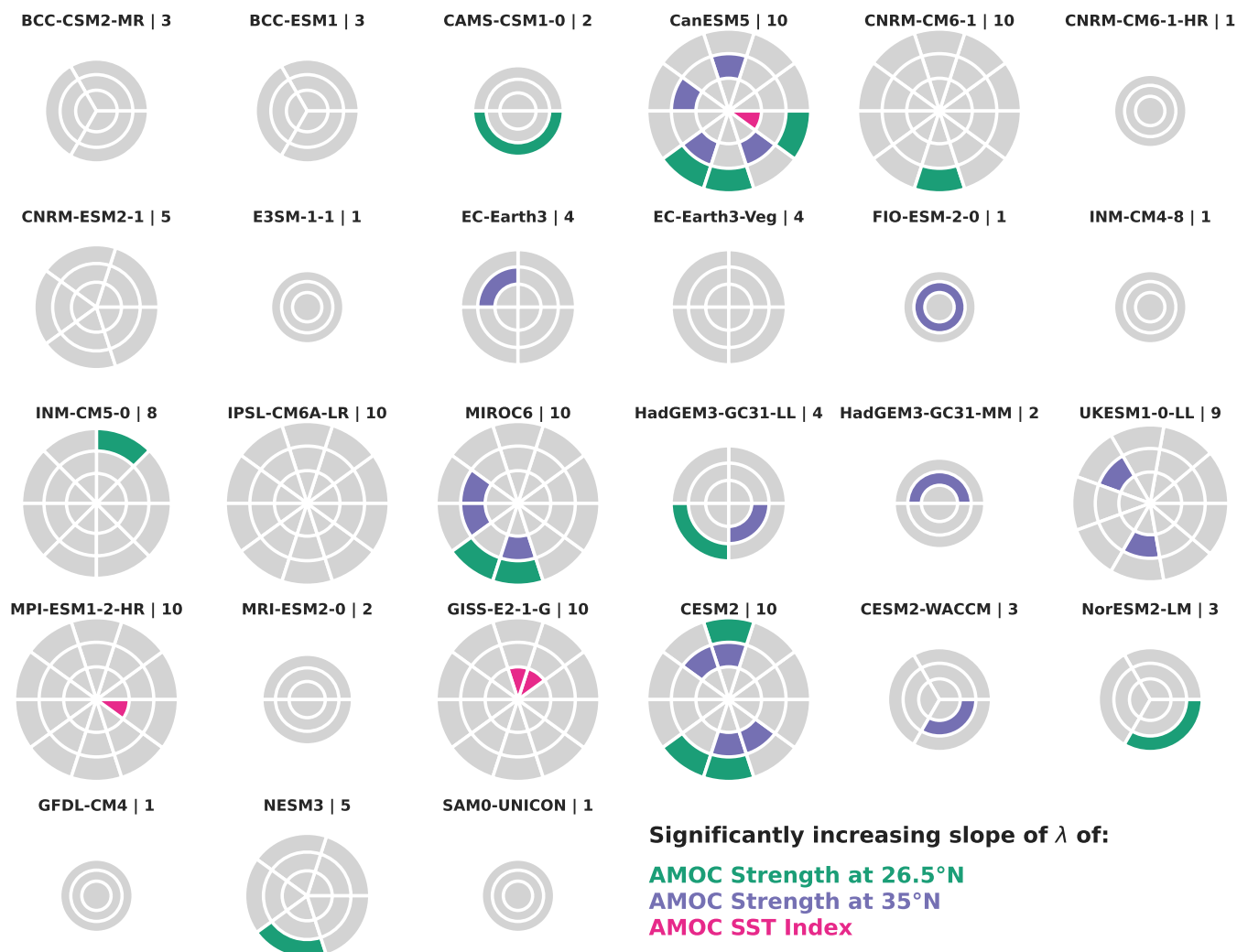


Figure 4. Overview of ensemble members and the occurrence of significant increases in λ in AMOC strength and the SST fingerprint. Each model is represented by a circle which is cut into a number of slices corresponding to the ensemble members analysed in this study. These slices are also cut into three layers, each representing one of either the AMOC streamfunction strength time series or the AMOC SST Index. So a model with 10 ensemble members will have $10 \times 3 = 30$ slice layers in the circle corresponding to AMOC time series. For each time series, if its λ is significantly increasing at 0.05 confidence level, the corresponding slice layer is coloured, otherwise it is left grey. The order from outside in is: S26.5N (turquoise), S35N (purple) and AMOC SST index (pink). The name of the model and the number of ensemble members are printed above each circle. The first ensemble member (r1) is always the one above the $+90^\circ$ line from the top (3 o'clock), and the ensemble members then proceed counterclockwise. For example, for CanESM5 the SST Index (pink) has a significant slope in λ in member r10, and the S26.5 index in members r7, r8 and r10. The corresponding list can be found in Table C4.

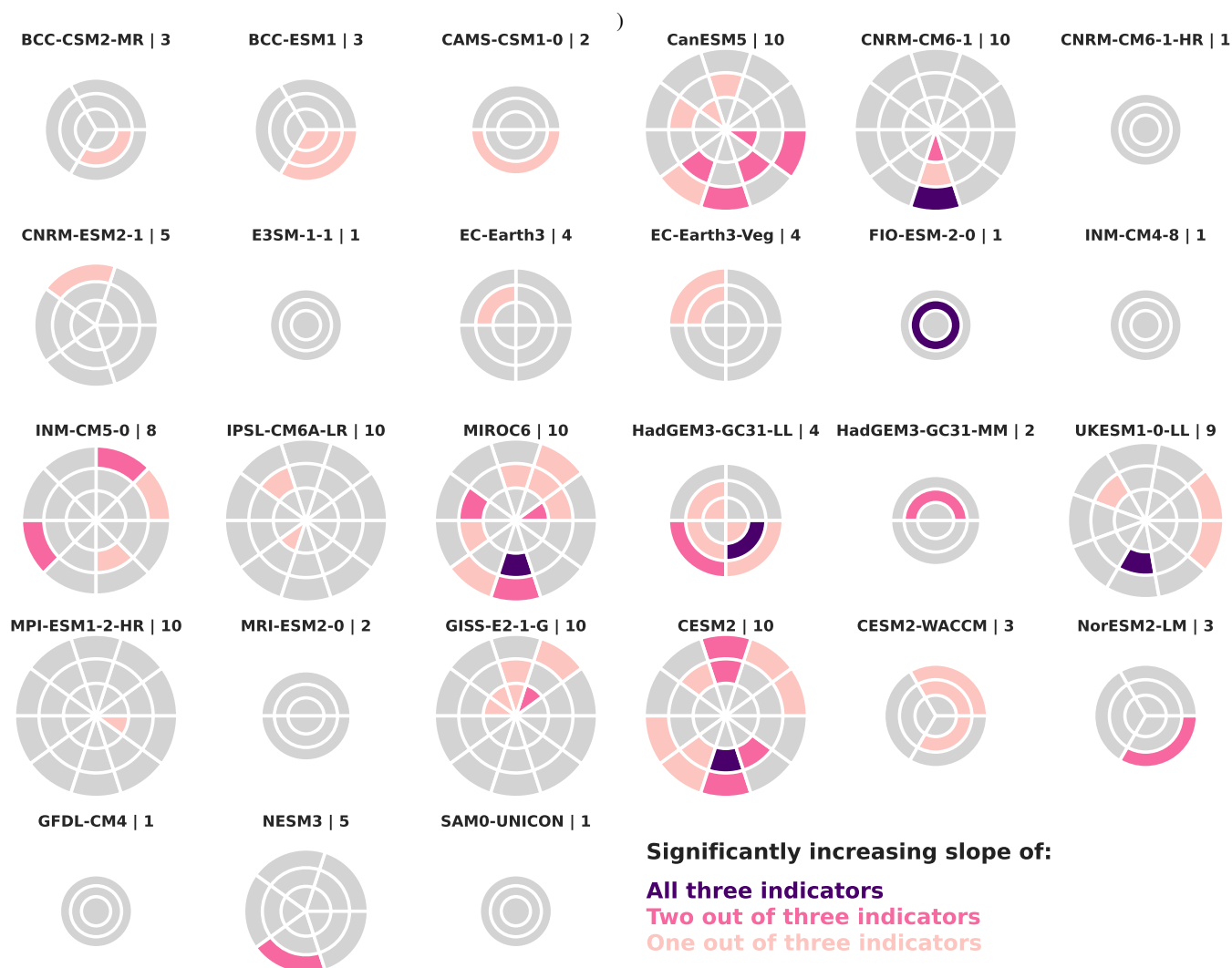


Figure 5. Number of indicators with a significantly increasing slope in the ensemble members' AMOC strengths and SST fingerprints. The structure of the figure is the same as Figure 4, but instead of representing a significant increase in λ , the colours now represent different numbers of indicators with a significant increase. These are all three (dark purple), two out of three (pink) and just one (light pink).

164 compare our results to the observed AMOC SST Index, where there is a significant increase in the three main CSD indicators
 165 (Ben-Yami et al., 2023; Boers, 2021). With regards to this comparison, the results are conclusive: no model agrees with the
 166 observed indication of CSD in the SST Index (Figs. 3, 5). There are only four significant increases in λ in the SST Index (out
 167 of 133 time series), and similarly the number of significant increases in the AR1 and variance are also not unlikely (Table B3).

168 The results are less conclusive for CSD in the streamfunction strengths. We should first note that none of the models used
 169 in this study have an abrupt collapse of the AMOC before 2100 in any of the future warming scenarios (Weijer et al., 2020).

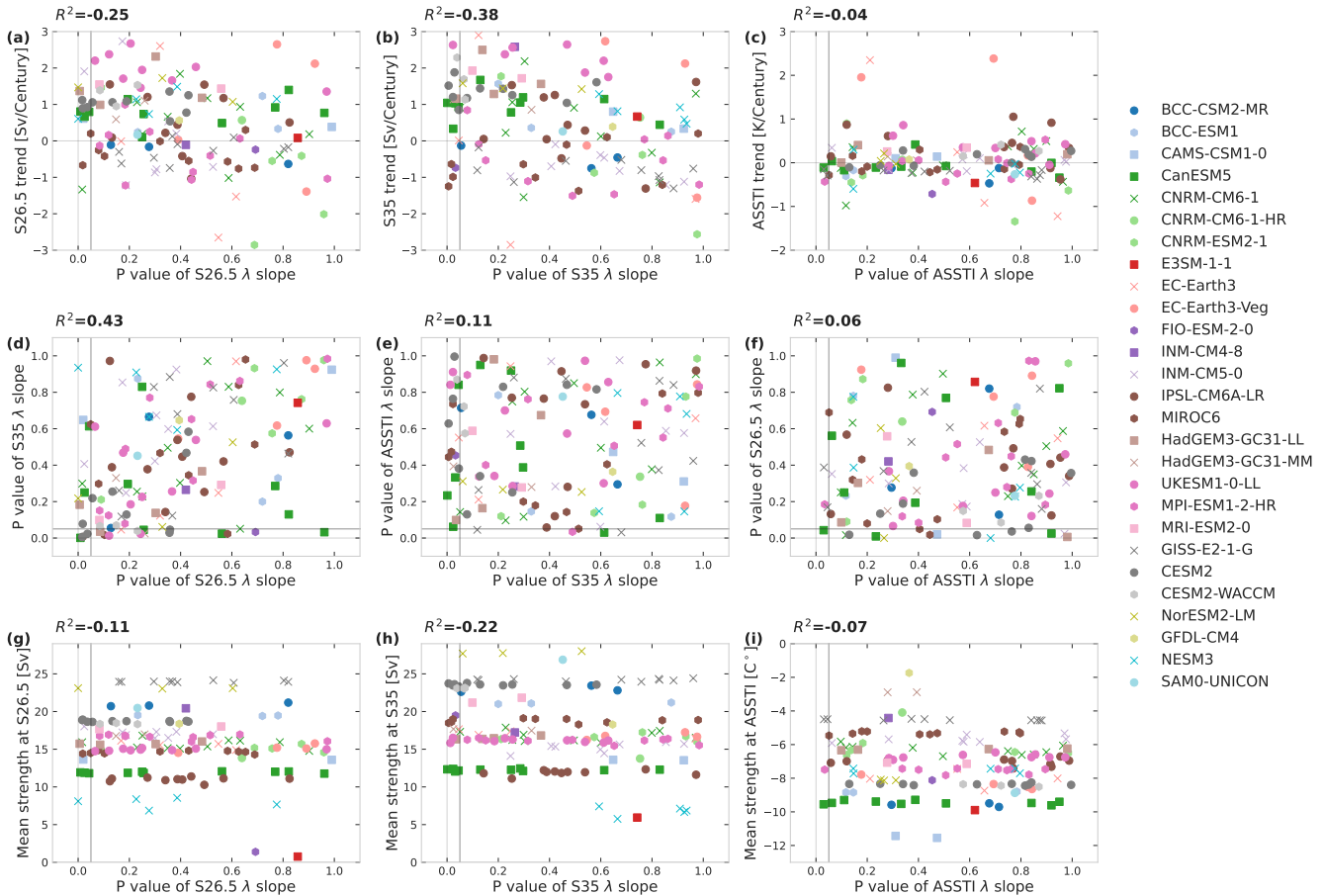


Figure 6. Correlations between the p-values of λ slopes and other quantities. S26.5, S35 and ASSTI stand for the AMOC streamfunction strengths at 26.5°N and 35°N and the AMOC SST Index, respectively. The x-axis is always the p-values of the slope of λ calculated from S26.5 (a,d,g), S35 (b,e,h) or ASSTI (c,f,i). This is plotted against the slope of the AMOC time series itself (a-c), the p-values of the slope of λ calculated from another of the time series (d-f) and the mean strength of the time series (g-i). All panels show the individual ensemble members of 27 CMIP6 models. For each model all ensemble member slopes are shown using the same coloured symbol. Where appropriate darker grey lines mark the 0.05 significance region.

170 Thus, even if the AMOC does collapse after 2100 (see Romanou et al.), we may still not be able to detect CSD 200 or more
 171 years in advance. We therefore would not necessarily expect to see CSD in the streamfunction strength time series, even if the
 172 AMOC in a given model is bistable and ultimately can undergo a warming-induced abrupt transition.

173 However, in some models we do see signs of CSD in the historical period, even if the results are not consistent enough to
 174 manifest a conclusive signal. Using a conservative probability calculation, we find four models that have a number of significant
 175 increases that are statistically significant: CanESM5, HadGEM3-GC31-LL, CESM2 and MIROC6 (Table B2). The latter two
 176 of these also have one ensemble member with what we would classically expect for an indication of CSD: significant increases



177 of all three CSD indicators in the streamfunction strength at 35°N, and an increase in two out of three CSD indicators in the
178 streamfunction strength at 26.5°N. Indeed, out of the six timeseries with significant increases in all three CSD indicators, five
179 are of the streamfunction strength at 35°N. If this is due to the AMOC truly destabilizing in these models, it would make sense
180 that the signal would be stronger in the more northern latitude, as the destabilization usually originates in the north Atlantic.

181 The situation is complex, but we can narrow our discussion to four major questions:

- 182 1. What is happening in the models with no indication of CSD?
- 183 2. What is happening in the models with a significant indication of CSD, but only in a fraction of the ensemble members
184 and time series?
- 185 3. Why is there no CSD in the modelled SST index, as opposed to the observed SST index?
- 186 4. What do our answers to questions (1-3) imply for the utility of the SST Index for detection of CSD?

187 Question (1) is the easiest to answer. We have identified four models where the signs of CSD are significant, but our
188 probability calculation is conservative, and for some models it is difficult to say if the indication of CSD is present or not.
189 But for around 20 out of the 27 models it seems like there's definitely no indication of CSD. The most straightforward reason
190 would be that there is simply no CSD occurring in these models, either because the abrupt transition is too far in the future,
191 or the AMOC is not destabilizing at all in these model runs. This does not mean that the models are correct - the AMOC is
192 notoriously hard to simulate in complex climate models (Jackson et al., 2023b), and the CMIP6 models may have AMOC
193 systems that are too stable (Valdes, 2011; Liu et al., 2017). In either case the models simply cannot help us confirm or reject
194 the validity of observed AMOC CSD.

195 Answering question (2) is much more difficult, because in those four models the signs of CSD are significant but not
196 conclusive. There are a few different possible explanations for this:

- 197 (a) In these models, the CSD is real only in the ensemble members that show a strong signal in the indicators. For example,
198 in a small number of ensemble members the phase of the internal variability is such that the AMOC is undergoing CSD,
199 and that is detected in the streamfunction strength. In the other ensemble members the AMOC is stable. This is unlikely,
200 since all ensemble members in a given model have the same forcing.
- 201 (b) In these models, the CSD is detected only in those ensemble members, but is real in all of the ensemble. As the AMOC
202 is a complex high-dimensional system, it is possible that the one-dimensional signals considered here cannot properly
203 capture the changes in its stability. Alternatively, the CSD signs could be masked by other, unrelated processes such as
204 effects of the atmospheric circulation. This could make it difficult to detect CSD even in the streamfunction strength time
205 series, and so only a small fraction of the time series would show a signal even if the AMOC is destabilizing in all the
206 members.
- 207 (c) There is no CSD in any of the model ensemble members. Since the signs of CSD are not conclusive, there is also the
208 possibility that something else is causing those signals in the models. The likelihood of so many signs of CSD in a



209 completely random system is very low, but it is possible that some underlying processes make it more likely to have
210 increases in variance, autocorrelation, etc. in those particular models. This would mean that our method of calculating
211 statistical significance lacks important priors.

212 There are two ways in which the above answers to question (2) would affect the AMOC SST Index:

213 – In cases (a) and (b), the indicator slopes in the streamfunction strength are at least a sufficient condition for CSD. Thus
214 if one of those cases is true, the SST Index would prove to be a bad measure of AMOC CSD, as it doesn't indicate any
215 CSD when the streamfunction strengths do. It is possible in that case that the SST Index is simply a weaker measure
216 influenced by too many other factors, not an invalid one - e.g. if the approaching collapse is after 2100, perhaps the signs
217 of CSD in the SST Index start to emerge later.

218 – Case (c) is in some ways more favourable for the utility of the SST Index. If the signs of CSD in the streamfunction
219 strengths are a false alarm, we would at least know that the SST index is not likely to show that particular type of false
220 positive.

221 It is not possible to discern which of the above cases is (part of) the correct answer. To do that would require knowing if the
222 AMOC is becoming less stable in these historical runs. Although there is no abrupt collapse in these models before 2100 in the
223 SSP scenarios, that is not a guarantee that it would not happen later. While one option is that the AMOC would collapse after
224 2100, it is also possible that the AMOC becomes unstable under 1850-2014 conditions, but then the specific SSP scenarios
225 avoid its collapse in the 21st century.

226 Whichever case is true, there is still one definite result for the AMOC SST Index. The streamfunction strengths are by far
227 the more direct measure of the AMOC. Thus, if we had seen a large number of significant increases in the SST Index but none
228 in the streamfunction strengths, we would have had to conclude that the SST Index has many spurious false positives. As it
229 is, in the CMIP6 models the SST Index follows the statistics we would expect from random time series. Although this does
230 not confirm that the SST Index would be able to detect AMOC CSD, it does show that it is not overly sensitive to other ocean
231 processes and could perhaps be quite a conservative fingerprint for measuring CSD. In addition, there is no indication of CSD
232 in the AMOC SST Index in the runs identified by Swingedouw et al. (2021) as having an abrupt collapse of the SPG in the
233 21st century (r1i1p1f1 from CESM2-WACCM, MRI-ESM2-0, NORESM2-LM and CESM2). This is an indication that using
234 the AMOC SST Index to detect CSD in the AMOC would not produce a false warning in case of a collapse of only the SPG.

235 Finally, it is not only uncertain how well the models represent the AMOC as a whole, but also how well they represent the
236 connection between its components (Jackson et al., 2023b). In particular, one of the reasons for using the AMOC SST Index
237 for the detection of CSD relies on a physical connection of those local SSTs to the stability of the AMOC. One of the main
238 identified restoring forces of SPG perturbations is the negative feedback cycle in which heat loss leads to increased deep water
239 formation, which in turns leads to a stronger circulation and more heat transport into the SPG (Menary et al., 2015; Sun et al.,
240 2021). This restoring force is intrinsically connected to the AMOC, and if the AMOC becomes less stable, it takes longer for
241 the ocean circulation transporting heat into the SPG to restore from perturbations, and thus the SSTs in that region exhibit CSD.
242 However, it is not certain that these physical mechanisms will be present in CMIP6 models. For example, the extent, location,



243 and even existence of deep water formation in the SPG differs between models (Jackson et al., 2023b). Conclusive statements
244 on the validity of the AMOC SST Index would thus require an in-depth analysis of the physical processes, which is beyond the
245 scope of this work.

246 Future work on AMOC CSD in complex climate models should focus on scenarios where the AMOC undergoes a warming-
247 induced abrupt transition. In those cases the utility of different indicators and fingerprints for CSD detection could be analysed
248 in detail, and would give important context to our results.

249 In conclusion, we have found that the historical CMIP6 experiments do not agree with the observational records on the
250 indication of CSD in the AMOC SST Index. There are some models that show significant signs of CSD in the AMOC stream-
251 function strength, but those signs are not consistent enough to make for a conclusive result. Whilst our results cannot confirm
252 or deny the utility of the AMOC SST Index for detection of AMOC CSD, they do indicate that the index is not prone to false
253 positives, which strengthens the significance of the observational AMOC CSD.

254 *Code availability.* All code used to analyse the data and generate figures will be uploaded at <https://github.com/mayaby>.

255 *Data availability.* All CMIP6 data used in this study is available online on the ESGF database (<https://aims2.llnl.gov>). The HadCRUT5
256 dataset is available at <https://www.metoffice.gov.uk/hadobs/>. The ERSSTv5 operational dataset is available at
257 <https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html>, and the ERSSTv5 uncertainties are available at
258 <https://www.ncei.noaa.gov/pub/data/cmb/ersst/v5/ensemble.1854-2017/>. No new data has been produced.

259 **Appendix A: Critical slowing down indicators and significance calculation**

The restoring rate λ , the variance and the autocorrelation are calculated in the same manner as in B21. Each time series is first
non-linearly detrended using a running mean with a 50-year window. The edges are not removed, so the detrending method is
less certain at the first and last 25 years of the time series. The CSD indicators are then calculated in 60-year running windows.
The restoring rate of the underlying continuous dynamics is defined by the equation

$$dX_t = \lambda' X_t dt + \sigma dW_t,$$

260 which results from linearizing the dynamics around a given stable equilibrium, where dW_t denotes the increments of a
261 Wiener process. Note that as in B21 the λ plotted in this study is the numerical result of the regression of the discrete increments
262 $\Delta X_t = X_{t+1} - X_t$ against X_t , and so is related to the analytical, continuous λ' by $\lambda = e^{\lambda'} - 1$ (when the timestep $\Delta t = 1$). As
263 the magnitude of λ is immaterial in this study and we are only concerned with its increase or decrease, both definitions behave
264 similarly and are thus interchangeable for our purposes. In order to test the statistical significance of the linear slopes of CSD
265 indicators, Fourier surrogates are created from the detrended time series. These are calculated by taking the discrete Fourier
266 transform of the time series, multiplying by random phases and then taking the inverse Fourier transform. By the Wiener-



267 Khinchin theorem, the variance and autocorrelation function of wide-sense-stationary random processes are specified by the
268 squared amplitudes of the (discrete) Fourier transform. Thus the Fourier surrogates preserve the variance and autocorrelation
269 function of the original time series. For the calculations in the main text that use a 60 year window, 1000 Fourier surrogates are
270 generated for each time series. For the checks on the 50 and 70 year window lengths, 100 Fourier time surrogates are generated
271 for each time series for reasons of computational expense.

272 **Appendix B: Conservative probability calculation**

273 Since we use $p \leq 0.05$ as the condition for significance for each individual time series, for each time series there is a 0.05 chance
274 that it will be significant. The probability for significance in a number of time series then follows a binomial distribution, for
275 which the probability mass function is:

$$276 \Pr_b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (\text{B1})$$

277 This is the probability that the event with probability p happens k times out of the n time series.

278 Out of the 1197 AMOC time series analysed in this study, 105 show a significant increase in a CSD indicator. If we regard
279 both the three CSD indicators and the three different AMOC strengths and fingerprints as independent, the probability of this
280 happening is $\Pr_b(105; 1197, 0.05) = 1.55e - 08$. However, neither the CSD indicators nor the AMOC strengths can be viewed
281 as independent (see Figs. C5, 6d, C6d and C7d). We thus have nine sets of 133 time series, each with their own number of
282 significant increases. For a set of 133 time series with m significant increases, we can use the binomial distribution to calculate
283 the probability p' of m significances out of 133 as if the other sets don't exist: $p' = \Pr_b(m; 133, 0.05)$. In reality, the probability
284 is likely slightly higher than p' because we have nine sets of time series, but these sets are dependent to an unknown extent, so
285 we cannot view the nine sets of time series as multiple experiments. The probability of the result for each strength-indicator or
286 index-indicator combination viewed alone are shown in Table B3.

287 We are also interested in the likelihood of the result for individual models. A model like CESM2 with 17 out of 90 significant
288 time series looks like it would be unlikely, but it is important to quantify this to some extent. Here again we consider the CSD
289 indicators and AMOC strengths and index separately, as there is no way to quantify their dependence. For a given CSD
290 indicator and AMOC strength we can then first calculate the probability for the model result considered alone. For model and
291 indicator-strength combination i this is $p'_i = \Pr_b(l_i; n_i, 0.05)$, where l_i is the number of significant increases out of n_i time
292 series. However, as noted before, an event that would be unlikely if we only had one model could be likely to happen at least
293 once for 27 models. We thus need to consider how likely it is to have at least one event of probability p'_i out of 27 instances. We
294 can do this here because for a given CSD indicator and AMOC strength we assume that the model results are independent. This
295 is again the binomial distribution from equation B1, but now with $n = 27$ and $p = p'_i$. We are now interested in the likelihood
296 of at least one occurrence of a p'_i probability event, which is: $\Pr_b(k \geq 1; 27, p'_i) = 1 - \Pr_b(0; 27, p'_i)$. So combining this with the
297 equation for p'_i , we get:

$$298 \Pr_i(l_i) = 1 - \Pr_b(0; 27, \Pr_b(l_i; n_i, 0.05)) \quad (\text{B2})$$



299 This is the probability that out of 27 models, we have at least one set with a probability p'_i event (which for model i is l_i out
 300 of n_i significant time series). Of course since the value of l_i , n_i and p'_i will be different for the models, this is a conservative
 301 calculation that applies only to each model individually - if we have multiple models with low $\text{Pr}_i(l_i)$ the result in total is even
 302 less likely. The model indicator-strength combinations for which $\text{Pr}_i(l_i)$ is less than 0.05 are shown in Table B2.

303 These calculations assume that the CMIP6 models are completely independent from each other. This is not strictly true, and
 304 in particular some dependence in the AMOC time series may arise from multiple models using the same ocean component.
 305 However, in our results there is no correlation between the ensemble members with lower p-values and the ocean component
 306 (Fig. C13). In addition, the four models in Table B2 all use different ocean components. We can therefore assume that for our
 307 purposes the CMIP6 models are independent.

	S26	S35	ASSTI	All three
At least one	28	32	13	73
One	17	21	9	47
Two	10	6	4	20
Three	1	5	0	6

Table B1. Number of time series per streamfunction or SST Index with significant increases in the CSD indicators. Columns are the streamfunction strength latitude, AMOC SST Index and all three. Rows are the number of CSD indicators with a significant increase. So for example, the first cell from the top left counts the number of S26 timeseries that have at least one CSD indicator with a significant increase.

Model	λ S35	AR1 S35	AR1 S26
CanESM5	2.57e-02		
CESM2	2.57e-02		2.57e-02
HadGEM3-GC31-LL		1.27e-02	
MIROC6		2.57e-02	

Table B2. Probability of outcome amongst all models, for models where for a given CSD indicator and AMOC time series the outcome in CMIP6 is statistically unlikely to occur amongst 27 models (see methods).

	S26	S35	ASSTI
λ	4.17e-03	8.31e-05	1.04e-01
AR1	2.45e-04	6.76e-04	1.52e-01
Variance	9.19e-02	4.17e-03	1.59e-01

Table B3. Probability that a given set of time series has m_i out of 133 significant increases, where m_i is the number of significant increases in the indicator-index combination. This is calculated without considering the rest of the sets of time series.



308 Appendix C: Extended figures and tables

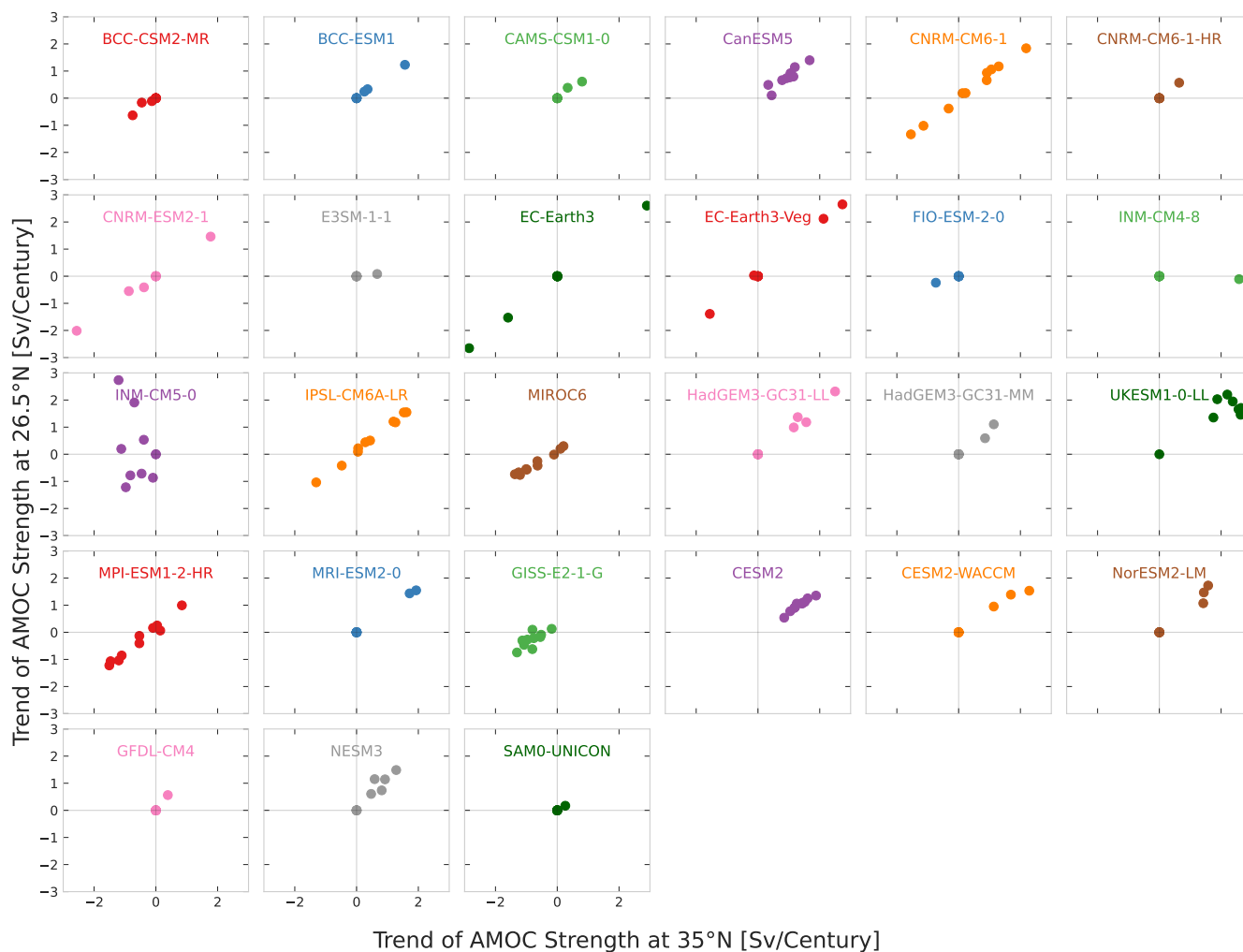


Figure C1. Same as Figure 1 but with the values for the ensemble members of each model plotted in a separate figure. Even though the aerosol forcing is the same within each model, there is still a large scatter of ensemble member slopes.

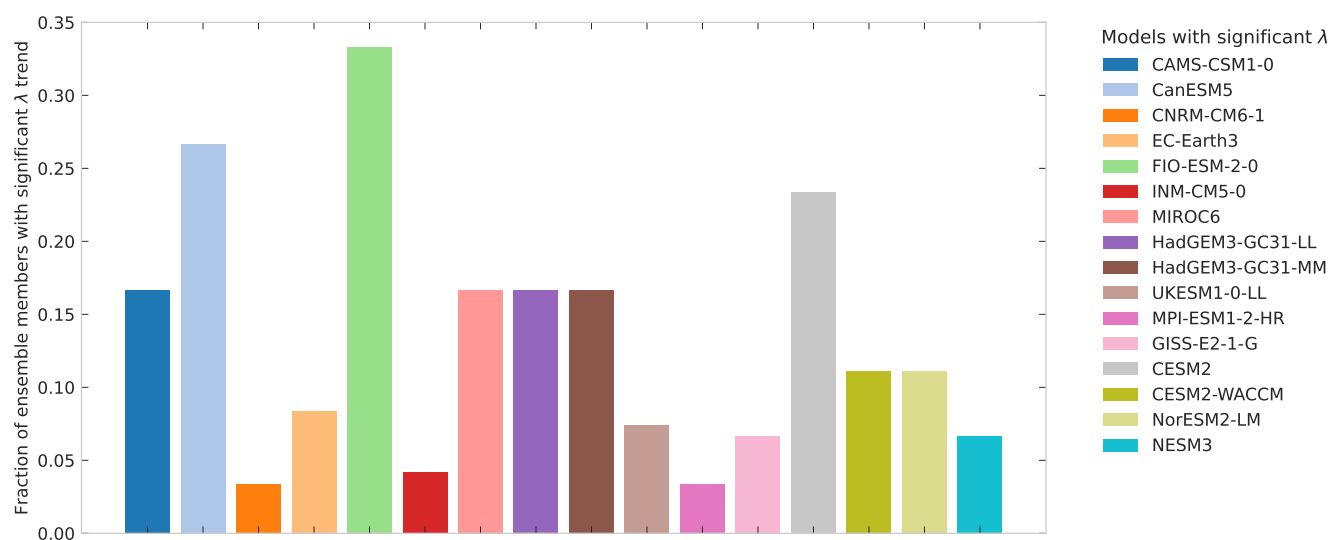


Figure C2. Fraction of all timeseries for a given model that have a significant increase in λ (so both in streamfunction strength and SST Index). Only the models that do have such significant increases are shown, and the full list of these can be found in Table C4. Note that the colours for the models are different from other figures.

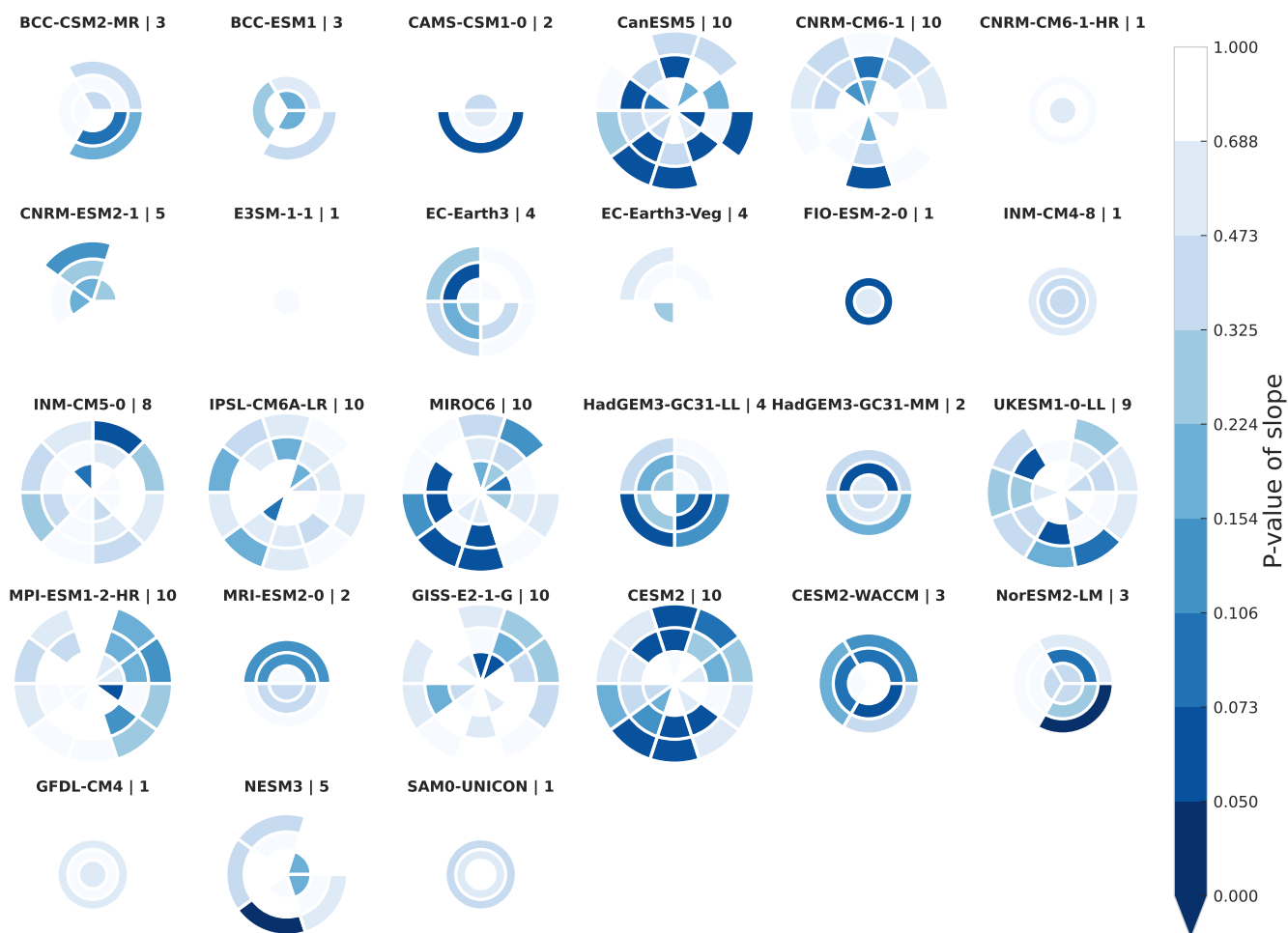


Figure C3. Same as Figure 4, but with the slices coloured according to the p-value of the corresponding λ . Darker blue is a smaller p-value and so more statistically significant.

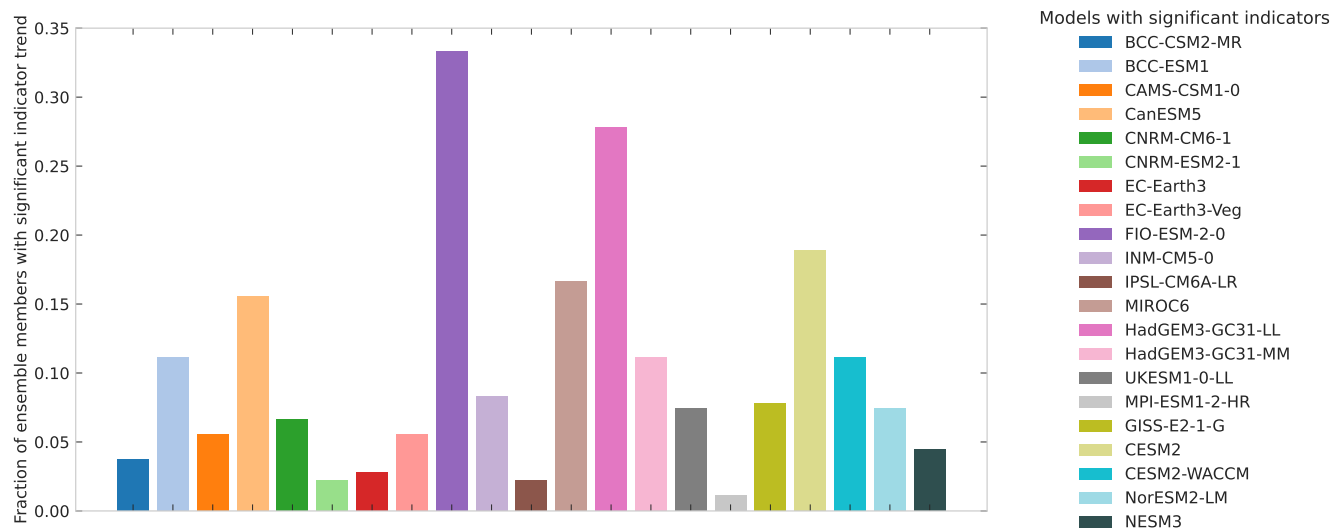


Figure C4. Fraction of timeseries for a given model that have a significant increase in any of the three CSD indicators. Only the models that do have such significant increases are shown, and the full list of these can be found in Tables C4, C5 and C6. Note that the colours for the models are different from other figures.

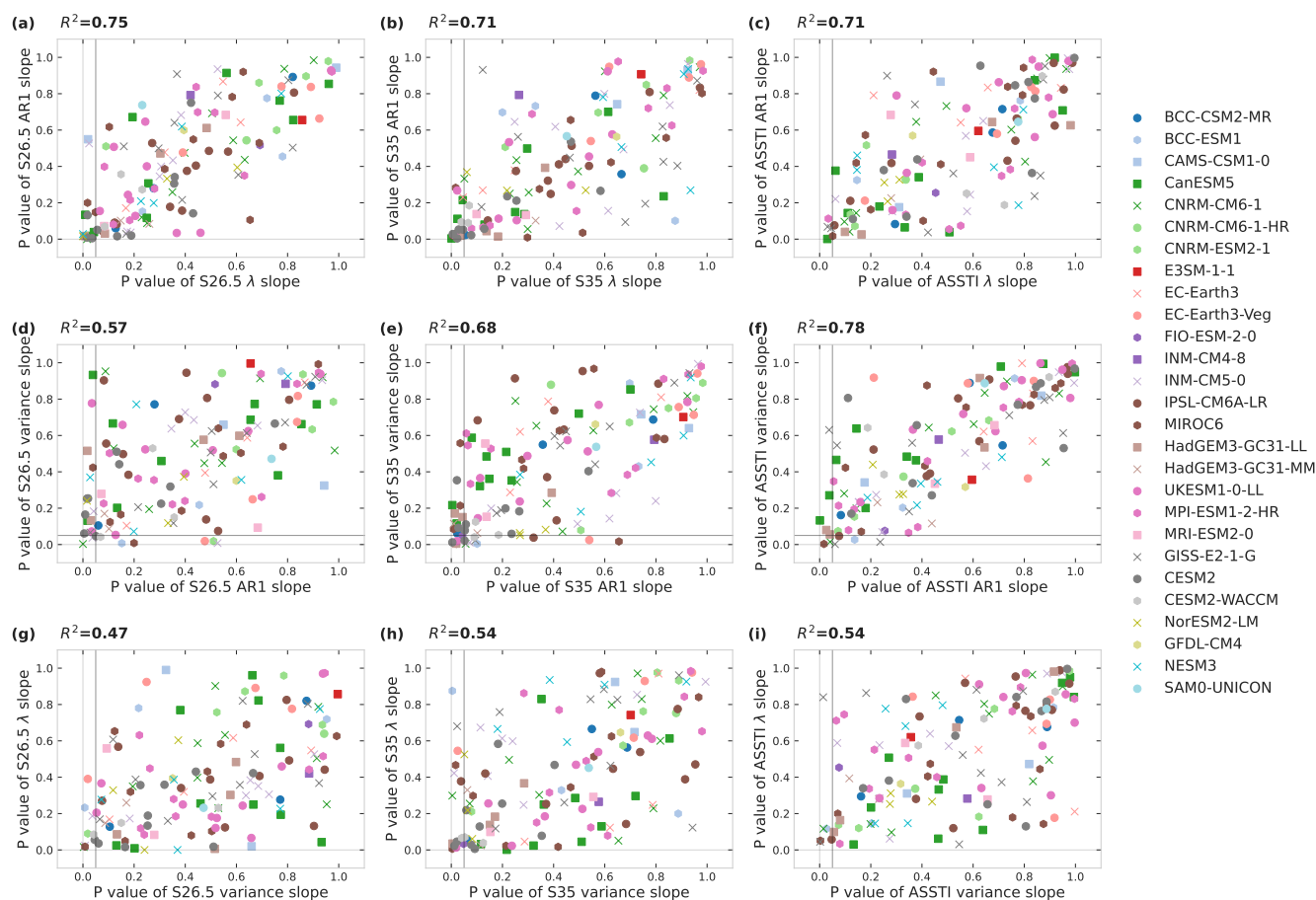


Figure C5. Same as Figure 6 but for the correlations between the p-values of different CSD indicators.

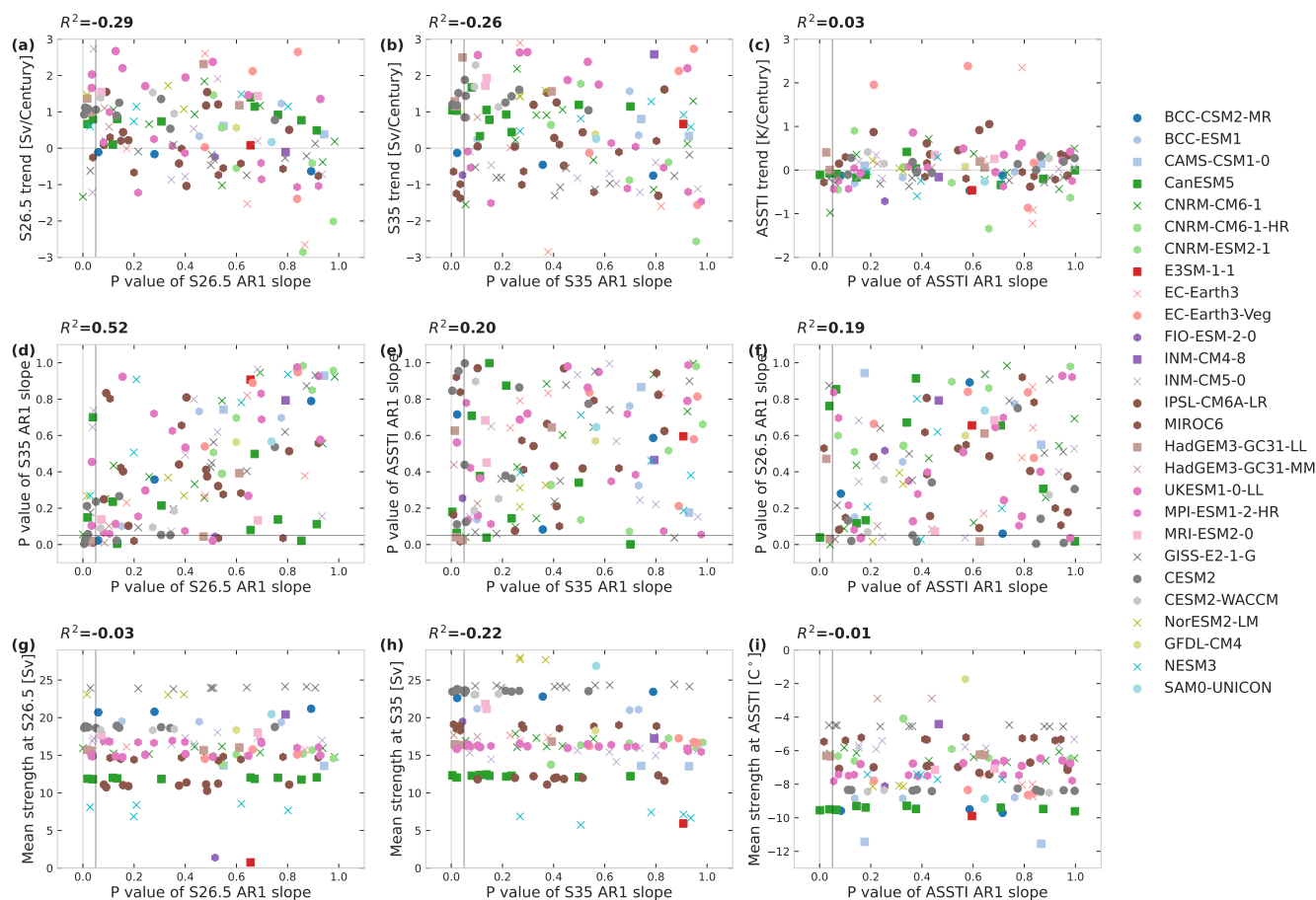


Figure C6. Same as Figure 6 for the AR1 instead of λ .

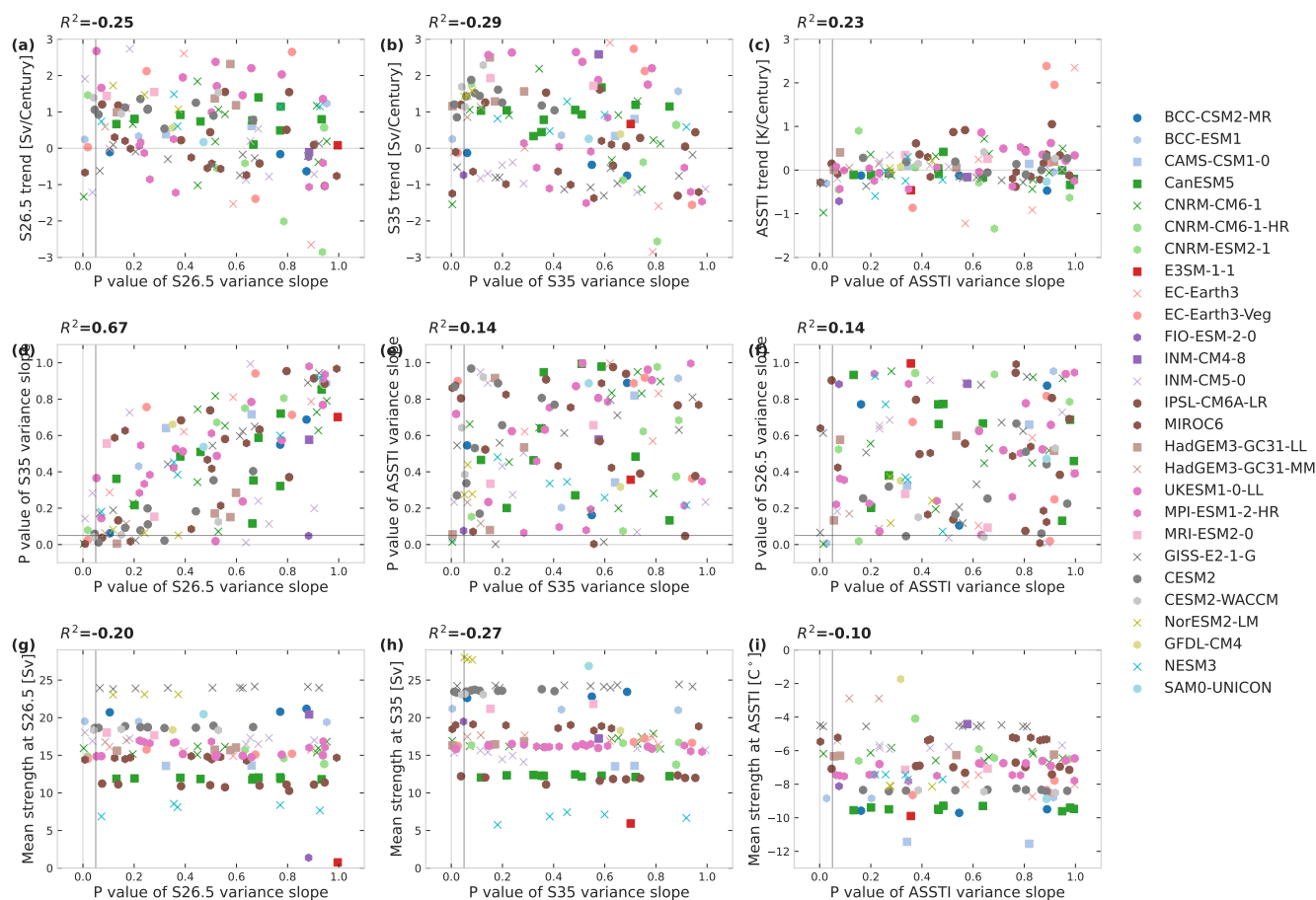


Figure C7. Same as Figure 6 for the variance instead of λ .

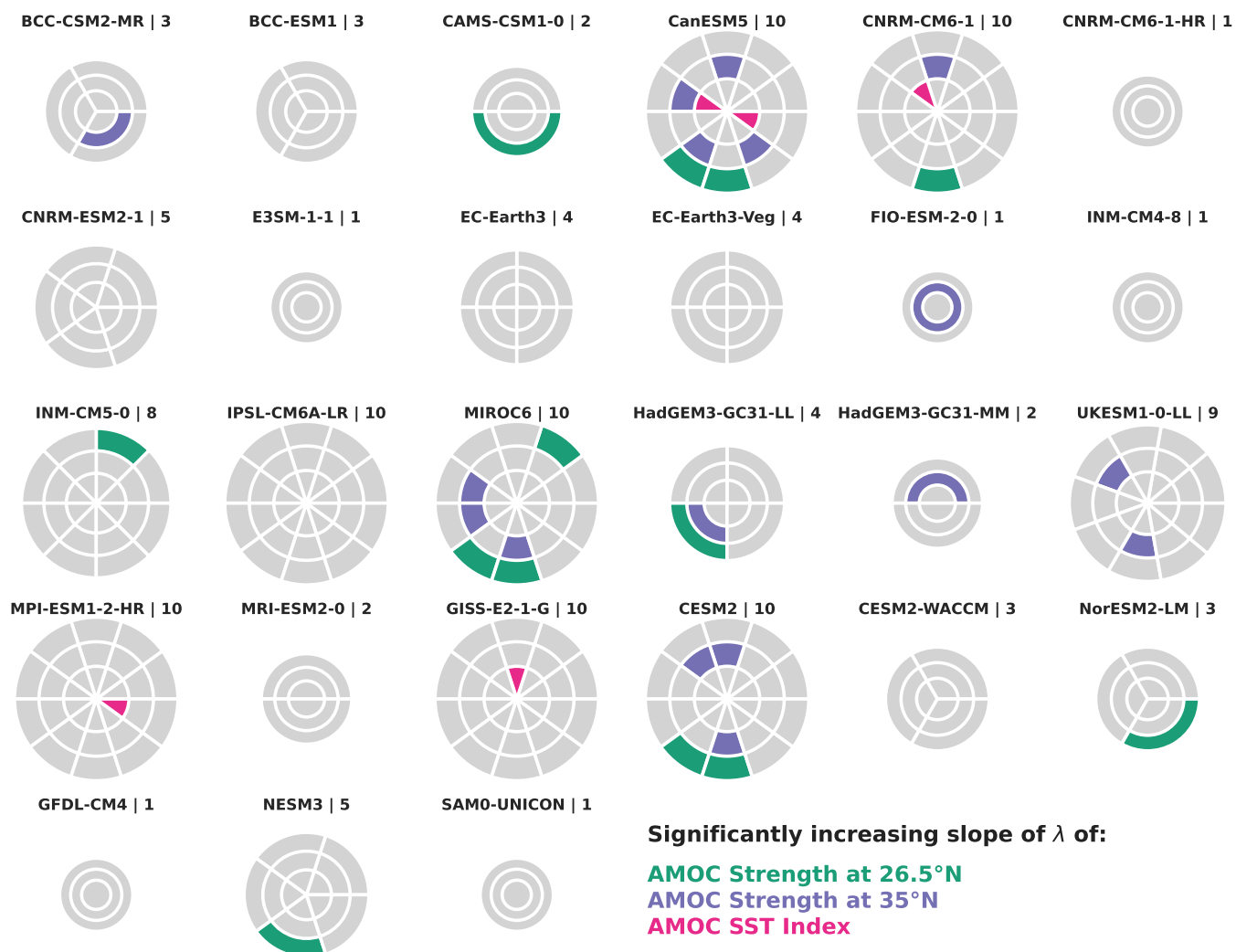


Figure C8. Same as Figure 4, but with a window size of 50 years.

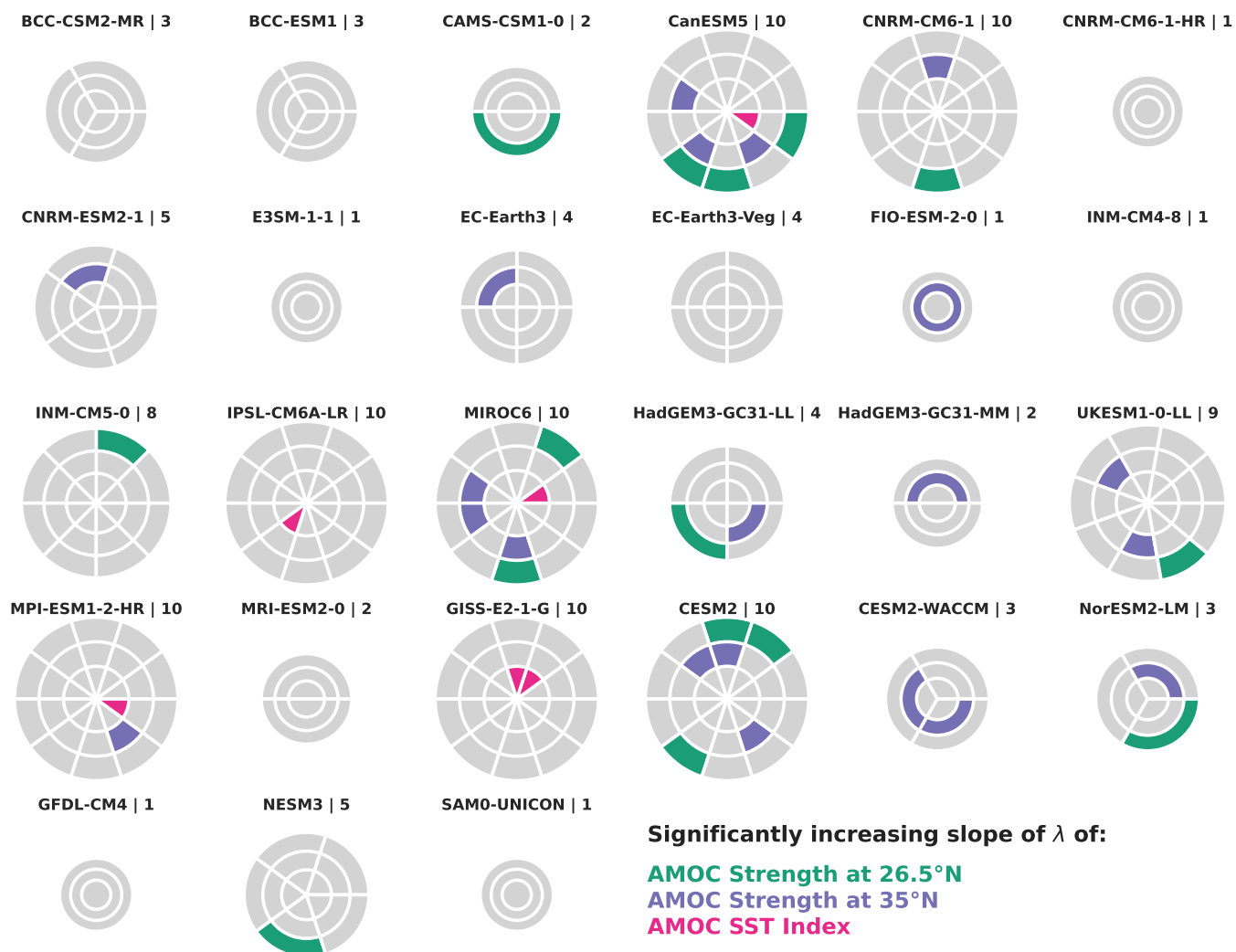


Figure C9. Same as Figure 4, but with a window size of 70 years.

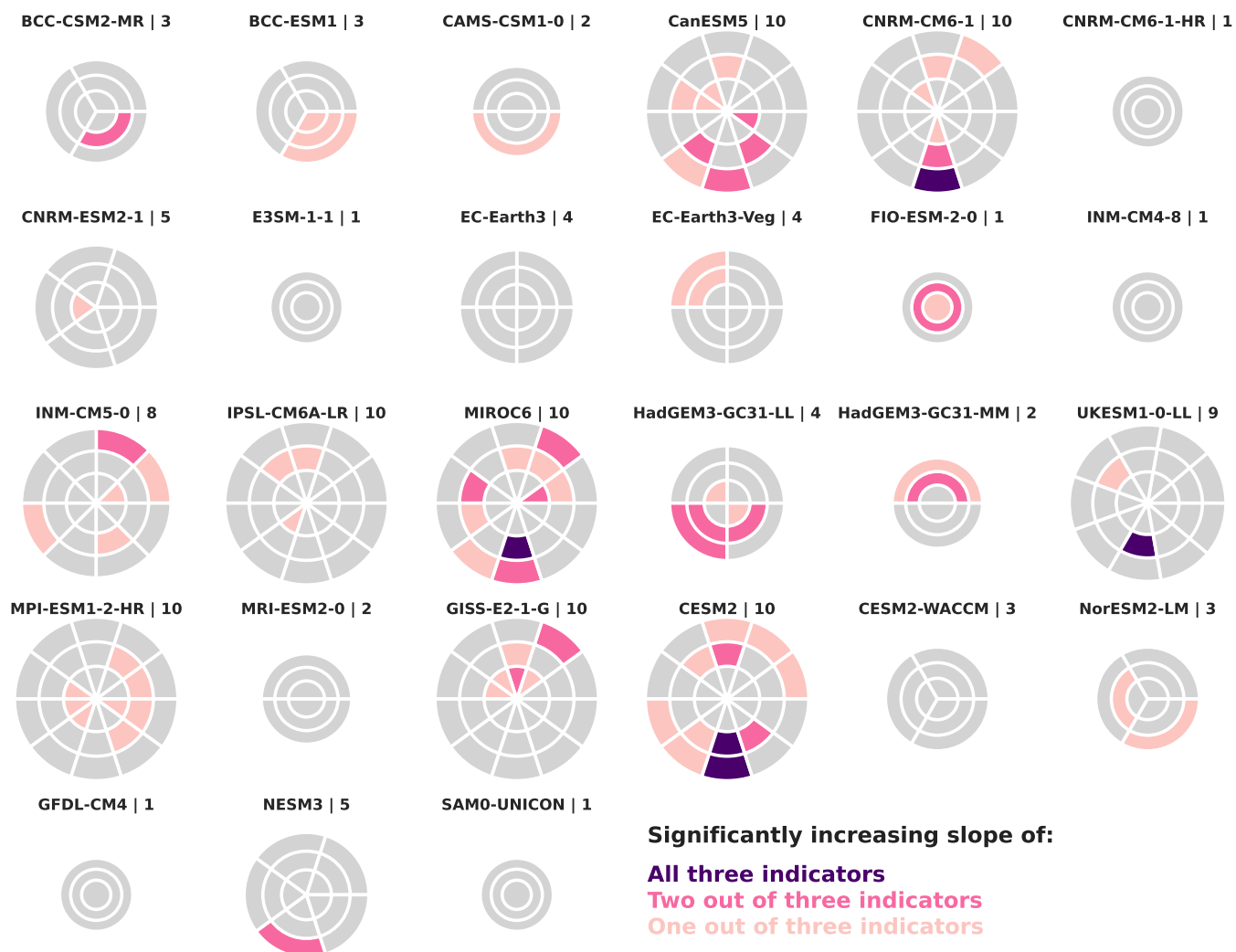


Figure C10. Same as Figure 5, but with a window size of 50 years.

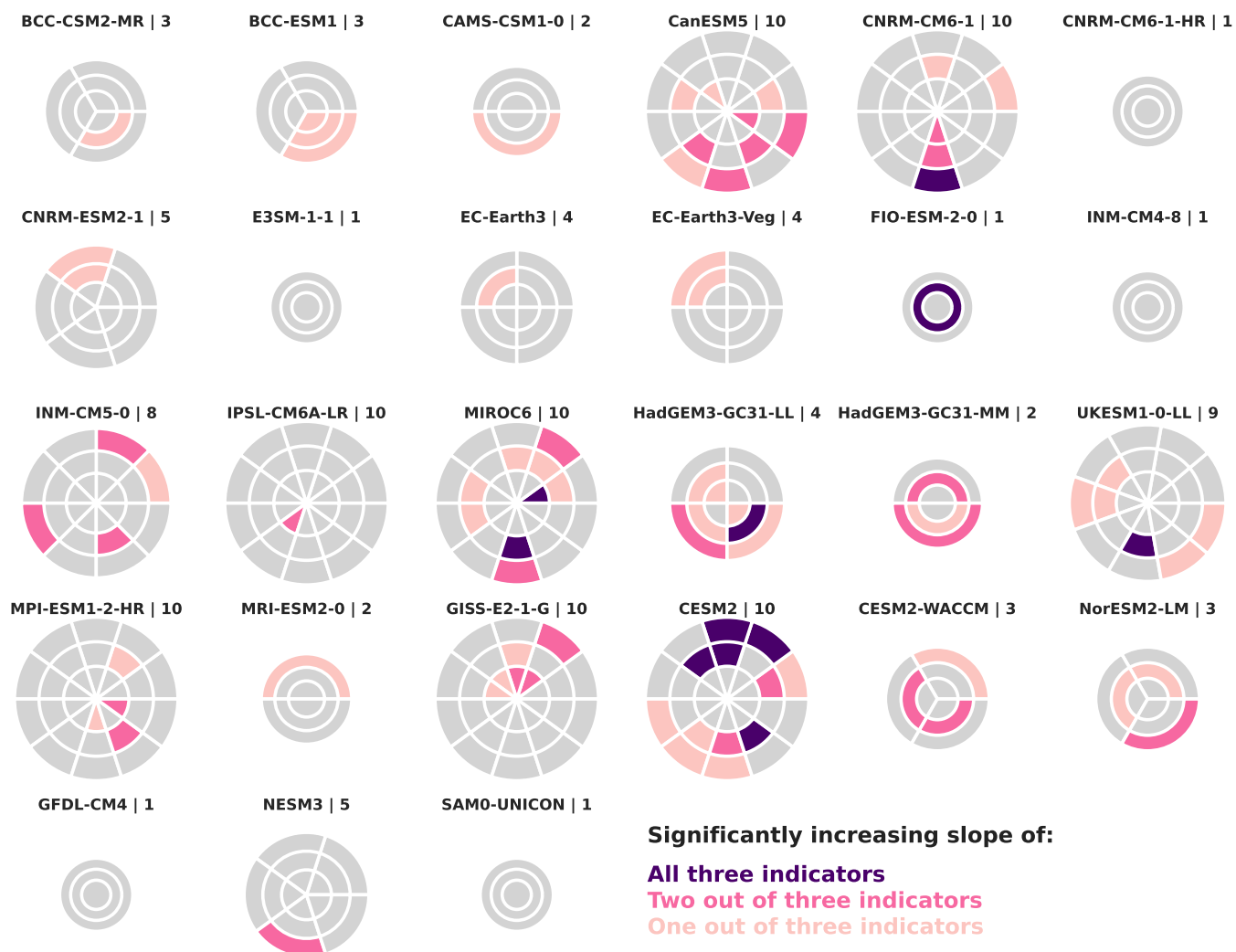


Figure C11. Same as Figure 5, but with a window size of 70 years.

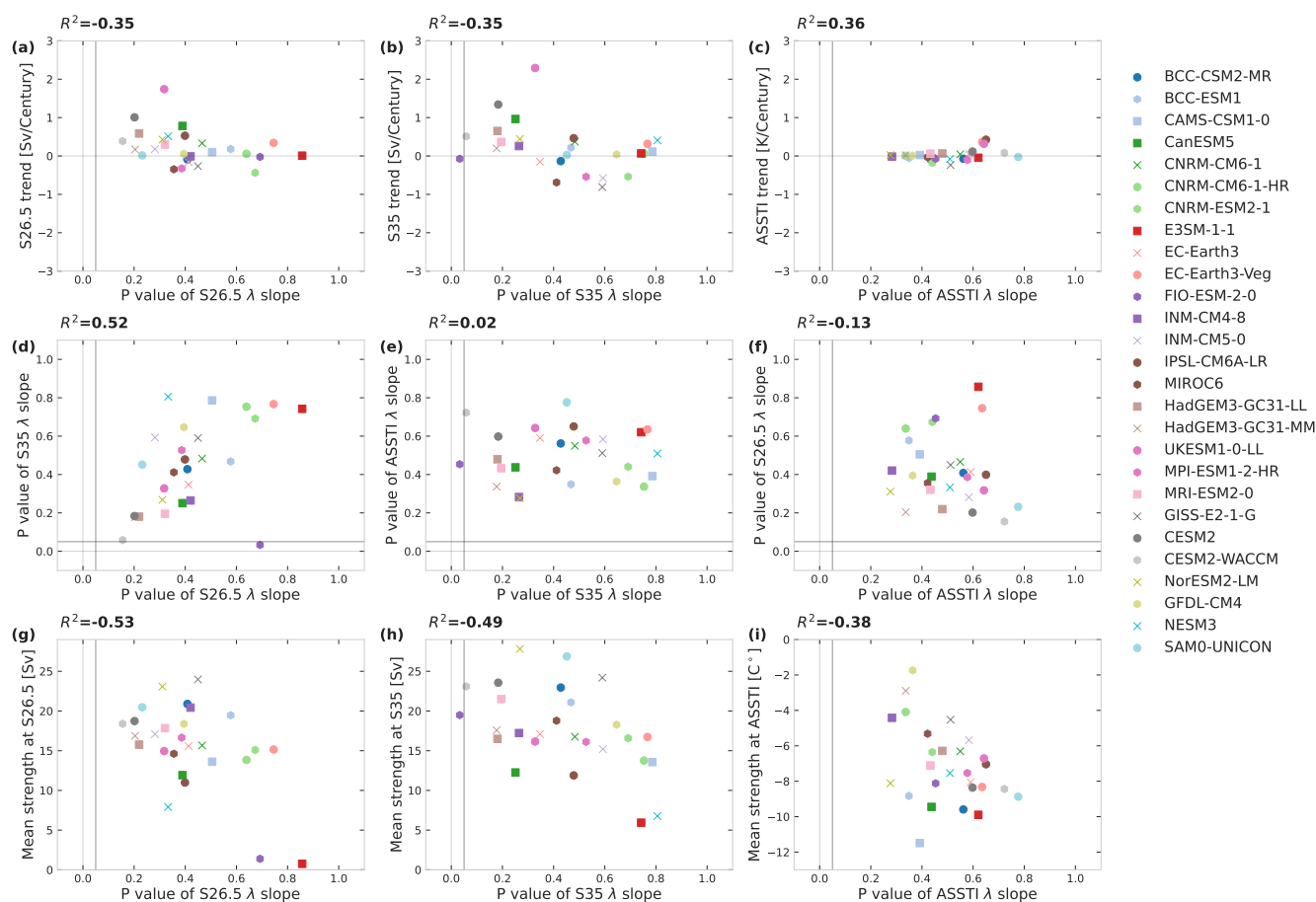


Figure C12. Same as Figure 6 for the ensemble mean values of each of the models.

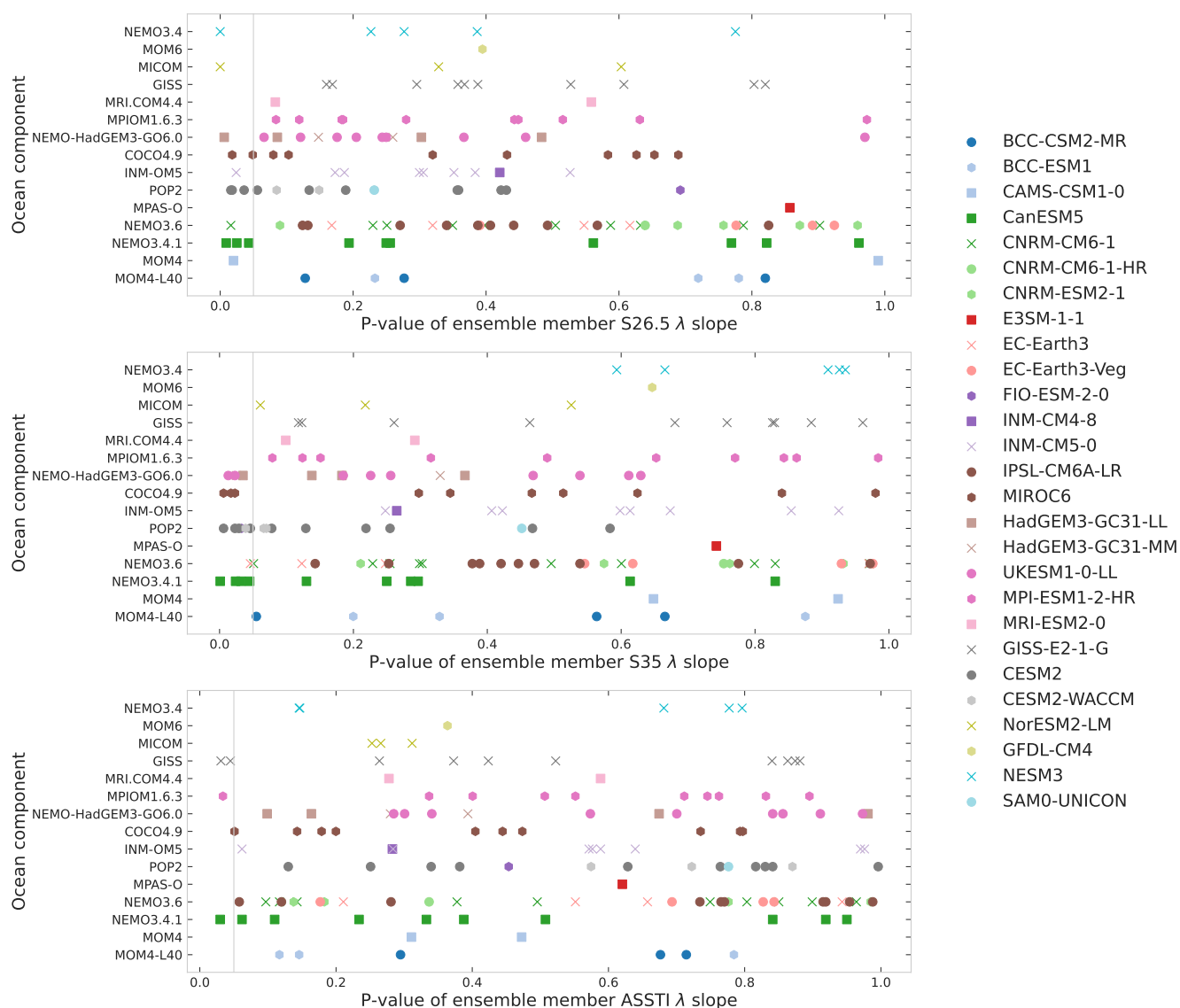


Figure C13. The distribution of the ensemble member p-values for the λ trend of the AMOC streamfunction strengths at 26.5°N and 35°N and the AMOC SST Index, plotted against the ocean component of the model. The models with lower p-values have a range of different ocean components.

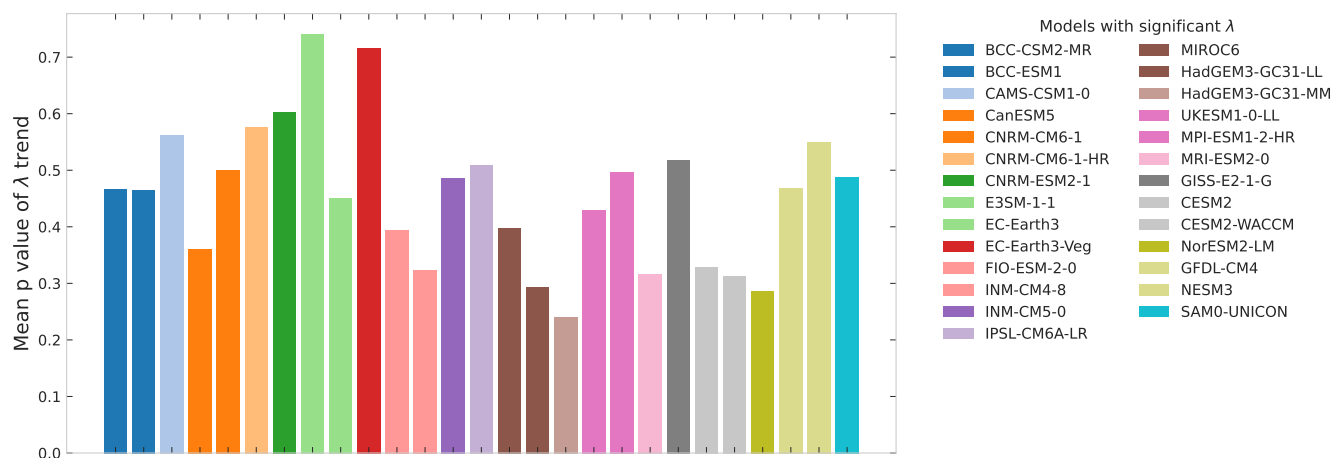


Figure C14. Mean p-value for all λ timeseries of a given model. Note that the colours and shading of the models are different from other figures. Note that the models with highest indication of CSD are those with the *lowest* p-values.

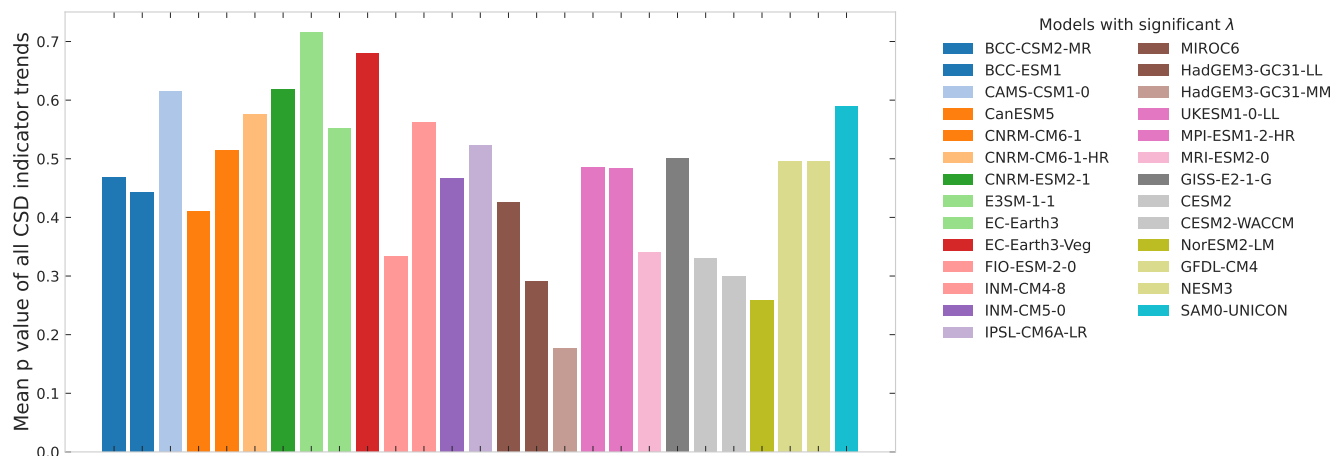


Figure C15. Mean p-value for all CSD indicator timeseries of a given model. Note that the colours and shading of the models are different from other figures.



Model	Ensemble members
BCC-CSM2-MR	r1 r2 r3
BCC-ESM1	r1 r2 r3
CAMS-CSM1-0	r1 r2
CanESM5	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
CNRM-CM6-1	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
CNRM-CM6-1-HR	r1
CNRM-ESM2-1	r1 r2 r3 r4 r5
E3SM-1-1	r1
EC-Earth3	r1 r2 r4 r7
EC-Earth3-Veg	r1 r2 r3 r4
FIO-ESM-2-0	r1
INM-CM4-8	r1
INM-CM5-0	r1 r2 r5 r6 r7 r8 r9 r10
IPSL-CM6A-LR	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
MIROC6	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
HadGEM3-GC31-LL	r1 r2 r3 r4
HadGEM3-GC31-MM	r1 r2
UKESM1-0-LL	r2 r3 r4 r5 r6 r7 r8 r9 r10
MPI-ESM1-2-HR	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
MRI-ESM2-0	r1 r5
GISS-E2-1-G	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
CESM2	r1 r2 r3 r4 r5 r6 r7 r8 r9 r10
CESM2-WACCM	r1 r2 r3
NorESM2-LM	r1 r2 r3
GFDL-CM4	r1
NESM3	r1 r2 r3 r4 r5
SAM0-UNICON	r1

Table C1. List of models and their ensemble members that are used in this study. All ensemble members for a listed rx are rxi1pf1 except for CNRM-CM6-1, CNRM-ESM2-1, and UKESM1-0-LL which are rxi1pf2 and HadGEM3-GC31-MM HadGEM3-GC31-LL which are rxi1pf3.



	S26	S35	SSTI	Any
Any	24.0	35.0	21.0	80.0
One	15.0	21.0	18.0	54.0
Two	7.0	11.0	3.0	21.0
Three	2.0	3.0	0.0	5.0

Table C2. Same as Table B1 for a window size of 50 years.

	S26	S35	SSTI	Any
Any	31.0	40.0	14.0	85.0
One	17.0	23.0	7.0	47.0
Two	11.0	10.0	6.0	27.0
Three	3.0	7.0	1.0	11.0

Table C3. Same as Table B1 for a window size of 70 years.



	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
BCC-CSM2-MR										
BCC-ESM1										
CAMS-CSM1-0		S26.5								
CanESM5			S35		S35		S26.5 S35	S26.5	S35	S26.5 ASSTI
CNRM-CM6-1								S26.5		
CNRM-CM6-1-HR										
CNRM-ESM2-1										
E3SM-1-1										
EC-Earth3		S35								
EC-Earth3-Veg										
FIO-ESM-2-0	S35									
INM-CM4-8										
INM-CM5-0		S26.5								
IPSL-CM6A-LR										
MIROC6					S35	S35	S26.5	S26.5 S35		
HadGEM3-GC31-LL			S26.5	S35						
HadGEM3-GC31-MM	S35									
UKESM1-0-LL					S35			S35		
MPI-ESM1-2-HR										ASSTI
MRI-ESM2-0										
GISS-E2-1-G		ASSTI	ASSTI							
CESM2			S26.5 S35	S35			S26.5	S26.5 S35	S35	
CESM2-WACCM			S35							
NorESM2-LM			S26.5							
GFDL-CM4										
NESM3				S26.5						
SAM0-UNICON										

Table C4. Significant increases of λ . The same abbreviations for the time series as Figure 6 are used. For each ensemble member the time series in which there is a significant increase is listed in the corresponding cell. Note that not all models have all 10 ensemble members - the list of members used can be seen in Table C1.



	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
BCC-CSM2-MR			S35							
BCC-ESM1										
CAMS-CSM1-0										
CanESM5				ASSTI			S35	S26.5	S35	S26.5 ASSTI
CNRM-CM6-1								S26.5 ASSTI		
CNRM-CM6-1-HR										
CNRM-ESM2-1										
E3SM-1-1										
EC-Earth3										
EC-Earth3-Veg										
FIO-ESM-2-0	S35									
INM-CM4-8										
INM-CM5-0	S26.5						S26.5			
IPSL-CM6A-LR										
MIROC6	S35 ASSTI	S26.5 S35			S35			S35		
HadGEM3-GC31-LL		S35 ASSTI	S26.5 S35	S26.5 S35 ASSTI						
HadGEM3-GC31-MM	S35									
UKESM1-0-LL		S26.5						S35		S26.5
MPI-ESM1-2-HR										
MRI-ESM2-0										
GISS-E2-1-G		S26.5		ASSTI						
CESM2	S26.5		S26.5 S35			S26.5	S35	S26.5 S35		
CESM2-WACCM										
NorESM2-LM			S26.5							
GFDL-CM4										
NESM3										
SAM0-UNICON				S26.5						

Table C5. Significant increases in the autocorrelation. Same structure as Table C4.



	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10
BCC-CSM2-MR										
BCC-ESM1			S26.5 S35 ASSTI							
CAMS-CSM1-0										
CanESM5										
CNRM-CM6-1								S26.5 S35 ASSTI		
CNRM-CM6-1-HR										
CNRM-ESM2-1		S26.5								
E3SM-1-1										
EC-Earth3										
EC-Earth3-Veg		S26.5 S35								
FIO-ESM-2-0	S35									
INM-CM4-8										
INM-CM5-0		S26.5					S26.5		S35	
IPSL-CM6A-LR				S35			ASSTI			
MIROC6	ASSTI		S35					S26.5 S35		
HadGEM3-GC31-LL				S35						
HadGEM3-GC31-MM										
UKESM1-0-LL								S35		
MPI-ESM1-2-HR										
MRI-ESM2-0										
GISS-E2-1-G		ASSTI	S35		ASSTI					
CEESM2		S26.5						S35	S35	
CEESM2-WACCM	S26.5 S35									
NorESM2-LM										
GFDL-CM4										
NESM3										
SAM0-UNICON										

Table C6. Significant increases in the variance. Same structure as Table C4.



309 *Author contributions.* MBY, LB and NB conceived the study and designed it with contributions from SB. MBY and LB carried out the
310 analysis. MBY, LB, SB and NB discussed results, and MBY wrote the paper with contributions from LB, SB and NB.

311 *Competing interests.* The authors declare that there are no competing interests.

312 *Acknowledgements.* MBY and NB acknowledge funding by the European Union's Horizon 2020 research and innovation programme under
313 the Marie Skłodowska-Curie grant agreement No.956170. NB and SB acknowledge funding by the Volkswagen foundation. This is ClimTip
314 contribution #X; the ClimTip project has received funding from the European Union's Horizon Europe research and innovation programme
315 under grant agreement No. 101137601. LB acknowledges funding via the Tipping Points in the Earth System (TiPES) project that has
316 received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 820970. In
317 addition, the authors would like to thank Dr Matthew Menary for providing the CMIP6 AMOC data.



318 References

- 319 Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G.-K., Rogelj, J., Rojas,
320 M., Sillmann, J., Storelmo, T., Thorne, P. W., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala,
321 R., Berger, S., Canadell, J. G., Cassou, C., Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J.,
322 Dereczynski, C., Di Luca, A., Diongue Niang, A., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer,
323 E., Forster, P., Fox-Kemper, B., Fuglested, J. S., Fyfe, J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi,
324 R., Hawkins, E., Hewitt, H. T., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Kosaka, Y., Kossin, J., Krakovska, S.,
325 Lee, J.-Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S.-K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I.,
326 Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J.-B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I.,
327 Sörensson, A. A., Szopa, S., Takayabu, I., Treguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., and
328 Zickfeld, K.: Technical Summary, book section 1, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,
329 https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_TS.pdf, 2021.
- 330 Bellomo, K., Angeloni, M., Corti, S., and von Hardenberg, J.: Future climate change shaped by inter-model differences in Atlantic meridional
331 overturning circulation response, *Nature Communications*, 12, 1–10, <https://doi.org/10.1038/s41467-021-24015-w>, 2021.
- 332 Ben-Yami, M., Skiba, V., Bathiany, S., and Boers, N.: Uncertainties in critical slowing down indicators of observation-based fingerprints of
333 the Atlantic Overturning Circulation, <http://arxiv.org/abs/2303.06448>, 2023.
- 334 Boers, N.: Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation, *Nature Climate
335 Change*, 11, 680–688, <https://doi.org/10.1038/s41558-021-01097-4>, 2021.
- 336 Boers, N., Ghil, M., and Stocker, T. F.: Theoretical and paleoclimatic evidence for abrupt transitions in the Earth system, *Environmental
337 Research Letters*, 17, <https://doi.org/10.1088/1748-9326/ac8944>, 2022.
- 338 Bonnet, R., Swingedouw, D., Gastineau, G., Boucher, O., Deshayes, J., Hourdin, F., Mignot, J., Servonnat, J., and Sima, A.: Increased risk
339 of near term global warming due to a recent AMOC weakening, *Nature Communications*, 12, 1–9, [https://doi.org/10.1038/s41467-021-
26370-0](https://doi.org/10.1038/s41467-021-
340 26370-0), 2021.
- 341 Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed fingerprint of a weakening Atlantic Ocean overturning circula-
342 tion, *Nature*, 556, 191–196, <https://doi.org/10.1038/s41586-018-0006-5>, 2018.
- 343 Caesar, L., McCarthy, G. D., Thornalley, D. J., Cahill, N., and Rahmstorf, S.: Current Atlantic Meridional Overturning Circulation weakest
344 in last millennium, *Nature Geoscience*, 14, 118–120, <https://doi.org/10.1038/s41561-021-00699-z>, 2021.
- 345 Cessi, P.: A Simple Box Model of Stochastically Forced Thermohaline Flow, *Journal of Physical Oceanography*, p. 1911–1920,
346 [https://doi.org/https://doi.org/10.1175/1520-0485\(1994\)024<1911:ASBMOS>2.0.CO;2](https://doi.org/https://doi.org/10.1175/1520-0485(1994)024<1911:ASBMOS>2.0.CO;2), 1994.
- 347 Frajka-Williams, E., Moat, B., Smeed, D., Rayner, D., Johns, W., Baringer, M., Volkov, D., and Collins, J.: Atlantic meridional overturning
348 circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary
349 Time Series) array at 26N from 2004 to 2020 (v2020.1), <https://doi.org/10.5285/cc1e34b3-3385-662b-e053-6c86abc03444>, 2021.
- 350 Hawkins, E., Smith, R. S., Allison, L. C., Gregory, J. M., Woollings, T. J., Pohlmann, H., and Cuevas, B. D.: Bistability of the Atlantic
351 overturning circulation in a global climate model and links to ocean freshwater transport, 38, 1–6, <https://doi.org/10.1029/2011GL047208>,
352 2011.
- 353 Henry, L. G., McManus, J. F., Curry, W. B., Roberts, N. L., Piotrowski, A. M., and Keigwin, L. D.: North Atlantic ocean circulation and
354 abrupt climate change during the last glaciation, *Science*, 353, 470–474, <https://doi.org/10.1126/science.aaf5529>, 2016.



- 355 Jackson, L. C. and Wood, R. A.: Hysteresis and Resilience of the AMOC in an Eddy-Permitting GCM, *Geophysical Research Letters*, 45,
356 8547–8556, <https://doi.org/10.1029/2018GL078104>, 2018.
- 357 Jackson, L. C. and Wood, R. A.: Fingerprints for early detection of changes in the AMOC, *Journal of Climate*, 33, 7027–7044,
358 <https://doi.org/10.1175/JCLI-D-20-0034.1>, 2020.
- 359 Jackson, L. C., Alastrué De Asenjo, E., Bellomo, K., Danabasoglu, G., Haak, H., Hu, A., Jungclaus, J., Lee, W., Meccia, V. L., Saenko, O.,
360 Shao, A., and Swingedouw, D.: Understanding AMOC stability: the North Atlantic Hosing Model Intercomparison Project, *Geoscientific*
361 *Model Development*, 16, 1975–1995, <https://doi.org/10.5194/gmd-16-1975-2023>, 2023a.
- 362 Jackson, L. C., Hewitt, H. T., Bruciaferri, D., Calvert, D., Graham, T., Guiavarc’h, C., Menary, M. B., New, A. L., Roberts, M., and Storkey,
363 D.: Challenges simulating the AMOC in climate models, *Philosophical transactions. Series A, Mathematical, physical, and engineering*
364 *sciences*, 381, 20220 187, <https://doi.org/10.1098/rsta.2022.0187>, 2023b.
- 365 Kilbourne, K. H., Wanamaker, A. D., Moffa-Sanchez, P., Reynolds, D. J., Amrhein, D. E., Butler, P. G., Gebbie, G., Goes, M., Jansen, M. F.,
366 Little, C. M., Mette, M., Moreno-Chamarro, E., Ortega, P., Otto-Bliesner, B. L., Rossby, T., Scourse, J., and Whitney, N. M.: Atlantic
367 circulation change still uncertain, *Nature Geoscience*, pp. 13–16, <https://doi.org/10.1038/s41561-022-00896-4>, 2022.
- 368 Latif, M., Sun, J., Visbeck, M., Bordbar, M. H., Hadi Bordbar, M., and Bordbar, M. H.: Natural variability has dominated Atlantic Meridional
369 Overturning Circulation since 1900, *Nature Climate Change*, 12, 455–460, <https://doi.org/10.1038/s41558-022-01342-4>, 2022.
- 370 Lee, J. Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J. P., Engelbrecht, F., Fischer, E., Fyfe, J. C., Jones, C., Maycock, A., Mutemi, J.,
371 Ndiaye, O., Panickal, S., and Zhou, T.: Future Global Climate: Scenario-Based Projections and Near-Term Information, book section 4,
372 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, [https://www.ipcc.ch/report/ar6/wg1/downloads/
373 report/IPCC_AR6_WGI_Chapter_04.pdf](https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_04.pdf), 2021.
- 374 Little, C. M., Zhao, M., and Buckley, M. W.: Do Surface Temperature Indices Reflect Centennial-Timescale Trends in Atlantic Meridional
375 Overturning Circulation Strength?, *Geophysical Research Letters*, 47, 1–10, <https://doi.org/10.1029/2020GL090888>, 2020.
- 376 Liu, W., Xie, S. P., Liu, Z., and Zhu, J.: Overlooked possibility of a collapsed atlantic meridional overturning circulation in warming climate,
377 *Science Advances*, 3, 1–8, <https://doi.org/10.1126/sciadv.1601666>, 2017.
- 378 Liu, W., Fedorov, A. V., Xie, S. P., and Hu, S.: Climate impacts of a weakened Atlantic meridional overturning circulation in a warming
379 climate, *Science Advances*, 6, 2–10, <https://doi.org/10.1126/sciadv.aaz4876>, 2020.
- 380 McKay, D. I., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rock-
381 ström, J., and Lenton, T. M.: Exceeding 1.5°C global warming could trigger multiple climate tipping points, *Science*, 377,
382 <https://doi.org/10.1126/science.abn7950>, 2022.
- 383 Mckay, D. I. A., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., and Lenton,
384 T. M.: multiple climate tipping points, 1171, <https://doi.org/10.1126/science.abn7950>, 2022.
- 385 Menary, M. B., Hodson, D. L., Robson, J. I., Sutton, R. T., and Wood, R. A.: A mechanism of internal decadal Atlantic Ocean variability in
386 a high-resolution coupled climate model, *Journal of Climate*, 28, 7764–7785, <https://doi.org/10.1175/JCLI-D-15-0106.1>, 2015.
- 387 Menary, M. B., Robson, J., Allan, R. P., Booth, B. B., Cassou, C., Gastineau, G., Gregory, J., Hodson, D., Jones, C., Mignot, J., Ringer, M.,
388 Sutton, R., Wilcox, L., and Zhang, R.: Aerosol-Forced AMOC Changes in CMIP6 Historical Simulations, *Geophysical Research Letters*,
389 47, <https://doi.org/10.1029/2020GL088166>, 2020.
- 390 Rahmstorf, S.: Ocean circulation and climate during the past 120,000 years, 419, 207–214, 2002.
- 391 Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J.: Exceptional twentieth-century
392 slowdown in Atlantic Ocean overturning circulation, *Nature Climate Change*, 5, 475–480, <https://doi.org/10.1038/nclimate2554>, 2015.



- 393 Robson, J., Menary, M. B., Sutton, R. T., Mecking, J., Gregory, J. M., Jones, C., Sinha, B., Stevens, D. P., and Wilcox, L. J.: The Role of
394 Anthropogenic Aerosol Forcing in the 1850–1985 Strengthening of the AMOC in CMIP6 Historical Simulations, *Journal of Climate*, 35,
395 3243–3263, <https://doi.org/10.1175/JCLI-D-22-0124.1>, 2022.
- 396 Romanou, A., Rind, D., Jonas, J., Miller, R., Kelley, M., Russell, G., Orbe, C., Nazarenko, L., Latto, R., and Schmidt, G. A.: Stochastic
397 Bifurcation of the North Atlantic Circulation Under A Mid-Range Future Climate Scenario With The NASA-GISS ModelE.
- 398 Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., and Sugihara,
399 G.: Early-warning signals for critical transitions, *Nature*, 461, 53–59, <https://doi.org/10.1038/nature08227>, 2009.
- 400 Stommel, H.: Thermohaline Convection with Two Stable Regimes of Flow, *Tellus*, 13, 224–230, <https://doi.org/10.3402/tellusa.v13i2.9491>,
401 1961.
- 402 Sun, J., Latif, M., and Park, W.: Subpolar gyre–AMOC–atmosphere interactions on multidecadal timescales in a version of the Kiel climate
403 model, *Journal of Climate*, 34, 6583–6602, <https://doi.org/10.1175/JCLI-D-20-0725.1>, 2021.
- 404 Swingedouw, D., Bily, A., Esquerdo, C., Borchert, L. F., Sgubin, G., Mignot, J., and Menary, M.: On the risk of abrupt changes
405 in the North Atlantic subpolar gyre in CMIP6 models, *Annals of the New York Academy of Sciences*, 1504, 187–201,
406 <https://doi.org/10.1111/nyas.14659>, 2021.
- 407 Valdes, P.: Built for stability, *Nature Geoscience*, 4, 414–416, <https://doi.org/10.1038/ngeo1200>, 2011.
- 408 Weijer, W., Cheng, W., Garuba, O. A., Hu, A., and Nadiga, B. T.: CMIP6 Models Predict Significant 21st Century Decline of the Atlantic
409 Meridional Overturning Circulation, *Geophysical Research Letters*, 47, <https://doi.org/10.1029/2019GL086075>, 2020.
- 410 Westen, R. M. V., Kliphuis, M., and Dijkstra, H. A.: Physics- based early warning signal shows that AMOC is on tipping course, pp. 1–11,
411 <https://doi.org/10.1126/sciadv.adk1189>, 2024.
- 412 Zhu, C., Liu, Z., Zhang, S., and Wu, L.: Likely accelerated weakening of Atlantic overturning circulation emerges in optimal salinity finger-
413 print, *Nature Communications*, 14, 1–9, <https://doi.org/10.1038/s41467-023-36288-4>, 2023.