Clim Dyn (2018) 50:4171–4188
DOI 10.1007/s00382-017-3867-7

North Atlantic observations sharpen meridional overturning projections

R. Olson[1] · S.-I. An[1] · Y. Fan[2] · J. P. Evans[3] · L. Caesar[4,5]

Received: 2 March 2017 / Accepted: 15 August 2017 / Published online: 23 August 2017

Abstract Atlantic Meridional Overturning Circulation (AMOC) projections are uncertain due to both model errors, as well as internal climate variability. An AMOC slowdown projected by many climate models is likely to have considerable effects on many aspects of global and North Atlantic climate. Previous studies to make probabilistic AMOC projections have broken new ground. However, they do not drift-correct or cross-validate the projections, and do not fully account for internal variability. Furthermore, they consider a limited subset of models, and ignore the skill of models at representing the temporal North Atlantic dynamics. We improve on previous work by applying Bayesian Model Averaging to weight 13 Coupled Model Intercomparison Project phase 5 models by their skill at modeling the AMOC strength, and its temporal dynamics, as approximated by the northern North-Atlantic temperature-based AMOC Index. We make drift-corrected projections accounting for structural model errors, and for the internal variability. Cross-validation experiments give approximately correct empirical coverage probabilities, which validates our method. Our results present more evidence that AMOC likely already started slowing down. While weighting considerably moderates and sharpens our projections, our results are at low end of previously published estimates. We project mean AMOC changes between periods 1960–1999 and 2060–2099 of −4.0 Sv and −6.8 Sv for RCP4.5 and RCP8.5 emissions scenarios respectively. The corresponding average 90% credible intervals for our weighted experiments are [−7.2, −1.2] and [−10.5, −3.7] Sv respectively for the two scenarios.

Keywords Atlantic Meridional Overturning Circulation · Climate modeling · Bayesian Model Averaging · Model structural error · Probabilistic projections

1 Introduction

Atlantic Meridional Overturning Circulation (AMOC) consists of a northward-flowing shallow current, and a southward return flow at depth. AMOC is thought to be influenced by vertical surface fluxes of heat and freshwater, by the Southern Ocean winds which drive deep upwelling there, as well as by vertical mixing (Kuhlbrodt et al. 2007; Kim and An 2013; Yeager and Danabasoglu 2014). The AMOC carries large amounts of heat northward (Buckley and Marshall 2016), which is thought to have a warming effect throughout the Northern Hemisphere, and in particular over the Arctic, the North Atlantic and Northwestern Europe (Manabe and Stouffer 1993; Broecker 1997; Vellinga and Wood 2002; Alley 2007; Kuhlbrodt et al. 2007). Moreover, AMOC is a key process for redistributing heat and carbon into the deep ocean (Stocker and Schmittner 1997; Buckley and Marshall 2016).

While potential future AMOC changes (Manabe and Stouffer 1993; Stocker and Schmittner 1997; Kim and An 2013) cannot "exceed the fertile imaginations of Hollywood

✉ S.-I. An
sian@yonsei.ac.kr

1 Department of Atmospheric Sciences, Yonsei University, 538 Sciences Building, Seoul, South Korea

2 School of Mathematics and Statistics, UNSW, Sydney, Australia

3 Climate Change Research Centre, ARC Centre for Excellence in Climate System Science, UNSW, Sydney, Australia

4 Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Berlin, Germany

5 Institute for Physics and Astronomy, University of Potsdam, Potsdam, Germany

writers" (Alley 2007), they can affect global temperatures, the North Atlantic CO_2 sink and sea level, the Atlantic rainfall, marine ecosystems and storm tracks, the Intertropical Convergence Zone (ITCZ), the land biogeochemistry, the Indian and Asian summer monsoons, and the El Niño-Southern Oscillation (ENSO) (Vellinga and Wood 2002; Stouffer et al. 2006; Alley 2007; Kuhlbrodt et al. 2007; Timmermann et al. 2007; Srokosz et al. 2012; Kim and An 2013; Stone et al. 2016; Buckley and Marshall 2016).

AMOC projections are uncertain (Stocker and Schmittner 1997; Schmittner et al. 2005; Goes et al. 2010; Urban and Keller 2010; Weaver et al. 2012; Chang et al. 2014; Schleussner et al. 2014; Reintges et al. 2016; Bakker et al. 2016). This uncertainty stems from several sources: the forcing uncertainty by greenhouse and other radiatively active gases, climate model uncertainty (including structural and parametric uncertainty), and internal climate variability. The model uncertainty is thought to play by far a dominant role (Reintges et al. 2016), which suggests that differentially weighing models by their skill at representing present-day observations has a strong potential to narrow the range of future projections.

Several previous attempts have been made to provide probabilistic AMOC projections using parsimonious climate models. Schmittner et al. (2005) derive weights for nine Global Climate Model (GCMs) runs from the Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset, based on their skill at representing global fields, North Atlantic sea surface temperature (SST), salinity (SSS), pycnocline depth, as well as measures of the overturning itself. They then construct a weighted mean AMOC projection for the year 2100. While an improvement on previous work, the study uses an old set of GCMs, does not provide probability distribution functions, and the weighing method is not based on established statistical theory. Goes et al. (2010) estimate two ocean parameters of the University of Victoria Earth System Climate Model (UVic ESCM) in the context of a multi-parameter ensemble, and use these estimates to provide future AMOC projections. Model likelihoods are based on model skill at representing global average tracer depth profiles, and account for cross-correlation of model errors between the tracers. The main improvement of that work is usage of well-established Bayesian statistical theory, which makes it possible to obtain probability distribution functions. Bhat et al. (2012) improve on Goes et al. (2010) by using 2D (lat × depth) fields of $\Delta^{14}C$ and CFC11 to constrain vertical diffusivity in UVic ESCM and to provide AMOC projections. They use a flexible hierarchical model to connect the two tracers, leverage kernel mixing to reduce the dimensionality of the data, and employ a Gaussian Process emulator to estimate model output at an arbitrary location and vertical ocean diffusivity setting. Chang et al. (2014) increase the dimensionality of observational constraints to 3D (lat × lon × depth), using potential temperatures in the North Atlantic to constrain vertical diffusivity in UVic ESCM. This is made possible through an advanced data reduction technique consisting of principal component decomposition and kernel mixing.

Schleussner et al. (2014) take a different approach. They fit linear response functions to AMOC output of several models from the fifth phase of Coupled Model Intercomparison Project (CMIP5). These simple linear response models are then integrated into the future under a large number of potential temperature change scenarios. Consequently, this work now incorporates both forcing uncertainties, as well as model structural and parametric uncertainties.

Recently, Bakker et al. (2016) have used the results from the AMOCMIP project, that uses 21 simulations from eight models to make probabilistic AMOC projections. The model runs include a temporary and spatially varying Greenland Ice Sheet (GIS) melt water forcing, as well as spatially varying iceberg discharge. The runs are complemented by a physically-based AMOC emulator. The probabilistic projections, based on this emulator, sample several relevant climate uncertainties, as well as the emulator error.

These studies break important new ground but they still suffer from a series of drawbacks. First, none of these studies fully incorporate uncertainty due to model internal variability in the projection stage. Second, no drift-correction is performed on AMOC output. Furthermore, they do not weight the models by their skill at representing temporal North Atlantic dynamics. Moreover, they typically do not fully account for model errors during the prediction stage. Specifically, AMOC projections from the models are usually taken at face value, without incorporating additional model error. Finally, no cross-validation is performed on the projections to confirm that the credible/confidence intervals have well-calibrated bounds.

We improve on previous probabilistic projections by (a) considering the largest to-date subset of the CMIP5 climate models, (b) weighting the models by their skill at representing present day AMOC, and its century scale trends as estimated by the AMOC Index, (c) cross-validating the projections, (d) fully incorporating model/observational error that provides correct coverage during the cross-validation experiments and (e) drift-correcting all variables using preindustrial control simulations. The rest of this paper is organized as follows: Sect. 2 describes the climate models, the observations used to weight them, the Bayesian Model Averaging procedure for making weighted projections, together with the cross-validation experiments to validate the methodology; Sect. 3 presents model weights and weighted AMOC projections, while Sect. 4 discusses the caveats. Finally, conclusions are provided in Sect. 5.

2 Data and methods

The method we use to obtain weighted probabilistic projections, Bayesian Model Averaging, requires a consistent set of model outputs, prior probabilities for these models, and observations to compare to the models and to weight them. Here we use a subset of CMIP5 models which we weight by their skill at modeling the AMOC Index—a temperature based AMOC proxy adapted from Rahmstorf et al. (2015). The priors for the models are based on their ability to reproduce the AMOC during years 1957–2004 (Kanzow et al. 2010). The following subsections describe our methodology in more detail.

2.1 Climate model output

2.1.1 CMIP5 models

We use 13 CMIP5 models that have pre-calculated AMOC streamfunctions (Table 1). For all models the first ensemble member is used. Our model choice is motivated by the availability of the outputs (particularly for the pre-calculated streamfunctions). Even though there are many other models that output velocity fields from which streamfunctions can be derived, similarly to previous work (Weaver et al. 2012), we do not use such models. This is because we find that the streamfunction obtained from these fields is highly sensitive to interpolation parameters, and ignores subgrid-scale/eddy transports.

2.1.2 AMOC Index

There is no consistent long-term record of AMOC strength over the past 100 or more years. However, such a long record is highly desirable for statistically sound Bayesian techniques. Hence, here we use an AMOC Index—a proxy of long-term AMOC strength trends which is based on temperatures in the North Atlantic region south of Greenland, in a slightly modified form from Rahmstorf et al. (2015).

There are a couple of mechanisms linking temperatures in northern North Atlantic to AMOC strength. First, a reduced AMOC would lead to less northward ocean heat transport by the North Atlantic current into northern North Atlantic, which is expected to cool that area (Alley 2007; Kim and An 2013). Second, northern North Atlantic temperatures can, in their turn, also affect the AMOC through their influence on the buoyancy forcing (Yeager and Danabasoglu 2014; Buckley and Marshall 2016), although previous work has focused on region to the west of the one discussed here (Yeager and Danabasoglu 2014).

These relationships have been explored in a number of studies. Stouffer et al. (2006) have subjected CMIP3 GCMs to a 0.1 Sv North Atlantic freshwater hosing experiment. In response to the resulting AMOC slowdown, the models showed a consistent cooling south of Greenland, which was strongly linearly related to the slowdown. AMOC strengthening has preceded the recently increased subpolar SSTs in a historically forced coupled GCM run (Yeager and Danabasoglu 2014). Drijfhout et al. (2012) and Rahmstorf et al. (2015) analyze observations and climate models to find that the region of conspicuous warming hole in the North Atlantic south of Greenland is associated with the AMOC decline throughout the last century. Furthermore Rahmstorf et al. (2015) find a very strong positive correlation ($r = 0.9$) between smoothed and detrended AMOC Index and AMOC in a state-of-the-art GCM with a realistic representation of the AMOC, which motivates them to use this index as a proxy for past AMOC Index reconstructions. As will be

Table 1 Information about the CMIP5 models used

Model number	Model name	Institution/country
1	ACCESS1.0	CSIRO and BOM/Australia
2	ACCESS1.3	CSIRO and BOM/Australia
3	CanESM2	Canadian Centre for Climate Modelling and Analysis/Canada
4	CCSM4	National Center for Atmospheric Research/USA
5	CESM1-BGC	National Science Foundation, Department of Energy, National Center for Atmospheric Research/USA
6	CESM1-CAM5	National Science Foundation, Department of Energy, National Center for Atmospheric Research/USA
7	GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory/USA
8	INM-CM4	Institute for Numerical Mathematics/Russia
9	MPI-ESM-MR	Max Planck Institute for Meteorology (MPI-M)/Germany
10	MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)/Germany
11	MRI-CGCM3	Meteorological Research Institute/Japan
12	NorESM1-M	Norwegian Climate Centre/Norway
13	NorESM1-ME	Norwegian Climate Centre/Norway

shown here, we find a moderate positive correlation between century-long changes in AMOC and the AMOC Index over the last century in the 13 CMIP5 models we use (Sect. 3.1). Significant North Atlantic cooling (up to 8 °C) in response to an AMOC collapse has also been reported by Vellinga and Wood (2002). Other work has shown that decreased meridional heat transport by the AMOC moderated subarctic North Atlantic warming in CMIP3 CO_2 doubling experiments (Kim and An 2013).

Here we use the AMOC Index in a slightly modified form to the one presented in Rahmstorf et al. (2015). As in Rahmstorf et al. (2015), the AMOC Index is the difference between mean SST in the "gyre" region south of Greenland and the mean Northern Hemispheric surface air temperature, however the gyre region is different. It is now the area between 40° and 80°N, and between 80° and 0°W, where the recent century-scale observed SST trends in ERSSTv4 observations (Huang et al. 2014; Liu et al. 2014) are consistently negative (Fig. 1). Specifically, the trends are obtained from start years ranging from 1880 to 1900, and the end year of 2015; and the mean linear trend is calculated for each grid cell. Earlier periods are not included due to the sparseness of the observations (NOAA 2017). The grid cells that show a negative mean trend are assigned to the "gyre" region. The annual AMOC Index anomalies are calculated for years 1880–2004. Prior to the calculation, modeled output is bilinearly interpolated to the $2\times2°$ lat/lon grid of the ERSSTv4 SST observations, and area-averaging is used for both gyre and Northern Hemispheric temperatures.

The AMOC Index is drift-corrected using information from preindustrial control runs. First, the branch time of the control runs is matched to the historical runs using information from CMIP5 errata table publicly provided by Gregory (2012). For National Center for Climate Research (NCAR) models (e.g., CCSM4, CESM1-BGC and CESM1-CAM5) this errata is out of date, and correct branch times are instead given by the NCAR errata page (NCAR 2017). For GFDL-ESM2M, branch time information is retrieved directly from

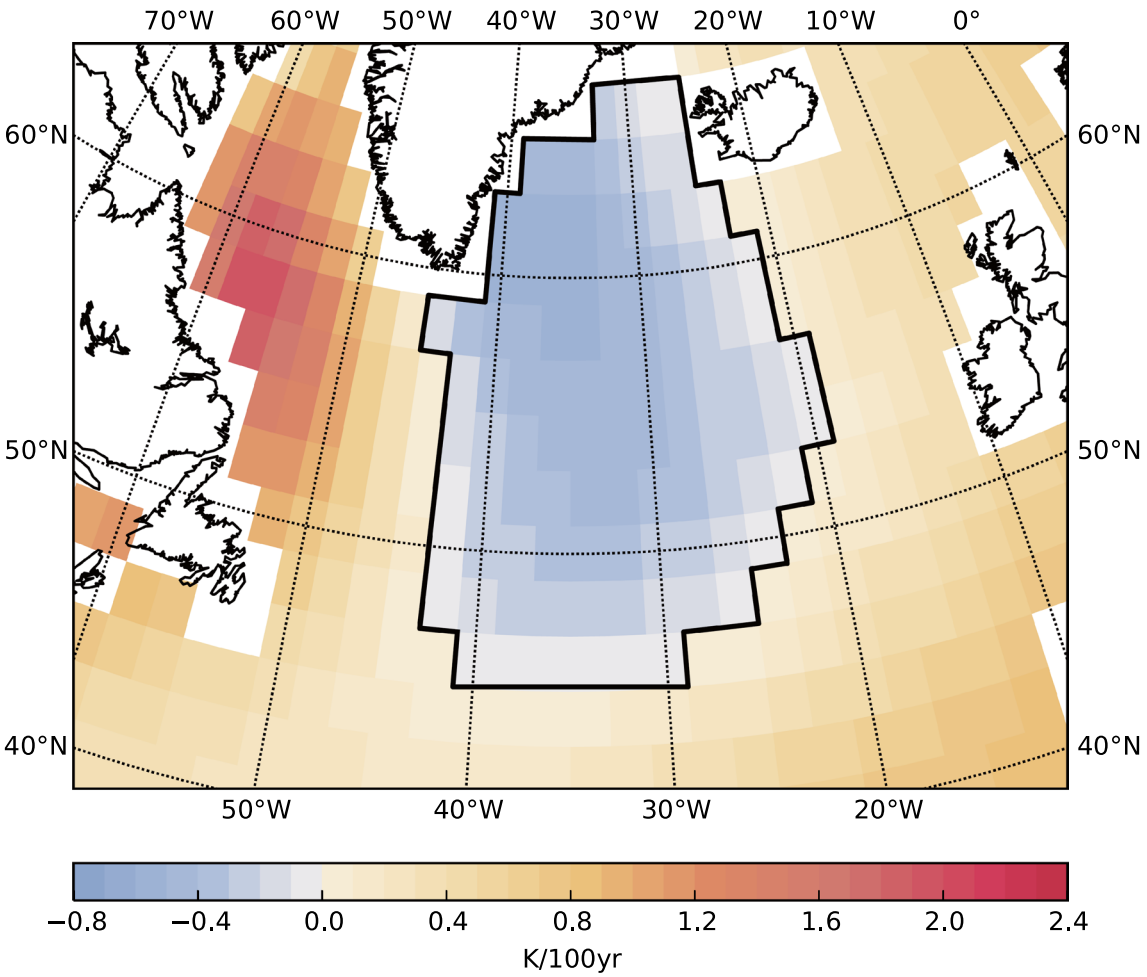

Fig. 1 Mean linear temperature trend for periods starting in 1880 through 1900 and ending in 2015 based on ERSSTv4 observations. The subpolar gyre is also shown

GFDL FAQ (Sentman 2016). For CESM1-CAM5, the control run ends in year 1944 and thus does not cover global warming simulations in full. We detect no considerable trend in the control run. Thus, no drift is assumed for this model. For other models, a quadratic fit is applied to the 1850–2100 control simulations and then subtracted from historical simulations. We note that considerable drifts have been found in two models (not shown).

The drift-corrected AMOC Index from the 13 GCMs, smoothed using robust locally weighted regression (Cleveland 1979) and demeaned (e.g., bias-corrected to have a zero mean over years 1880–2004), is shown in Fig. 2.

2.1.3 AMOC

Modelled annual-mean AMOC strength is extracted for years 1880–2004, and 2060–2099. If the northward overturning streamfunction has been pre-calculated, it is used. Otherwise, if just the Y-overturning meridional streamfunction (on the model's native grid) is available, it is used, as long as the model's grid is deemed to be sufficiently similar to a latitude-longitude grid over the area of AMOC calculation south of 65°N. We convert mass streamfunctions into volume streamfunctions [Sv] using a constant ocean density $\rho = 1035$ kg m^{-3}, a typical value for the upper 3000 m of the ocean (Pawlowicz 2013). We define AMOC strength as the maximum of the streamfunction between 20° and 65°N, above 3000 m. From here on, unless we specifically say so, we are referring to this definition. However when assigning model priors, for better comparison with observations we use a quantity AMOC_{26}, which is the maximum of the Atlantic streamfunction just at 26°N, above 3000 m. For the future, we use outputs from RCP4.5 and RCP8.5 emissions scenarios (Moss et al. 2010). Both AMOC and AMOC_{26} are drift corrected similarly to the AMOC Index. We note that one of the models experiences a positive drift of almost 2.5 Sv over the 1850–2100 period (not shown), which highlights the importance of drift-correction for AMOC and related variables. The future drift-corrected AMOC for the two scenarios is shown in Fig. 3.

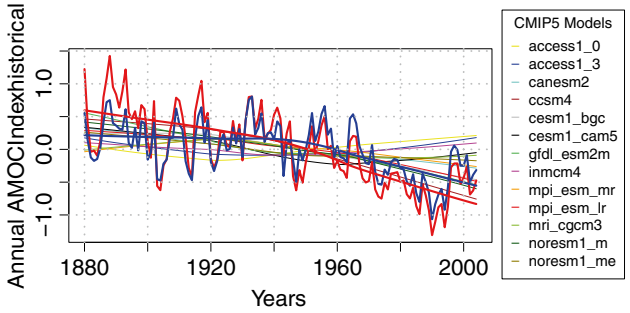

Fig. 2 Model output of present-day drift-corrected lowess-smoothed demeaned modeled AMOC Index compared to observations [K]. *Thick red lines* un-smoothed and smoothed ERSSTv4 observations. *Thick blue lines* COBE-SST2 raw and smoothed observations. *Thin colored lines* smoothed model output

2.2 Observations

The demeaned AMOC Index is also calculated for the observations, for years 1880–2004 (Fig. 2). For the AMOC Index calculations, the "gyre" SSTs come from two sources: ERSSTv4 dataset (Liu et al. 2014; Huang et al. 2015), and COBE-SST2 dataset (Hirahara et al. 2013). The ERSSTv4 dataset is on the $2\times2°$ grid, while the COBE-SST2 dataset is interpolated to the $2\times2°$ grid using bilinear interpolation. Northern Hemisphere temperatures are obtained from the GISTEMP dataset (Hansen et al. 2010).

2.3 Bayesian model averaging and statistical model

Bayesian model averaging (BMA) is a technique to weight models and provide weighted probabilistic projections (Hoeting et al. 1999; Raftery et al. 2005; Montgomery and Nyhan 2010; Bhat et al. 2011; Terando et al. 2012). At the core of BMA is a formula for the probability density function (pdf) of a projected quantity Δ from K models given observations D and models $M_1, \ldots, M_K$.

$$p(\Delta|D) = \sum_{i=1}^{K} p\big(\Delta|M_i, D\big)p\big(M_i|, D\big). \quad (1)$$

We use the BMA implementation of Olson et al. (2016) with important additions. The original methodology (1)

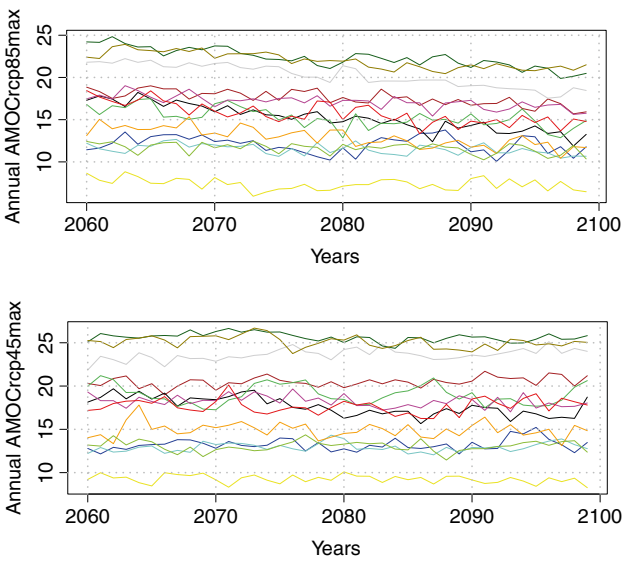

Fig. 3 Model output of 2060–2099 AMOC strength for RCP8.5 (*top*) and RCP4.5 (*bottom*) scenarios. *Color* convention the same as in Fig. 2

does not account for auto-correlation of model-observational residuals and (2) does not include model/observational error beyond a simple time-independent bias term. We improve on that work by incorporating auto-correlation and a more robust representation of model/observational error. This requires modifications to our statistical model that describes the relationship between model output and observations.

Before the BMA, we smooth the historical model AMOC Index using robust locally-weighted "lowess" regression (Cleveland 1979) in an attempt to remove the effects of internal variability. We set the smoother span parameter f to 2/3, which appears to remove inter-annual and decadal variability but preserves century-scale trends. After the smoothing, the data are demeaned to remove any biases that are introduced by smoothing (Fig. 2). For the future 2060–2099 AMOC we use a different smoothing method. Nonlinearities in the time series of just 40 years are likely mostly caused by internal variability, hence we assume a linear trend and use Theil–Sen slopes (Sen 1968) for smoothing. Theil–Sen slopes are less sensitive to outliers than standard linear regression slopes.

Our statistical model for the historical demeaned AMOC Index first connects the "best" but imperfect smoothed ith model output $\boldsymbol{x_i}$ to the observed climate trend $\boldsymbol{y}$ (which differs from the actual climate trend due to presence of long-term observational error, Fig. 2), and then the observed trend to observations $\boldsymbol{y'}$ [primed quantities refer to un-smoothed quantities, all bold quantities are vectors: $\boldsymbol{y} = (y_1, \ldots, y_N)$ where $N = 125$ is the number of years in period 1880–2004]:

$$\begin{cases} \boldsymbol{y} = \boldsymbol{x_i} + \boldsymbol{\varepsilon_D} \\ \boldsymbol{y'} = \boldsymbol{y} + \boldsymbol{\varepsilon_{NV}} \end{cases}. \tag{2}$$

Here $\boldsymbol{\varepsilon_D}$ is discrepancy which includes long-term model and observational errors, and $\boldsymbol{\varepsilon_{NV}}$ models short-term internal variability and observational error. The discrepancy $\boldsymbol{\varepsilon_D}$ contains two components: $\boldsymbol{\varepsilon_{D,M}}$ (model trend error) and $\boldsymbol{\varepsilon_{D,O}}$ (observed trend error). $\boldsymbol{\varepsilon_{D,M}}$ is a potentially non-stationary multivariate distribution $\boldsymbol{\varepsilon_{D,M}} \sim \boldsymbol{D}\left(\boldsymbol{\mu_{D,M}}, \boldsymbol{\Sigma_{D,M}}\right)$. We do not estimate any parameters of this distribution. Rather, we assume that samples from this distribution are directly provided by $\boldsymbol{d_j}$, the scaled differences between each smoothed model output $\boldsymbol{x_j}$, and the next closest output $\boldsymbol{x_k}$, where k is chosen to minimize the l_1 norm of the differences $\boldsymbol{x_k} - \boldsymbol{x_{j_1}}$. The reasoning for this is as follows. Imagine that a particular smoothed model k represents the "true" climate trend, and 12 other models are available. If we choose the "best" model as the one with the minimal l_1 difference with the "true" climate, this difference would represent model error. Finally, $\boldsymbol{d_j}$ also includes a vector of zeros, which indicates our prior belief that a model might not have any structural error. This model error implementation is inspired by that of Sexton et al. (2011). $\boldsymbol{\varepsilon_{D,O}}$ is difficult to quantify and is likely persistent (Liu et al. 2014, our Fig. 2). We assume that it is similar to model error in terms of structure and magnitude. This may be a reasonable assumption, given the two observational trends that we consider (Fig. 2). Thus we take discrepancy samples (which combine long-term model and observational error) as $2\boldsymbol{d_j}$.

$\boldsymbol{\varepsilon_{NV}}$ is modeled as an AR(1) process with autocorrelation ρ and innovation standard deviation σ. We define parameter vector $\boldsymbol{\theta} = (\sigma, \rho)$. We use the AR(1) process because it provides a simple method to account for time dependence. Some support for this is provided by the fact that spectra of historical AMOC Index variability around the long term trends for observations and models typically compare well to an AR(1) process (Fig. 4).

In this formulation, model weights can be calculated as:

$$p(M_i|\boldsymbol{y'}) \propto p\left(M_i\right) \iint p\left(\boldsymbol{y'}|\boldsymbol{y}, \boldsymbol{\theta}, M_i\right) p(\boldsymbol{\theta}) p(\boldsymbol{y}|M_i) d\boldsymbol{y} d\boldsymbol{\theta}. \tag{3}$$

Here, $p(M_i)$ is the climate model prior, discussed in Sect. 2.4. The first term under the integral is the likelihood of observations given the long-term observed trend. The next two terms are the prior for the statistical parameters of the short-term internal variability/observational error, and the conditional distribution for the long-term observed trend. Thus, the overall weight depends on the likelihood of the observations given the model, accounting for the uncertainties in properties of internal variability, and in the long-term observational trend. This integral can be approximated by Monte Carlo sampling. Specifically, $\boldsymbol{\theta}$ and $\boldsymbol{y}$ can be sampled from their prior/conditional distributions using Monte Carlo, and the integral can be calculated as the average of posterior probabilities at the sampled values. We use 1,000,000 samples for the Monte Carlo sampling.

For the future, we assume that AMOC projections for 2060–2099 are:

$$\boldsymbol{y'(f)} = \boldsymbol{x_i^{(f)}} + \boldsymbol{b^{(f)}} + \boldsymbol{\varepsilon_{NV,i}^{(f)}}, \tag{4}$$

where $\boldsymbol{b^{(f)}}$ is time-independent future model error (bias) and $\boldsymbol{\varepsilon_{NV,i}^{(f)}}$ is internal variability of the ith model. Our motivation for the time-independent error stems from the fact that future inter-model AMOC differences are very persistent and can, to the first order, be approximated by a constant bias (Fig. 3). We assume that $\boldsymbol{b^{(f)}} = b^{(f)}\mathbf{1}$, a scaled vector of ones of length $N_f = 40$ which is the length of the future period 2060–2099. Furthermore, $b^{(f)} \sim N\left(0, \sigma_b^{(f)}\right)$. $\sigma_b^{(f)}$ is obtained as standard deviation of period mean of future differences $\bar{\boldsymbol{d}}_j^{(f)}$ which are defined analogously to present-day. As previously (Olson et al. 2016), $\boldsymbol{\varepsilon_{NV,i}^{(f)}}$ is simply resampled using bootstrapping

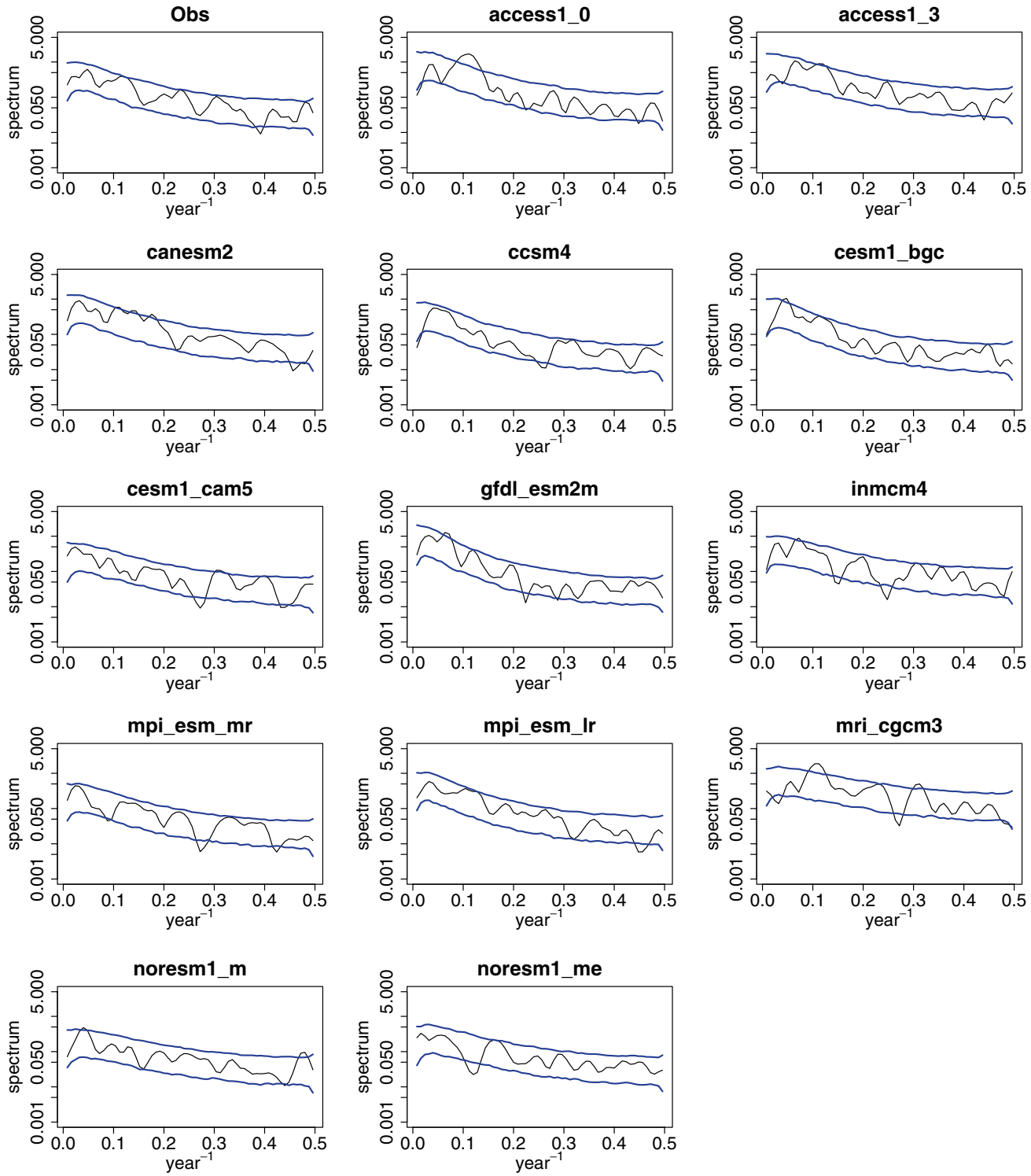

Fig. 4 AMOC Index spectra plots for ERSSTv4 observational and model residuals from long-term robust locally weighted regression trendlines (Cleveland 1979) for the years 1880–2004. *Blue lines* 90% confidence intervals for spectra of AR(1) processes that were fit to observed and modeled anomalies, based on 1000 random realizations

from the residuals between the raw and the smoothed models. We use this simple method because these residuals tend to be only mildly auto-correlated for both future scenarios (not shown). To obtain future AMOC change pdf, we generate weighted samples for future AMOC using Eqs. (1) and (4), and subtract corresponding 1960–1999 modelled

AMOC, with a total number of 100,000 samples. The BMA code is reproducible and is available from the authors upon request.

2.4 Model priors

Our method requires a specification of model priors—a prior degree of belief in each of the models. We calculate the priors based on the models' representation of historical 1957–2004 $AMOC_{26}$ estimates from Bryden et al. (2005), corrected for seasonal cycle by Kanzow et al. (2010). To this end, we first calculate the ratio of modeled to observed $AMOC_{26}$ for each model i, S_i. The prior for the ith model is specified as the following function (Olson et al. 2016):

$$p(M_i) = P(S_i) = Gamma(\alpha, \beta), \tag{5}$$

where $\alpha = \left(1 + ln\frac{16}{9}\right) \Big/ \left(1 - ln\frac{9}{4}\right)$, and $\beta = \frac{1}{\alpha - 1}$. This ensures that a model that gets $AMOC_{26}$ perfectly will receive the highest prior, while a model that overestimates $AMOC_{26}$ by 50% will get down-weighted by a factor of 2 compared to the perfect model. Relative model priors are plotted in Fig. 5. After the calculation, model priors are re-scaled to sum to 1. Most models can capture historical AMOC magnitude relatively well, however, the two Norwegian Climate Centre models overestimate it by more than 50%.

Note that recently more continuous AMOC observations have become available from 2004 onwards as a result of the RAPID mooring array project (Cunningham et al. 2007; Srokosz et al. 2012). However, they mostly cover the "future" period of the CMIP5 simulations. Thus, to the extent the model response differs between the two scenarios, using these observations would result in two different sets of prior weights. Hence, we refrain from using RAPID observations at the moment.

For the parameters of the statistical model, we use uniform priors for ρ on an interval [0, 1], and for σ on [0, 5]. The ρ prior reflects our belief that the sum of short-term observational error and model internal variability is positively autocorrelated. Some support for this is provided by higher internal AMOC Index variability at longer time scales (Fig. 4).

2.5 One-at-a-time cross-validation experiments

To validate our method, we perform one-at-a-time cross-validation experiments. We assume there are 13 worlds. In each of these worlds, one of the models provides "true" observations of climate, and "true" future projections. First, we use our method to assign weights to all 13 models based on their ability to simulate the AMOC and the AMOC Index from that model. The model associated with "true" observations always receives the highest weight (Fig. 6). The AMOC-based prior typically does not vary drastically between the models, with an exception of the two rightmost models NorESM1-M and NorESM1-ME. This indicates that information from the AMOC itself provides only a weak constraint on model skill for most models. Instead, the models are distinguished mostly by their ability to simulate the temporal change in the AMOC Index. The large variation in posterior weights between models illustrates that the method is very powerful at separating "good" and "poor" models.

After the weight estimation, the model giving the "true" projection is excluded from the analysis. The weights from the remaining 12 models are re-scaled to sum up to 1. The probabilistic AMOC change projections between 1960 and 1999 and 2060–2099 from the 12 remaining models are compared to the "true" projections in each world (Figs. 7, 8). The "true" projection falls outside the 90% credible interval 2 out of 26 times, indicating a reasonable 92% coverage. The pdfs often extend considerably beyond the ensemble range, indicating the importance of sampling the uncertainty in internal variability and model error. While the pdfs are usually smooth and unimodal, bimodality is evident in some cases. Bimodality is sometimes present in probabilistic climate model projections (Tomassini et al. 2007; Olson et al. 2016). There are several possible reasons for bimodality. One is clustering of model output, in which bimodality may already appear before the weighting. Another is due to

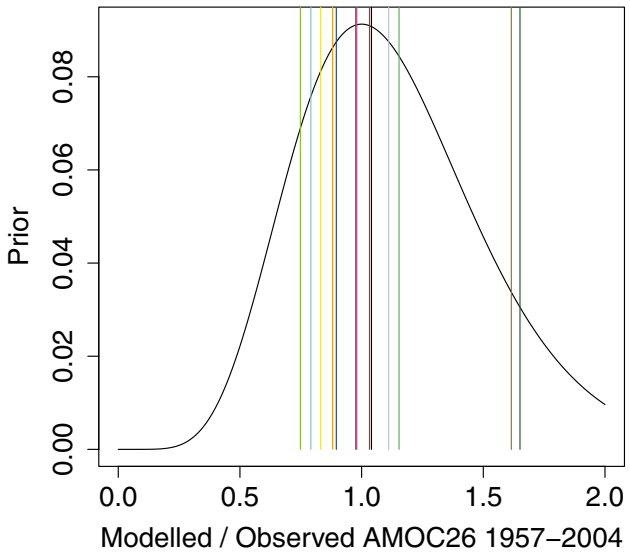

Fig. 5 AMOC priors. *Black line* adopted prior form. *Colored lines* the ratios of mean modeled to observed AMOC at 26°N for years 1957–2004 for the CMIP5 models. *Coloring* convention the same as in Fig. 2

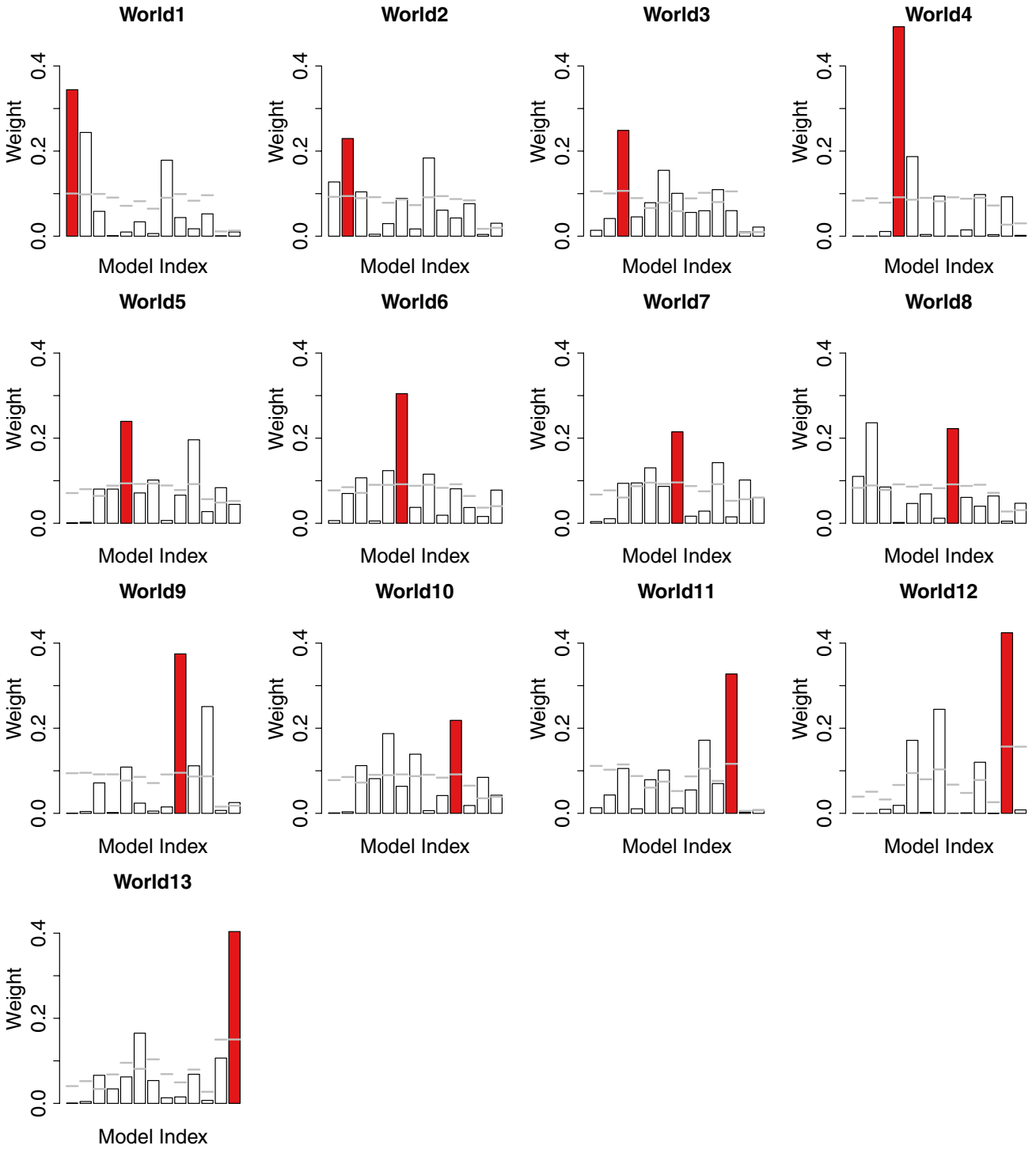

Fig. 6 Model weights for one-at-a-time cross-validation. Model index corresponds to Table 1 and increases from 1 to 13 *left to right. Red bar* denotes the model associated with "true" pseudo-observations. *Horizontal grey lines* model priors

highly weighted models generating drastically different projections (with less likely models in between). This kind of bimodality is introduced during the weighting procedure. It is possible that both of these reasons contribute to the bimodality seen in our experiments.

2.6 Real estimation experiments

We conduct four future projection experiments to test the sensitivity of our results to methodological choices. In the "ERSST" experiment we use ERSSTv4 SST

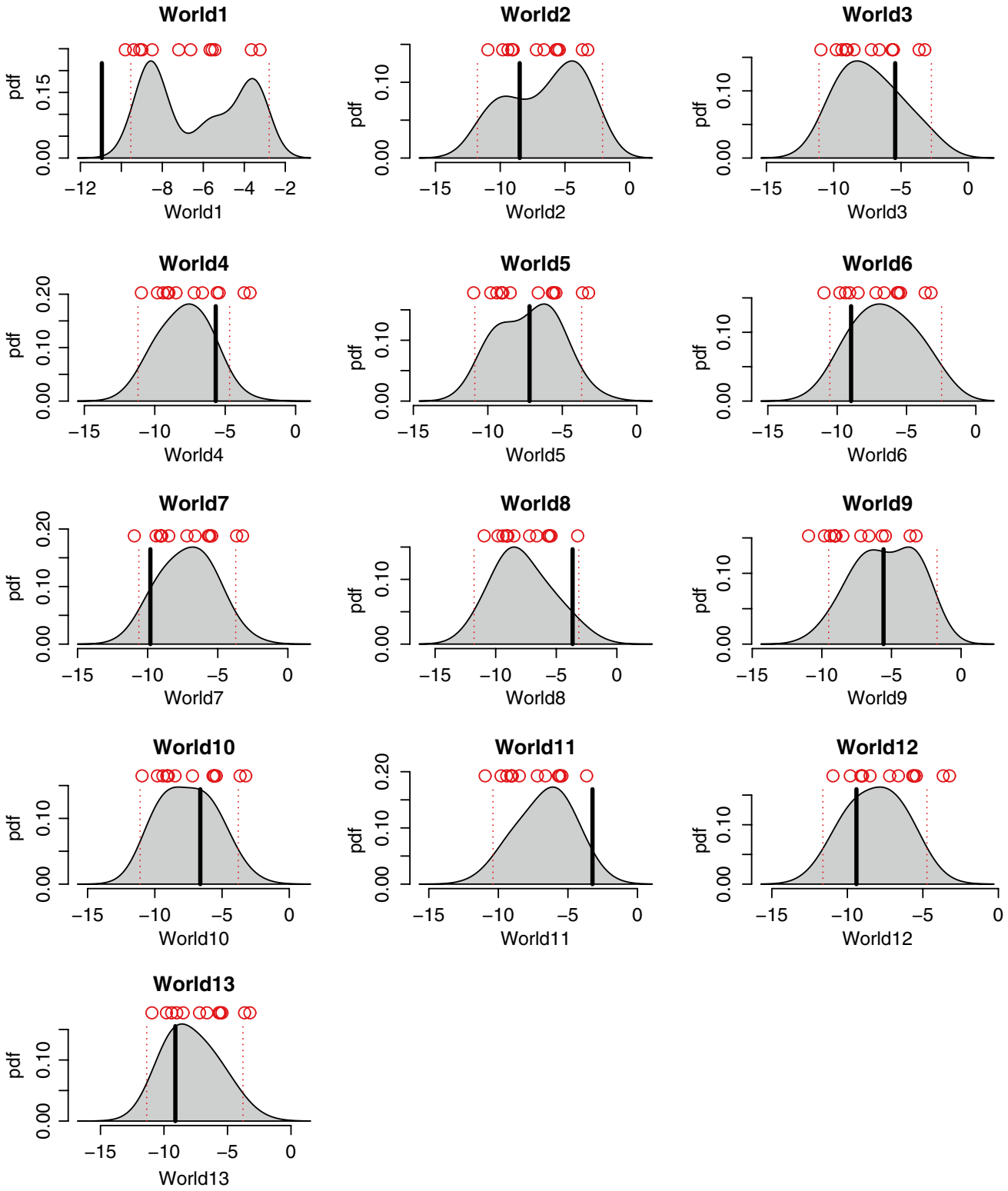

Fig. 7 AMOC change 1960–1999 to 2060–2099 projection pdfs for cross-validation experiments under the RCP8.5 emissions scenario. *Red dotted lines* 90% posterior credible intervals. *Thick black lines* "true" changes. *Circles* changes from individual models

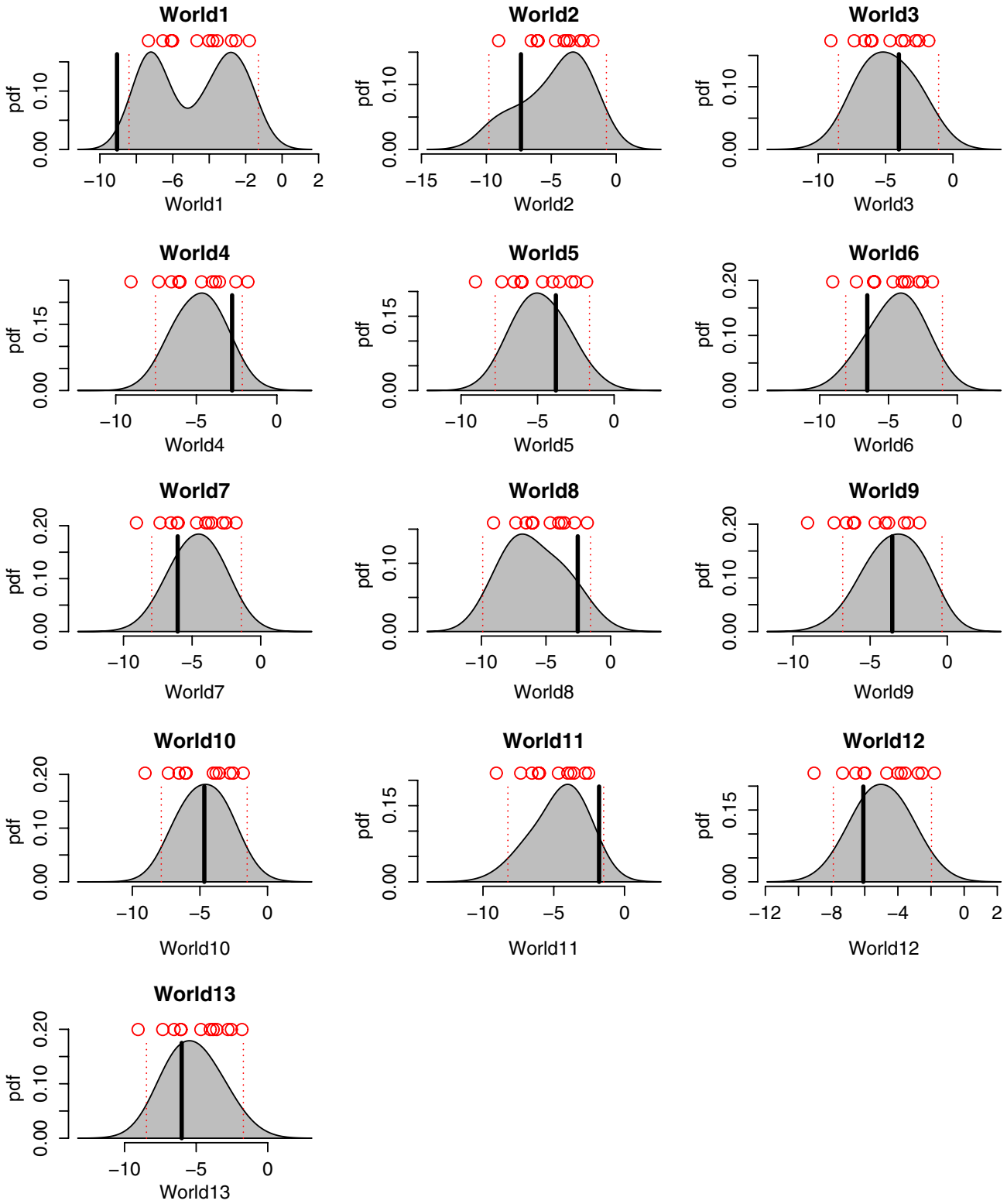

Fig. 8 Same as in Fig. 8, but under the RCP4.5 emissions scenario

dataset to calculate North Atlantic gyre temperatures used for the AMOC Index, and smooth the Index using the lowess method. In the "ERSST-LIN" experiment we substitute Theil–Sen slopes for lowess smoothing of the modelled AMOC Index to generate a trend that does not include nonlinearities in model output with respect to time. The "COBE" experiment uses COBE-SST2 SST dataset for the gyre temperatures, and is otherwise identical to the "ERSST" experiment. Finally, we test the sensitivity to the lowess smoother span parameter f used during AMOC Index smoothing. In the experiment "COBE-F55" we use the COBE-SST2 dataset and smooth the modeled AMOC Index output using $f = 0.55$, which produces weaker smoothing. We choose this value, because using lower settings does not fully eliminate internal variability. Also, when using a higher value of 0.8 we obtain similar mean and confidence intervals for the projections compared to f = 2/3 (for the ERSSTv4 observations), so we do not further report on that experiment here.

3 Results and discussion

3.1 The relationship between AMOC Index and AMOC strength

We find a moderate positive linear relationship (p value = 0.0077) between the century scale AMOC Index change and AMOC change in the 13 models (Fig. 9). Moderate correlation also occurs with smoothed output, and with an end period of 1965–1984 (not shown). This strongly suggests, that on the century scale AMOC changes are an important driver of northern North Atlantic SST anomalies with respect to the hemispheric average, through the AMOC's effect on the meridional heat flux.

The models that project a strong decrease in the AMOC Index that is within the observed estimates tend to also showcase a considerable weakening of the AMOC. One model—NorESM1-M—is an exception to this. While projecting strong AMOC Index decline, it is exhibiting slight AMOC strengthening. If the modeled linear relationship exists in the real world, the observed relative gyre cooling implies that the AMOC has likely already started slowing down (Fig. 9). These results are in line with the emerging new paradigm that AMOC is likely already responding to climate change (Drijfhout et al. 2012; Rahmstorf et al. 2015).

The linear trendline (Fig. 9) predicts a minor relative gyre cooling without any long-term AMOC change. This leaves the door open for an additional process for the cooling. This is consistent with results of Drijfhout et al. (2012), and Kim and An (2013) which suggest that Ekman advection changes may play a role. More detailed examination of the driving processes is beyond the scope of this paper.

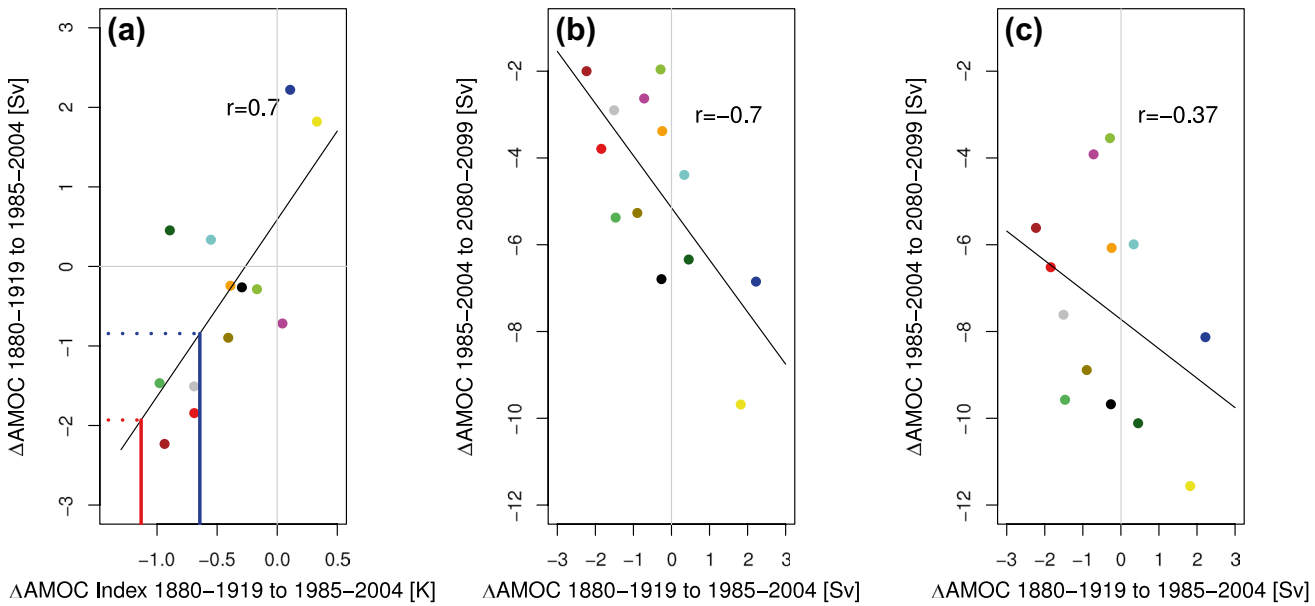

Fig. 9 **a** Correlation between AMOC Index change and AMOC change between years 1880–1919 and 1985–2004 in CMIP5 models, together with a simple linear regression trendline. *Color* convention as in Fig. 2. *Solid red* and *blue lines* represent observed AMOC Index changes from ERSSTv4 and COBE-SST2 SST datasets, respectively. *Dotted lines* show corresponding expected AMOC changes projected by the simple linear regression. **b** Correlation between AMOC changes from years 1880–1919 to 1985–2004, and corresponding future changes from 1985–2004 to 2080–2099 under the RCP4.5 scenario. **c** Same as **b** but under the RCP8.5 scenario

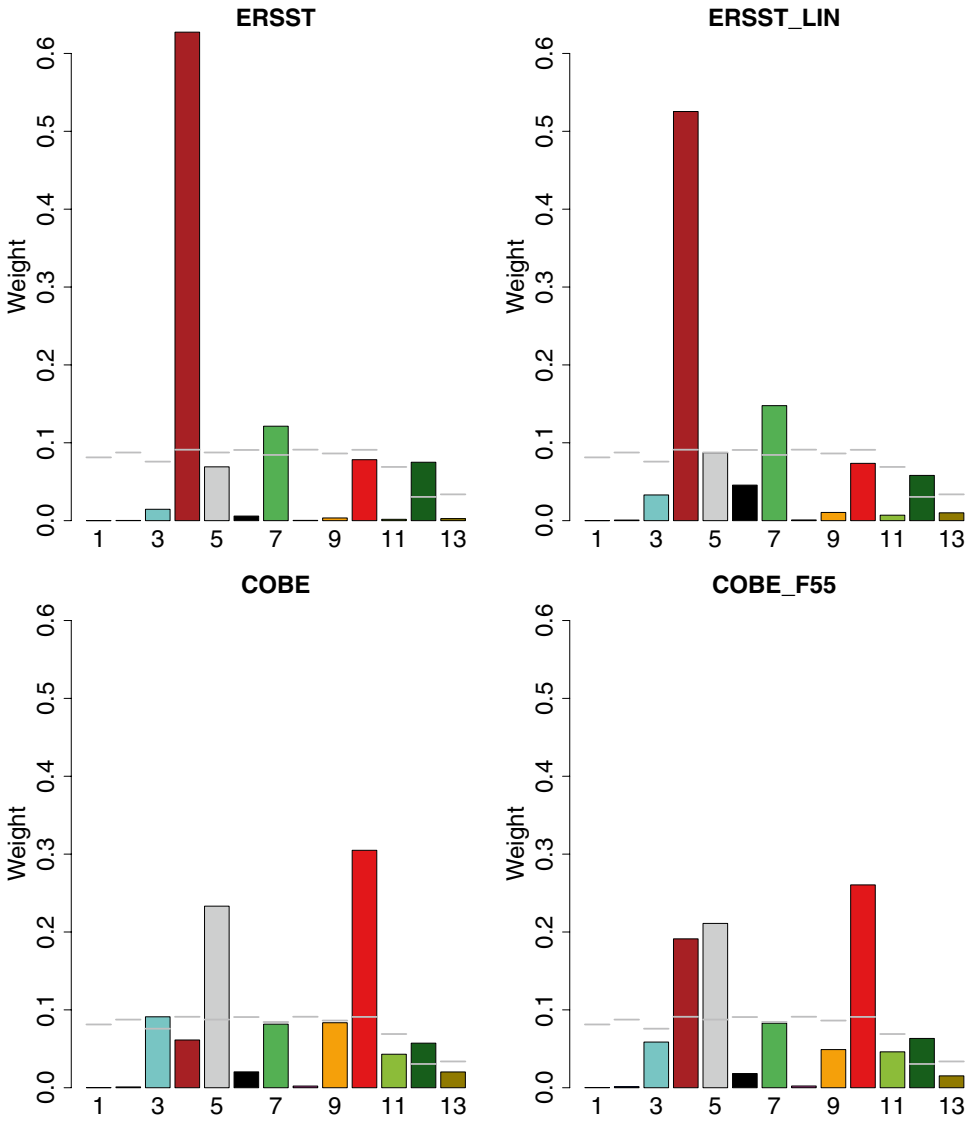

Fig. 10 Posterior model weights for the four experiments. Model indices correspond to Table 1. *Horizontal grey lines* model priors. *Color* scheme as in Fig. 2

3.2 Model weights

Model weights (and priors) are presented in Fig. 10. Generally, priors have a relatively small spread, with an exception of the two rightmost models. This indicates that most models have similar skill at modeling the absolute value of AMOC. However, the posterior weights clearly distinguish good from poor models. This indicates that in our study model skill at capturing the time evolution of the AMOC Index is the key to weighting the models. Specifically, many models are unable to capture the considerable decline in the AMOC Index (Fig. 2). Model AMOC errors can be influenced by a variety of sources, such as potential errors in the vertical diffusivity parameterizations (e.g., Goes et al. 2010), the Southern Ocean winds (Meijers 2014), as well as the errors more directly related to the North Atlantic climate (Zappa et al. 2014; Drews and Greatbatch 2016).

Models weights are strongly dependent on the dataset used, and are less sensitive to the smoothing method for model output. The ERSSTv4 observations show much more relative gyre cooling compared to the COBE-SST2 dataset (Fig. 9). CCSM4 has a strong decreasing AMOC Index trend consistent with the ERSSTv4 observations, therefore it gets by far the highest weight in the ERSST and ERSST-LIN experiments. The COBE-SST2 dataset shows considerably smaller decreases (Fig. 9). Two models that fit this dataset well are MPI-ESM-LR and CESM1-BGC (Fig. 10), although when smaller smoothing is used in the COBE-F55 experiment, CCSM4 also gets a considerable weight.

3.3 AMOC change projections

Our future projection pdfs for both scenarios are characterized by smooth, unimodal distributions (Fig. 11). One of the reasons for this is the large structural model error

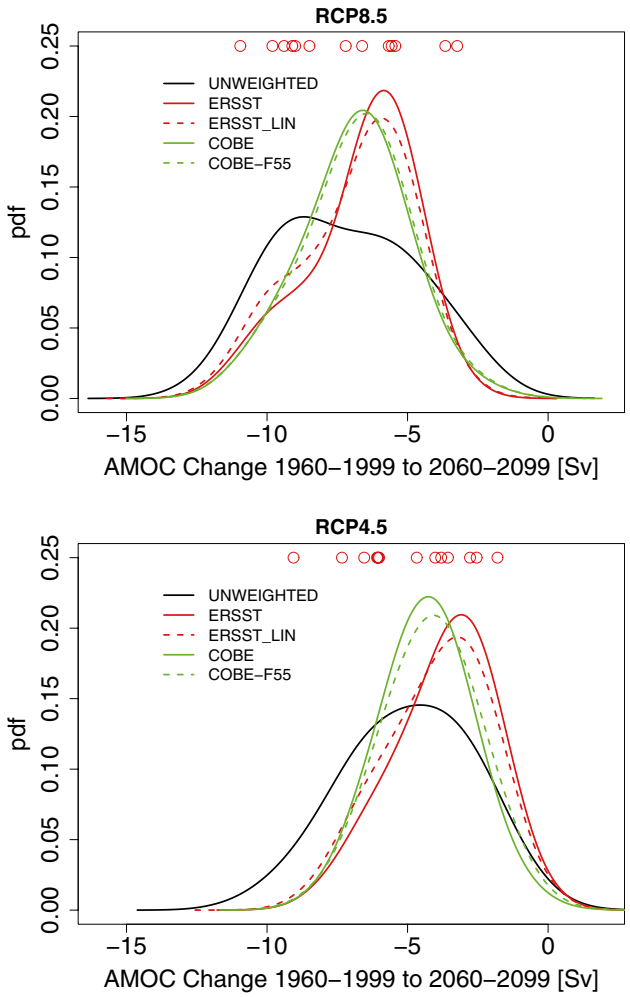

Fig. 11 Probabilistic AMOC change projections 1960–1999 to 2060–2099 [Sv] under RCP8.5 and RCP4.5 emissions scenarios for the weighted and un-weighted experiments. *Red circles* changes from individual models

Table 2 Summary of AMOC change projections for 1960–1999 to 2060–2099 under RCP4.5 emissions scenario

Experiment	Mean [Sv]	Median [Sv]	Mode [Sv]	90% credible interval [Sv]
Unweighted	**−4.9**	**−4.9**	**−4.6**	**−9.1, −1.1**
Mean of weighted	**−4.0**	**−3.9**	**−3.7**	**−7.2, −1.2**
ERSST	−3.7	−3.5	−3.1	−7.1, −0.87
ERSST-LIN	−3.9	−3.8	−3.2	−7.4, −0.97
COBE	−4.3	−4.3	−4.2	−7.2, −1.5
COBE-F55	−4.2	−4.1	−4.1	−7.2, −1.2

(as well as internal climate variability) that is added onto the projections, thereby broadening them. The presence of considerable error also explains why the tails of the distributions generally go beyond the range of the model ensemble. This indicates that bounding future projection pdfs to ensemble range can result in overconfidence.

Under the RCP4.5 scenario the AMOC is projected to weaken by 4 Sv, with a [−7.2, −1.2] 90% posterior credible interval (Table 2). Assuming the 1960–1999 mean can be well represented by 18.3 Sv (Kanzow et al. 2010), the mean decline is −22%, with the 90% interval from −6.6 to −39%. The lower bound is somewhat lower than the one shown in Schleussner et al. (2014). The projections are also lower compared to Bakker et al. (2016). The decrease of the lower bound can be attributed to using more models, and more thorough sampling of model structural error and internal variability. At the same time, as in Bakker et al. (2016), given the large model errors and presence of internal variability, the possibility of intensification cannot be completely discounted.

For the RCP8.5 scenario, the projections indicate larger declines compared to the RCP4.5 (Tables 2, 3). Under this scenario, a mean slowdown of around −6.8 Sv is expected, with a 90% posterior credible interval of [−10.5, −3.7] Sv. In percentage terms, the mean decline is −37%, with the 90% interval from −20 to −57% of the late twentieth century values. The effects of decrease by more than a half—a distinct possibility—can be extensive enough to modulate the annual cycle in the eastern equatorial Pacific and to lead to ENSO changes (Timmermann et al. 2007). These decreases are much stronger and wider compared to Chang et al. (2014) who use an ensemble of Earth model of intermediate complexity (EMIC) runs that varies vertical ocean diffusivity. Their analysis shows that increasing the dimensionality of the observational constraints from 1D to 3D mediates the projected AMOC decreases and reduces the uncertainty [but see Schmittner et al. (2009) for a case when 1D observations can better reduce model uncertainties]. However, they use just one EMIC, and the projections do not fully include model errors and internal variability. Our mean/median projections are in agreement with two recent studies (Weaver et al. 2012; Bakker et al. 2016). However, the median (−6.6 Sv/−36%) appears slightly lower than in Schleussner et al. (2014). Also, the

Table 3 Summary of AMOC change projections for 1960–1999 to 2060–2099 under RCP8.5 emissions scenario

Experiment	Mean [Sv]	Median [Sv]	Mode [Sv]	90% credible interval [Sv]
Unweighted	**−7.2**	**−7.4**	**−8.7**	**−11.3, −2.8**
Mean of weighted	**−6.8**	**−6.6**	**−6.2**	**−10.5, −3.7**
ERSST	−6.6	−6.3	−5.9	−10.5, −3.8
ERSST-LIN	−6.9	−6.5	−5.9	−10.7, −3.8
COBE	−6.9	−6.8	−6.6	−10.3, −3.7
COBE-F55	−6.8	−6.7	−6.5	−10.4, −3.6

upper bound of the 90% credible interval is lower than in previous studies (Schleussner et al. 2014; Bakker et al. 2016). Thus, the weighting cuts the upper tail of the projections (Fig. 11). It is difficult to compare our projections with other studies due to use of different forcing scenarios. Defining AMOC collapse as attaining a 0 Sv circulation (a decline of 18.3 Sv from present; Urban and Keller 2010; Bakker et al. 2016), we do not have any future projection samples showing a collapse. Same result is obtained when redefining it as a decline of more than 90% (or 16.5 Sv) from present-day climatology (Zickfeld et al. 2007). However, using <5 Sv (severely reduced circulation with a strength of about a quarter of present-day value) results in a collapse probability of 4.8×10^{-4}. To calculate this probability, we pool all posterior samples from the four RCP8.5 weighted experiments together. In light of poor knowledge of the model structural errors, we assess the collapse as exceptionally unlikely over the twenty-first century, where "exceptionally unlikely" is used in accordance with the Intergovernmental Panel on Climate Change (IPCC) guidance for treatment of uncertainty (Mastrandrea et al. 2010). This assessment is in agreement with previous work (Bakker et al. 2016).

Comparing weighted with un-weighted distributions illustrates the sharpening of the pdfs when the new information from observations is taken into account (Fig. 11; Tables 2, 3). There is a strong shift of the mode of the changes from −8.7 Sv for the unweighted RCP8.5 case to −6.2 Sv for the mean of the weighted cases. Also, cutting off of the high-impact low tail is a very robust feature and illustrates the advantages of the weighting. This is especially prominent in the RCP4.5 case: the low bound of the 90% credible interval shrinks by almost 2 Sv after the weighting. Moreover, the central tendencies of the weighted changes are considerably higher. Specifically, for RCP4.5 both mean and median projections are around 1 Sv higher than in the unweighted case. To analyze this behavior, we correlate past century-scale modeled AMOC changes with future changes (Fig. 9). Perhaps surprisingly, the models that project larger past declines in AMOC tend to project smaller future slowdown. The correlation is moderate in the RCP4.5 case, and weak in the RCP8.5 case. We test the sensitivity of the correlations to some assumptions. When we use 1965–1984 and 2060–2079 period for the recent past and the future respectively, the RCP4.5 correlation remains the same at −0.7, while the RCP8.5 correlation increases to −0.56 (moderate correlation). Finally, moderate and weak negative correlations for the RCP4.5 and RCP8.5 respectively remain after smoothing model output. The reason for such behavior is the subject of future work. Overall, the models that correctly capture the strong relative gyre cooling tend to exhibit stronger past AMOC declines, and those are the models that tend to produce smaller future declines. We also stress that the magnitude of the century scale model decline may not provide the whole story—the shape of the time-series may also be important for providing model weights. The relationships between past and future long term AMOC trends in the models, as well as the strong constraint provided by the time-resolved AMOC proxy, underscore the importance of continuous North Atlantic monitoring projects like RAPID (Smeed et al. 2014).

Finally, our results are not very sensitive to the choice of gyre SST temperature dataset, or the smoothing method (Fig. 11). Using ERSSTv4 observations, which exhibit stronger relative gyre cooling, we find somewhat weaker future declines. This is consistent with the correlation relationships described above.

4 Caveats

Our study is subject to several caveats. First, we use only a subset of available models, and just two observational datasets. Second, our observation system simulation experiment set-up is highly simplified and does not account for model errors that are common to all models. Third, we do not fully explore key parametric uncertainties, specifically in climate sensitivity (Stocker and Schmittner 1997; e.g.; Forest et al. 2006; Libardoni and Forest 2011; Olson et al. 2012). Stocker and Schmittner (1997) showed the paramount effect of climate sensitivity on the potential for future AMOC collapse using an EMIC. Fourth, we parameterize future structural model bias using a normal distribution. Quantifying model errors is an ongoing research area (Sexton et al. 2011), and in reality the error could follow a different distribution. Fifth, CMIP5 models used here do not include dynamic ice sheet models. This likely leads to underestimation of AMOC slowdown, depending in magnitude on the rate of future Greenland ice loss. There is considerable uncertainty both regarding this future mass loss and the sensitivity of the AMOC to freshwater input (Stouffer et al. 2006; Gierz et al. 2015; Bakker et al. 2016; Böning et al. 2016; Kandiano et al. 2016).

An important caveat concerns the possible abrupt response of the AMOC, related to its Stommel bifurcation point (Stommel 1961). AMOC stability (the existence of one or multiple equilibria under a given freshwater forcing) appears to be related to F_{ov}, the freshwater transport by the AMOC into the Atlantic basin across approximately 30–35°S (Alley 2007; Huisman et al. 2010; Hawkins et al. 2011; Srokosz et al. 2012). The proximity of the AMOC bifurcation threshold (if it exists) is only very poorly known and, due to high computational cost, has thus far been systematically investigated mostly in intermediate complexity models (Rahmstorf et al. 2005). However, a number of recent studies point to a bias of some current climate models

towards stable conditions too far away from this potential threshold (Hofmann and Rahmstorf 2009; Huisman et al. 2010; Weaver et al. 2012; Liu and Liu 2014). Here, out of the two models that get the highest weight in our experiments (CCSM4 and MPI-ESM-LR), CCSM4 operates in a monostable regime (for the RCP4.5 scenario), while MPI-ESM-LR has a bistable regime during the experiments and time periods considered (Weaver et al. 2012). Observations suggest a bistable regime (e.g., Huisman et al. 2010; Bryden et al. 2011). Thus, including AMOC stability or other physical AMOC-related variables such as freshwater fluxes (Kim and An 2013) into model weighing is the subject of future research.

Finally, we do not explicitly include any paleo-constraints on our models. Since abrupt AMOC-related changes are believed to have occurred in the past (Broecker 1997; Alley 2007; Negre et al. 2010; Stone et al. 2016), such constraints have the potential to shed new light onto future projections.

5 Conclusions

We weight 13 CMIP5 models by their skill at modeling late twentieth century Atlantic Meridional Overturning Circulation (AMOC), and the temporal evolution of a sea-surface temperature based AMOC Index for the years from 1880 to 2004. We show that century-scale changes in this Index are related to AMOC changes in the models. We use these weights to make probabilistic projections of AMOC change between the years 1960–1999 and 2060–2099. We improve on previous work through the use of time-dependent observational constraints, addition of drift-correction, as well as extensive implementation of model error, and internal variability. Furthermore, we achieve an approximately correct projection coverage during cross-validation.

Models consistent with observed AMOC Index decreases tend to project AMOC weakening over last century. Thus, our results are consistent with the emerging line of thought that AMOC has likely already started slowing down.

Time series of the AMOC Index provides a powerful constraint that separates the poor from the good models. Weighting the models considerably sharpens and moderates future AMOC change projections, while strongly reducing the high-impact low tails, compared to the un-weighted results.

Despite the sharpening after the weighting, the projections are at the low end of previously published estimates. This is the joint effect of using more models, and of more thoroughly accounting for internal variability, model and observational errors. The mean declines between 1960 and 1999 and 2060–2099 are 4.0 and 6.8 Sv, respectively, under the RCP4.5 and RCP8.5 scenarios. Under the RCP8.5 scenario, the emerging picture is that of near-certainty in the AMOC weakening, with a distinct possibility of over 50% slowdown by the end of the century. At the same time, we find that, given the caveats of this study, a chance of AMOC collapse is exceptionally unlikely. However, previous work indicates that the probability of triggering a collapse beyond the twenty-first century (e.g., committing to a future collapse) might be orders of magnitude higher than the probability of actually experiencing a collapse (Urban and Keller 2010). Further work includes improving the statistical method and the cross-validation, the choice and dimensionality of the observational constraints, as well as enlarging model ensembles and/or complementing them with perturbed parameter ensembles.

Acknowledgements For their roles in producing, coordinating, and making available the CMIP5 model output, we acknowledge the climate modeling groups (listed in Table 1), the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM), and the Global Organization for Earth System Science Portals (GO-ESSP). Some data used here has been downloaded from the German Climate Computing Centre (DKRZ), with funding from the Federal Ministry for Education and Research. Jong-Soo Shin and Eun-Sook Heo provided technical assistance with downloading model output, and code debugging, respectively. Fruitful conversations with Stefan Rahmstorf and Axel Timmermann are gratefully acknowledged. S.-I. An and R. Olson were supported by the National Research Foundation of Korea Grant funded by the Korean Government (MEST, NRF-2009-0093069).

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Alley RB (2007) Wally was right: predictive ability of the North Atlantic "conveyor belt" hypothesis for abrupt climate change. Annu Rev Earth Planet Sci 35:241–272. doi:10.1146/annurev.earth.35.081006.131524

Bakker P, Schmittner A, Lenaerts JTM et al (2016) Fate of the Atlantic Meridional overturning circulation: strong decline under continued warming and Greenland melting. Geophys Res Lett 2016GL070457. doi:10.1002/2016GL070457

Bhat KS, Haran M, Terando A, Keller K (2011) Climate projections using Bayesian Model averaging and space-time dependence. J Agric Biol Environ Stat 16:606–628. doi:10.1007/s13253-011-0069-3

Bhat KS, Haran M, Olson R, Keller K (2012) Inferring likelihoods and climate system characteristics from climate models and multiple tracers. Environmetrics 23:345–362. doi:10.1002/env.2149

Böning CW, Behrens E, Biastoch A et al (2016) Emerging impact of Greenland meltwater on deepwater formation in the North Atlantic Ocean. Nat Geosci 9:523–527. doi:10.1038/ngeo2740

Broecker WS (1997) Thermohaline circulation, the achilles heel of our climate system: will man-made CO_2 upset the current balance? Science 278:1582–1588. doi:10.1126/science.278.5343.1582

Bryden HL, Longworth HR, Cunningham SA (2005) Slowing of the Atlantic meridional overturning circulation at 25°N. Nature 438:655–657. doi:10.1038/nature04385

Bryden HL, King BA, McCarthy GD (2011) South Atlantic overturning circulation at 24°S. J Mar Res 69:39–56

Buckley MW, Marshall J (2016) Observations, inferences, and mechanisms of the Atlantic meridional overturning circulation: A review. Rev Geophys 54:2015RG000493. doi:10.1002/2015RG000493

Chang W, Haran M, Olson R, Keller K (2014) Fast dimension-reduced climate model calibration and the effect of data aggregation. Ann Appl Stat 8:649–673. doi:10.1214/14-AOAS733

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836. doi:10.1080/01621459.1979.10481038

Cunningham SA, Kanzow T, Rayner D et al (2007) Temporal variability of the Atlantic meridional overturning circulation at 26.5°N. Science 317:935–938. doi:10.1126/science.1141304

Drews A, Greatbatch RJ (2016) Atlantic multidecadal variability in a model with an improved North Atlantic current. Geophys Res Lett 43:2016GL069815. doi:10.1002/2016GL069815

Drijfhout S, van Oldenborgh GJ, Cimatoribus A (2012) Is a decline of AMOC causing the warming hole above the north Atlantic in observed and modeled warming patterns? J Clim 25:8373–8379. doi:10.1175/JCLI-D-12-00490.1

Forest CE, Stone PH, Sokolov AP (2006) Estimated PDFs of climate system properties including natural and anthropogenic forcings. Geophys Res Lett 33:L01705. doi:10.1029/2005GL023977

Gierz P, Lohmann G, Wei W (2015) Response of Atlantic overturning to future warming in a coupled atmosphere-ocean-ice sheet model. Geophys Res Lett 42:2015GL065276. doi:10.1002/2015GL065276

Goes M, Urban NM, Tonkonojenkov R et al (2010) What is the skill of ocean tracers in reducing uncertainties about ocean diapycnal mixing and projections of the Atlantic meridional overturning circulation? J Geophys Res Oceans. doi:10.1029/2010JC006407

Gregory J (2012) CMIP5 Model Ancestry Info from Jonathan Gregory, University of Reading. http://cmip-pcmdi.llnl.gov/cmip5/errata/jgregory_cmip5ancestry.txt. Accessed 9 Feb 2017

Hansen J, Ruedy R, Sato M, Lo K (2010) Global surface temperature change. Rev Geophys 48:RG4004. doi:10.1029/2010RG000345

Hawkins E, Smith RS, Allison LC et al (2011) Bistability of the Atlantic overturning circulation in a global climate model and links to ocean freshwater transport. Geophys Res Lett 38:L10605. doi:10.1029/2011GL047208

Hirahara S, Ishii M, Fukuda Y (2013) Centennial-scale sea surface temperature analysis and its uncertainty. J Clim 27:57–75. doi:10.1175/JCLI-D-12-00837.1

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Stat Sci 14:382–417

Hofmann M, Rahmstorf S (2009) On the stability of the Atlantic meridional overturning circulation. Proc Natl Acad Sci 106:20584–20589. doi:10.1073/pnas.0909146106

Huang B, Banzon VF, Freeman E et al (2014) Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part I: upgrades and intercomparisons. J Clim 28:911–930. doi:10.1175/JCLI-D-14-00006.1

Huang B, Thorne PW, Smith TM et al (2015) Further Exploring and quantifying uncertainties for extended reconstructed sea surface temperature (ERSST) version 4 (v4). J Clim 29:3119–3142. doi:10.1175/JCLI-D-15-0430.1

Huisman SE, den Toom M, Dijkstra HA, Drijfhout S (2010) An indicator of the multiple equilibria regime of the Atlantic meridional overturning circulation. J Phys Oceanogr 40:551–567. doi:10.1175/2009JPO4215.1

Kandiano ES, van der Meer MTJ, Bauch HA et al (2016) A cold and fresh ocean surface in the Nordic Seas during MIS 11: significance for the future ocean. Geophys Res Lett 43:2016GL070294. doi:10.1002/2016GL070294

Kanzow T, Cunningham SA, Johns WE et al (2010) Seasonal variability of the Atlantic meridional overturning circulation at 26.5°N. J Clim 23:5678–5698. doi:10.1175/2010JCLI3389.1

Kim H, An S-I (2013) On the subarctic North Atlantic cooling due to global warming. Theor Appl Climatol 114:9–19. doi:10.1007/s00704-012-0805-9

Kuhlbrodt T, Griesel A, Montoya M et al (2007) On the driving processes of the Atlantic meridional overturning circulation. Rev Geophys 45:RG2001. doi:10.1029/2004RG000166

Libardoni AG, Forest CE (2011) Sensitivity of distributions of climate system properties to the surface temperature dataset. Geophys Res Lett 38:L22705. doi:10.1029/2011GL049431

Liu W, Liu Z (2014) Assessing the stability of the Atlantic meridional overturning circulation of the past, present, and future. J Meteorol Res 28:803–819. doi:10.1007/s13351-014-4006-6

Liu W, Huang B, Thorne PW et al (2014) Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): part II. Parametric and structural uncertainty estimations. J Clim 28:931–951. doi:10.1175/JCLI-D-14-00007.1

Manabe S, Stouffer RJ (1988) Two stable equilibria of a coupled ocean-atmosphere model. J Clim 1:841–866. doi:10.1175/1520-0442(1988)001<0841:TSEOAC>2.0.CO;2

Manabe S, Stouffer RJ (1993) Century-scale effects of increased atmospheric CO_2 on the ocean–atmosphere system. Nature 364:215–218. doi:10.1038/364215a0

Mastrandrea MD, Field CB, Stocker TF et al (2010) Guidance note for lead authors of the IPCC Fifth assessment report on consistent treatment of uncertainties. Intergovernmental Panel on Climate Change (IPCC)

Meijers AJS (2014) The Southern Ocean in the Coupled Model Intercomparison Project phase 5. Philos Trans R Soc Lond Math Phys Eng Sci 372:20130296. doi:10.1098/rsta.2013.0296

Montgomery JM, Nyhan B (2010) Bayesian model averaging: theoretical developments and practical applications. Polit Anal 18:245–270. doi:10.1093/pan/mpq001

Moss RH, Edmonds JA, Hibbard KA et al (2010) The next generation of scenarios for climate change research and assessment. Nature 463:747–756. doi:10.1038/nature08823

NCAR (2017) Correct "branch_time": information for all NCAR CMIP5 datasets. http://www.cesm.ucar.edu/CMIP5/errata/branch_times.html. Accessed 9 Feb 2017

Negre C, Zahn R, Thomas AL et al (2010) Reversed flow of Atlantic deep water during the Last Glacial Maximum. Nature 468:84–88. doi:10.1038/nature09508

NOAA (2017) Extended reconstructed sea surface temperature (ERSST) v4. https://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v4

Olson R, Sriver R, Goes M et al (2012) A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. J Geophys Res Atmosp 117:D04103. doi:10.1029/2011JD016620

Olson R, Fan Y, Evans JP (2016) A simple method for Bayesian model averaging of regional climate model projections: application to southeast Australian temperatures. Geophys Res Lett 43:2016GL069704. doi:10.1002/2016GL069704

Pawlowicz R (2013) Key physical variables in the ocean: temperature, salinity, and density. Nat Educ Knowl 4:1:13

Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. Mon Weather Rev 133:1155–1174. doi:10.1175/MWR2906.1

Rahmstorf S, Crucifix M, Ganopolski A et al (2005) Thermohaline circulation hysteresis: a model intercomparison. Geophys Res Lett 32:L23605. doi:10.1029/2005GL023655

Rahmstorf S, Box JE, Feulner G et al (2015) Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. Nat Clim Change 5:475–480. doi:10.1038/nclimate2554

Reintges A, Martin T, Latif M, Keenlyside NS (2016) Uncertainty in twenty-first century projections of the Atlantic Meridional Overturning Circulation in CMIP3 and CMIP5 models

Schleussner C-F, Levermann A, Meinshausen M (2014) Probabilistic projections of the Atlantic overturning. Clim Change 127:579–586. doi:10.1007/s10584-014-1265-2

Schmittner A, Latif M, Schneider B (2005) Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations. Geophys Res Lett 32:1–4. doi:10.1029/2005GL024368

Schmittner A, Urban NM, Keller K, Matthews D (2009) Using tracer observations to reduce the uncertainty of ocean diapycnal mixing and climate–carbon cycle projections. Glob Biogeochem Cycles 23:GB4009. doi:10.1029/2008GB003421

Sen PK (1968) Estimates of the regression coefficient based on Kendall's Tau. J Am Stat Assoc 63:1379–1389. doi:10.2307/2285891

Sentman L (2016) ESMs CMIP5 FAQs. https://www.gfdl.noaa.gov/esms-cmip5-faqs/. Accessed 9 Feb 2017

Sexton DMH, Murphy JM, Collins M, Webb MJ (2011) Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. Clim Dyn 38:2513–2542. doi:10.1007/s00382-011-1208-9

Smeed DA, McCarthy GD, Cunningham SA et al (2014) Observed decline of the Atlantic meridional overturning circulation 2004–2012. Ocean Sci 10:29–38. doi:10.5194/os-10-29-2014

Srokosz M, Baringer M, Bryden H et al (2012) Past, present, and future changes in the Atlantic Meridional Overturning Circulation. Bull Am Meteorol Soc 93:1663–1676. doi:10.1175/BAMS-D-11-00151.1

Stocker TF, Schmittner A (1997) Influence of CO_2 emission rates on the stability of the thermohaline circulation. Nature 388:862–865. doi:10.1038/42224

Stommel H (1961) Thermohaline convection with two stable regimes of flow. Tellus 13:224–230. doi:10.1111/j.2153-3490.1961.tb00079.x

Stone EJ, Capron E, Lunt DJ et al (2016) Impact of meltwater on high-latitude early last interglacial climate. Clim Past 12:1919–1932. doi:10.5194/cp-12-1919-2016

Stouffer RJ, Yin J, Gregory JM et al (2006) Investigating the cause of the response of the thermohaline circulation to past and future climage changes. J Clim 19:1365–1387. doi:10.1175/JCLI3689.1

Terando A, Keller K, Easterling WE (2012) Probabilistic projections of agro-climate indices in North America. J Geophys Res Atmos. doi:10.1029/2012JD017436

Timmermann A, Okumura Y, An S-I et al (2007) The influence of a weakening of the Atlantic meridional overturning circulation on ENSO. J Clim 20:4899–4919. doi:10.1175/JCLI4283.1

Tomassini L, Reichert P, Knutti R et al (2007) Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods. J Clim 20:1239–1254. doi:10.1175/JCLI4064.1

Urban NM, Keller K (2010) Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: a Bayesian fusion of century-scale observations with a simple model. Tellus A 62:737–750. doi:10.1111/j.1600-0870.2010.00471.x

Vellinga M, Wood RA (2002) Global climatic impacts of a collapse of the Atlantic thermohaline circulation. Clim Change 54:251–267. doi:10.1023/A:1016168827653

Weaver AJ, Sedláček J, Eby M et al (2012) Stability of the Atlantic meridional overturning circulation: a model intercomparison. Geophys Res Lett 39:L20709. doi:10.1029/2012GL053763

Yeager S, Danabasoglu G (2014) The origins of late-twentieth-century variations in the large-scale North Atlantic circulation. J Clim 27:3222–3247. doi:10.1175/JCLI-D-13-00125.1

Zappa G, Masato G, Shaffrey L et al (2014) Linking Northern hemisphere blocking and storm track biases in the CMIP5 climate models. Geophys Res Lett 41:135–139. doi:10.1002/2013GL058480

Zickfeld K, Levermann A, Morgan MG et al (2007) Expert judgements on the response of the Atlantic meridional overturning circulation to climate change. Clim Change 82:235–265. doi:10.1007/s10584-007-9246-3