Climate Dynamics
https://doi.org/10.1007/s00382-018-4093-7

On the spectral characteristics of the Atlantic multidecadal variability in an ensemble of multi-century simulations

Irene Mavilia[1] · Alessio Bellucci[2] · Panos J. Athanasiadis[2] · Silvio Gualdi[2,3] · Rym Msadek[4] · Yohan Ruprich-Robert[5]

Received: 9 May 2017 / Accepted: 3 January 2018

Abstract

The Atlantic multidecadal variability (AMV) is a coherent pattern of variability of the North Atlantic sea surface temperature field affecting several components of the climate system in the Atlantic region and the surrounding areas. The relatively short observational record severely limits our understanding of the physical mechanisms leading to the AMV. The present study shows that the spatial and temporal characteristics of the AMV, as assessed from the historical records, should also be considered as highly uncertain. Using 11 multi-century preindustrial climate simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5) database, we show that the AMV characteristics are not constant along the simulation when assessed from different 200-year-long periods to match the observed period length. An objective method is proposed to test whether the variations of the AMV characteristics are consistent with stochastic internal variability. For 7 out of the 11 models analysed, the results indicate a non-stationary behaviour for the AMV time series. However, the possibility that the non-stationarity arises from sampling errors can be excluded with high confidence only for one of the 7 models. Therefore, longer time series are needed to robustly assess the AMV characteristics. In addition to any changes imposed to the AMV by external forcings, the detected dependence on the time interval identified in most models suggests that the character of the observed AMV may undergo significant changes in the future.

1 Introduction

In this study we focus on the Atlantic multidecadal variability (AMV). The AMV is defined as a coherent pattern of pseudo-oscillatory changes in the North Atlantic sea surface temperature record (0.4 °C range variation) with a dominant period in the 60–80 years range (Schlesinger and Ramankutty 1994; Kushnir 1994). These multidecadal variations have been observed in instrumental data (Enfield et al. 2001; Trenberth and Shea 2006; Deser and Blackmon 1993) and paleo-proxy records (Delworth and Mann 2000; Gray et al. 2004; Kilbourne et al. 2014; Saenger et al. 2009).

In past studies (e.g., Schlesinger and Ramankutty 1994) but also more recently (e.g., Latif et al. 2006 and; Ting et al. 2009), the AMV has also been considered to be part of the internal variability of the climate system, largely reflecting the intrinsic multidecadal variability of the North Atlantic meridional overturning circulation (Delworth and Mann 2000; Knight et al. 2005; Zhang 2008; Medhaug and Furevik 2011). However, other studies indicate that external factors, such as changes in greenhouse gas concentrations, anthropogenic aerosol concentrations, volcanic activity and solar insolation, could also play a role in forcing such variability (Otterå et al. 2010; Saenger et al. 2009; Booth et al. 2012; Swingedouw et al. 2013).

The idea of a mixed origin of AMV, due to the combination of internal climate variability and external forcing associated with human activity (aerosols, CO_2), is gaining ground (Rotstayn and Lohmann 2002; Terray 2012; Ting et al. 2014; Bellucci et al. 2017).

A number of studies have demonstrated the connection between the AMV and other components of the climate system both within and outside the Atlantic region. The AMV

✉ Irene Mavilia
i.mavilia@isac.cnr.it

1 Istituto di Scienze dell'Atmosfera e del Clima (CNR-ISAC), Bologna, Italy

2 Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy

3 Istituto Nazionale di Geofisica e Vulcanologia (INGV), Bologna, Italy

4 Centre National de la Recherche Scientifique (CNRS)/CERFACS, Toulouse, France

5 Barcelona Supercomputing Center (BSC), Barcelona, Spain

can, for instance, affect climate multidecadal variability over both North America and western Europe (Sutton and Hodson 2005; Knight et al. 2006; Sutton and Dong 2012; Ting et al. 2011; Zhang and Delworth 2006); it can be linked to precipitation changes over different regions (e.g. Sahel, India, North America and North Eastern Brazil), to variability in the Atlantic hurricane activity (Enfield et al. 2001; Goldenberg et al. 2001; Latif and Keenlyside 2011), and to Mediterranean sea surface temperature variability (Mariotti and Dell'Aquila 2012; Marullo et al. 2011).

Besides, the AMV has alternately obscured or amplified the global warming signal, which makes the attribution of global warming more difficult to ascertain (e.g. DelSole et al. 2011).

Despite its relevance for climate variability over the Atlantic sector and adjacent regions, the exact nature of the AMV needs to be understood, as its underlying mechanisms have not been adequately explained. Although past studies have shown similar ultra-low variability in models, an in-depth examination of the AMV long-term behaviour is currently missing.

Due to the shortness of the observational records, variability over multidecadal time scales is poorly understood. Since most of the related observational records go back no more than 150 years, too few multidecadal variation cycles of AMV can be detected; as a result, the multidecadal component of the variability spectrum is only marginally resolved. In addition, reliable deep ocean observations are even shorter, which makes the assessment of the role of ocean dynamics more difficult. In the present paper, we show that the period and amplitude of the AMV as assessed from the observational records should be considered uncertain. Special attention is paid to the possibility that AMV undergoes different regimes with different time scales and amplitude, compared to the present ones. The aim of this study is to analyse the possible existence of non-stationary aspects of the AMV.

In the climate science context, continuous modifications in the processes influencing a variable potentially result in a non-stationary behaviour where the statistical properties depend on the particular period in question. In literature, various studies, using both models and data, reveal this aspect. Based on a 2000-year simulation, performed with the global coupled GFDL-CM2.1 model, Wittenberg (2009) identifies different regimes in El Niño/Southern Oscillation (ENSO) modulations. Occasional epochs mimic detailed temporal sequences of observed ENSO events, but a different 150-year interval may not contain the same oscillation, suggesting that the observed ENSO may not be truly representative of ENSO longer-term behaviour, since it is not sufficient to resolve the temporal variability of ENSO beyond the most recent period. Similarly, Hurrell and Van Loon (1997) notice a non-stationary behaviour in the observed evolution of winter North Atlantic Oscillation (NAO) over the 1865–1994 period: a large fraction of the NAO variance around the biennial period comes from the early part of the record, while the variability between 6 and 10 years is more prominent over the second half of the twentieth-century. The possible non-stationarity of the AMV could have important consequences in terms of observational uncertainties and predictability, yet it has not been the focus of any previous study. Comparing different coupled general circulation models, Ting et al. (2011) find that the simulated AMV does not have a well-defined periodicity but varies across a range of decadal time scales: some models do not exhibit significant oscillatory behaviour, while for models with significant autocorrelation above the noise level the periods of oscillation range from 20 to 80 years. Such wide range of variability is also seen in the multi-model AMV spectra analysed by Ba et al. (2014), where most AMV indices exhibit a similar red noise character and show power spectral density on multidecadal time scales, but with different periodicity. In the present study we examine in depth this aspect by analysing the AMV behaviour in different epochs of the model simulation and applying a statistical test to evaluate that behaviour. Moreover, a multi-model approach for AMV analysis allows us to investigate similarities and differences across a large ensemble of 11 state-of-the-art coupled climate models.

The paper is organised as follows: the models, simulations and observations that are used are described in Sect. 2. The AMV spatial and temporal patterns, as represented by the models examined in this study, are presented in Sect. 3, together with the phenomenological evidence of the non-stationary AMV behaviour, using a time-slice approach for spectral analysis. Section 4 focuses on the statistical method applied to test AMV stationarity (with further details given in the Appendix). Finally, a summary and discussion of the results are presented in Sect. 5.

2 Models and data

To explore the stationarity of the AMV we analysed a multi-model set of simulations, mostly performed under the Coupled Model Intercomparison Project Phase 5 (CMIP5) framework. In order to remove any non-stationarity associated with the external forcing, we used preindustrial simulations. These experiments reveal the unforced variability of the model, imposing constant pre-1850 conditions: they have prescribed non-evolving atmospheric concentrations of all well-mixed gases (including CO_2), natural aerosols and some short-lived (reactive) compounds, while land use is unperturbed with a fixed agriculture disturbance mask (Taylor et al. 2012). To properly resolve the multidecadal range of the full variability spectrum, long multi-century simulations were required. Based on these criteria, 11 model simulations were selected (Table 1). Out of these, ten models are from the CMIP5 archive. In addition, a 3500-year-long integration performed with the GFDL-CM2.1 model, not available in the CMIP5 archive, was analysed. The CMIP5 protocol for preindustrial (piControl) experiments requires that an output of 500 years is supplied after a 100 years' spin-up, in order to avoid initial transients. For the same purpose, the first 500 years of GFDL-CM2.1 data are not used in the analysis. The Met Office Hadley Centre's sea ice and sea surface temperature (SST) data set, HadISST (Rayner et al. 2003), from 1870 to 2014 was used to validate the model results.

3 Simulated Atlantic multidecadal variability

3.1 Spatial and temporal patterns

For each model the AMV index is calculated as the annual mean SST area-averaged over the region [0°–60°N and 75°–7.5°W], as in Sutton and Hodson (2005). Prior to the analysis, the long-term mean is removed and the resulting time series of SST anomalies are linearly detrended in order to remove spurious long-term changes associated with residual model drifts, which are not physically meaningful. In order to isolate the low-frequency variability, SSTs are low-pass filtered with a 10-year running mean (low-pass Butterworth filters of different orders are also tested to filter the time series, with and without zero-phase filtering, yet the presented results are not affected by such choices). The observed AMV index is calculated in the same manner as for the models; nevertheless, the comparison with models is not completely consistent because the observed record is comparatively shorter and, more importantly, it also accounts for external forcings, which are excluded in the model simulations. The applied detrending partially removes the warming trend associated with changes in the radiative forcing on the observed SSTs, yet the likely influence of other non-linear forcings on the SST cannot be completely filtered out from observations. In the following, we use the general term NASST index to indicate the unfiltered, detrended time series and keep the term AMV index to refer to the low-pass filtered NASST index.

In this section we characterise the variability of the NASST, with a spectral analysis of its unfiltered annual mean time series, performed via a multi-taper method (MTM; Thomson 1982). This methodology does not prescribe an a priori model for the process generating the time series that are analysed, while it attempts to reduce the variance of spectral estimates by using a small set of tapers rather than the unique data taper, or spectral window. In Fig. 1 the MTM power spectra of the simulated and observed NASST time series are shown in colour with the corresponding red noise spectra overlapped as thin black lines. The frequency bands exhibiting power above the red noise level identify the time scales for which the NASST deviates from the red noise null-hypothesis, which in turn postulates that the low-frequency variance is entirely determined by the integration of the atmospheric white noise (Hasselmann 1976). Focusing on periodicities longer than 10 years, at multidecadal time scales, spectral peaks above the red noise threshold can be seen in all the models and in observations. However, for some models (especially for the ones with the shortest simulations) there are no spectral peaks exceeding the 90% confidence level, though the 80% confidence level is still exceeded by some peaks, which is

Table 1 Eleven multi-century preindustrial simulations that participate in this study

Model name	Length (years)	Ocean grid		Atmospheric grid	
		(Lat × Ion, degree)	Vertical levels	(Lat × Ion, degree)	Vertical levels
MPI-ESM-P	1156	1.5 × ~ 1.5	L40	1.8653 × 1.875	L47
CCSM4	1051	0.27, 0.54 × 1.11	L60	0.942 × 1.250	L26
MPI-ESM-LR	1000	1.5 × ~ 1.5	L40	1.8653 × 1.875	L47
CanESM2	996	0.9303, 1.1407 × 0.40625	L40	2.7906 × 2.8125	L35
NorESMl-M	501	~ 1 × ~ 0.5	L53	1.8947 × 2.5	L26
GFDL-ESM2M	500	0.334, 1 × 1	L50	2.0225 × 2.5	L24
MRI-CGCM3	500	0.5, 0.5 × 1	L50	1.12148 × 1.125	L48
ACCESS1-0	500	0.334, 1 × 1	L50	1.25 × 1.875	L38
ACCESS1-3	500	0.334, 1 × 1	L50	1.25 × 1.875	L38
CESM1-BGC	500	0.27, 0.53 × 1.125	L60	0.9424 × 1.25	L26
GFDL-CM2.1	3500	0.334, 1 × 1	L50	2.0225 × 2.5	L24

In case of atmospheric grid and its latitude, the tabulated resolution is only valid for the equator region. For higher latitudes, deviations may occur. If two values are given for latitude resolution of the ocean grid, the first value is that for the equator and the second for the poles

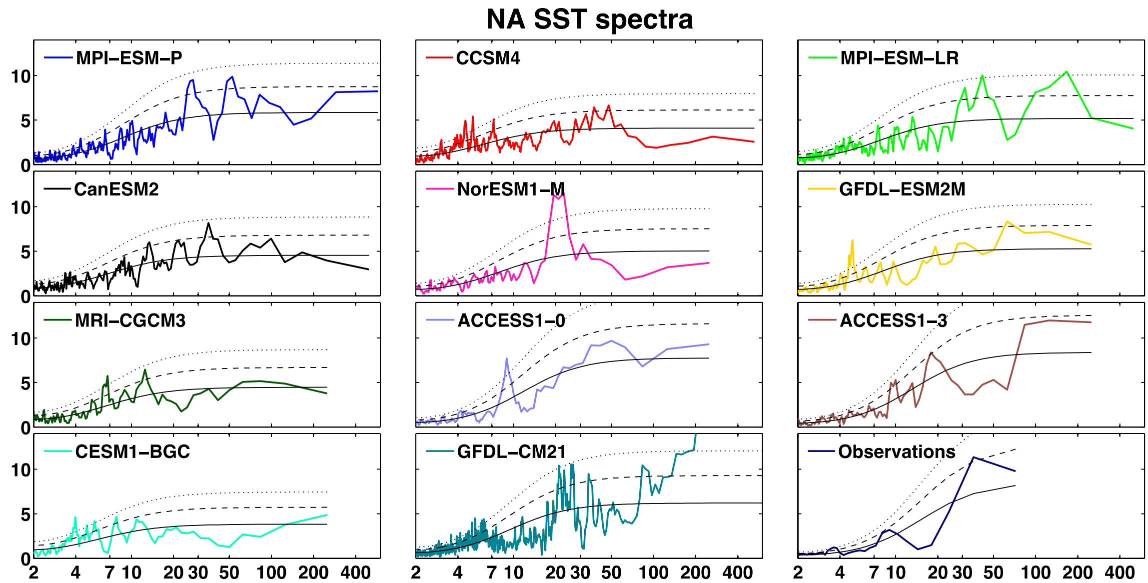

Fig. 1 Multi-taper (MTM) power spectral density estimates of North Atlantic SST for the analysed models and the HadISST observations from 1870 to 2014. The black thin lines are the AR1 red noise fits, dashed and dotted lines are the 80 and 90% confidence levels, respectively

consistent with similar findings of previous multi-model assessments (Zhang and Wang 2013; Zanchettin et al. 2014). The observations present more power than a red noise process for periods longer than 30 years, but the observed spectrum is not resolved sufficiently well due to the shortness of the data record. Flat low-frequency tails of the spectra are seen in CCSM4, MRI-CGCM3 and CESM1-BGC models, while peaks are more distinguishable for the rest of models. Depending on the model considered, the power spectrum of NASST index shows peaks at different frequencies and with different amplitudes. This wide range of variability is also found in Ba et al. (2014) multi-model comparison.

Figure 2 illustrates the spatial patterns of SST associated with the AMV index for the models and the HadISST observations over the North Atlantic region. The point-wise regression patterns between the global SST field and the AMV index are shown. Although there are some differences among the models, their spatial SST patterns resemble the typical basin-wide, comma-shaped pattern described by Ting et al. 2011. The majority of the models also show low or slightly negative anomalies in the Gulf Stream and the Nordic Seas areas. Most of the commonalities among models and observations stem from the slightly positive regression loadings seen in the tropical Atlantic (except for the MRI-CGCM3 model) while north of 30°N the patterns show generally larger differences.

Next, we inspect the individual AMV indices: the amplitudes (standard deviation) are shown in Fig. 3a and the corresponding time series in Fig. 3b; note that the GFDL-CM2.1 AMV index is divided into 1200-year long segments in order to better discern its features and facilitate the comparison with the other models. A visual inspection suffices for spotting that the modelled AMV can exhibit significant changes from period to period. Focusing on the MPI-ESM-P simulation, as an example, different AMV characters can be grossly grouped into 4 categories: (a) mostly warm-skewed, (b) nearly-harmonic, (c) intense with longer periods, and (d) small amplitude AMV evolution (Fig. 4). Besides the AMV time scale, the amplitude of the AMV signal is also subject to variations from one period to another. Different epochs with different characteristics and duration can be identified in the evolution of AMV index time series.

3.2 Evidence for non-stationary behaviour

Stationarity is a form of temporal homogeneity used in time series analysis, which is defined as time-invariance of the whole probability distribution of the data generating process (strict stationarity), or just of its first two moments (weak stationarity) (e.g. Brockwell and Davis 1991).

From now on, the word "stationarity" refers to the weak stationarity that requires the process to have the same mean and variance at each time and an autocorrelation function that depends on the time lag only. A time series is only one possible outcome of the underlying stochastic process that generated the data part. In practice, we can only have data for finite time spans; therefore, even to check these definitions, we have to make approximations (Huang et al. 1998).

In light of Fig. 4, North Atlantic SST variability is analysed for different epochs. An objective method is used to select epochs, which consists in dividing the AMV index time series into equal-length intervals. The length of these

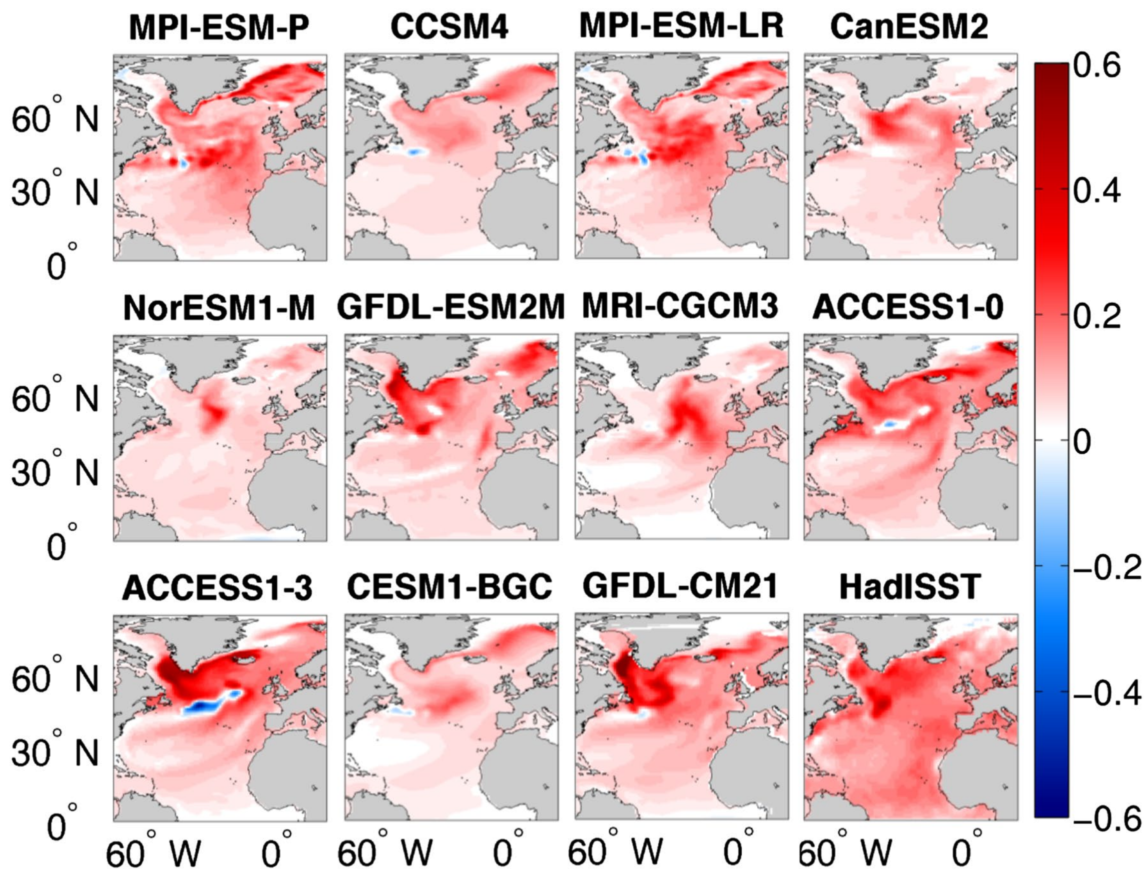

Fig. 2 Spatial patterns of SST variations associated with the AMV index for the models and the HadISST observations from 1870 to 2014. Shown are the regression coefficients (°C per standard deviation) obtained by regressing the 10-year running mean filtered and detrended SST data on a normalised (unit variance) version of the index

intervals is chosen to be 200 years to be approximately comparable to the length of the observational record. In order to underline the AMV non-stationary behaviour, multi-taper spectral analysis is carried out for chunks of consecutive, overlapping 200-year-long intervals, shifted by 50 years; for both models and observations. Figure 5 shows the results for the longest CMIP5 models, for illustrative purposes, while the other models present an analogous non-stationary behaviour. Spectral analysis is performed for each interval, for the entire time series and for observations. Each power spectral density is scaled by its maximum. The first interval refers to the period from year 1 to year 200, the second interval refers to the period [50–250] years, the third to [100–300], and so on.

This time-slice/moving-window view of AMV spectral features highlights a strong dependence on the selected time interval. For example, looking at the MPI-ESM-P model, we can see that the AMV during the [450–650] time interval, with central year 550, displays almost no variability for periods shorter than 20 years, while during the [900–1100] time interval the maximum variability is found at 20 years. Therefore, the dominant periods (inverse frequencies) can differ a lot depending on the selected intervals. The spectrum of the observed index resembles, for example, the spectrum of the period [200–400], as both show enhanced power in the 30–50 years band. However, the observed spectrum differs substantially, for example, from the spectrum of the period [900–1100], which presents the highest power at higher frequencies (periods shorter than 30 years). It is therefore plausible that the observed AMV spectrum represents only one of the several possible characters that emerge in long model data sets. If that is so, in the future the AMV could enter a regime remarkably different from that of the twentieth century. A similar approach is used by Saenger et al. (2009) in a 440-year-long proxy-record of Atlantic SST anomalies. The authors find that some periods show multidecadal power similar to the currently observed AMV, whereas no significant multidecadal power is evident in other past epochs.

Transitions among these behaviours of different variability also emerge from the AMV autocorrelation, diagnosed with a similar moving window (Fig. 6). For all the analysed models, the AMV time series is split again in 200-year-long intervals, shifted by 50 years, and the autocorrelation function is computed for each interval.

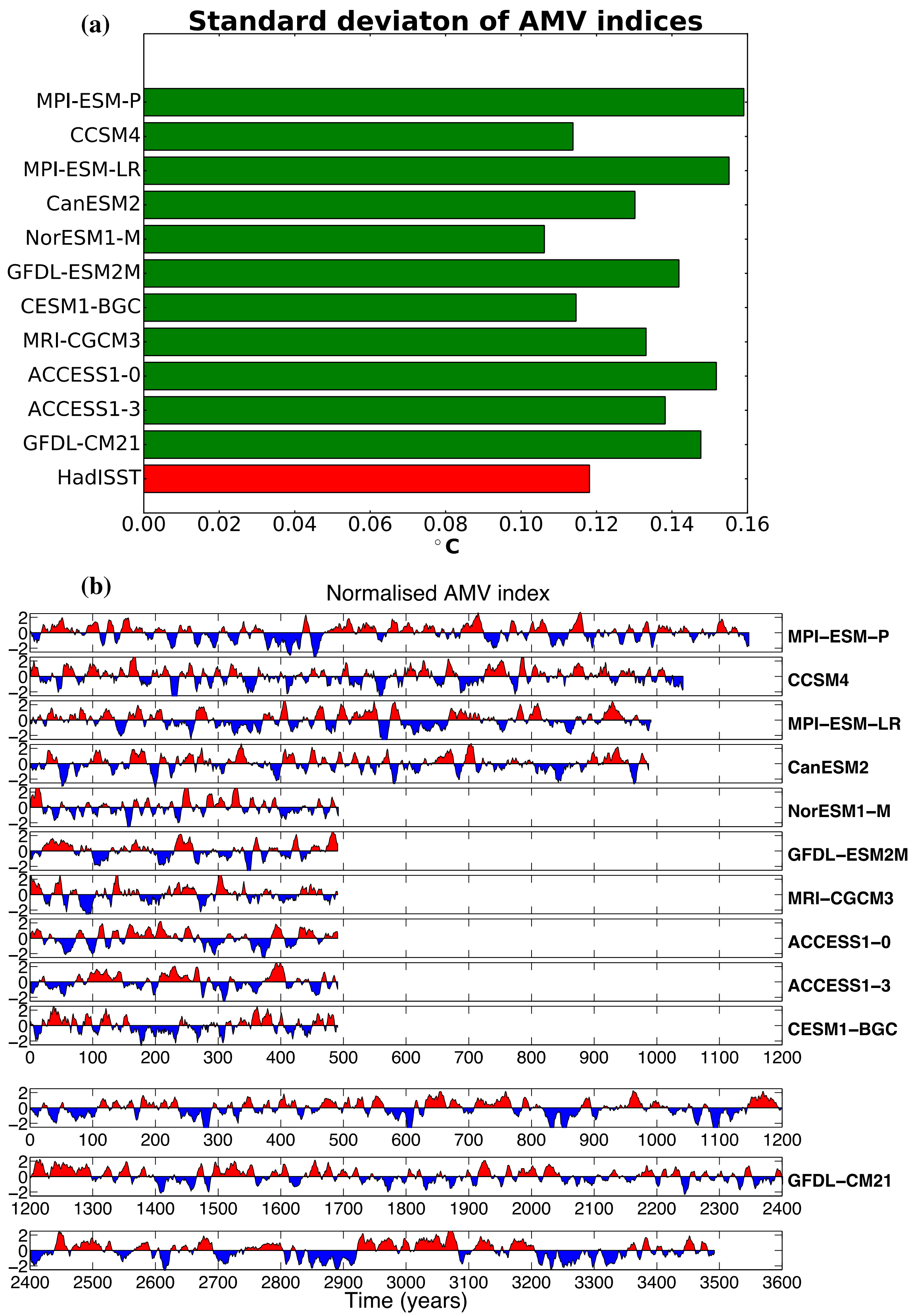

Fig. 3 **a** Standard deviations of AMV indices for both models (in green) and HadISST data from 1870 to 2014 (in red). **b** Normalised AMV index in the CMIP5 preindustrial simulations (top) and in the GFDL-CM2.1 model (bottom)

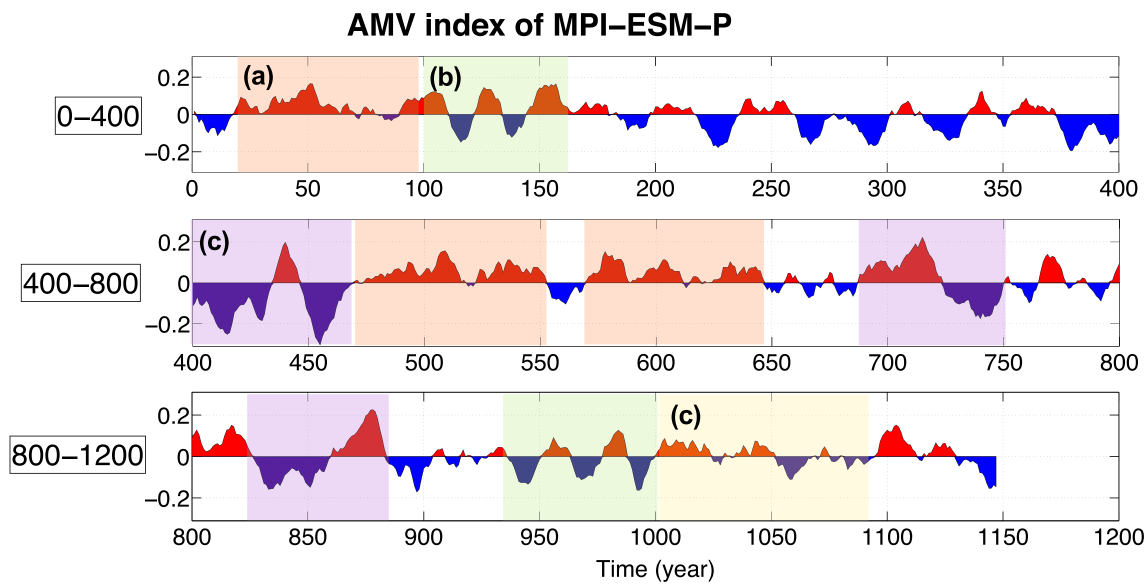

Fig. 4 AMV modulation shows epochs with different characteristics. (a) RED: mostly warm-skewed events; (b) GREEN: moderate, nearly-harmonic events; (c) PURPLE: intense events with longer periods; (d) YELLOW: small amplitude events

For most models, the dominant time scale of the autocorrelation function changes with the time interval, revealing a modulation of AMV that is irregular and interval-dependent. For the MPI-ESM-LR model, a ~ 50-year time scale characterises the initial part of the record, later shortening down to ~ 25 years, as indicated by the sloping contours, where time scale can be seen as the lag distance between autocovariance crests or troughs. Other models display no preferred time scales for long segments of their evolution (see in particular CCSM4 and ACCESS1-3 models). On the other hand, NorESM1-M model shows an almost time-invariant autocorrelation throughout the whole length of the model record. The present results are consistent with Zanchettin et al. (2010) who show that observed and simulated time series of a climate index covering several centuries are often characterised by fluctuations on time scales that are not homogeneously distributed in time: they appear within irregularly intermittent temporal intervals, whose duration varies, in general, with the signal fluctuation frequency.

To conclude, the dominant periods/frequencies depend not only on the model, but also on the period considered. Does this modulation arise by chance? In other words, does the AMV show these variations either because of its random behaviour or because we only consider a limited time interval? We address these questions in the next section, by objectively analysing the non-stationary behaviour of the AMV in the 11 simulations, using suitable statistical tests.

4 A statistical test for AMV stationarity

The analysis of AMV in the multi-model set of preindustrial integrations, illustrated in the previous section, revealed the succession of epochs characterised by different spectral characteristics.

Here, a statistical test to rigorously check the AMV stationarity is described. The null hypothesis (H_0) is that the AMV is statistically stationary, and any differences that we see between different epochs are random and emerge due to undersampling. In such a system, apparent "transitions" among different regimes of variability would have no underlying "cause". It would be equivalent to a long run of heads or tails when flipping a fair coin. Theoretically, stationarity is defined as a quality of a process in which statistical parameters (mean and standard deviation) do not change with time. The most important property of a stationary process is that the autocorrelation function depends only on the lag alone and thus it does not change with the time at which the function is calculated. Therefore, a practical test for stationarity is to verify whether there is any interval for which the statistical properties (e.g. autocovariance) of the entire time series are significantly different. Jointly with that, we need also to exclude the possibility that the AMV non-stationarity in the models we analyse might arise from sampling errors. In fact, a rejection of H_0 suggests that the AMV is not stationary, in other words that the related differences are so large that cannot be explained by randomness and undersampling.

The resampling methods are useful to generate random time series that have similar properties to the members of the population from which the simulated AMV time series is

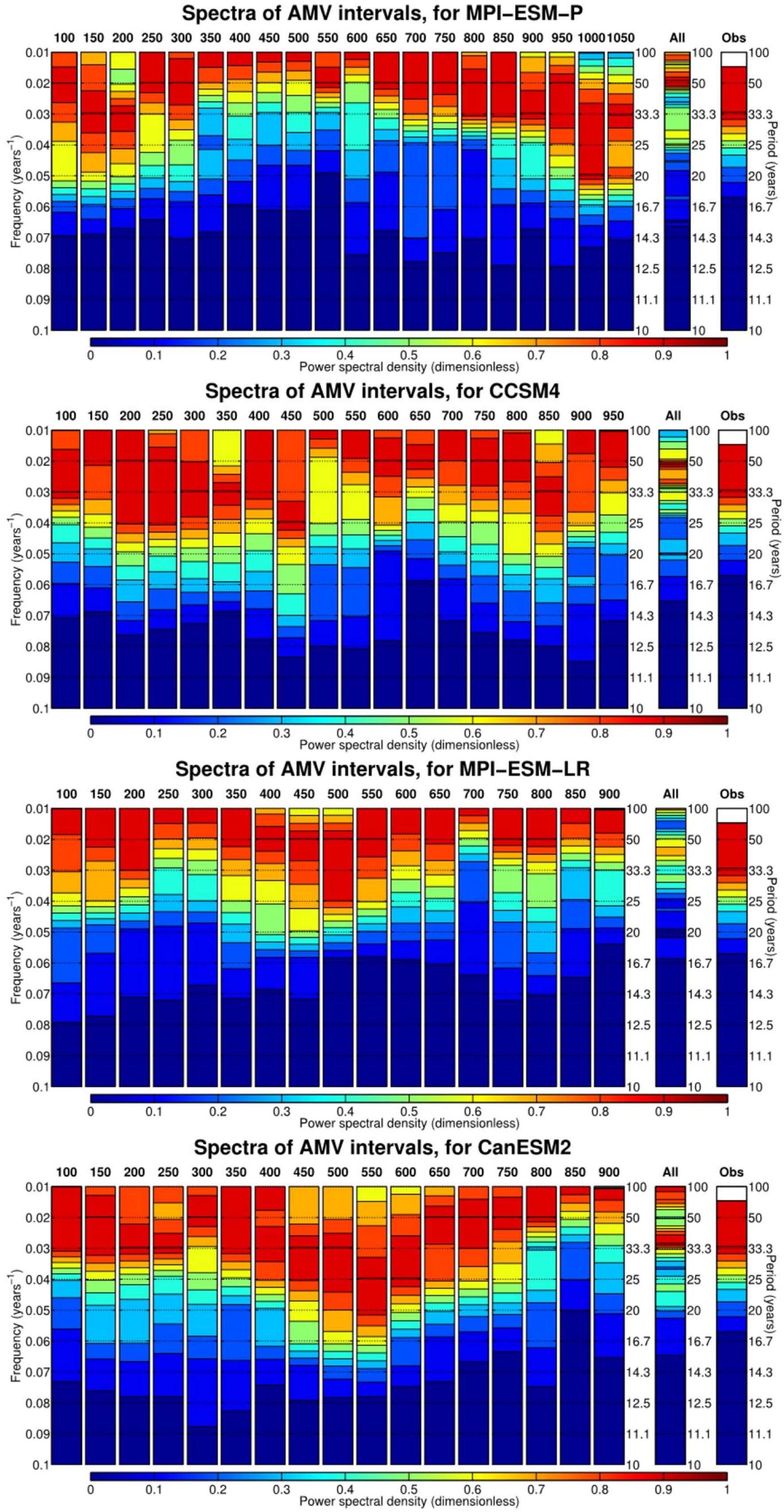

Fig. 5 Spectral signature of AMV index by using the multitaper method for 200-year-long intervals labelled with the central year (shown on top) and for the entire time series (indicated with "All" on top), for MPI-ESM-P, CCSM4, MPI-ESM-LR and CanESM2. "Obs" bin refers to HadISST data from 1870 to 2014

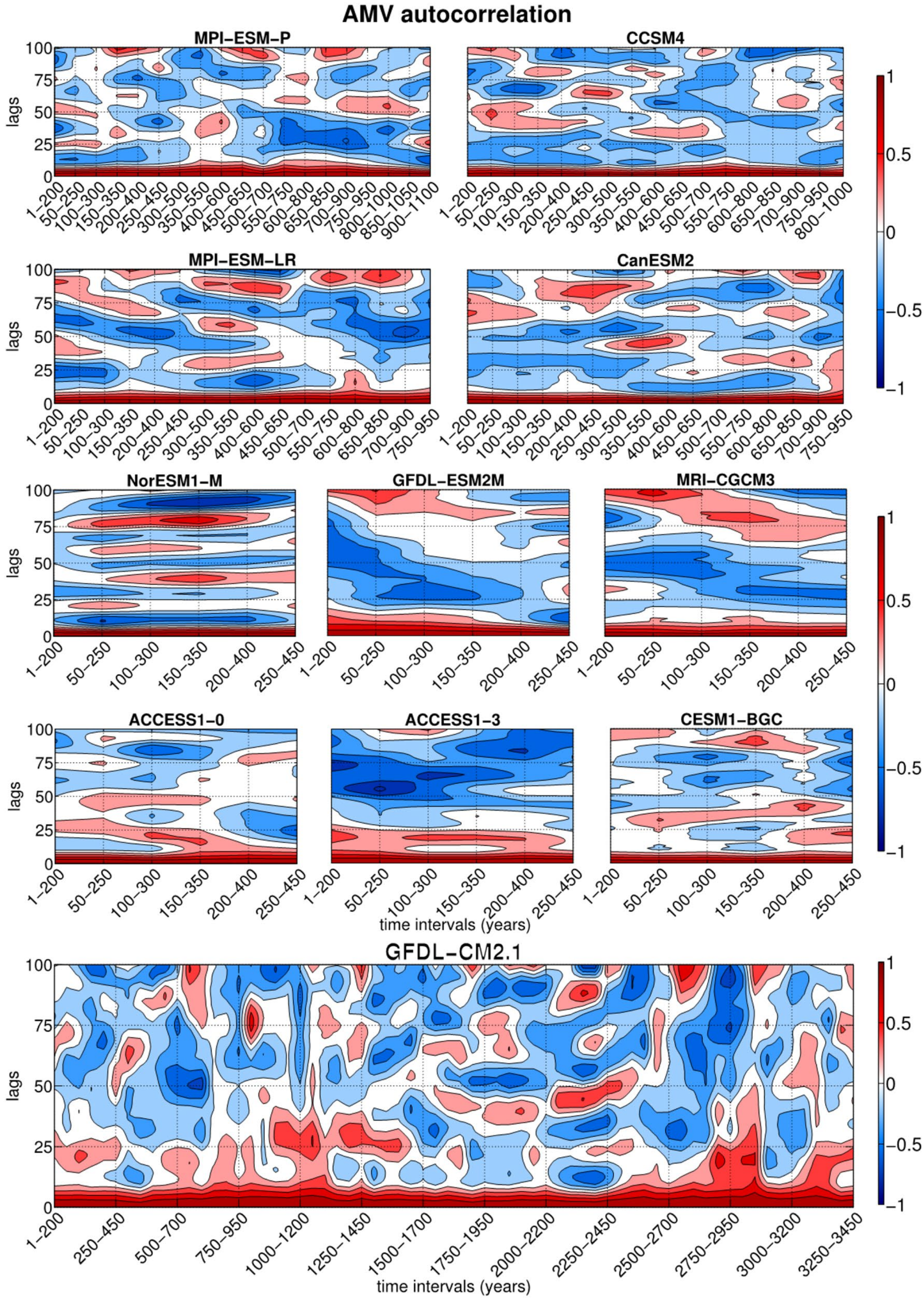

Fig. 6 Autocorrelation function of AMV index for moving and overlapped 200-year-long time windows, for the analysed models

taken. “Random-phase” test is applied to the analysed AMV time series. This technique is considered “resampling” in the frequency domain, as it does not preserve the distribution of values but rather the power spectrum. Therefore, the resampled series retains the same autocorrelation as the model series (Ebisuzaki 1997). First, the Fourier Transform (FT) of the model time series is computed. Then, the phases of the individual components of the FT are perturbed a number of times by adding a random number. This step creates a large number of new Fourier series with random phases at each frequency. Since the amplitude at each frequency is kept as in the model time series, the power spectrum is preserved. Then, back to the time domain with the Inverse Fourier Transforms (IFT), the new synthetic time series are obtained: they differ from the model series because they have modified phases in each of their Fourier components, but they have the same length and the same power spectrum (autocorrelation) of the model NASST time series. Once a large number of such synthetic series are created, the latter are used to compute the autocovariance range that contains 90% of the respective distribution, taken symmetrically around the respective mode (the 90% central range). Then, making use of this range, computed separately for each model, the respective stationarity is robustly assessed. Below, a concise description of the method used to test the AMV stationarity is provided.

The test applied to check the stationarity of the NASST and AMV time series involves the following steps:

1. 1000 synthetic time series are generated with the same length and the same spectrum as the model time series (bootstrap method) (Ebisuzaki 1997).
2. The 1000 synthetic time series are divided in 200-year-long intervals (length similar to the observations) shifted by 1 year.
3. The autocovariance function is computed for each of the 200-year-long intervals and for each of the 1000 synthetic time series, with a maximum lag equal to 60 years.
4. The distribution of the autocovariance values of the synthetic time series is used to compute the respective 90% central range. This range is defined by the 5th and the 95th percentiles.
5. The relative frequency of occurrence of autocovariance values of the original model time series (not synthetic) falling outside the defined range is determined (these counts are indicated with circles and triangles in Fig. 7).
6. The same counts as step 5 are computed also for the 1000 synthetic time series themselves, in order to show the range of variability covered by the 1000 counts and to see whether the model time series counts fall inside or outside of this range of variability (this range of variability is accounted by the error bars in Fig. 7).

In the Appendix, a more detailed explanation of the method is given using the MPI-ESM-P model as an example. Here, concise results are presented for the entire set of models.

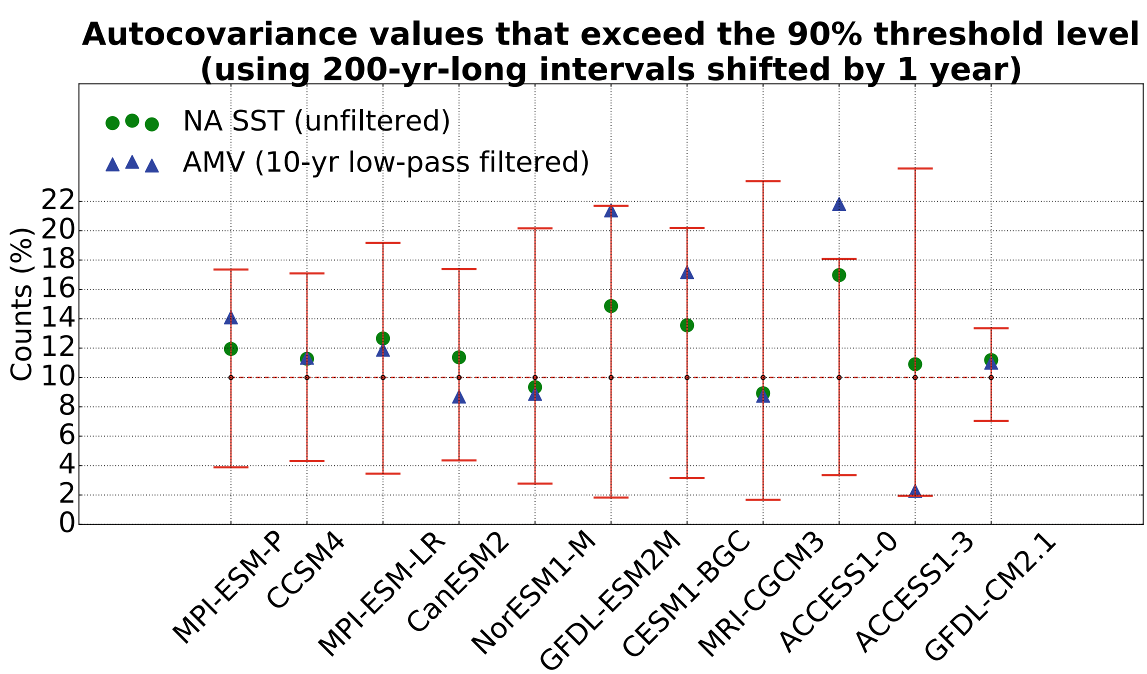

Fig. 7 Percentages of autocovariance values that exceed the 90% central range computed from the distribution of 1000 synthetic time series, for all the NASST and AMV time series. When the autocovariance function of the intervals departs from the autocovariance of the entire time series more than 10% of times (red dashed line), the indices present a non-stationary behaviour. The red error bars on top of the red dashed line represent the undersampling error due to the limited length of the time series

For each model, interval and lag, the respective AMV autocovariance value is compared to the corresponding 90% central range for that lag (the grey shading envelope in Fig. 10). The number of values found outside this range is counted. As a result, taking into account the total number of autocovariance values for the model time series (60 lags multiplied by the number of intervals), all models present AMV autocovariance values non contained in the 90% central range. This means that the autocovariance function of certain intervals differs significantly from the autocovariance function of the entire time series. For 9 out of 11 models, considering the unfiltered NASST (green circles in Fig. 7) the respective counts exceed 10% of the total autocovariance values, while for the actual AMV index (10-year low-pass filtered) (blue triangles in Fig. 7) the same occurs for 7 out of 11 models. This difference may be understood by considering the sub-decadal variance eliminated by the low-pass filter. There is a substantial difference in the behaviour of the ACCESS model in its 1.0 and 1.3 versions. From Fig. 7, it seems that the non-stationary behaviour of the 1.3 version might be largely due to the high-frequency component of the variability; whereas the 1.0 version appears to be quite unaffected by the low-pass filtering of the SST.

The rejection of H_0 for MPI-ESM-P, CCSM4, MPI-ESM-LR, GFDL-ESM2M, CESM1-BGC, ACCESS1-0 and GFDL-CM2.1 models, suggests a non-stationary behaviour for these AMV indices. However, the red dashed line at 10% in Fig. 7 indicates the number of autocovariance values that should be outside of the 90% central range for a stationary process of infinite time length. Since we are not dealing with infinite time series, we have to estimate the respective error due to the limited sampling. The sampling error is represented by the error bars in Fig. 7: for each of the 1000 synthetic time series we counted how many autocovariance values exceed the 90% central range (as we did for the model time series) then the inferior and superior limits of the error bars are the 5th and the 95th percentile of these 1000 counts.

The error bars are quite big for almost all the models, being the smallest for the longest time series (GFDL-CM2.1, 3500-year-long). Only the AMV index simulated by the ACCESS1-0 model stands outside the error bar, meaning that the variations of the AMV characteristics cannot be explained solely by randomness and undersampling. This implies that the non-stationarity of the AMV must arise from a physical mechanism. For the remaining 10 models, it cannot be confidently excluded that the apparent non-stationarity is the result of undersampling a stochastic process. The length of the model records affects the statistical test since short records provide fewer intervals. The test is also sensitive to the number of lags used for computing the autocovariance, this number is limited by the length of the intervals. Each model may "require" different number of lags (not necessarily 60 years) depending on the corresponding dominant variability time scale, as shown by the big differences in the power spectra in Fig. 1.

The reduction of the sampling error with the time series length is highlighted in Fig. 8. Here, the longest AMV index (from GFDL-CM2.1 model, 3500 years) is made longer in order to answer the question: how long should our simulated or observed time series be in order to properly estimate the AMV autocovariance robustly?

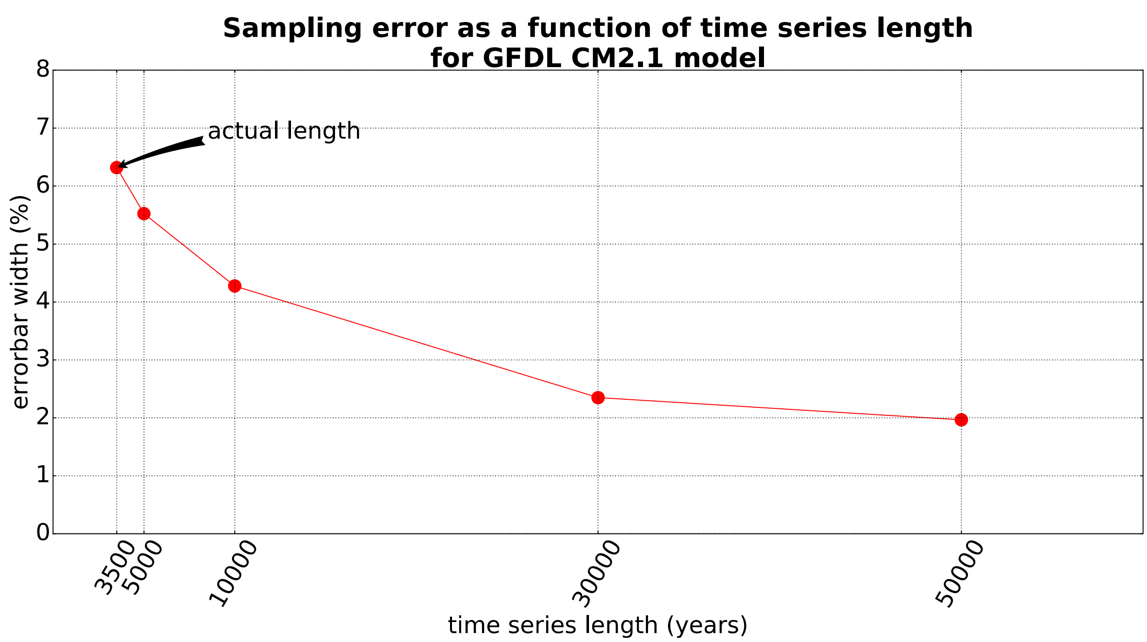

Fig. 8 Sampling error as a function of the length of GFDL-CM2.1 AMV time series

The length of the synthetic time series is related to the frequencies resolved. Therefore, the length of the time series can be increased by adding frequency components in the Fourier space and setting the extra frequencies to zero power. This way, going back to the time domain, a longer time series is produced. This "extended time series" still keep the same spectrum as the model time series (except for the extra frequencies carrying zero power). The analysis has been repeated using 200-year-long intervals shifted by 1 year as before. It was found that the sampling error (the red error bar in Fig. 7) decreases from a width of 6.3–2.0% as the length of the time series increases from 3500 years (actual length of the GFDL-CM2.1 AMV time series) to 50,000 years. This artificial length increase provides an indication about the sensitivity of the presented assessment to the length of the model simulations.

5 Conclusions

The properties of the AMV were examined in a set of multi-century preindustrial climate simulations performed with an ensemble of state-of-the-art coupled general circulation models. Different regimes of variability were identified in the modelled AMV time series, providing phenomenological evidence of a non-stationary behaviour. In order to rigorously assess this behaviour, a statistical test for AMV stationarity was developed starting from the null hypothesis of a statistically stationary AMV, whose modulation is entirely random. Most of the models display epochs characterised by an autocovariance that differs significantly from the one featured by the whole time series, therefore indicating the non-stationarity of the AMV time series. This test appears to be sensitive to the time filtering applied to the North Atlantic SSTs, with 9 (7) out 11 models presenting non-stationary behaviour, when unfiltered (low-pass filtered) SSTs are used. This suggests that the non-stationarity of unfiltered North Atlantic SSTs is coming from both high- and low-frequency variability. However, given the limited length of the model simulations, the uncertainty associated with undersampling allows concluding with high confidence that AMV is not-stationary only for one (ACCESS1-0) out of the 11 models. Nevertheless, there is such an indication for most of the models. The interval-dependent behaviour identified in most models suggests that the character of the observed AMV may undergo significant changes in the future similarly to what seen in the models (dominant periods and amplitude of the AMV changes from one epoch to another).

Our study points also to the large differences across models, which emerge also among the five CMIP5 models analysed by Deshayes et al. (2014), in particular in the representation of freshwater budget and circulation in the North Atlantic. Similar disagreements are found in Ba et al. (2014), Menary et al. (2015) and Deshayes et al. (2014), all showing that biases in the simulated mean climate state influence the characteristics of decadal variability.

To conclude, explaining multidecadal variability is challenging, not only because of the lack of observational data and the multitude of possible processes that may be involved and that may not be limited to North Atlantic itself (Frankcombe et al. 2010), but also because of model diversity and the associated uncertainty.

Appendix

A detailed explanation of the method for testing AMV stationarity

A detailed explanation of the method for testing AMV stationarity is given here using the MPI-ESM-P model example. The starting point is the generation of 1000 synthetic

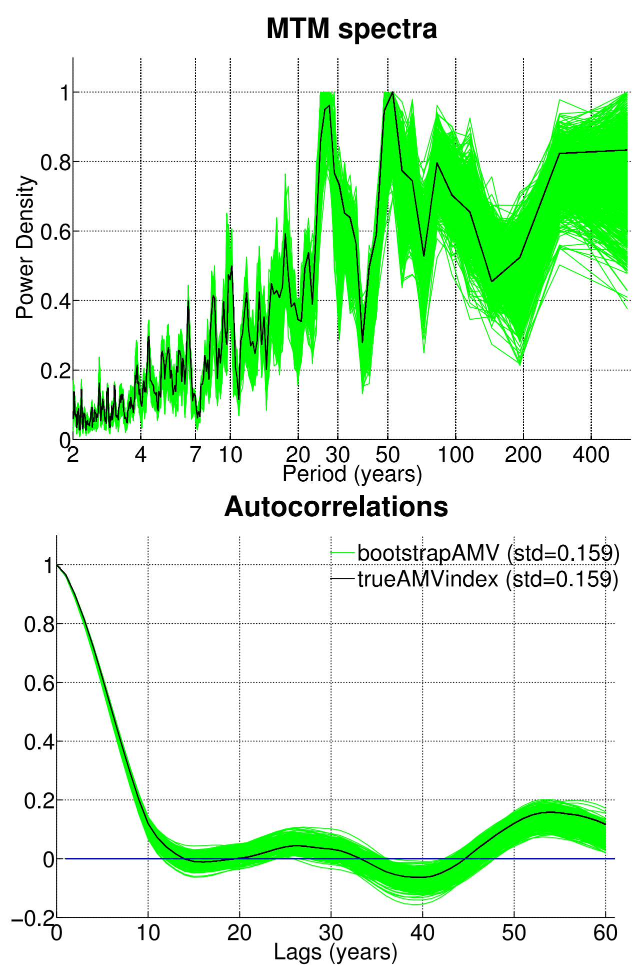

Fig. 9 (Top) model NASST spectrum in black and 1000 spectra associated to the 1000 synthetic NASST time series in green lines, via multi-taper method. (Bottom) model AMV autocorrelation in black and 1000 autocorrelations associated to the 1000 synthetic AMV time series in green lines

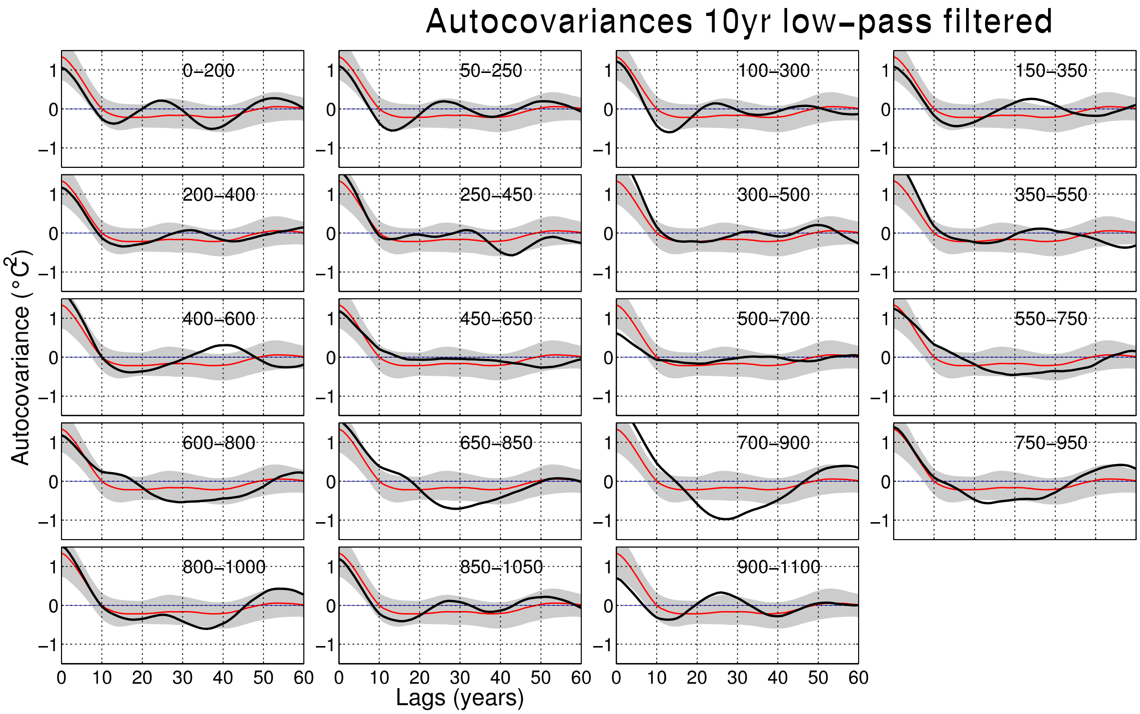

Fig. 10 Autocovariances for the 10-year low-pass filtered AMV time series split in 200-year-long intervals shifted by 50 years (black thick lines). The grey shading and the red curve are the same in all panels: the grey shading indicates the 90% central range of the autocovariance values for the entire time series and the red curve represents the average of the model autocovariance functions of the panels

time series with the same length and the same spectrum as the model NASST time series (1156 years for the MPI-ESM-P model). In Fig. 9, the 1000 synthetic time series, in green, share the same spectrum (top plot) and autocorrelation (bottom plot) as the model time series (superimposed as a thick black line). Having verified this, it is possible to compute the 90% central range for the autocovariance function that represents the 90% range of variation of autocovariance values for each lag. Each of the 1000 synthetic AMV time series is split in 200-year-long intervals shifted by 1 year, obtaining $956 (= 1156 - 200)$ intervals for MPI-ESM-P model. Therefore, there are 956,000 intervals for all the 1000 synthetic time series. The autocovariance function is computed for all these intervals, spanning 60 lags. For each lag, the 90% central range (i.e. the range between the 5th percentile and the 95th percentile) of the 956,000 autocovariance values is computed.

In parallel, also the model AMV time series is divided in 956 intervals and the autocovariance is computed for all of them. For each lag, the 956 autocovariance values are compared with the inferior and superior limits of the 90% central range and we count how many values exceed the range.

In order to have a reasonable number of intervals to be shown together, in Fig. 10 we plot the intervals shifted by 50 years instead of 1 year. Each plot refers to a different interval, the black line is the model autocovariance function for that interval, the grey shading is the 90% central range and the red curve represents the mean behaviour of the model intervals. Even from a visual inspection is possible to find periods where the autocovariance function falls outside the 90% central range (grey shading), showing that they do not have the same statistical characteristics as the whole time series. A more rigorous approach consists in counting the number of autocovariance values that fall outside the confidence interval. Taking into account all of the 956 intervals, there are 57,360 values (that is 956 intervals multiplied by 60 lags) in total. The result shows 8088 values outside, which corresponds to 14.1% of the total, meaning that, for the MPI-ESM-P model, 14.1% of the total amount of values fall outside the shading, as represented in Fig. 7 (blue triangle).

References

Ba J, Keenlyside NS, Latif M, Park W, Ding H, Lohmann K et al (2014) A multi-model comparison of Atlantic multidecadal variability. Clim Dyn 43(9–10):2333–2348

Bellucci A, Mariotti A, Gualdi S (2017) The role of forcings in the 20th century North Atlantic multi-decadal variability: the 1940–1975 North Atlantic cooling case study. J Clim 30:7317–7337

Booth BBB, Dunstone NJ, Halloran PR, Andrews T, Bellouin N (2012) Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. Nature 484:228–232

Brockwell PJ, Davis RA (1991) Time series: theory and methods. Springer, New York

DelSole T, Tippett MK, Shukla J (2011) A significant component of unforced multidecadal variability in the recent acceleration of global warming. J Clim 24:909–926

Delworth TL, Mann ME (2000) Observed and simulated multidecadal variability in the Northern Hemisphere. Clim Dyn 16:661–676

Deser C, Blackmon ML (1993) Surface climate variations over the North Atlantic Ocean during winter: 1900–1989. J Clim 6:1743–1753

Deshayes J, Curry R, Msadek R (2014) CMIP5 model intercomparison of freshwater budget and circulation in the North Atlantic. J Clim 27:3298–3317

Ebisuzaki W (1997) A method to estimate the statistical significance of a correlation when the data are serially correlated. J Clim 10:2147–2153

Enfield DB, Mestas-Nuñez AM, Trimble PJ (2001) The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophys Res Lett 28(10):2077–2080

Frankcombe LM, von der Heydt A, Dijkstra HA (2010) North Atlantic multidecadal climate variability: an investigation of dominant time scales and processes. J Clim 23(13):3626–3638

Goldenberg SB, Landsea CW, Mestas-Nuñez AM, Gray WM (2001) The recent increase in atlantic hurricane activity: causes and implications. Science 293:474

Gray ST, Graumlich LJ, Betancourtv JL, Pederson GT (2004) A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 A.D. Geophys Res Lett 31:L12205

Hasselmann K (1976) Stochastic climate models, Part I. Theory. Tellus 28:473–485

Huang J, Higuchi K, Shabbar A (1998) The relationship between the North Atlantic oscillation and El Niño-Southern oscillation. Geophys Res Lett 25:2707–2710

Hurrell JW, Van Loon H (1997) Decadal variations in climate associated with the North Atlantic oscillation. Clim Change 36:301–326

Kilbourne KH, Alexander MA, Nye JA (2014) A low latitude paleoclimate perspective on Atlantic multidecadal variability. J Mar Syst 133:4–13

Knight JR, Allan RJ, Folland CK, Vellinga M, Mann ME (2005) A signature of persistent natural thermohaline circulation cycles in observed climate. Geophys Res Lett 32:L20708

Knight JR, Folland CK, Scaife AA (2006) Climate impacts of the Atlantic multidecadal oscillation. Geophys Res Lett 33:L17706

Kushnir Y (1994) Interdecadal variations in North Atlantic Sea surface temperature and associated atmospheric conditions. J Clim 7:141–157

Latif M, Collins M, Pohlmann H, Keenlyside NS (2006) A review of predictability studies of Atlantic sector climate on decadal time scales. J Clim 19:5971–5987

Latif M, Keenlyside NS (2011) A perspective on decadal climate variability and predictability. Deep Sea Res. Part II 58(17–18):1880–1894

Mariotti A, Dell'Aquila A (2012) Decadal climate variability in the Mediterranean region: roles of large-scale forcings and regional processes. Clim Dyn 38:1129–1145

Marullo S, Artale V, Santoleri R (2011) The SST multidecadal variability in the Atlantic–Mediterranean region and its relation to AMO. J Clim 24:4385–4401

Medhaug I, Furevik T (2011) North Atlantic 20th century multidecadal variability in coupled climate models: sea surface temperature and ocean overturning circulation. Ocean Sci 7:389–404

Menary MB, Hodson DLR, Robson JI, Sutton RT, Wood RA, Hunt JA (2015) Exploring the impact of CMIP5 model biases on the simulation of North Atlantic decadal variability. Geophys Res Lett 42:5926–5934

Otterå OH, Bentsen M, Drange H, Suo L (2010) External forcing as a metronome for Atlantic multidecadal variability. Nat Geosci 3:688–694

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108:4407

Rotstayn LD, Lohmann U (2002) Tropical rainfall trends and the indirect aerosol effect. J Clim 15:2103–2116

Saenger C, Cohen AL, Oppo DW, Halley RB, Carilli JE (2009) Surface-temperature trends and variability in the low-latitude North Atlantic since 1552. Nat Geosci 2:492–495

Schlesinger ME, Ramankutty N (1994) An oscillation in the global climate system of period 65–70 years. Nature 367:723–726

Sutton RT, Dong B (2012) Atlantic Ocean influence on a shift in European climate in the 1990s. Nat Geosci 5:788–792

Sutton RT, Hodson DLR (2005) Atlantic Ocean forcing of North American and European summer climate. Science 309:115

Swingedouw D, Mignot J, Labetoulle S, Gilardi E, Madec G (2013) Initialisation and predictability of the AMOC over the last 50 years in a climate model. Clim Dyn 40:23981–23990

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498

Terray L (2012) Evidence for multiple drivers of North Atlantic multi-decadal climate variability. Geophys Res Lett 39:L19712

Thomson DJ (1982) Spectrum estimation and harmonic analysis. Proc IEEE 70(9):1055–1096

Ting M, Kushnir Y, Seager R, Li C (2009) Forced and internal twentieth-century SST trends in the North Atlantic. J Clim 22:1469–1481

Ting M, Kushnir Y, Seager R, Li C (2011) Robust features of Atlantic multi-decadal variability and its climate impacts. Geophys Res Lett 38:L17705

Ting M, Kushnir Y, Li C (2014) North Atlantic multidecadal SST oscillation: external forcing versus internal variability. J Mar Syst 133:27–38

Trenberth KE, Shea DJ (2006) Atlantic hurricanes and natural variability in 2005. Geophys Res Lett 33:L12704

Wittenberg AT (2009) Are historical records sufficient to constrain ENSO simulations? Geophys Res Lett 36:L12702

Zanchettin D, Rubino A, Jungclaus JH (2010) Intermittent multidecadal-to-centennial fluctuations dominate global temperature evolution over the last millennium. Geophys Res Lett 37:L14702

Zanchettin D, Bothe O, Müller W, Bader J, Jungclaus JH (2014) Different flavors of the Atlantic multidecadal variability. Clim Dyn 42(1–2):381–399

Zhang R (2008) Coherent surface-subsurface fingerprint of the Atlantic meridional overturning circulation. Geophys Res Lett 35:L20705

Zhang R, Delworth TL (2006) Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. Geophys Res Lett 33:L17712

Zhang L, Wang C (2013) Multidecadal North Atlantic sea surface temperature and Atlantic meridional overturning circulation variability in CMIP5 historical simulations. J Geophys Res Oceans 118(10):5772–5791