



# Quantifying the rarity of extreme multi-decadal trends: how unusual was the late twentieth century trend in the North Atlantic Oscillation?

R. Eade<sup>1,2</sup> · D. B. Stephenson<sup>2</sup> · A. A. Scaife<sup>1,2</sup> · D. M. Smith<sup>1</sup>

Received: 22 April 2021 / Accepted: 21 September 2021 / Published online: 3 October 2021  
© Crown 2021

## Abstract

Climate trends over multiple decades are important drivers of regional climate change that need to be considered for climate resilience. Of particular importance are extreme trends that society may not be expecting and is not well adapted to. This study investigates approaches to assess the likelihood of maximum moving window trends in historical records of climate indices by making use of simulations from climate models and stochastic time series models with short- and long-range dependence. These approaches are applied to assess the unusualness of the large positive trend that occurred in the North Atlantic Oscillation (NAO) index between the 1960s to 1990s. By considering stochastic models, we show that the chance of extreme trends is determined by the variance of the trend process, which generally increases when there is more serial correlation in the index series. We find that the Coupled Model Intercomparison Project (CMIP5 + 6) historical simulations have very rarely (around 1 in 200 chance) simulated maximum trends greater than the observed maximum. Consistent with this, the NAO indices simulated by CMIP models were found to resemble white noise, with almost no serial correlation, in contrast to the observed NAO which exhibits year-to-year correlation. Stochastic model best fits to the observed NAO suggest an unlikely chance (around 1 in 20) for there to be maximum 31-year NAO trends as large as the maximum observed since 1860. This suggests that current climate models do not fully represent important aspects of the mechanism for low frequency variability of the NAO.

**Keywords** North Atlantic · North Atlantic Oscillation · Extreme trends · Multi-decadal variability · Climate modelling · Stochastic modelling

## 1 Introduction

Climate trends over multiple decades are often estimated using moving window linear trend analysis, for example, trends in the North Atlantic Oscillation (Deser et al. 2017; Scaife et al. 2008, 2009; Semenov et al. 2008; Raible et al. 2005) and in the North Atlantic jet characteristics (Bracegirdle et al. 2018). Moving window trend analysis has been used to detect significant changes in other variables such as the onset of spring (Ge et al. 2014) and European winter precipitation (Matti et al. 2009), but also to investigate the likelihood of periods of zero trend, such as the slowdown

in the rise of global mean temperature at the start of this century (Shi et al. 2016). The trend slope in these studies is typically estimated using ordinary least squares (OLS) regression. In what follows, we will adopt such an approach to define *multi-decadal trends* using a 31-year window and investigate their distribution in climate time series.

In this study we shall use trends in the winter NAO as an exemplar because they have been particularly prominent and are known to have a strong impact on European and N. American regional climate. The NAO is a large-scale mode of atmospheric variability in the North Atlantic region that exhibits strong persistent multi-decadal variations (Hurrell et al. 2003). It is manifested by changes in the North–South pressure gradient related to the variability in the Azores high and Icelandic low pressure systems, and to changes in the mean wind flow and storm tracks (Wanner et al. 2001; Greatbatch 2000; Woollings et al. 2015; Hurrell 1995). The NAO is known to influence regional climate in the surrounding continents of Europe and North America, particularly on

---

✉ R. Eade  
rosie.eade@metoffice.gov.uk

<sup>1</sup> Met Office, Exeter, UK

<sup>2</sup> Department of Mathematics, University of Exeter, Exeter, UK

seasonal timescales (Scaife et al. 2014), but has also been shown to have far reaching teleconnections such as an influence on the variability of surface temperatures in South-Central China (Zuo et al. 2016). Trends in the winter NAO over multiple decades have a large impact on temperature in the whole of the Northern Hemisphere extra-tropics, with a large positive trend accounting for at least half of the winter warming from the 1960s to 1990s while a large negative trend more than halved the winter warming from the 1920s to 1960s (Iles and Hegerl 2017; Scaife et al. 2005).

The winter NAO is related to variations in temperature, rainfall and drought that have impacts on industries such as agriculture, fishing and water management (Hurrell et al. 2001). In the energy sector, the NAO affects energy demand through winter temperature variability and energy supply from wind, solar and hydropower (Jerez et al. 2013; Uvo and Berndtsson 2002; Thornton et al. 2017). The winter NAO has a large effect on crop yields in Europe and the USA, and planning for food imports (Kim and McCarl 2005). The NAO governs flooding and associated insurance losses (Zanardo et al. 2019). Understanding trends in the NAO over multiple decades, especially for winters, is therefore important for long term planning within all these sectors and needs to be taken into account when planning for regional climate change adaptation.

The 1960s to 1990s increasing trend mentioned above has gained considerable attention in the scientific literature. In the early 1990s it was identified as the largest multi-decadal trend in the observed NAO record and was speculated to be a response to increased greenhouse gases (e.g. Gillett et al. 2003; Shindell et al. 1999), however the negative trend that followed makes this speculation unlikely (e.g. Pinto and Raible 2012; Hanna et al. 2015). There is still a lack of consensus as to whether the increasing 1960s to 1990s NAO trend can be explained by internal climate variability, relating to the aggregation of shorter timescale atmospheric variability with the possibility of some longer term memory element such as that estimated using stochastic processes (e.g. Wunsch 1999), or whether the magnitude of variability is unusual in this context (e.g. Feldstein 2002).

General Circulation Model (GCM) historical simulations from the Coupled Model Intercomparison Project Phase 5, CMIP5 (Taylor et al. 2012) very rarely simulate a 30-year trend in the winter NAO or Atlantic jet strength of magnitude comparable to that observed in the 1960s to 1990s anywhere in their 150 year period of simulation (Bracegirdle et al. 2018). These CMIP5 simulations lack the low frequency variability of the NAO found in observations, despite successfully representing the interannual variability (Kravtsov 2017; Wang et al. 2017). It is not immediately clear whether the failure of GCMs to simulate the magnitude of the observed trend is due to the unlikelihood of occurrence, a missing or incorrect forcing element, or some model deficiencies in simulating

relevant processes (e.g. Osborn 2004). For example, in GCMs, the NAO has been shown to be weakly forced by Sea Surface Temperature (SST), but the models appear to underestimate the strength of the real-world trend (Scaife et al. 2009; Simpson et al. 2018). If GCMs are underestimating the magnitude of NAO trends, then this could have important implications for attributing the contribution of external forcing to past observed trends (Hegerl et al. 2007) and forecasting future regional climate trends over the next few decades (e.g. Lowe et al. 2018) as projections of trends in temperature and rainfall in the northern extra-tropics are dominated by the internal variability of the NAO on these time scales (Deser et al. 2017).

This study estimates how unusual are notable extreme multidecadal trends using simulations from state of the art GCMs and from stochastic time series models. Our focus is on understanding the large positive 1963–1993 trend in winter NAO, which is the maximum 31-winter trend in the historical NAO record. When considering the chance of exceeding this trend value, we shall take into account the important fact that it has been selected because of it being the maximum of all the values in the historical time series. A similar issue is discussed in Percival and Rothrock (2005) in the context of trends at the end of a time series where the start point (or window length) from which the trend is calculated may have been chosen after “eyeballing” a period of interest. They account for this in their probability estimate by simulating a large sample of stochastic time series for a range of window lengths, which encompass the length which has essentially been “eyeballed”. We instead approach this issue using extreme value distributions, building on the distribution of moving window linear trends to estimate the distribution of the maximum trend value in the time series of moving window slope estimates.

## 2 Definitions of multi-decadal trends

Multi-decadal trends in a climate index can be quantified by using slope parameter estimates obtained from linear regression of the index on time. Moving window trends can be obtained by shifting a window along the index time series year-by-year and calculating the linear trend estimate within each window. For a regular time series  $\{Y_1, Y_2, \dots, Y_n\}$  of length  $n$ , the OLS slope parameter estimate of the trend ( $Z_i$ ) in a window of length  $2K + 1$  centered at time  $i = 1 + K, 2 + K, \dots, n - K$  is given by

$$Z_i = \frac{1}{h^2} \sum_{j=-K}^K j Y_{i+j}, \quad (1)$$

where  $h^2 = \sum_{j=-K}^K j^2 = K(K + 1)(2K + 1)/3$ .

The window length  $2K + 1$  is chosen to be an odd number to avoid the central point being half-way between two years. From this processed time series the maximum trend can be

identified, which we will hereon refer to as an “extreme trend”.

From Eq. (1), it can be seen that the resulting trend values are a linear combination of the index values: trend time series  $Z$  is a moving average filtered version of index series  $Y$  with the  $2K + 1$  filter weights  $\{\frac{-K}{h^2}, \frac{-(K-1)}{h^2}, \dots, \frac{K-1}{h^2}, \frac{K}{h^2}\}$ . Hence, if  $Y$  can be represented as a  $p$ 'th order auto-regressive process  $AR(p)$  then  $Z$  is an auto-regressive moving average process  $ARMA(p, 2K + 1)$  (see Appendix). The Moving Average  $MA(2K + 1)$  filter is non-invertible since it has a unit root (i.e., the filter applied to a constant series gives 0) but this is not a problem here since we are not concerned with estimating the ARMA process from a given trend series—the MA coefficients are the known filter weights determined by our choice of  $K$ .

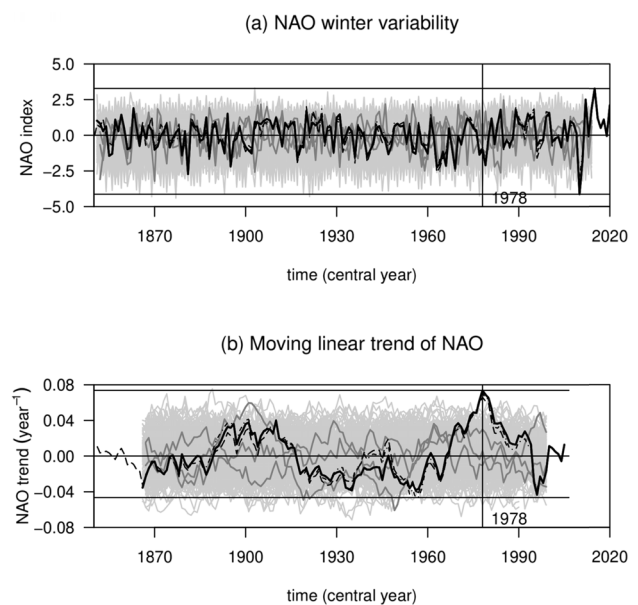
In this study, we choose a fixed window length of  $2K + 1 = 31$  winters ( $\sim 3$  decades) as this is within the standard 10–30 year time scale considered to be decadal variability (Meehl et al. 2009) while still having enough data points to robustly estimate a linear trend. Shpakova et al. (2020) suggest that window length should be around one third of the length of the time series but also long enough to take into account climate factors such as solar irradiance, thus  $\sim 3$  decades seems a sensible window length for century long station based climate time series.

### 3 Data

#### 3.1 NAO observations

The winter NAO index is defined here as a standardised difference in the seasonal DJF (December to February) mean sea level pressure (MSLP) at the two main nodes of NAO variability, that is the Azores (37.7 N, 25.7 W) minus Iceland (65.0 N, 22.8 W). We use the HadSLP2r gridded observation dataset available for 1850–2020 (Allan and Ansell 2006) and the twentieth century reanalysis data (“C20C”) available for 1872–2012 (Compo et al. 2011). The nearest grid box is used to represent each of the NAO nodes, and the individual time series for each node are first standardised so that neither node dominates the variability. This definition is chosen to match the NAO reconstructed index from Luterbacher et al. (1999, 2001), which is available for 1659–2001 (“L99”), based on a mix of station data (pressure, temperature and precipitation) and proxy data. It also makes it easier to compare all model and observation time series simultaneously with the same stochastic processes, without having to consider differences in the interannual variance of MSLP fields. The climate period for standardisation was chosen to be 1862–2005 to match the period used for the GCM experiments.

All three NAO datasets show consistent interannual variability over the common period of 1872–2001, with a prominent shift from large negative values in the 1960s to large positive values in the 1990s (Fig. 1a). The 31-winter moving window trend time-series (Fig. 1b) shows that the maximum 31-winter trend occurs for the winters 1963–1993, i.e. the window centred on 1978 with years referring to the January in DJF. The 1978 centred window has the maximum 31-winter trend compared to any other window in any of the observation or reanalysis datasets including the whole 343 years of L99 reconstruction, i.e. this is an extreme trend. This trend has magnitude  $0.0737 \text{ year}^{-1}$  in HadSLP2r gridded observations,  $0.0709 \text{ year}^{-1}$  in C20C reanalysis gridded observations and  $0.0659 \text{ year}^{-1}$  in L99 reconstruction. Over the 3 decades this is equivalent to a total shift of  $\sim 2$  standard deviations of the winter mean NAO interannual variability. The HadSLP2r maximum trend value,  $0.0737 \text{ year}^{-1}$ , is used for the exceedance threshold in our subsequent analysis as trend magnitudes are somewhat sensitive to the period used for standardisation of the NAO time series and this data set covers the same period as the GCMs.



**Fig. 1** Low frequency variability of the NAO and its trend in observations and GCMs. **a** The standardised DJF mean North Atlantic Oscillation (NAO) index based on HadSLP2r gridded observation data (black solid line), twentieth century reanalysis data (black dot-dash line) and proxy-reconstructed L99 NAO data (black dash line). **b** The 31-winter moving linear trend estimate (per year). The maximum and minimum observed/reanalysis values are shown as horizontal black lines and the window containing the maximum 31-winter moving linear trend, centred on 1978, is marked with a vertical black line. The individual standardised NAO time series from 281 general circulation model (GCM) simulations are shown in gray, with the 3 simulations from CMIP6-HadGEM3-GC31-MM highlighted in darker gray as an example of the variability in individual simulations

### 3.2 NAO simulations from general circulation models

To further understand the distribution of extreme trends in the NAO we assess historical simulations from state of the art coupled GCMs from the Coupled Model Inter-comparison Project Phase 6, CMIP6 (Eyring et al. 2016), alongside those from CMIP5 (Taylor et al. 2012) that were assessed in Bracegirdle et al. (2018). These are continuous transient simulations with external forcings (solar, volcanic, and anthropogenic) and we focus on the common period of 1862–2005. From CMIP6 we assess 178 simulations across 32 models (Table 1), while from CMIP5 we assess 103 simulations across 42 models (Table 2) making a total of 40,464 years (281 simulations  $\times$  144 years). For each individual simulation we calculate the standardised NAO index as for the observations, using the MSLP output and the nearest grid box definition. We then filter the NAO time series in the same way as for the observed NAO to calculate the 31-winter moving window trend (Fig. 1b).

For the *absolute* NAO index (difference in MSLP anomalies without standardisation) the GCMs simulate interannual variability in reasonable agreement with observations: the average interannual standard deviation of the winter NAO in GCM simulations is 7.8 hPa (with a standard deviation across members of 0.7 hPa) while the observed HadSLP2r and C20C values are 5.9 hPa and 8.0 hPa for the common period 1862–2005, noting that HadSLP2r is known to underestimate interannual variability (e.g. Semenov et al. 2008). For the rest of this paper the *standardised* NAO index is used (Fig. 1a) to enable comparison with single stochastic models (with variance 1) and isolate differences in behaviour due to features in autocorrelation structure rather than due to sample variance.

**Table 1** GCM historical experiments used to model the NAO with columns showing the institution, model name and the number of distinct simulations (ensemble members)

Institution	Model name	Number of simulations
AS-RCEC	TaiESM1	1
AWI	AWI-CM-1-1-MR; AWI-ESM-1-1-LR	5; 1
BCC	BCC-CSM2-MR; BCC-ESM1	3; 3
CAMS	CAMS-CSM1-0	3
CAS	CAS-ESM2-0; FGOALS-g3	4; 6
CCCma	CanESM5 (ripf code p1; p2)	25; 40
CCCR-IITM	IITM-ESM	1
CMCC	CMCC-CM2-HR4; CMCC-CM2-SR5	1; 1
CSIRO-ARCCSS	ACCESS-CM2	3
E3SM-Project	E3SM-1-0; E3SM-1-1	5; 1
FIO-QLNM	FIO-ESM-2-0	3
HAMMOZ-Consortium	MPI-ESM-1-2-HAM	2
INM	INM-CM4-8; INM-CM5-0	1; 10
KIOST	KIOST-ESM	1
MIROC	MIROC-ES2L	10
MOHC	HadGEM3-GC31-LL; HadGEM3-GC31-MM	4; 3
MPI-M	MPI-ESM1-2-LR	10
MRI	MRI-ESM2-0	5
NASA-GISS	GISS-E2-1-G-CC; GISS-E2-1-G	1; 3
NCAR	CESM2-FV2; CESM2-WACCM-FV2; CESM2-WACCM; CESM2	1; 1; 2; 3
NCC	NorCPM1; NorESM2-LM; NorESM2-MM	2; 2; 1
NIMS-KMA	KACE-1-0-G	2
NUIST	NESM3	5
SNU	SAM0-UNICON	1
UA	MCM-UA-1-0	2

All experiments are from the CMIP6 database <https://esgf-index1.ceda.ac.uk/search/cmip6-ceda> (Eyring et al. 2016)

**Table 2** GCM historical experiments used to model the NAO with columns showing the institution, model name and the number of distinct simulations (ensemble members)

Institution	Model name	Number of simulations
BCC	BCC-CSM1-1; BCC-CSM1-1-m	3; 3
BNU	BNU-ESM	1
CCCma	CanESM2	5
CMCC	CMCC-CM; CMCC-CESM; CMCC-CMS	1; 1; 1
CNRM-CERFACS	CNRM-CM5-2; CNRM-CM5	1; 5
CSIRO-BOM	ACCESS1-0; ACCESS1-3	2; 1
CSIRO-QCCCE	CSIRO-Mk3-6-0	5
FIO	FIO-ESM	1
INM	INMCM4	1
IPSL	IPSL-CM5A-LR; IPSL-CM5A-MR; IPSL-CM5B-LR	5; 3; 1
LASG-CESS	FGOALS-g2	1
MIROC	MIROC5; MIROC-ESM; MIROC-ESM-CHEM	5; 3; 1
MOHC	HadGEM2-ES; HadGEM2-CC; HadCM3	4; 1; 1
MPI-M	MPI-ESM-LR; MPI-ESM-MR; MPI-ESM-P	3; 3; 1
MRI	MRI-CGCM3; MRI-ESM1	3; 1
NASA-GISS	GISS-E2H	5
NCAR	CCSM4	5
NCC	NorESM1-M; NorESM1-ME	3; 1
NIMR-KMA	HadGEM2AO	1
NOAA-GFDL	GFDL-ESM2M; GFDL-ESM2G; GFDL-CM2p1; GFDL-CM3	1; 1; 10; 5
NSF-DOE-NCAR	CESM1-BGC; CESM1-CAM5; CESM1-FASTCHEM; CESM1-WACCM	1; 1; 1; 1

All experiments are from the CMIP5 database <https://esgf-node.llnl.gov/search/cmip5> (Taylor et al. 2012)

## 4 Stochastic time series models

### 4.1 Autocorrelation in the observed and simulated NAO indices

We can attempt to understand the differences between the properties of the observed and the GCM simulated NAO indices by considering the stochastic properties of the series. The most common stochastic models used to represent the NAO are short-range dependence red noise processes (e.g. Wunsch 1999; Feldstein 2000; Thompson et al. 2015) and long-range dependence processes (Stephenson et al. 2000). A short-range first order auto-regressive AR(1) process has the form  $Y_{i+1} = \rho Y_i + \varepsilon_{i+1}$ , where  $\rho$  is the lag-1 year autocorrelation parameter and  $\varepsilon_{i+1}$  are independent Gaussian random variables. A long-range fractional difference (FD) process has the form  $(1 - B)^d Y_{i+1} = \varepsilon_{i+1}$ , where  $d$  is the difference parameter and  $B$  is the backward shift operator such that  $BY_t = Y_{t-1}$ . For the AR(1) process, the autocorrelations at time lag  $k$  have a known parametric form  $\rho_k = \rho^{|k|}$ . For the FD process, the lag autocorrelations decay slower than exponential i.e.  $\rho_{k+1} = \rho_k (k+d)/(k+1-d)$  for  $k=0, 1, 2, \dots$  (Hosking 1981).

Higher order auto-regressive processes were also considered but are not included in subsequent analysis as our observed NAO time series shows no significant autocorrelation beyond lag-1 year, and hence fitted higher order auto-regressive models show no real improvement over an AR(1) model. Stephenson et al. (2000) found that the best fitted stochastic models were a fractional difference process and an AR(10) process, however as they gave very similar results the fractional difference process is preferred here due to its relative simplicity and having fewer parameters to estimate.

The HadSLP2r winter NAO time series has a lag-1 year autocorrelation  $\rho_1$  estimate of 0.169 for the period 1850–2020 ( $n=170$  years). There is considerable uncertainty in this estimate due to the short length of the time series, with an approximate 95% confidence interval of  $[0.0212, 0.317]$  based on the Bartlett formula  $\rho_1 \pm 1.96 \sqrt{\frac{(1-\rho_1^2)}{n}}$  (Bartlett 1946). Similar lag-1 year autocorrelation values of 0.142 and 0.161 are obtained for C20C and L99 within the same time period. Fitting a long-range FD process to the HadSLP2r winter NAO time series has a difference parameter ( $d$ ) estimate of 0.123. Some studies have suggested non-stationarity in the NAO behaviour (e.g. Hurrell and Van Loon 1997; Pozo Vazquez et al.



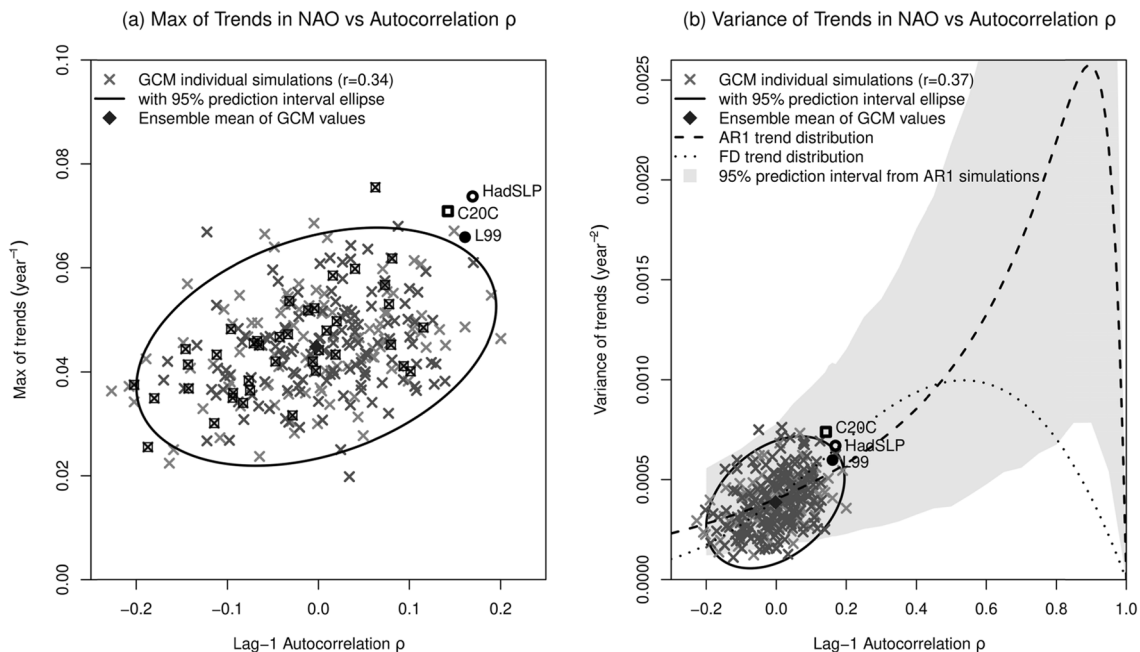
2001), however such trends in autocorrelation would be difficult to detect due to the large amount of sampling uncertainty. We assume that the NAO multi-decadal variability can be reasonably represented by a stationary stochastic process that represents the aggregation of atmospheric noise (Wunsch 1999; Feldstein 2002) with some year-to-year memory deriving from either internal atmospheric processes or external drivers such as ocean variability or external forcing.

It is notable that the observed estimates of lag-1 year autocorrelation values are substantially larger than estimates from GCMs (Fig. 2a). Only 13 out of the total of 281 GCM simulations (a proportion of 5%) are found to have lag-1 year autocorrelation estimates lying outside the 95% confidence interval ( $0.0 \pm 0.163$ ) for zero autocorrelation, i.e. a white noise process of matching length. This selection of 13 simulations is not dominated by any particular models so is not highlighting any single model as having more or less realistic autocorrelation than the rest of the CMIP models. This suggests that the CMIP GCM simulated NAO is consistent with white noise, i.e. having no serial correlation, as was found for the jet stream variability in Simpson et al. (2018).

## 4.2 NAO simulations from stochastic time series models

To investigate the behaviour of maximum moving window trends in stochastic models, discussed in Sect. 5, we generated 5000 random stochastic simulations of length  $n = 144$  from both short-range and long-range Gaussian stochastic processes, and then computed the moving window trends (window length 31) for each simulation to match the handling of the GCM NAO simulations. We estimate the distribution of maximum moving window trends for a specified stochastic model by using the sample of 5000 maximum trends, one from each of the 5000 simulated moving window trend series.

A set of stochastic simulations was generated from short-range AR(1) processes for lag-1 year autocorrelation parameters  $\rho$  in the range  $-0.2$  to  $1.0$ . Another set of simulations was generated for long-range FD processes using a set of difference parameters matched to the lag-1 year autocorrelation parameters, noting that  $\rho_1(d) = d/(1 - d)$  (Hosking 1981), to enable a direct comparison of the two types of stochastic process in Sect. 5



**Fig. 2** The relationship of maximum trends and variance of moving window trends with autocorrelation in the NAO. **a** The maximum moving window trend (from the 31-winter moving window trends in Fig. 1) vs the lag-1 year autocorrelation estimate. Observed values are shown for HadSLP2r (black circle), twentieth century reanalysis (black square) and proxy-reconstructed NAO data (black filled circle). CMIP6 and CMIP5 historical simulations are differentiated by dark and light gray crosses respectively, with the 40 simulations from CMIP6-CanESM5(p2) highlighted by black squares around the

crosses. The 95% prediction interval ellipse (black solid line) and ensemble mean value (black diamond) represent the combined multi-model ensemble. The correlation  $r$  for the cloud of CMIP simulation values is shown in the legend for each scatter plot. **b** Shows the relationship of the variance of moving window trends with autocorrelation in the same manner as (a), with the expected variance for an autoregressive (black dashed curve, 95% prediction interval in gray shading) and fractional difference process (black dotted curve) as described in Sect. 5.2

## 5 Distribution of trends

### 5.1 Distribution of maximum trends simulated by the General Circulation Models

Assessing the GCMs for the specific time window 1963–1993, none of the individual simulations (ensemble members) contain a trend as large as that observed (Fig. 1b). The multi-model ensemble mean has only a very weak positive trend of  $0.00183 \text{ year}^{-1}$  compared to the observed value of  $0.0737 \text{ year}^{-1}$ . The ensemble members are spread fairly symmetrically about zero with a 95% confidence interval of  $-0.0366$  to  $0.0360 \text{ year}^{-1}$ . This suggests that NAO variability in the GCMs for this period is primarily caused by internal variability and is not responding strongly to their common boundary condition forcing such as greenhouse gases.

By taking the maximum 31-winter trend from each GCM simulation, over the period 1862–2005 (144 years), we can explore the distribution of extreme trends in the GCMs. Over the full length of the GCM simulations there is no consistency in the timing of maximum linear trends across the models, and their individual maximum trend windows are spread fairly evenly across the whole time period (Fig. 1b). Figure 2a shows the maximum NAO trends from all the CMIP5 and CMIP6 simulations. The distribution of maximum trends from CMIP6 simulations is broadly similar to the distribution from CMIP5 (Fig. 2a). Just 1 out of the 281 simulations has a maximum trend greater than  $0.0737 \text{ year}^{-1}$ , giving an exceedance probability of  $3.56 \times 10^{-3}$  (Table 3), i.e. it is a *very unlikely* occurrence in GCMs (IPCC likelihood scale, Mastrandrea et al. 2010). This simulation is from the CMIP6 model CanESM5(p2), however the remaining 39 simulations from this ensemble have a similar spread in maximum trend and lag-1 year autocorrelation estimates to

**Table 3** Probabilities (in 1000ths) of trend exceedance for a single 31-winter trend  $p = \Pr(Z_i \geq 0.0737 \text{ year}^{-1})$  using all available data (trend series of length 140 from HadSLP2r observation dataset) and for the maximum 31-winter trend  $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \geq 0.0737 \text{ year}^{-1})$  in a trend series of length 114 years, calculated using GCMs and fitted stochastic models

Model	p	q
Empirical HadSLP2r	7.14 (= 1/140)	Not available
GCMs	0.0312 (= 1/32034)	3.56 (= 1/281)
White noise	0.121	6.00
AR(1) $\rho = 0.169$	0.822 (0.159, 3.09)	34.6 (7.80, 123.0)
FD $d = 0.123$	1.41 (0.199, 5.84)	60.6 (13.4, 194.0)

Lower and upper bounds for the stochastic models are shown in brackets, computed using the 95% confidence interval for the lag-1 year autocorrelation parameter  $\rho$ , [0.0212, 0.317] (Bartlett 1946)

that of the whole CMIP ensemble (Fig. 2a), so it does not appear that this model is more realistic than the rest of the CMIP models.

Semenov et al. (2008) used two GCMs to deduce that the distribution of 30-year NAO trends in GCMs is not significantly different to the observed distribution. However, their assessment is for the distribution of *all* moving window trends in the series rather than the distribution of *maximum* trends. They suggest that any differences between the observed distribution and the GCM distribution can be explained by the short length of the observed trend series. However, they state that their 3150 year simulation (moving window trend series of length 3121) only exceeds the maximum observed 30-year NAO trend once, i.e. the exceedance probability estimate would be  $1/3121 = 3.20 \times 10^{-4}$ , compared to our CMIP estimate of  $1/(281 \times 114) = 3.12 \times 10^{-5}$ . In our case it cannot be argued that the lack of extreme multi-decadal NAO trends in GCM simulations is due to the GCM sample being too small. Using the values from Semenov et al. (2008), the approximate probability of a maximum trend exceeding the maximum observed 30-winter NAO trend would be  $\sim 1/21 = 0.0476$  (sub-setting the 3150 years into 21 intervals of length 144 to match the CMIP simulations), which is consistent with our result that the 1960s to 90s trend is a very unlikely occurrence in GCMs. The following two sections will attempt to understand these results by considering the distribution of extreme trends in stochastic processes.

### 5.2 Distribution of the trend in a predefined year

The probability distribution of the moving window trend *in any predefined year* is determined by the moving window filter weights and the distribution of the underlying index series. Firstly, since any linear combination of Gaussian variables is Gaussian, the trend series will be Gaussian if the index series is Gaussian (which is often a reasonable approximation for climate indices). Furthermore, for index series that are 1st order stationary (i.e.  $E[Y_k]$  does not depend on year  $k$ ), the expectation of the trend is zero,

$$E[Z_i] = \frac{1}{h^2} \sum_{j=-K}^K E[Y_{i+j}]j = \frac{E[Y]}{h^2} \sum_{j=-K}^K j = 0.$$

Hence, the trend in any year  $i$  will be Gaussian distributed  $Z_i \sim N(0, \sigma_z^2)$ , where  $\sigma_z^2$  is the variance of the trend process. The probability of the trend in year  $i$  exceeding a threshold value of  $z$  is then simply given by  $\Pr(Z_i \geq z) = 1 - \Phi(z / \sigma_z)$ , where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal. For a given threshold  $z$ , the probability of exceedance is solely determined by the variance of the trend process  $\sigma_z^2$ .

By substitution of Eq. (1), the variance of the trend process is given by

$$\begin{aligned}\sigma_z^2 &= \frac{1}{h^4} \sum_{j=-K}^K \sum_{k=-K}^K jk \text{Cov}[Y_{i+j}Y_{i+k}] \\ &= \frac{\sigma_y^2}{h^4} \sum_{j=-K}^K \sum_{k=-K}^K jk \rho_{k-j}\end{aligned}\quad (2)$$

where  $\rho_{k-j}$  is the autocorrelation function and  $\sigma_y^2$  is the variance of the original index process. Woodward and Gray (1993) used sample moment estimates of  $\rho_1, \rho_2, \dots$  and a similar equation to estimate the trend variance. Alternatively, one can obtain more precise estimates by assuming that the lag autocorrelations have a known parametric form e.g.  $\rho_{k-j} = \rho^{|k-j|}$  if the index can be well represented by an AR(1) process, as is often assumed for climate indices (e.g. Santer et al. 2000; Thompson et al. 2015).

The black dashed curve in Fig. 2b shows the trend variance calculated numerically using Eq. (2), for an AR(1) process having variance  $\sigma_y^2 = 1$ , over a range of values of the lag-1 autocorrelation  $\rho$ . It captures the mean variance  $\sigma_z^2$  dependence on  $\rho$ , increasing slowly with  $\rho$  for small values, but then more rapidly as  $\rho$  increases. Interestingly, as  $\rho$  approaches one, the variance drops sharply back to zero. This is a consequence of using an OLS estimator of slope that tends to zero for large values of  $\rho$  when consecutive values in the index series become very similar to one another.

For typical  $\rho < 0.3$  for the NAO, similar variance results are obtained with different choices of time series model. The black dotted curve in Fig. 2b shows results using the long-range FD process. The variance for moving window trends is slightly higher for a FD process than for an AR(1) process for lag-1 autocorrelation values relevant for the NAO, but for larger lag-1 autocorrelation the variance increases more slowly and peaks at a considerably lower value for the FD process than for the AR(1) process. The moving window trend filter is a high-pass filter and so the resulting trend is not unduly sensitive to the low-frequency properties of the index series i.e., how the weak red noise is modelled.

In Fig. 2b we show the variance of moving window trends estimated directly from GCM NAO trend series. The GCM spread in autocorrelation and variance estimates is consistent with being that of a white noise process. The GCM estimate of the total variance of trends over all simulations combined is  $3.89 \times 10^{-4} \text{ year}^{-2}$  (the average for individual simulations is  $3.87 \times 10^{-4} \text{ year}^{-2}$ , black diamond in Fig. 2b), which is very close to the white noise estimate of  $4.03 \times 10^{-4} \text{ year}^{-2}$ . To estimate the uncertainty in the variance of moving window trends we used the stochastic simulations from Sect. 4.2, computing the variance of 31-winter moving window trends for each individual simulation. The 95% prediction interval for the variance estimate is shaded

gray in Fig. 2b using the 2.5% and 97.5% variance percentiles from the AR(1) simulations. The individual GCM simulation estimates for the variance of moving window trends are consistent with a white noise model as the 95% prediction interval for a white noise process encapsulates 96% of the GCM values in terms of the variance of trends (Fig. 2b, gray shading at  $\rho = 0.0$ ).

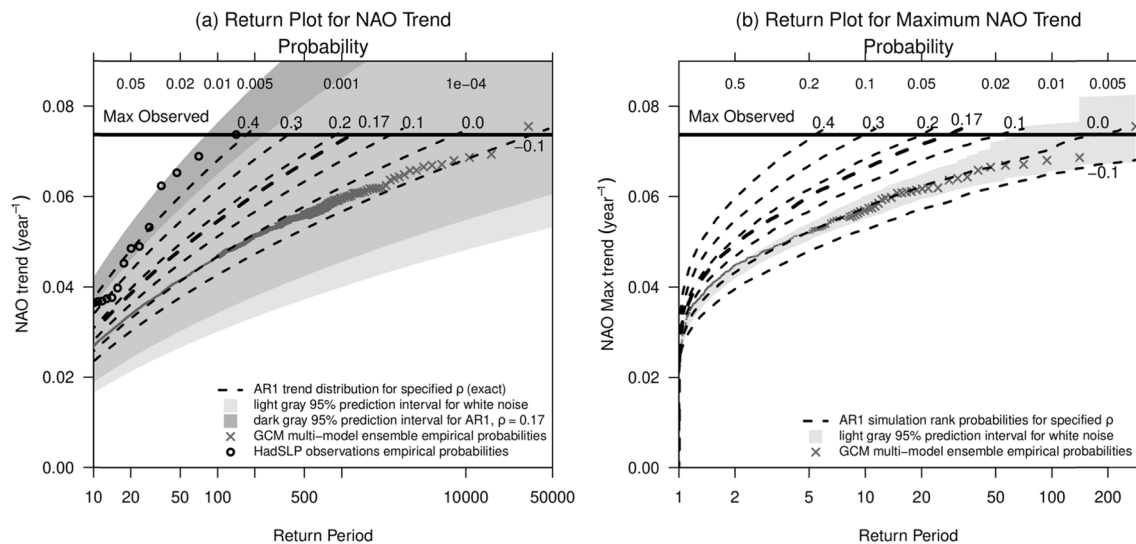
Within the CMIP ensemble of simulations there is a fairly strong relationship between the maximum trend and the variance of moving window trends (correlation 0.65). However, the GCM spread does not show any systematic improvement of one climate model over another as multiple members from the same model show a similar range of values to the multi-model ensemble (e.g. CMIP6-CanESM5(p2) in Fig. 2a). The large spread in stochastic model variance estimates for single simulations (Fig. 2b) shows that the observed time series is too short to confidently estimate the distribution of trends directly.

Figure 3a shows the distribution of moving window trends in terms of the exceedance probability  $p = \Pr(Z_i \geq z)$  and return period ( $1/p$ ). Gaussian distributions with mean zero and variance  $\sigma_z^2$  (Eq. 2) for a set of stochastic processes are used to show how the exceedance probability and return period vary with threshold  $z$  and the autocorrelation function of the original time series. The 95% prediction interval for the white noise model probabilities is shown, calculated using the 95% prediction interval for the white noise model variance of trends (Fig. 2b). The same technique is used for the fitted AR(1) process (Fig. 3a) using the best parameter estimate from HadSLP2r of  $\rho = 0.169$ .

The empirical probabilities based on sample rank values from the multi-model ensemble of GCM NAO moving window trend are indistinguishable from a white noise process as the probabilities follow the curve for  $\rho = 0.0$  (Fig. 3a). In contrast, the observations are outside the white noise 95% prediction interval for large trend thresholds, suggesting the tail of the distribution may be significantly different to that of a white noise model. The observed probability curves for trend exceedance are consistent with the fitted AR(1) process, however they are towards the upper edge of the 95% prediction intervals (Fig. 3a dark gray shading). Similar results are found for a fitted FD processes with the difference parameter estimate  $d = 0.123$  (not shown).

Using the HadSLP2r period 1851–2020 (moving window trend series of length 140), the empirical probability of a trend exceeding the maximum observed trend is given by  $1/140 = 7.14 \times 10^{-3}$ , whereas the AR(1) and FD fitted models give considerably smaller probabilities for  $\Pr(Z_i \geq 0.0737 \text{ year}^{-1})$  of  $8.22 \times 10^{-4}$  and  $1.41 \times 10^{-3}$  respectively (Table 3). The uncertainty on these estimates calculated using the Bartlett 95% confidence interval for lag-1 year autocorrelation [0.0212, 0.317] gives a probability 95% confidence interval of  $[1.59 \times 10^{-4}, 3.09 \times 10^{-3}]$





**Fig. 3** Return plots and exceedance probabilities for trends and maximum trends in the NAO, CMIP GCMs and stochastic models. The return period (lower x-axis), i.e. the reciprocal of the exceedance probabilities (upper x-axis), is shown for a range of trend thresholds (y-axis) for **a** the distribution of moving window linear trends (see Sect. 5.2) and **b** the distribution of maximum trends (see Sect. 5.3). The empirical probabilities based on GCM sample rank values are shown as gray crosses for all  $114 \times 281$  trend values (**a**) and all 281 maximum trend values (**b**). The empirical probabilities from the HadSLP2r observed NAO trend series are shown as black circles (**a**) and the maximum observed trend  $0.0737 \text{ year}^{-1}$  is marked as a horizontal black line. The expected exceedance probabilities for stochastic mod-

els are shown as black dashed lines for the  $\rho$  values labelled using an AR(1) process. In **a** the 95% prediction intervals are shown as light gray shading for the white noise model and dark gray shading for the fitted AR(1) model with  $\rho=0.169$  (labelled “0.17”) representing the uncertainty due to time series of length 144 years (medium gray shading where they overlap, such that the lower edge marks the lower boundary of the dark gray and the upper edge marks the upper boundary of the light gray). In **b** the 95% prediction interval is shown for a white noise model (gray shading) representing the uncertainty due to having a limited sample of 281 GCM simulations, using random bootstrapped samples from white noise simulations of maximum trends

for the AR(1) model and  $[1.99 \times 10^{-4}, 5.84 \times 10^{-3}]$  for the FD model. Compared to these confidence intervals, the empirical probability  $3.12 \times 10^{-5}$  estimated from the GCMs is significantly lower, and is even smaller than the estimate for a white noise process ( $1.21 \times 10^{-4}$ ). The observed estimate is not very robust as it is totally dependent on the length of the time series, however, using the whole L99 trend series still leads to a probability estimate that is higher than the stochastic and climate model best estimates:  $1/313 = 3.19 \times 10^{-3}$ . This suggests that the extreme trends are more likely in the observations than one might expect from natural variability represented by an AR(1) or a FD process. Furthermore, the observed trends are far more likely than is simulated by GCMs. This suggests a role for external drivers that may not be adequately represented in the models, for example Sun et al. (2015) show, using observation based analysis with a delayed oscillator model, that coupling with the Atlantic Ocean SST and Atlantic Meridional Overturning Circulation (AMOC) is key to modelling the NAO multi-decadal variability.

### 5.3 Distribution of the maximum in a trend series

Rather than considering trends in a predefined year, one is often more interested in particularly unusual trends that have been selected because they are the maximum seen in a given record e.g., the 1963–1993 trend in NAO observations. To assess how unusual such trends are, we need an estimate of the exceedance probability of *maximum* trends

$$q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \geq z).$$

Generalised extreme value (GEV) distributions can be used to relate the distribution of the maximum trend to properties of the index series of random variables. It would be convenient to apply extreme value theory at this point to find the theoretical relationship, but unfortunately such asymptotic behaviour is not achieved for our trend series due to the shortness of the series and the strong serial dependence due to the moving window filter. If the trend series  $\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\}$  were independent Gaussian variables (white noise), the maximum value distribution would asymptotically

converge to the Gumbel distribution, a special case of the GEV distribution with a shape parameter of zero (Coles et al. 2001 chapter 3; Kinnison 1985 chapter 7; Wilks 2006 chapter 4). However, GEV fits to the stochastic simulations of maximum trends from Sect. 4.2 find that negative shape parameter GEV distributions (Weibull distributions) provide better fits.

In Fig. 3b we estimate the probability distribution for maximum moving window trends by using the stochastic simulations of maximum trends from Sect. 4.2 and calculating the empirical probabilities based on sample rank values. Figure 3b shows the probability  $q$  for a range of thresholds  $z$  using simulations from the fitted AR(1) process with lag-1 year autocorrelation 0.169. The return period is simply  $1/q$ , but as the event refers to the maximum trend in a series generated from 144 years, the units relate to the number of such series, rather than the total number of years. Probability curves are also shown for the white noise process ( $\rho = 0.0$ ) and for AR(1) processes with  $\rho = -0.1, 0.1, 0.2, 0.3, 0.4$ . The probability  $q$  is higher for a FD process than an AR(1) process for  $\rho$  values relevant for the NAO index, but the FD probability curves are less sensitive to  $\rho$  as  $\rho$  increases, due to the relationship of the variance of trends to  $\rho$  shown in Fig. 2b.

The AR(1) and FD fitted models give exceedance probabilities ( $q$ ) of 0.0346 and 0.0606 respectively for trend threshold  $0.0737 \text{ year}^{-1}$  (Table 3). The 95% confidence intervals for these exceedance probabilities, based on the interval for  $\rho$  found in Sect. 4.1, are [0.00780, 0.123] for the AR(1) model and [0.0134, 0.194] for the FD model. In contrast, the GCM exceedance probability  $q$  is  $1/281 = 0.00356$  which is significantly smaller than both the fitted stochastic model estimates. The empirical probabilities calculated from the GCM maximum trend sample lie close to the white noise probability curve and are well within the white noise 95% prediction interval for the extreme trend values of the NAO we are interested in (Fig. 3b). This 95% probability prediction interval has been calculated using 5000 bootstrap samples of size 281 (the number of GCM simulations) taken from the white noise sample of 5000 maximum trend values. This provides evidence that the maximum observed NAO trend is more consistent with the fitted stochastic models than the GCMs. It therefore seems that the GCMs underestimate the likelihood of the 1963–1993 observed trend occurring, and the lack of multi-decadal variability in the GCMs can be partially explained by a lack of year-to-year memory.

To gain some insight into the role of ocean SST in driving NAO trends, the extreme trend analysis has been repeated for an ensemble of ERA-20CM atmospheric model simulations which cover the winters 1900/1901 to 2009/2010 (Hersbach et al. 2015). There are 10 simulations, each driven by observations of SST/sea ice and the same radiative forcing as used in CMIP5. None of these simulations reproduce

the magnitude of the observed maximum NAO trend and the average lag-1 year autocorrelation is  $-0.0287$ . The timing of the maximum trends in individual ERA-20CM simulations are spread fairly evenly throughout the period, as for the CMIP simulations. This suggests that the models are missing processes that lead to persistence within the atmosphere itself or that the model atmosphere is responding too weakly to boundary forcings that may themselves provide some year-to-year memory, such as Atlantic SST and the AMOC (Sun et al. 2015).

To test the sensitivity of our results to the choice of winter NAO definition, the analysis has been repeated, using HadSLP2, GCMs and AR1 models, for the extended winter season December-to-March (DJFM) and for an Empirical Orthogonal Function (EOF) based NAO index using the leading principal component in the North Atlantic (Hurrell 1995). For the DJFM season, the results are consistent with those for DJF. The probability of a maximum NAO trend exceeding the maximum observed ( $0.0758 \text{ year}^{-1}$ ) is 0.0176 when estimated by the fitted AR1 process, whereas none of the GCM simulations reproduce this magnitude of trend. It has been shown that there is greater multi-decadal variability for the observed NAO in late winter than early winter (Simpson et al. 2018), which may lead to this larger discrepancy between observations and CMIP models. The probability of a maximum NAO EOF index trend (DJF) exceeding the maximum observed ( $0.0701 \text{ year}^{-1}$ ) is 0.00356 when estimated by the GCMs, which is the same value as for the grid point index (i.e. just 1 simulation reproduces this magnitude of trend). In this case the average autocorrelation for the GCMs is close to zero ( $\rho = -0.00849$ ) with 15 out of 281 GCM simulations having lag-1 year autocorrelation estimates lying outside the 95% confidence interval for zero autocorrelation, i.e. very similar results to those for the grid point based index (Sect. 4.1).

As mentioned in the introduction, it is important to remember that the NAO trend from 1963 to 1993 (or close to this time interval) has been a feature of interest in the climate literature as it has already been identified as unusual in the historical record. Percival and Rothrock (2005) account for this in their analysis by considering a range of different window lengths. Repeating our analysis for window lengths 15 years and 45 years leads to results that are consistent with those of the 31 year window in terms of the GCMs likely underestimating the magnitude of multi-decadal NAO trends. The probability of a maximum NAO trend exceeding the maximum observed trend from HadSLP2 ( $0.177 \text{ year}^{-1}$  for a 15 year window and  $0.0402 \text{ year}^{-1}$  for a 45 year window) is significantly lower when estimated by the GCMs (probabilities 0.114 and 0.0107 respectively) than when estimated by the fitted AR1 process (probabilities 0.269 and 0.0306 respectively), however the AR1 model probabilities are only 2 and 3 times that of the GCM probabilities whereas

for the 31-year window this ratio is 10 times. Using the fitted AR1 model, the maximum observed 45-year trend would still be considered a *very unlikely* occurrence whereas the maximum observed 15-year trend would be characterised as *unlikely*, according to the IPCC likelihood scale (Mastrandrea et al. 2010).

## 6 Conclusions

This study has explored the use of time series modelling to quantify the behaviour of extreme trends in a moving window trend series. If the time series can be well represented by a stationary Gaussian stochastic process, such as an AR(1) or FD process, then the distribution of the moving window trend in any given year is solely determined by the variance of moving window trends ( $\sigma_z^2$ ). The distribution of maximum moving window trends has been found empirically using a large set of stochastic simulations and is not overly sensitive to the choice of stochastic model for time series with some weak year-to-year memory.

This stochastic modelling approach has been compared with a climate modelling approach to address the question of how likely it is for multi-decadal trends (linear-in-time trends) to exceed high thresholds. The focus has been to quantify the chance of seeing larger positive trends in the NAO than have occurred in historical observations, specifically the maximum seen for the 31-winter window 1963–1993 ( $0.0737 \text{ year}^{-1}$ ). The exceedance probability for a moving window trend *in any predefined year*,  $p = \Pr(Z_i \geq 0.0737 \text{ year}^{-1})$ , estimated from the fitted stochastic models is  $8.22 \times 10^{-4}$  for the AR(1) model and  $1.41 \times 10^{-3}$  for the FD model (Table 3). Both estimates are significantly higher than the GCM probability of  $3.12 \times 10^{-5}$ .

The exceedance probability for the maximum trend in a trend series of length 114 (i.e. for  $n = 144$  years, the length of the coincident CMIP historical period and observed record),  $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \geq 0.0737 \text{ year}^{-1})$ , estimated from the fitted stochastic models is 0.0346 for the AR(1) model and 0.0606 for the FD model. The 1963–1993 NAO trend was therefore a very unlikely occurrence (around 1 in 20 chance) with respect to these stochastic models (IPCC likelihood scale, Mastrandrea et al. 2010) and significantly less likely in the GCMs with a probability  $q = 0.00356$  (around 1 in 200 chance). Furthermore, we have found no improvement in CMIP6 compared to CMIP5 models in reproducing the observed low frequency variability of the NAO.

CMIP GCMs underestimate the multi-decadal variability in the NAO and this appears as a lack of serial correlation compared to the observed NAO. This lack of serial correlation in CMIP GCMs is closely linked to the low signal-to-noise ratio in GCMs (Zhang and Kirtman 2019), whereby the variance

of the ensemble mean (signal) in predictions of the NAO is weaker than would be expected given the correlation with observations (Scaife et al. 2014; Eade et al. 2014; Scaife and Smith 2018; Smith et al. 2020). The CMIP 1963–1993 multi-model ensemble mean trend is very weak ( $0.00183 \text{ year}^{-1}$ ) relative to the observed trend, and there is no consistency in the timing of maximum NAO trends in the individual GCM simulations. This suggests that the extreme multi-decadal NAO trends in GCMs are dominated by internal variability and are not responding strongly to their common boundary condition forcing (e.g. greenhouse gases).

The extreme NAO trend from the 1960s to 1990s accounted for at least half of the winter warming in the northern extratropics (Iles and Hegerl 2017; Scaife et al. 2005), yet the GCM estimate of the likelihood of this trend is only about 10% of that from a fitted AR(1) or FD model. Policy makers and planners need to be aware of this deficiency in model simulations of multi-decadal variability when using climate projections of the NAO for adaptation in order to avoid underestimating the risk of extreme trends. We have found that both AR(1) and FD processes are valuable additional models for estimating the distribution of *extreme* multi-decadal trends and they will be a useful tool to better quantify the uncertainty in multi-decadal or century long projections.

## Appendix: Proof that the trend series is the realization of an ARMA(p, 2K + 1) process

If the index series is an AR process, then the proof below shows that the moving window trend series is an ARMA process (see Wilks 1995 chapter 8 for an introduction to AR and MA processes). From Eq. (1), it can be noted that the trend series is a MA(2K + 1) filtered version of the index series:

$$Z_i = \sum_{j=-K}^K \theta_j Y_{i+j},$$

having filter weights given by  $\theta = [\frac{-K}{h^2}, \frac{-(K-1)}{h^2}, \dots, \frac{K-1}{h^2}, \frac{K}{h^2}]$ . Note that this MA filter has a unit root since the filter weights add up to 0. If the index series  $Y$  is a  $p$ 'th order autoregressive process AR( $p$ ) with parameters  $\varphi = [\varphi_1, \varphi_2, \dots, \varphi_p]$  then  $Y_j = \sum_{l=1}^p \varphi_l Y_{j-l} + \varepsilon_j$  (where  $\varepsilon_j$  are independent Gaussian variables). Substitution of this into the Equation above and rearranging the sums gives

$$\begin{aligned}
Z_i &= \sum_{j=-K}^K \theta_j Y_{i+j} \\
&= \sum_{j=-K}^K \theta_j \left( \sum_{l=1}^p \varphi_l Y_{i+j-l} + \varepsilon_{i+j} \right) \\
&= \sum_{l=1}^p \varphi_l \left( \sum_{j=-K}^K \theta_j Y_{i+j-l} \right) + \sum_{j=-K}^K \theta_j \varepsilon_{i+j} \\
&= \sum_{l=1}^p \varphi_l Z_{i-l} + \sum_{j=-K}^K \theta_j \varepsilon_{i+j} \\
\Rightarrow Z_i - \sum_{l=1}^p \varphi_l Z_{i-l} &= \sum_{j=-K}^K \theta_j \varepsilon_{i+j}.
\end{aligned}$$

The  $\varepsilon_{i+j}$  are independent and identically distributed and therefore the trend series  $Z$  is a realization of an ARMA( $p$ ,  $2K + 1$ ) process.

**Acknowledgements** R. Eade, A. A. Scaife and D. M. Smith were supported by the Met Office Hadley Centre Climate Programme funded by BEIS. This work and its contributors (R. Eade and A. A. Scaife) were supported by the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory. The authors would like to thank the anonymous reviewers for their helpful comments which have improved the quality of this paper.

**Author contributions** Not applicable.

**Funding** R. Eade, A. A. Scaife and D. M. Smith were supported by the Met Office Hadley Centre Climate Programme funded by BEIS. This work and its contributors (R. Eade and A. A. Scaife) were supported by the UK-China Research & Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund.

**Availability of data and material** The climate model datasets analysed in this study are available from the CMIP data archive: <https://esgf-node.llnl.gov/search/cmip5/> and <https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/>. Hadley Centre Sea Level Pressure dataset (HadSLP2r), a near-real-time update of HadSLP2, provided by the Met Office Hadley Centre from their Web site <https://www.metoffice.gov.uk/hadobs> (Allan and Ansell 2006). 20th Century Reanalysis V2 data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their Web site at <https://psl.noaa.gov/> (Compo et al. 2011). Reconstructed monthly NAO index provided by the Climatic Research Unit, University of East Anglia from their Web site at <https://crudata.uea.ac.uk/cru/data/paleo/naojurg/> (Luterbacher et al. 1999, 2001).

**Code availability** Not applicable.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Allan RJ, Ansell TJ (2006) A new globally complete monthly historical mean sea level pressure data set (HadSLP2): 1850–2004. *J Clim* 19(22):5816–5842. <https://doi.org/10.1175/JCLI13937.1>
- Bartlett MS (1946) On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl J R Stat Soc* 8(1):27–41. <https://doi.org/10.2307/2983611>
- Bracegirdle T, Lu H, Eade R, Woollings T (2018) Do CMIP5 models reproduce observed low-frequency north atlantic jet variability? *Geophys Res Lett*. <https://doi.org/10.1029/2018GL078965>
- Coles S, Bawa J, Trenner L, Dorazio P (2001) An introduction to statistical modeling of extreme values, vol 208. Springer, London
- Compo GP et al (2011) The Twentieth Century Reanalysis Project. *Q J R Meteorol Soc* 137(654):1–28. <https://doi.org/10.1002/qj.776>
- Deser C, Hurrell JW, Phillips AS (2017) The role of the North Atlantic Oscillation in European climate projections. *Clim Dyn* 49:3141–3157. <https://doi.org/10.1007/s00382-016-3502-z>
- Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L, Robinson N (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys Res Lett*. <https://doi.org/10.1002/2014GL061146>
- Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev* 9:1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Feldstein SB (2000) The timescale, power spectra, and climate noise properties of teleconnection patterns. *J Clim* 13(24):4430–4440. [https://doi.org/10.1175/1520-0442\(2000\)013<4430:TTPSAC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<4430:TTPSAC>2.0.CO;2)
- Feldstein SB (2002) The recent trend and variance increase of the annual mode. *J Clim*. [https://doi.org/10.1175/1520-0442\(2002\)015<0088:TRTAVI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0088:TRTAVI>2.0.CO;2)
- Ge Q, Wang H, Zheng J, This R, Dai J (2014) A 170 year spring phenology index of plants in eastern China. *J Geophys Res Biogeosci* 119(3):301–311
- Gillett NP, Graf HF, Osborn TJ (2003) Climate change and the North Atlantic oscillation. *Geophys Monogr Ser*. <https://doi.org/10.1029/134GM09>
- Greatbatch RJ (2000) The North Atlantic Oscillation. *Stoch Environ Res Risk Assess* 14:213–242. <https://doi.org/10.1007/s004770000047>
- Hanna E, Cropper TE, Jones PD, Scaife AA, Allan R (2015) Recent seasonal asymmetric changes in the NAO (a marked summer



- decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index. In *J Climatol*. <https://doi.org/10.1002/joc.4157>
- Hegerl G, Zwiers F, Braconnot P, Gillett NP, Luo YM et al (2007) Understanding and attributing climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, hal-03026507
- Hersbach H, Peubey C, Simmons A, Berrisford P, Poli P, Dee D (2015) ERA-20CM: a twentieth-century atmospheric model ensemble. *Q J R Meteorol Soc* 141:2350–2375. <https://doi.org/10.1002/qj.2528>
- Hosking JRM (1981) Fractional differencing. *Biometrika* 68(1):165–167
- Hurrell JW (1995) Decadal trends in the North Atlantic oscillation: regional temperatures and precipitation. *Science*. <https://doi.org/10.1126/science.269.5224.676>
- Hurrell JW, Kushnir Y, Visbeck M (2001) The North Atlantic Oscillation. *Science*. <https://doi.org/10.1126/science.1058761>
- Hurrell JW, Kushnir Y, Ottersen G, Visbeck M (2003) An overview of the north Atlantic oscillation. *Geophys Monogr Ser*. <https://doi.org/10.1029/134GM01>
- Hurrell JW, Van Loon H (1997) Decadal variations in climate associated with the North Atlantic Oscillation. In: Diaz HF, Beniston M Bradley, RS (eds) *Climatic change at high elevation sites*. Springer, Dordrecht. [https://doi.org/10.1007/978-94-015-8905-5\\_4](https://doi.org/10.1007/978-94-015-8905-5_4)
- Iles C, Hegerl G (2017) Role of the North Atlantic Oscillation in decadal temperature trends. *Environ Res Lett*. <https://doi.org/10.1088/1748-9326/aa9152>
- Jerez S, Trigo RM, Vicente-Serrano SM, Pozo-Vázquez D, Lorente-Plazas R, Lorenzo-Lacruz J, Santos-Alamillos F, Montávez JP (2013) The impact of the north Atlantic oscillation on renewable energy resources in Southwestern Europe. *J Appl Meteorol Climatol*. <https://doi.org/10.1175/JAMC-D-12-0257.1>
- Kim MK, McCarl BA (2005) The agricultural value of information on the North Atlantic oscillation: yield and economic effects. *Clim Change*. <https://doi.org/10.1007/s10584-005-5928-x>
- Kinnison RR (1985) *Applied extreme value statistics*. MacMillan, New York
- Kravtsov S (2017) Pronounced differences between observed and CMIP5-simulated multidecadal climate variability in the twentieth century. *Geophys Res Lett* 44:5749–5757. <https://doi.org/10.1002/2017GL074016>
- Lowe JA et al (2018) UKCP18 Science Overview Report November 2018. <https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf>. Accessed Mar 2019
- Luterbacher J, Schmutz C, Gyalistras D, Xoplaki E, Wanner H (1999) Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophys Res Lett*. <https://doi.org/10.1029/1999GL900576>
- Luterbacher J et al (2001) Extending North Atlantic Oscillation reconstructions back to 1500. *Atmos Sci Lett*. <https://doi.org/10.1006/asle.2001.0044>
- Mastrandrea MD, Field CB, Stocker TF, Edenhofer O, Ebi KL, Frame DJ, Held H, Kriegler E, Mach KJ, Matschoss PR, Plattner G-K, Yohe GW, Zwiers FW (2010) Guidance note for lead authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). <http://www.ipcc.ch>
- Matti C, Pauling A, Küttel M, Wanner H (2009) Winter precipitation trends for two selected European regions over the last 500 years and their possible dynamical background. *Theor Appl Climatol*. <https://doi.org/10.1007/s00704-007-0361-x>
- Meehl GA et al (2009) Decadal prediction: can it be skillful? *Bull Am Meteorol Soc* 90:1467–1485. <https://doi.org/10.1175/2009BAMS2607.1>
- Osborn TJ (2004) Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing. *Clim Dyn* 22(6–7):605–623
- Percival DB, Rothrock DA (2005) Eyeballing trends in climate time series: a cautionary note. *J Clim* 18(6):886–891. <https://doi.org/10.1175/JCLI-3300.1>
- Pinto JG, Raible CC (2012) Past and recent changes in the North Atlantic oscillation. *Wiley Interdiscip Rev Clim Change*. <https://doi.org/10.1002/wcc.150>
- Pozo-Vázquez D, Esteban-Parra MJ, Rodrigo FS, Castro-Díez Y (2001) A study of NAO variability and its possible non-linear influences on European surface temperature. *Clim Dyn*. <https://doi.org/10.1007/s003820000137>
- Raible CC, Stocker TF, Yoshimori M, Renold M, Beyerle U, Casty C, Luterbacher J (2005) Northern Hemispheric trends of pressure indices and atmospheric circulation patterns in observations, reconstructions, and coupled GCM simulations. *J Clim*. <https://doi.org/10.1175/JCLI3511.1>
- Santer BD, Wigley TM, Boyle JS, Gaffen DJ, Hnilo JJ, Nychka D, Parker DE, Taylor KE (2000) Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J Geophys Res Atmos*. <https://doi.org/10.1029/1999JD901105>
- Scaife AA, Smith D (2018) A signal-to-noise paradox in climate science. *NPJ Clim Atmos Sci*. <https://doi.org/10.1038/s41612-018-0038-4>
- Scaife AA, Knight JR, Vallis GK, Folland CK (2005) A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophys Res Lett* 32:L18715. <https://doi.org/10.1029/2005GL023226>
- Scaife AA, Folland CK, Alexander LV, Moberg A, Knight JR (2008) European climate extremes and the North Atlantic Oscillation. *J Clim*. <https://doi.org/10.1175/2007JCLI1631.1>
- Scaife AA et al (2009) The CLIVAR C20C project: Selected twentieth century climate events. *Clim Dyn*. <https://doi.org/10.1007/s00382-008-0451-1>
- Scaife AA et al (2014) Skillful long-range prediction of European and North American winters. *Geophys Res Lett*. <https://doi.org/10.1002/2014GL059637>
- Semenov VA, Latif M, Jungclauss JH, Park W (2008) Is the observed NAO variability during the instrumental record unusual? *Geophys Res Lett* 35:L11701. <https://doi.org/10.1029/2008GL033273>
- Shi Y, Zhai P, Jiang Z (2016) Multi-sliding time windows based changing trend of mean temperature and its association with the global-warming hiatus. *J Meteorol Res*. <https://doi.org/10.1007/s13351-016-5093-3>
- Shindell DT, Miller RL, Schmidt GA, Pandolfo L (1999) Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature*. <https://doi.org/10.1038/20905>
- Shpakova RN, Kusatov KI, Mustafin SK (2020) Spatiotemporal trends in changes in the river water contents in the Sakha Republic (Yakutia). *IOP Conf Ser Earth Environ Sci*. <https://doi.org/10.1088/1755-1315/459/5/052062>
- Simpson IR, Deser C, McKinnon KA, Barnes EA (2018) Modeled and observed multidecadal variability in the North Atlantic jet stream and its connection to sea surface temperatures. *J Clim*. <https://doi.org/10.1175/JCLI-D-18-0168.1>
- Smith DM et al (2020) North Atlantic climate far more predictable than models imply. *Nature* 583(7818):796–800. <https://doi.org/10.1038/s41586-020-2525-0>
- Stephenson DB, Pavan V, Bojariu R (2000) Is the North Atlantic oscillation a random walk? *Int J Climatol*. <https://doi.org/10.1002/joc.4157>

- 1002/(SICI)1097-0088(200001)20:1%3c1::AID-JOC456%3e3.0.CO;2-P
- Sun C, Li J, Jin FF (2015) A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. *Clim Dyn* 45:2083–2099. <https://doi.org/10.1007/s00382-014-2459-z>
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc.* <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Thompson DW, Barnes EA, Deser C, Foust WE, Phillips AS (2015) Quantifying the role of internal climate variability in future climate trends. *J Clim.* <https://doi.org/10.1175/JCLI-D-14-00830.1>
- Thornton HE, Scaife AA, Hoskins BJ, Brayshaw DJ (2017) The relationship between wind power, electricity demand and winter weather patterns in Great Britain. *Environ Res Lett.* <https://doi.org/10.1088/1748-9326/aa69c6>
- Uvo CB, Berndtsson R (2002) North Atlantic Oscillation; a climatic indicator to predict hydropower availability in Scandinavia. *Nordic Hydrol.* <https://doi.org/10.2166/nh.2002.0016>
- Wang X, Li J, Sun C, Liu T (2017) NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations. *J Geophys Res Atmos* 122:4202–4227. <https://doi.org/10.1002/2016JD025979>
- Wanner H, Brönnimann S, Casty C, Gyalistras D, Luterbacher J, Schmutz C, Stephenson DB, Xoplaki E (2001) North Atlantic oscillation—concepts and studies. *Surv Geophys.* <https://doi.org/10.1023/A:1014217317898>
- Wilks DS (2006) *Statistical Methods in the atmospheric sciences* second edition. International geophysics series, p 91
- Woodward WA, Gray HL (1993) Global warming and the problem of testing for trend in time series data. *J Clim.* [https://doi.org/10.1175/1520-0442\(1993\)006%3c0953:GWATPO%3e2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006%3c0953:GWATPO%3e2.0.CO;2)
- Woollings T, Franzke C, Hodson DL, Dong B, Barnes EA, Raible CC, Pinto JG (2015) Contrasting interannual and multidecadal NAO variability. *Clim Dyn* 45:539–556. <https://doi.org/10.1007/s00382-014-2237-y>
- Wunsch C (1999) The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull Am Meteorol Soc.* [https://doi.org/10.1175/1520-0477\(1999\)080%3c0245:TIOSCR%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080%3c0245:TIOSCR%3e2.0.CO;2)
- Zanardo S, Nicotina L, Hilberts AG, Jewson SP (2019) Modulation of economic losses from European Floods by the North Atlantic Oscillation. *Geophys Res Lett.* <https://doi.org/10.1029/2019GL081956>
- Zhang W, Kirtman B (2019) Understanding the signal-to-noise paradox with a simple markov model. *Geophys Res Lett.* <https://doi.org/10.1029/2019GL085159>
- Zuo J, Ren H-L, Li W, Wang L (2016) Interdecadal variations in the relationship between the Winter North Atlantic oscillation and temperature in South-Central China. *J Clim.* <https://doi.org/10.1175/JCLI-D-15-0873.1>

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