



Recalibration of missing low-frequency variability and trends in the North Atlantic Oscillation

R. Eade^{1,2} · D. B. Stephenson² · A. A. Scaife^{1,2} · D. M. Smith¹

Received: 12 March 2024 / Accepted: 11 June 2024 / Published online: 22 June 2024
© Crown 2024

Abstract

Multi-decadal trends in the wintertime North Atlantic Oscillation (NAO) are under-represented by coupled general circulation models (CGCMs), consistent with a lack of autocorrelation in their NAO index series. This study proposes and tests two simple “reddening” approaches for correcting this problem in simulated indices based on simple one parameter short-term (AR; Auto-Regressive order 1) and long-term (FD; Fractional-Difference) time series filters. Using CGCMs from the Coupled Model Intercomparison Project Phase 6 (CMIP6), the FD filter successfully improves the autocorrelation structure of the NAO, and in turn the simulation of extreme trends, while the AR filter is less successful. The 1963–1993 NAO trend is the maximum 31-year trend in the historical period. Raw CGCMs underestimate the likelihood of this trend by a factor of ten but this discrepancy is corrected after reddening. CMIP6 future projections show that long-term (2024–2094) NAO ensemble mean trends systematically increase with the magnitude of radiative forcing: -2.4 to 3.5 hPa/century for low-to-high forcing after reddening (more than double the range using raw output). The related likelihood of future maximum 31 year trends comparable to 1963–1993 ranges from 3 to 7% whereas none of these CMIP6 projections simulate this without reddening. Near-term projections of the next 31 years (2024–2054) are less sensitive than long term trends to the future scenario, showing weak-to-no forced trend. However, reddening increases the ensemble range by 74% (to +/-1 standard deviation/decade), which could increase/decrease regional climate change signals in the Northern Hemisphere by magnitudes that are underestimated when using raw CGCM output.

Keywords North Atlantic Oscillation · Recalibration · Future projections · Extreme trends · Multi-decadal variability · Climate modelling · Stochastic modelling

1 Introduction

The North Atlantic Oscillation (NAO) is the dominant mode of variability in the North Atlantic on interannual and longer timescales. It is often defined as the first Empirical Orthogonal Function (EOF) of monthly mean sea level pressure (MSLP) variability in the North Atlantic region, with physically consistent patterns of variability in other atmospheric variables (Hurrell 1995; Ambaum et al. 2001). This paper focusses on the winter NAO which has been shown to impact many aspects of society in Europe and the USA.

The NAO can impact agriculture, fishing and water management through its effect on temperature, extreme rainfall and drought (Hurrell 2001) which in turn affects planning for food import needs (Kim and McCarl 2005). The energy sector is influenced by the NAO in terms of energy demand relating to winter temperatures and in terms of energy supply relating to wind, solar and hydropower (Jerez et al. 2013; Uvo and Berndtsson 2002; Thornton et al. 2017). The NAO can also impact the insurance industry, related to damage from extreme events such as high winds and flooding (Zanardo et al. 2019), and transport via snow and ice (Palin et al. 2016). Temperature related impacts in Eurasia may also be apparent beyond winter into spring and summer due to the interaction between the NAO and sea surface temperature in the North Atlantic (Wu and Chen 2020; Chen et al. 2020). Understanding multi-decadal variability of the NAO is therefore important for all these sectors when planning for future adaptation.

✉ R. Eade
rosie.eade@metoffice.gov.uk

¹ Met Office, Exeter, UK

² Department of Mathematics and Statistics, University of Exeter, Exeter, UK

The multi-decadal variability of the NAO dominates northern hemisphere regional rainfall and temperature projections in winter (Deser et al. 2017). The NAO was shown to be responsible for the tendency for cold UK winters in the 1960s moving towards milder and wetter winters in the 1990s (Hurrell 1995). The large positive trend in the winter NAO index from 1963 to 1995 accounted for at least half of the winter warming in the Northern Hemisphere extratropics (Iles and Hegerl 2017; Scaife et al. 2005), while the large negative trend from 1920 to 1971 more than halved the winter warming (Iles and Hegerl 2017). These time windows, on which the linear trends are calculated, actually contain the *maximum* 33-year trend and *minimum* 52-year trend in the NAO historical record. It is therefore of interest to quantify the likelihood of such extreme NAO trends occurring in the future.

Previous studies have found that externally forced coupled general circulation models (CGCMs), such as those from the Coupled Model Intercomparison Project (CMIP), are generally unable to simulate the full magnitude of multi-decadal NAO trends observed in the historical record. This issue was originally noticed after the large positive NAO trend was observed around the 1960s to 1990s and both CMIP Phase 2 (CMIP2) CGCMs and atmosphere-only simulations underestimated the magnitude of this trend (e.g. Stephenson et al. 2006; Scaife et al. 2009). Davini and Cagnazzo (2013) show that CMIP5 models are fairly good at simulating the observed NAO pattern of interannual variability, but some of these models do not correctly capture the related dynamics and physical processes particularly the observed coupling to the Greenland blocking frequency. More recently, the CMIP Phase 5 (CMIP5) and Phase 6 (CMIP6) simulations have been used to show that climate models still underestimate the NAO multi-decadal variability (Lee et al. 2021; O'Reilly et al. 2021) and extreme 31-year linear trends (Bracegirdle et al. 2018; Eade et al. 2022; Bracegirdle 2022), despite having interannual variability similar that observed (Lee et al. 2021; Eade et al. 2022). The recent longer-term NAO trend 1951–2020 is also underestimated in CMIP6 experiments (combining historical experiments with near-term future projections) which so happens to be the maximum 70-year NAO trend in the observational record (Blackport and Fyfe 2022). The lack of multi-decadal variability in CGCMs in the historical experiments suggests that future projections likely also underestimate the magnitude of NAO multi-decadal variability (O'Reilly et al. 2021).

The underestimation of multi-decadal NAO trends in CGCMs has been shown to be consistent with a lack of autocorrelation in the NAO index series compared to the observed NAO (Eade et al. 2022). The most common stochastic models used to represent the NAO are short-range dependence first order auto-regressive (AR(1)) red noise processes (e.g. Wunsch 1999; Feldstein 2000; Thompson

et al. 2015) and long-range dependence (FD) processes (Stephenson et al. 2000). These simple stochastic models, using parameter estimates fitted to the observed winter mean NAO index, have been shown to better simulate the magnitude of the observed extreme NAO trends compared to CGCMs which in turn behave more like a white noise process i.e. with zero autocorrelation (Eade et al. 2022). The source of the observed NAO autocorrelation may be related to the interaction with North Atlantic sea surface temperature (Mosedale et al. 2006) and ocean reemergence (e.g. Timlin et al. 2002; Buchan et al. 2014). It could also be related to the response to external forcings such as solar variability (Scaife et al. 2013; Gray et al. 2013) or aerosol forcing (e.g. Watanabe and Tatebe 2019). The CGCM lack of autocorrelation and similarity to a white noise process has been noted in terms of jet stream variability (Simpson et al. 2018), and also for North Atlantic variability more generally (O'Reilly et al. 2021). The CGCM lack of autocorrelation and underestimation of extreme trends is not sensitive to the definition of the NAO index, as Eade et al. (2022) found the same results when using an Empirical Orthogonal Function based index. The cause of the lack of autocorrelation in CGCMs is unclear but could be related to differences in the atmospheric internal variability, the response to external forcings and the strength of ocean-atmosphere coupling (Gastineau and Frankignoul 2015; Blackport and Fyfe 2022; Bracegirdle 2022). The lack of autocorrelation is also likely related to the signal-to-noise paradox (Zhang et al. 2021; Strommen and Palmer 2018; Strommen 2020). This paradox has been identified in recent studies whereby the magnitude of the CGCM NAO forced signal is weaker than expected from the magnitude of the correlation with observations (Scaife et al. 2014; Eade et al. 2014; Scaife and Smith 2018; Smith et al. 2020; Klavans et al. 2021). The cause of this is not yet fully explained but may be related to a lack of eddy feedback in models (Hardiman et al. 2022).

A new recalibration reddening method is proposed here to increase the autocorrelation of the NAO index series in CGCMs and thereby improve the simulation of extreme multi-decadal NAO trends. The NAO index definition and the reddening method are described in Section 2. The reddening method is tested on the historical simulations from CMIP6 in Section 3 and then applied to future projections in Section 4 to forecast potential changes in NAO multi-decadal variability under different future scenarios.

2 Data and methods

2.1 Observation and model data

The winter NAO index is calculated using Mean Sea Level Pressure (MSLP) averaged over the winter season December

to February (DJF), where January is used to define the year. The HadSLP2r gridded observation dataset (Allan and Ansell 2006) is used to represent the observed NAO as it is available for a long historical period of 171 years (1851–2021).

The CGCM NAO index series are computed using the multi-model ensemble output from CMIP6 (Eyring et al. 2016) for sets of historical experiments (1850–2014) and future projections (2015–2100). Experiments from CMIP6 use a single set of historical external forcings for the period up to 2014 representing past *natural* changes such as solar irradiance and volcanic aerosol due to past eruptions and also past *anthropogenic* changes such as emissions of greenhouse gases and aerosols. From 2015 to 2100 multiple future scenarios have been defined to determine the levels of external forcing applied. These scenarios are known as Shared Socio-economic Pathways (SSP) and represent a range of possible futures with different anthropogenic forcings. Four scenarios are analysed in this paper: SSP126, SSP245, SSP370 and SSP585 going from optimistic reductions of greenhouse gas and aerosol emissions (compatible with a 2 °C target level of global mean warming by 2100) to the more pessimistic “business as usual” scenario where emissions continue to increase (O’Neill et al. 2016). All future scenarios have the same natural external forcings with a regular 11-year solar cycle and a background level of volcanic forcing without future eruptions. All scenarios show an increase of the anthropogenic effective radiative forcing out to 2100 except SSP126 which has a slight decrease in the second half of the 21st century but levels are still higher than historical values (Arias et al. 2021). Models are included where they have the same set of members available for the historical experiments and all four future scenario experiments. This results in a set of 36 models with a total of 258 members (Appendix 1).

2.2 Definition of the NAO trend series

A *standardised* NAO index is defined as the standardised difference of the standardised MSLP time series at the two main nodes of NAO variability, that is Ponta Delgada, Azores at 37.7 N, 25.7 W minus Reykjavík, Iceland at 65.0 N, 22.8 W (Hurrell 1995), using the nearest grid boxes in the winter mean MSLP gridded observation datasets and CGCM output. This NAO index is referred to in units of “sd” meaning standard deviations. The climate period for standardisation is defined to be 1862–2005. A *non-standardised* version of this index, the straight MSLP difference between Ponta Delgada and Reykjavík, is also sometimes referred to so that results can be discussed in units of hPa, and results are not sensitive to the choice of definition.

A *multi-decadal trend* is here defined as an Ordinary Least Squares (OLS) trend in time for a window of fixed

length, and is the slope parameter estimate obtained from linear regression of the index on time. This NAO trend is referred to in units of “sd/decade”, meaning standard deviations per decade. Moving window trend series are obtained by shifting a window along the index time series year-by-year and calculating the linear trend estimate within each window, resulting in a moving average filtered version of the original index series (Eade et al. 2022). A 31-year window is chosen to match results from Eade et al. (2022) but other window lengths are also considered. An *extreme trend* is defined as the block maximum (or minimum) value in the moving window trend series.

2.3 Distribution of moving window trends

In Eade et al. (2022) it was shown that the distribution of moving window trends can be explicitly determined if the original index is a Gaussian stationary stochastic process. In this case, the moving window trends have a Gaussian distribution with a mean of zero and variance

$$\sigma_z^2 = \frac{\sigma_Y^2}{\left(\sum_{j=-K}^K j^2\right)^2} \sum_{j=-K}^K \sum_{k=-K}^K jk \rho_{k-j} \quad (1)$$

where $2K+1$ is the window length, and σ_Y^2 and ρ_{k-j} are respectively the variance and the lag $k-j$ autocorrelation of the index process. It can be seen from Eq. (1) that, for a given window length, the variance of the moving window trends is related to the magnitude of the variance σ_Y^2 and autocorrelation ρ_{k-j} of the original index process. For example, if the lag-1 year autocorrelation is increased then the multi-decadal variability of the index series is generally increased, described here by the variance of the moving window trend series (σ_z^2), and thus the likelihood of large trends is also increased, whether the window length represents a short- or a long-term trend.

Discrepancies between CGCMs and observations may thus be approached by correcting the variance and/or the autocorrelation structure of the CGCM index series. In Eade et al. (2022) it was shown that the CMIP5 and CMIP6 models underestimate the autocorrelation of the winter mean NAO index series with lag-1 year autocorrelation coefficients that are approximately zero and behaviour similar to that of a white noise process, while the observed estimate of lag-1 year autocorrelation is around 0.17. For the *standardised* NAO index definition used here, the CGCM variance has already been corrected to match that observed (i.e. a variance of one). For the *non-standardised* version of the NAO index, CMIP5 and CMIP6 models are also reasonable at simulating the interannual variability (Eade et al. 2022). Thus, it is just the discrepancy in autocorrelation structure that is considered here.

2.4 Reddening methods

To correct for the low level of autocorrelation in the CGCM NAO index series, a parsimonious low-pass linear time series filter approach is proposed to increase the autocorrelation. The filters are based on a Fractional first order Auto-Regressive FAR(1) process which involves two parameters, the auto-regressive parameter $\varphi \in (-1, 1)$ and the difference parameter $d \in (-0.5, 0.5)$. Given an original index series Y , the filtered or “reddened” series Y^R using a FAR(1) process is given by

$$Y^R = \gamma(1 - \varphi B)^{-d} Y \quad (2)$$

(Hosking 1981) where B is the backward shift operator such that $BY(t) = Y(t-1)$ and γ is a constant chosen to retain the variance of the original series. The parameters for this reddening filter can be determined such that the resulting reddened CGCM NAO index series will have the same fitted FAR(1) parameters as for the observed NAO index series.

Because of the shortness of the observed index series, it is not possible to reliably estimate all three of the parameters in this (or more complex) models. Therefore, we consider instead these simple cases: Short-term AR(1) filter and long-term FD filter. These processes have been used previously to model the NAO (Wunsch 1999; Feldstein 2000; Stephenson et al. 2000).

If it is assumed that the observed NAO index can be approximated as an AR(1) process then it is appropriate to apply an AR(1) filter to increase the level of short-range dependence in the CGCM index (Y). For the AR(1) filter, the new reddened index Y^R is defined by

$$Y^R = \gamma(1 - \varphi B)^{-1} Y \quad (3)$$

If Y is assumed to be a memory-less Gaussian white noise process, Y^R is then an AR(1) process and φ can be simply estimated to be equal to the lag-1 sample autocorrelation of the observed index series:

$$\varphi = \rho_R. \quad (4)$$

The scaling parameter γ is chosen to ensure that the expected variance of Y^R equals the expected variance of Y . It can be shown that the variance of the AR(1) process Y^R is given by

$$\text{Var}(Y^R) = \frac{\gamma^2 \text{Var}(Y)}{1 - \varphi^2} \quad (5)$$

(Hamilton 1994). To make these variances equal, the scaling parameter γ is given by:

$$\gamma^2 = 1 - \varphi^2. \quad (6)$$

If it is assumed that the observed NAO index series can be approximated as a FD process, then it is appropriate to apply a FD filter to increase the level of long-range dependence in the CGCM index (Y). For the FD filter, the new reddened index Y^R is defined by

$$Y^R = \gamma(1 - B)^{-d} Y. \quad (7)$$

For the simple case where Y is a Gaussian white noise process, Y^R is then an FD process having lag-1 autocorrelation $\rho = \frac{d}{1-d}$. For given values of the difference parameter d_R estimated from the observed index series, the level of long-range dependence in this new FD process Y^R is controlled by setting

$$d = d_R. \quad (8)$$

The scaling parameter γ is chosen to ensure that the expected variance of Y^R equals the expected variance of Y . It can be shown that the variance of the FD process Y^R is given by

$$\text{Var}(Y^R) = \frac{\gamma^2 \text{Var}(Y) \Gamma(1 - 2d_R)}{(\Gamma(1 - d_R))^2} \quad (9)$$

where Γ is the Gamma function (Hosking 1981). To make these variances equal, the scaling parameter γ is given by

$$\gamma^2 = \frac{(\Gamma(1 - d_R))^2}{\Gamma(1 - 2d_R)}. \quad (10)$$

The reddening methods outlined above assume that the CGCM NAO index series can be modelled as a white noise process, which appears to be justifiable from the lag-1 autocorrelations (Section 3.1; Eade et al. 2022). An extension of these methods to account for a non-zero level of autocorrelation or long-range dependence in CGCM NAO index series is outlined in Appendix 2, but the results show little sensitivity compared to the original method (Section 3).

2.5 Autocorrelation in the observed NAO index

The observed winter mean NAO index has a lag-1 year autocorrelation parameter of $\rho = 0.17$ estimated from the Had-SLP2r NAO index 1851–2020 (Eade et al. 2022). For the convenience of quantitatively comparing the AR(1) and FD reddening methods, the difference parameter is estimated from ρ using

$$d(\rho) = \frac{\rho}{1 + \rho} \quad (11)$$

(Hosking 1981) such that $d = 0.15$. The autocorrelation parameter estimate has been shown to be somewhat sensitive to the time period used (e.g. Hurrell and van Loon

1997), but this could easily be due to the uncertainty from using relatively short observation datasets. The distribution of the lag-1 year autocorrelation parameter can be assumed to be Gaussian such that the 95% confidence interval (i.e. 2.5th to 97.5th percentiles) for the lag-1 year autocorrelation parameter ρ for an index series of length n is given by the Bartlett formula:

$$\rho \pm 1.96 \sqrt{\frac{(1 - \rho^2)}{n}} \quad (12)$$

(Bartlett 1946) where 1.96 is the threshold such that the probability $P(|X| < 1.96) = 0.95$ for the standard Gaussian variable X . For the HadSLP2r lag-1 year autocorrelation estimate of $\rho = 0.17$ with $n = 170$ (time series length), the Bartlett 95% confidence interval for ρ is (0.021, 0.32) and the equivalent interval for d is (0.021, 0.24). These confidence intervals are considered when reddening the CGCM NAO series, however the lower values are so close to zero that there will be little-to-no effect from reddening.

3 Improved distribution of historical NAO trends in climate models

3.1 Autocorrelation of NAO index series

Figure 1a shows the effect of the reddening process on the distribution of the lag-1 year autocorrelation parameter across all the 258 CGCM NAO simulations. The distribution for the raw CGCM output is centred on zero, as found in Eade et al. (2022). The observed level of lag-1 year autocorrelation is 0.17 and significantly different to that of a white noise process when compared to the related 95% confidence interval (Fig. 1a). After applying the AR(1) and FD reddening methods with observed estimates of autocorrelation and fractional difference parameters ($\rho_R = 0.17$ and $d_R = 0.15$), these distributions are shifted up to the observed values as expected and retain roughly the same width as for the raw output. The 95% confidence interval for the estimate of the autocorrelation in the observed NAO index series shows considerable uncertainty (Section 2.5), thus this interval is also considered when reddening the CGCMs. The lower values in these intervals are so close to zero that there will be little-to-no effect from reddening so only the upper values are shown in subsequent analysis. Figure 1b shows the

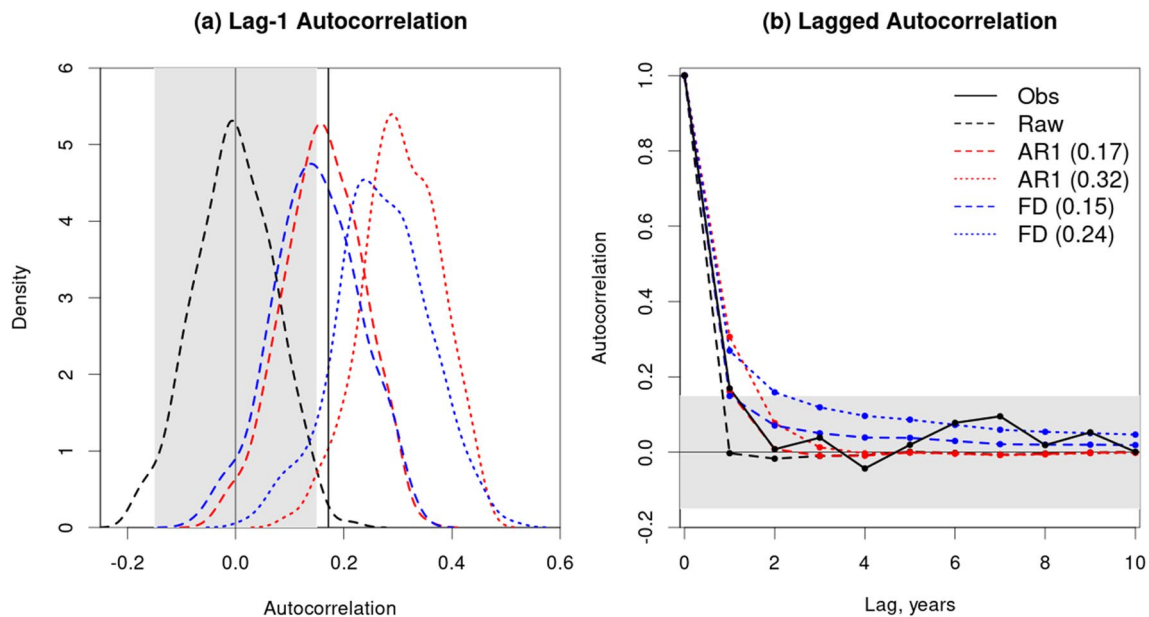


Fig. 1 Reddening effect on autocorrelation function of CGCM NAO index. **a** The distribution of lag-1 year autocorrelation parameter estimates for all the individual raw CGCM historical simulations (1851–2021) using SSP585 beyond 2014 (black dashed line). The distributions are also shown for the reddened CGCM output using the AR(1) method with stochastic parameters shown in brackets ($\rho_R = 0.17, 0.32$; $d_R = 0.15, 0.24$). The HadSLP2r observed estimate is shown by the

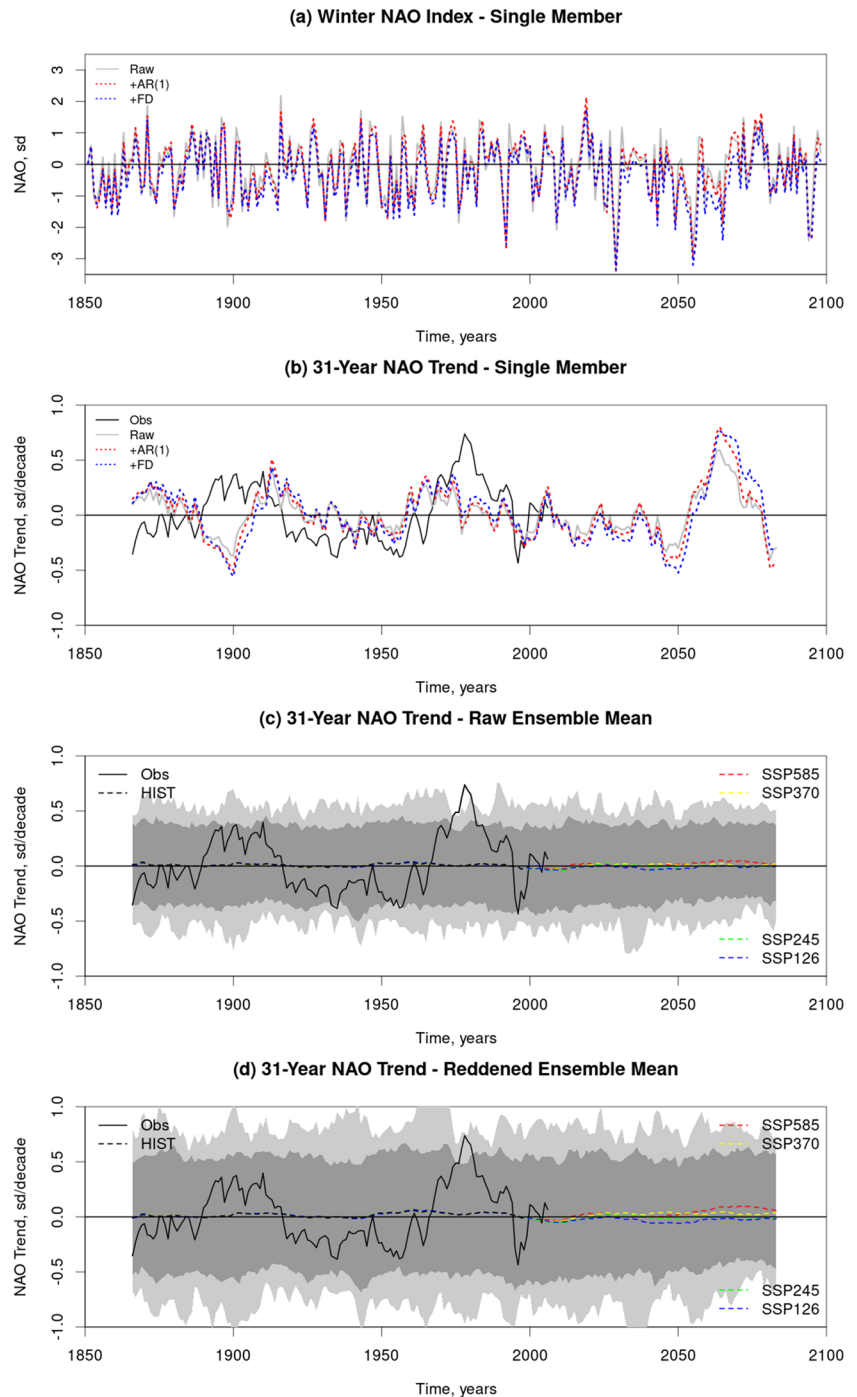
thick black vertical line ($\rho = 0.17$). **b** The average lag autocorrelation value across all individual CGCM ensemble members for lags zero to ten years for the raw and reddened CGCM output as in (a). The observed estimates are shown by the thick black solid line. Gray shading represents the 95% confidence interval for a white noise process using the Bartlett formula (Eq. 12) with $\rho = 0$

multi-model ensemble mean of lagged autocorrelation estimates for lags up to ten years. An AR(1) process can only model short-range dependence, thus the weak long-range dependence apparent in the observed NAO series is better captured by the FD reddened CGCMs than the AR(1) reddened, especially if the upper parameter estimate $d_R = 0.24$ is used.

3.2 Variability of NAO multi-decadal trend series

Figure 2a shows the effect of the reddening process on a single CGCM simulation of the NAO index (UKESM1-0-LL) using the AR(1) and FD methods with upper estimates of the stochastic distribution parameters $\rho_R = 0.32$ and $d_R = 0.24$. Both reddening methods lead to accentuated clusters of positive or negative phases of CGCM NAO index values while retaining the original timing. The reddening methods

Fig. 2 Reddening effect on CGCM NAO 31-year trend series. **a** NAO index and **(b)** 31-year trend (relative to the central year) for a single CGCM simulation (UKESM1-0-LL) using raw (black) model output and AR(1) (red) and FD (blue) reddened output with $\rho_R = 0.32$ and $d_R = 0.24$. **c** and **(d)** Dashed curves show CMIP6 multi-model ensemble mean NAO 31-year trend series using historical (black) and future (colours) experiments for **(c)** raw model output and **(d)** FD reddened output ($d_R = 0.24$). Light and dark gray shading shows the ensemble range and empirical 95% prediction interval respectively (historical and SSP585). The observed 31-year trend series is shown for HadSLP2r in a-c (black solid curve)



lead to a widening of the CGCM 31-year moving window trend distribution in both directions and an increase in the absolute magnitude of both maximum and minimum 31-year NAO trends (Fig. 2b).

Figures 2c and 3a show the time evolution of the raw CGCM 31-year and 71-year NAO trend series in terms of the multi-model ensemble mean, ensemble range, and 95% prediction interval (based on the empirical 2.5th and 97.5th percentiles across all 258 single members in the CMIP6 ensemble), using future scenario SSP585 beyond 2014. 31-year and 71-year windows are chosen to enable comparisons to other recent results for extreme NAO trends in Section 3.3. The multi-model ensemble mean trend series exhibit multi-decadal variability but this is much weaker than the observed variability as expected. Over the historical period, the observed *moving window* trend series are mainly contained within the CGCM ensemble 95% prediction interval but the observed *extreme* trends are generally outside of the prediction interval and in many cases outside of the CGCM total ensemble range. After applying the FD reddening method with $d_R = 0.24$ (Figs. 2d and 3b), the variance of the CGCM moving window NAO trends is increased (as expected from Section 2.3 and Eq. (1) where the reddening method has increased the magnitude of the autocorrelation function) and thus the 95% prediction interval and range are inflated to better encompass the full range of the observed multi-decadal variability.

The standard deviation of the observed 31-year trend series for the historical period 1851–2014 is about 35% larger than the total standard deviation of the raw CGCM ensemble (0.264 sd/decade vs. 0.195 sd/decade, Table 1 row a). The upper level of reddening ($\rho_R = 0.32$ or $d_R = 0.24$) is needed for the CGCMs to have a total standard deviation of 31-year trends comparable to that observed, and this also leads to an inflation of the standard deviation of the ensemble mean by 33% and 52% respectively (Table 1 row b). The estimate of the standard deviation of the observed trend series is not very robust given that there are very few independent 31-year windows in the 164-year historical record (just five non-overlapping windows in period 1851–2014) so this analysis is not repeated for longer windows.

The cumulative distribution function for *moving window* trends is given by one minus the exceedance probability $p(z)$ of the trend in year i (Z_i) exceeding a threshold value of z :

$$p(z) = \Pr(Z_i \geq z). \quad (13)$$

For the probability of trends more negative than a threshold, we will use the terms *negative-exceedance* probability and $p'(z)$:

$$p'(z) = \Pr(Z_i \leq z). \quad (14)$$

The effect of the reddening methods on CGCM 31-year moving window trend exceedance probabilities is shown in

Fig. 3 Reddening effect on CGCM NAO 71-year trend series. **a** and **(b)** Dashed curves show CMIP6 multi-model ensemble mean NAO 71-year trend series using historical (black) and future (colours) experiments for **(a)** raw model output and **(b)** FD reddened output ($d_R = 0.24$). Light and dark gray shading shows the ensemble range and empirical 95% prediction interval respectively (historical and SSP585). The observed 31-year trend series is shown for HadSLP2r (black solid curve)

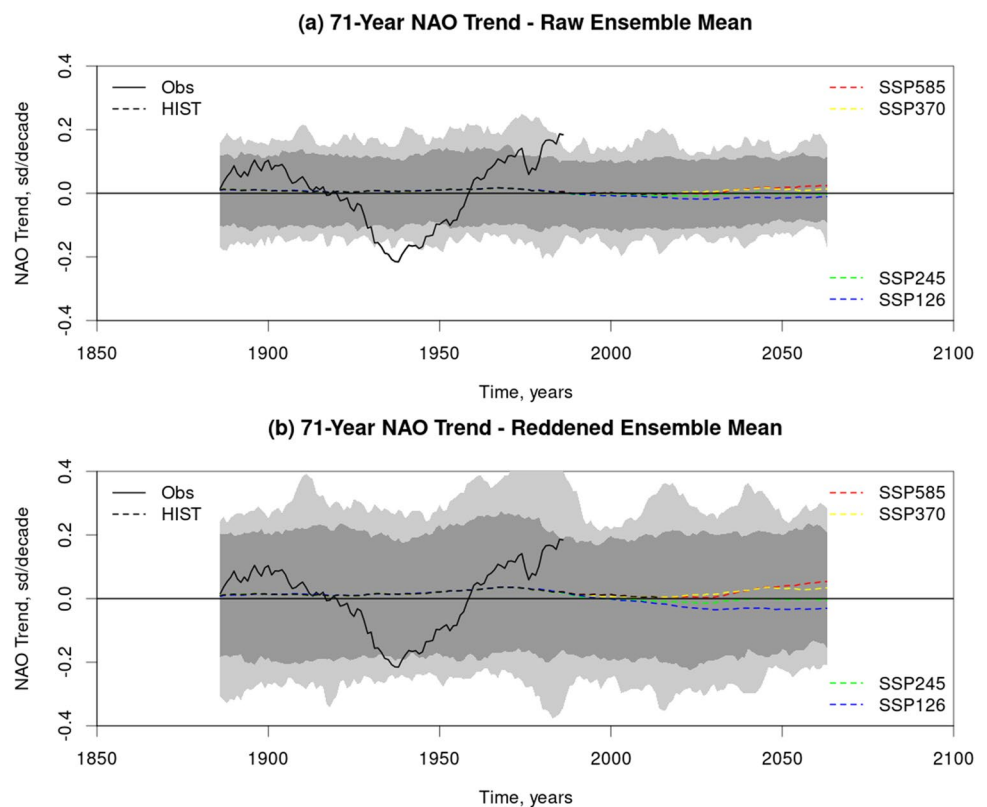


Table 1 Effect of the reddening methods on the CGCM historical distribution of 31-year moving window and extreme trends

	Raw CMIP	AR(1) reddened CGCM		FD reddened CGCM	
	($\rho_R = 0.0$)	$\rho_R = 0.17$	$\rho_R = 0.32$	$d_R = 0.15$	$d_R = 0.24$
(a) Total S.D.	0.195	0.227 (17%)	0.262 (34%)	0.248 (27%)	0.284 (46%)
(b) Ens. Mean S.D.	0.012	0.014 (16%)	0.016 (33%)	0.016 (26%)	0.019 (52%)
(c) Z_{1978} Range	1.108	1.310 (18%)	1.526 (38%)	1.505 (36%)	1.740 (57%)
(d) CGCM q	0.388%	2.71%	11.6%	6.59%	15.5%
(e) Stoch. q	0.600%	3.46%	12.30%	6.06%	19.40%
(f) CGCM q ext.	0.388%	2.33%	9.69%	6.20%	16.7%

CGCM statistics are shown using the raw NAO index series and for the AR(1) and FD reddened NAO index series with autocorrelation and difference parameters as indicated by ρ_R and d_R in the column headings. (a) The total standard deviation of 31-year moving window NAO trend series from CMIP6 simulations over the historical period 1851–2014 (in units of sd/decade) and (b) the standard deviation of the CGCM ensemble mean. The standard deviation of the observed trend series using HadSLP2r over the same period is 0.264 sd/decade. (c) The ensemble range for the linear trend on the specific time window centred on 1978 (Z_{1978} , 1963–1993). The percentage increase in standard deviation and range values for the reddened CGCMs relative to the raw CGCM values are shown in brackets. (d) CGCM empirical probabilities (shown as percentages) of trend exceedance $q = \Pr(\max \text{ trend} \geq z)$ in the historical period 1862–2005, where z is the maximum observed 31-year trend (0.737 sd/decade). (e) Fitted stochastic model exceedance probability estimates of q taken from Eade et al. (2022; Appendix Table 3), using the same stochastic models as used in the reddening methods and using a white noise process for the Raw CMIP column. (f) as in (d) but the reddening methods use the sample estimates of the CGCM lag-1 year autocorrelation rather than assuming these are identically equal to zero. Values are shown to three significant figures (except percentage increases shown to two)

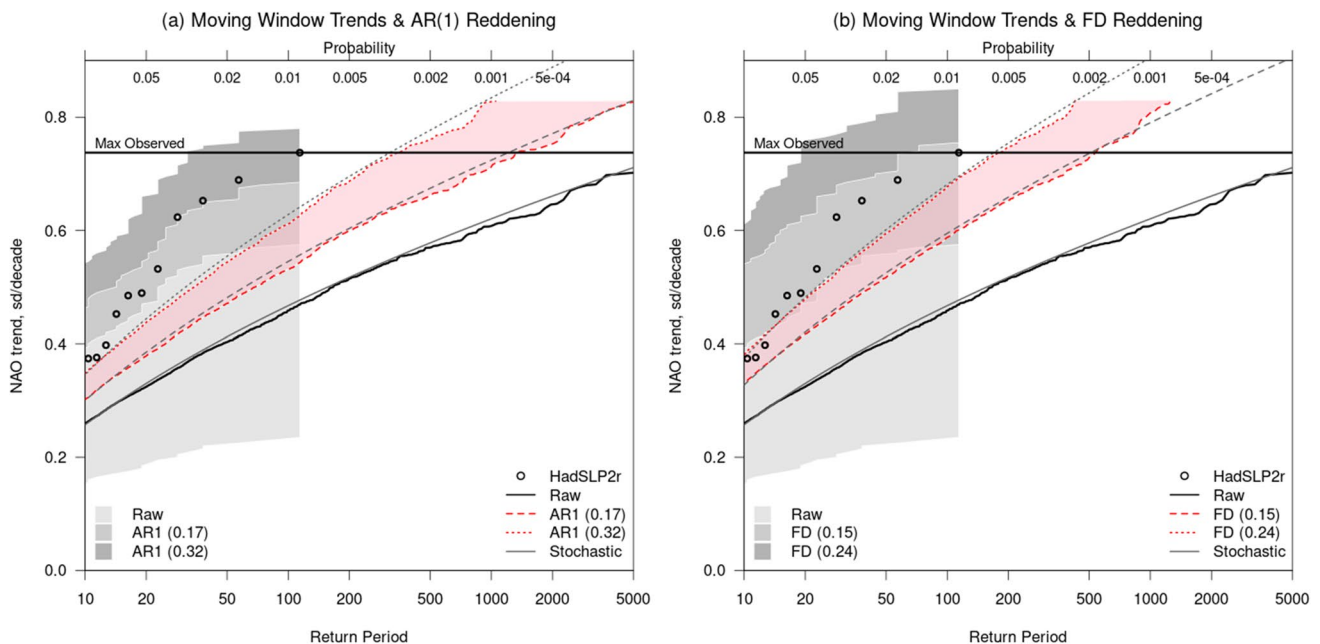


Fig. 4 Return Period Plots for NAO 31-year moving window trends and extreme trends for reddened CGCMs. The empirical exceedance probabilities (top axis) for 31-year *moving window* NAO trend series (return period on the bottom axis) are shown for the CGCMs using raw model output (black solid line) and observations (black open circles, using HadSLP2r) with the maximum observed trend identified by the horizontal black line. The pink shading shows the range of probabilities after applying (a) the AR(1) reddening method with ρ_R in range (0.17, 0.32) and (b) the FD reddening method with d_R in range (0.15, 0.24), bounded by the red curves identified in the key by solid and dashed lines respectively. The gray curves show the theoretical probability curves assuming a Gaussian distribution with

mean zero and variance computed using Eq. (1) with $\sigma_Y^2 = 1$ for a white noise process (solid) and FD processes with $d_R = 0.15$ (dashed) and 0.24 (dotted). The gray shading represents the CGCM sampling uncertainty due to short time series (sample size 114), using the 95% prediction interval for empirical exceedance probabilities from each of the individual CGCM NAO simulations, for raw and reddened output as shown in the key. These intervals overlap so they are plotted such that the upper bound of each interval is visible. The short time series mean that the lowest non-zero probability is 1/114 thus the gray shading does not extend across the full range of return periods shown

Fig. 4a and b. The raw CGCM empirical rank exceedance probabilities are shown using the total multi-model ensemble, treated as one large sample (114 trend points \times 258 simulations = 29,412 data points). The observation based empirical rank exceedance probabilities (using HadSLP2r) are considerably underestimated by the raw CGCM probability curve (as shown in Eade et al. 2022). The pink shading in Fig. 4 represents an uncertainty interval for reddened CGCM exceedance probabilities related to the uncertainty in the observed estimate of the ρ_R and d_R parameters used in the reddening filter. For clarity, only the upper half of the parameter confidence intervals from Section 2.5 are shown (observed estimate to 97.5th percentile) as the observed probabilities would clearly be outside of the range based on lower parameter estimates. Theoretical probability curves for $p(z)$ are shown (Fig. 4, gray curves) assuming a Gaussian distribution with mean zero and variance computed using Eq. (1) with $\sigma_Y^2 = 1$ and the autocorrelation functions for AR(1) and FD processes (as in Eade et al. 2022) with the same ρ and d parameters as used in the reddening filters. The raw CGCM probabilities are close to the Gaussian curve for a white noise process, and the reddened CGCM probabilities are close to the Gaussian curves for the associated AR(1) and FD processes with autocorrelation and difference parameters matching ρ_R and d_R , i.e. the reddened CGCM probabilities are increased relative to the raw CGCM probabilities by the amount expected based on the theory of Eade et al. (2022).

The gray shading in Fig. 4 represents an uncertainty interval for CGCM exceedance probabilities related to the sampling uncertainty for the raw and reddened CGCM NAO simulations. These prediction intervals represent the uncertainty due to having only 1 sample of length 144 years (31-year trend series of length 114), equivalent to the observed NAO series. 95% prediction intervals are constructed by computing the empirical exceedance probabilities for each of the 258 individual CGCM NAO simulations relative to a sequence of NAO trend thresholds, and then computing the 2.5th and 97.5th percentiles for each trend threshold. For moderate trend thresholds, the observed exceedance probabilities are within the 95% prediction intervals for the reddened CGCMs, whereas they are mainly outside of the prediction interval for the raw CGCMs. For more extreme trend thresholds, the upperestimate of reddening ($\rho_R = 0.32$ and $d_R = 0.24$) better captures the observed empirical rank probabilities, especially for the AR(1) reddening method for which the observed probabilities are outside of the uncertainty range when $\rho_R = 0.17$ is used. Both the AR(1) and FD reddening methods lead to a substantial improvement over using the raw CGCM NAO output which considerably underestimates the exceedance probabilities. The FD method is more effective, leading to probabilities that are more consistent with the observed distribution.

3.3 Distribution of NAO extreme multi-decadal trends

The cumulative distribution function for *extreme* trends, i.e. the *maximum* moving window trend in a series, is given by one minus the exceedance probability $q(z)$:

$$q(z) = Pr(\max\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \geq z) \quad (15)$$

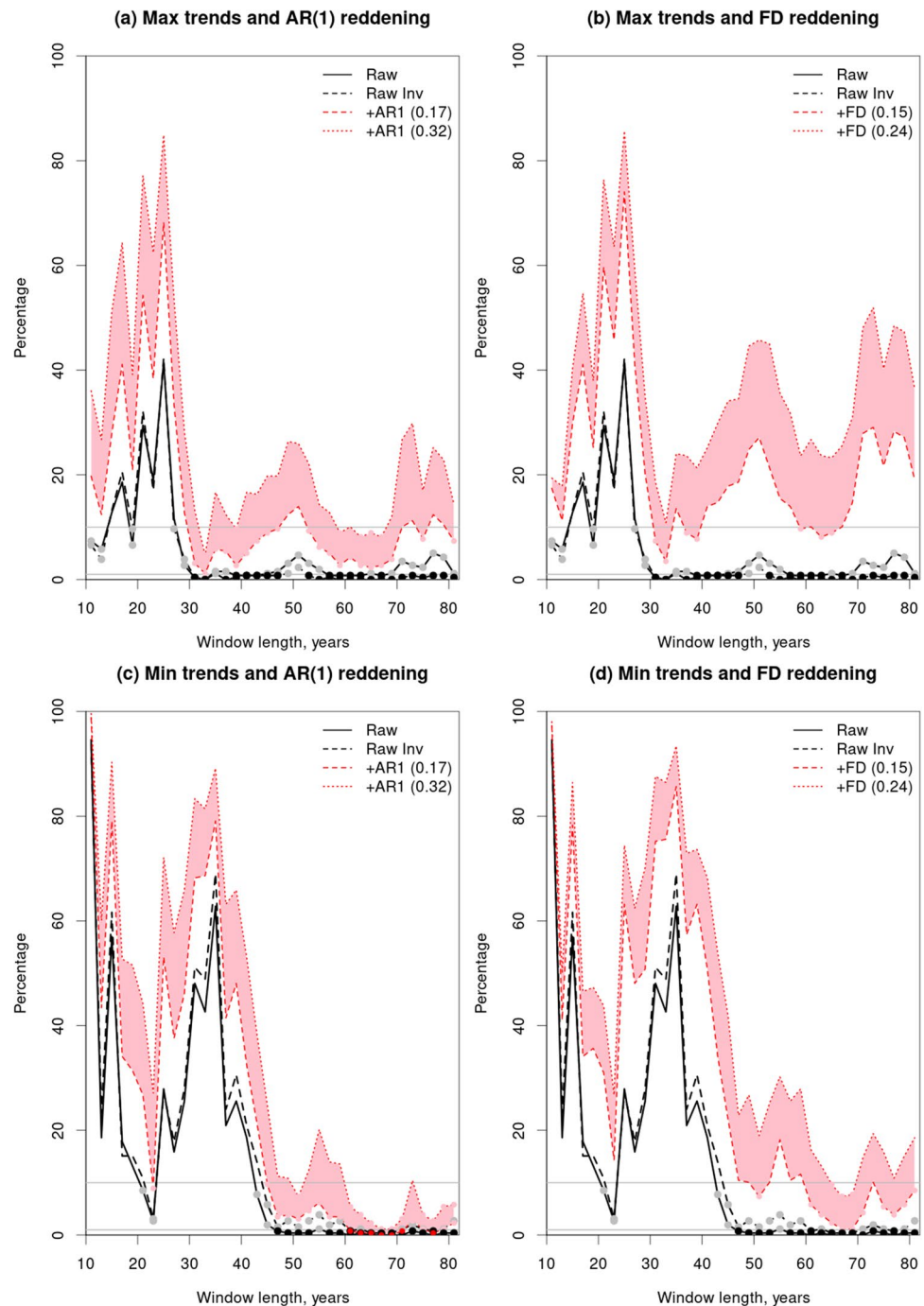
where Z_{1+K} is the moving window trend for a window centred on time step $1+K$, n is the length of the original index series and the window length is $2K+1$. For the probability of minimum trends more negative than those observed, the *negative-exceedance* probability is termed $q'(z)$:

$$q'(z) = Pr(\min\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \leq z). \quad (16)$$

The exceedance probability $q(z)$ is shown in Fig. 5a and b for trends of 11 to 81 years. The raw CGCMs estimate that the observed maximum NAO trends are rare ($q(z) < 10\%$) for windows of length 27 years and beyond, and in many cases these probabilities are less than 1%. The raw CGCMs estimate that the observed minimum NAO trends have negative-exceedance probabilities less than 10% for windows of length 43 years and beyond (Fig. 5c and d), and in most cases these probabilities are also less than 1%. The differences in rarity of extreme maximum and minimum NAO trends with respect to window length does not necessarily mean that CGCMs are more or less proficient at simulating specific types of extreme, as the probabilities are sensitive to the specific exceedance thresholds derived from the small sample of observations being more or less extreme. Some specific cases are chosen for comparison later in this section. If the CGCM moving window trend series are inverted (i.e. multiplied by -1), the exceedance and negative-exceedance probabilities are virtually identical to the original probabilities (black dashed versus black solid curves in Fig. 5) as the CGCM distributions of moving window trends are fairly symmetrical even in the tails, meaning that CGCMs are equally proficient at simulating maximum and minimum trends. The observed extreme NAO trends are consistently found to be rare in CGCMs for window lengths of three or four decades and beyond.

Application of the reddening methods leads to increased exceedance probabilities and negative-exceedance probabilities for all the 11 to 81 year windows tested (Fig. 5, pink shading). This suggests that the CGCMs consistently underestimate the distribution of minimum and maximum NAO trends for multi-decadal window lengths. This CGCM deficiency has also been noted in terms of multi-year mean NAO variability by Kravtsov (2017) who finds that the standard deviation of low-pass filtered winter NAO index series is consistently underestimated by CMIP5 models for filter timescales up to at least 61 years. The lack of autocorrelation

Fig. 5 Likelihood of extreme NAO trends versus window length. CGCM estimates of block maxima exceedance probabilities $q(z)$ (a and b) and block minima negative-exceedance probabilities $q'(z)$ (c and d) for NAO trends on window lengths 11 to 81 years in the historical period 1851–2021 using the empirical rank probabilities. Probabilities are shown for raw CGCM output (black solid line) relative to the observed (Had-SLP2r) maximum or minimum trend thresholds z . Pink shading represents the probabilities for reddened CGCM output using the AR(1) method with ρ_R in the range 0.17 to 0.32 (a and c) and the FD method with d_R in the range 0.15 to 0.24 (b and d). The black dashed line represents the probabilities using the raw CGCM inverted NAO trend series. Probability thresholds 10% and 1% are shown as gray horizontal lines, representing very unlikely and exceptionally unlikely categories of the IPCC likelihood scale (Mastrandrea et al. 2010), and points below these lines are marked as gray and black dots respectively for the raw model output and pink and red for the reddened model output



impacts extreme trend exceedance probabilities across this whole range of window lengths. The FD reddening method leads to a larger increase in probabilities than the AR(1) method, especially for the longer window lengths due to the long-range dependence characteristics of the FD process. Even after reddening, some observed extreme NAO trends are still rare in the CGCMS, specifically the 31-year and 33-year maximum trends and the longer-term minimum trends (61-years and beyond) where the pink shading in Fig. 5 dips below the 10% probability curve (pink dots).

31-year and 71-year windows are chosen for further discussion as comparisons can be made to other recent results.

The *maximum* observed 31-year trend (0.737 sd/decade) centred on 1978 (1963–1993) is clearly unusual as it is 85% larger in magnitude than the previous peak trend in the historical NAO observation record (0.398 sd/decade, centred on 1910). It is also very unusual with respect to the CGCM model range (Fig. 2c), with only 1 member out of 258 having a comparable maximum trend ($q(0.737)=0.4\%$), as found in previous studies (Bracegirdle et al. 2018; Eade et al. 2022;

Bracegirdle 2022). In contrast, the *minimum* 31-year trend is not such an extreme trend (-0.435 sd/decade, window 1981–2011 centred on 1996): It is just 13% larger in absolute magnitude than the previous negative peak trend (-0.385 sd/decade centred on 1935) and it is easily reproduced by the CGCM ensemble ($q'(-0.435) = 48\%$). After applying the FD reddening method with $d_R = 0.24$ (Fig. 2d), the ensemble range for the trend in the specific period 1963–1993 is inflated by 57% (Table 1 row c) but the observed maximum 31-year trend is still only just within the model spread for this period.

The *maximum* 71-year trend (0.186 sd/decade) centred on 1985 (1950–2020) is also unusual, being 78% larger in magnitude than the previous peak trend in the historical NAO record (0.104 sd/decade, centred on 1896). It is outside the CGCM model range for this specific period (Fig. 3a), as found in Blackport and Fyfe (2022), but it is within the range of CGCM max trends throughout the historical period ($q(0.186) = 3.5\%$). For the *minimum* observed 71-year trend (-0.215 sd/decade) centred on 1938 (1903–1973) there are no previous negative peaks in the relatively short observation record to compare this with, but it appears more unusual than the maximum 71-year trend as it is larger in absolute magnitude and is outside the entire CGCM model range. After applying the FD reddening method with $d_R = 0.24$ (Fig. 3b), the 95% prediction interval and range for 71-year trends are inflated to better encompass the observed multi-decadal variability, but the observed minimum 71-year trend is still only just within the model range for that specific period. Blackport and Fyfe (2022) suggest that climate model inadequacies may be due to underestimating the response to human emissions or underestimating the magnitude of internal multi-decadal variability, but the former suggestion is more difficult to reconcile with the underestimation of the observed *minimum* 71-year trend as trends in the effective radiative forcing have been positive throughout the 20th century (Arias et al. 2021).

After reddening, the CGCM estimates of the exceedance probability $q(0.737)$, i.e. the probability of a maximum 31-year trend greater than that observed, are very similar to the stochastic model estimates from Eade et al. (2022) as expected (Table 1 rows d and e), while the raw CGCM estimates are similar to the white noise model estimates. Reddening the CMIP6 CGCMs using the observed parameter estimates ($\rho_R = 0.17$ or $d_R = 0.15$) leads to an average estimate of $q(0.737) = 4.65\%$ (Table 1 row d), i.e. a 1 in 20 chance, while using raw CGCMs underestimates $q(0.737)$ by a factor of ten, as found in Eade et al. (2022). Relative to the CGCM distribution of maximum 31-year trends and minimum 71-year trends, the observed extreme trends are still very rare after reddening the CGCMs (q and q' less than 10%) unless the upper level of reddening is applied, and for the minimum 71-year trend only the FD reddening method

($d_R = 0.24$) is sufficient to achieve a value of q' greater than 10% (Fig. 5).

The reddening methods described in Section 2.4 assume that the raw CGCM NAO index series have a lag-1 year autocorrelation parameter (and difference parameter) of zero. The reddening methods can be extended to account for non-zero autocorrelation or long-range dependence characteristics in the original raw index series (see Appendix 2 for details). Repeating the analysis in this section to use the individual (non-zero) model sample estimates of lag-1 year autocorrelation (ρ_Y) does not substantially change the results but generally leads to slightly lower exceedance probabilities for extreme trends (Table 1 row f). The distribution of CGCM NAO autocorrelation estimates (ρ_Y) is centred around zero but with a fairly large range (Fig. 1a). Under the assumption that $\rho_Y = 0$, the new range of autocorrelation estimates from the reddened simulations has a similar width to the original range of ρ_Y sample estimates as all simulations are reddened by the same magnitude (Fig. 1a). In contrast, if the individual model estimates of ρ_Y are used for the reddening of the individual model NAO simulations, the range of autocorrelation estimates from the reddened simulations will be much narrower than the original range of ρ_Y sample estimates as each simulation is reddened just enough to achieve an autocorrelation of ρ_R . This means that the distribution of maximum trends is likely to be narrower using the ρ_Y sample estimates than when assuming that $\rho_Y = 0$. Despite these subtle differences, the assumption that $\rho_Y = 0$ is deemed valid for analysis of historical variability and future projections as the sensitivity of exceedance probabilities to this assumption is small compared with the sensitivity to the choice of reddening method (AR(1) versus FD) and to the strength of reddening applied (magnitude of ρ_R or d_R).

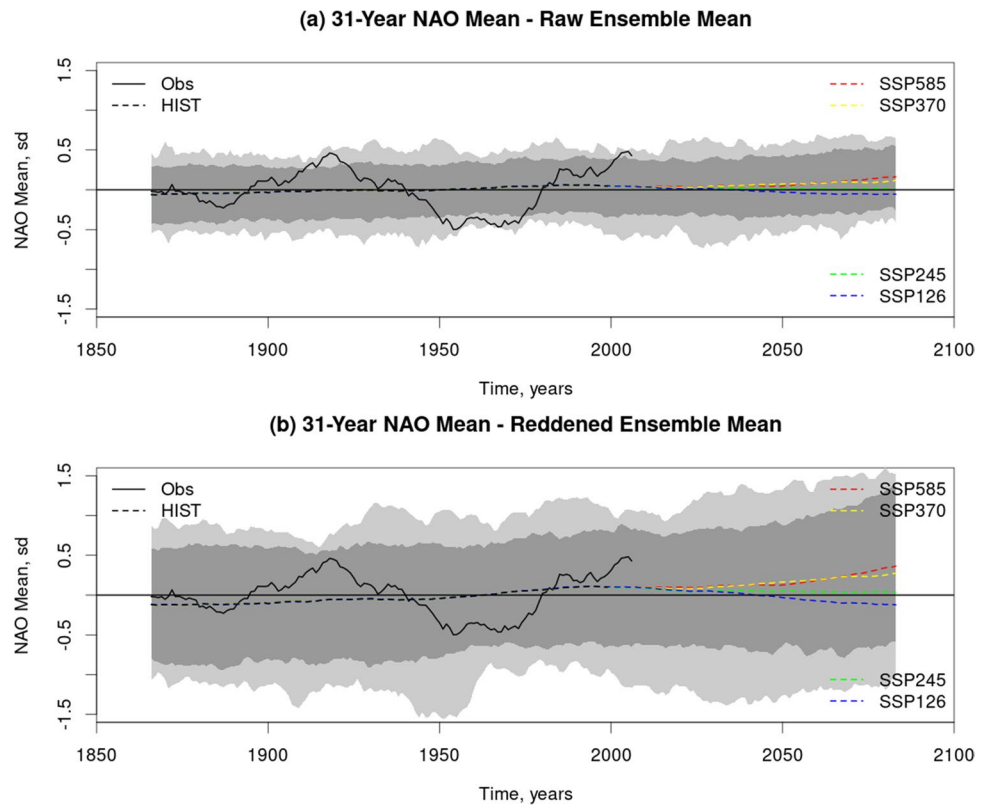
The AR(1) and FD reddening methods are useful recalibration tools to make the CGCM NAO multi-decadal variability consistent with that observed. The FD reddening method with the upper estimate of the difference parameter ($d_R = 0.24$) is needed for the CGCMs to be able to simulate the observed maximum 31-year trend and minimum 71-year trend for the same periods (Figs. 2d and 3b), and this also leads to the best representation of the observed exceedance probabilities for moving window trends (Fig. 4). The FD reddening method with $d_R = 0.24$ is chosen for application in the remainder of this paper.

4 Future distributions of NAO trends

4.1 Future projections of NAO multidecadal variability

Figure 6a shows the multi-model ensemble mean 31-year moving window mean series for the raw CMIP6 CGCM

Fig. 6 Reddening effect on CGCM NAO 31-year mean series. **a** and **b** Dashed curves show CMIP6 multi-model ensemble mean NAO 31-year moving average series using historical (black) and future (colours) experiments for **(a)** raw model output and **(b)** FD reddened output ($d_R = 0.24$). Light and dark gray shading shows the ensemble range and empirical 95% prediction interval respectively (historical and SSP585). The observed 31-year mean series is shown for Had-SLP2r (black solid curve)



historical experiments and four future experiments with different external forcing scenarios (Section 2.1). Despite the seemingly weak model response to external forcings, there are clear differences between the future scenarios with increased forcing scenarios leading to increased long-term trends in the NAO out to the end of this century (Fig. 3a).

SSP585 NAO simulations have a positive trend similar to that in the historical experiments whereas SSP126 simulations have a negative trend back towards early 20th century values (Figs. 3a and 6a). For the final 71-year trend (2028–2098), all scenarios have significantly different trend values at the 5% level when using a standard paired hypothesis test (Wilks 2006), except for SSP245 vs. SSP126. The differences between the future scenarios are clearer after the FD reddening method is applied (Figs. 2d, 3b and 6b). The FD reddening method also inflates the multi-decadal variability of the NAO (as discussed in Section 2.3), seen in Fig. 6b as a widening of the 95% prediction interval and range (of 31-year NAO mean) to better encompass the full range of the observed multi-decadal variability, as for the trend series in Figs. 2 and 3.

The projection of an increase in the NAO index under higher emissions scenarios is also noted in the recent IPCC report (Lee et al. 2021) but they state that it is likely that any such forced response will be smaller than the magnitude of natural internal variability. Under the high-emissions scenario, the projected future increase in winter mean NAO

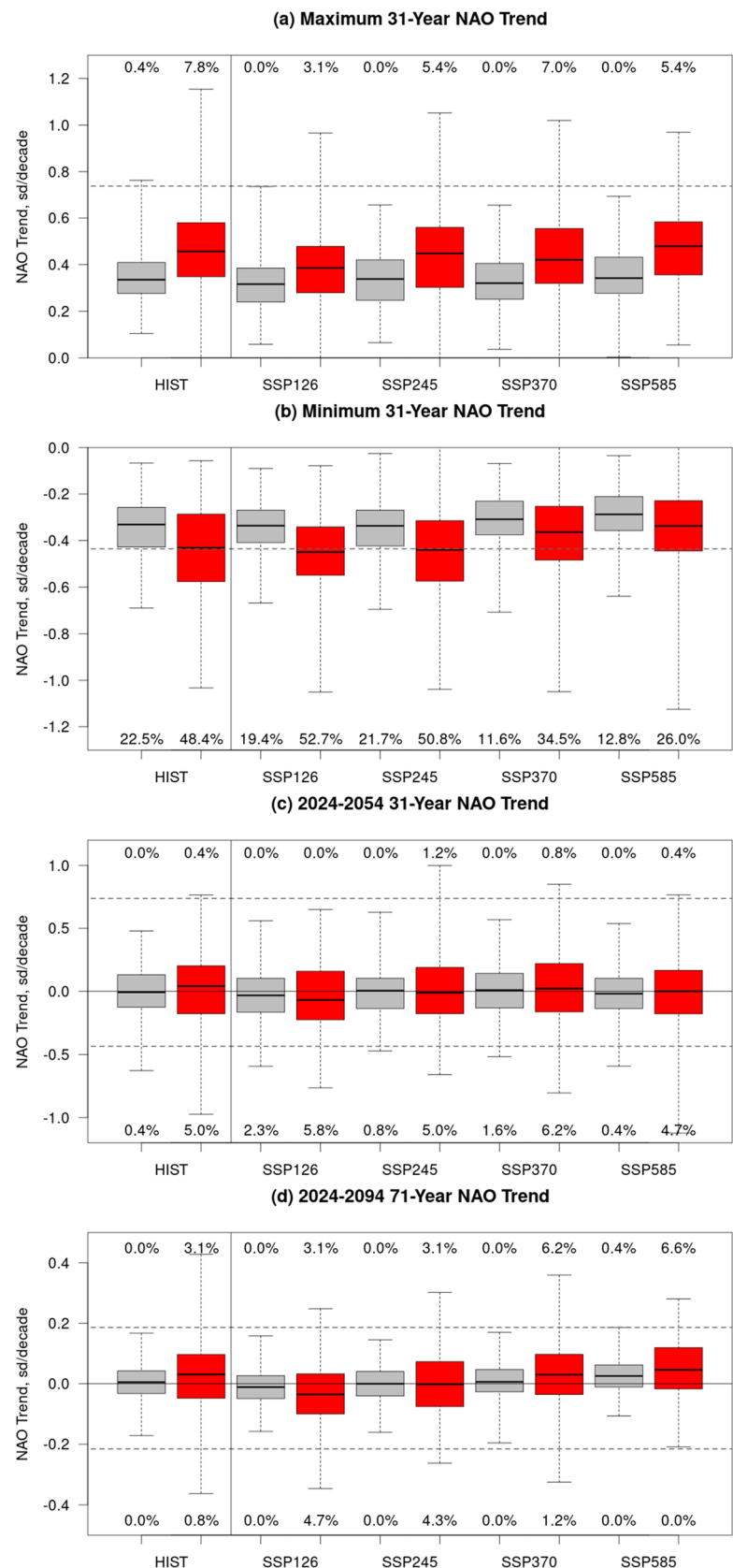
generally leads to an increase in extreme positive NAO winters and a related increase in the severity of precipitation impacts in Europe such as flooding and drought (McKenna and Maycock 2022). These studies use the raw CGCM output, thus they are likely to underestimate future impacts given that the CGCMs underestimate the multi-decadal trends in the winter NAO series.

4.2 Distribution of NAO extreme trends in future climate projections

The CGCM ensemble spread and 95% prediction intervals for trend series are continued into the future using the SSP585 scenario in Figs. 2c and d and 3a and b. As discussed in Section 3, the observed maximum 31-year trend and minimum 71-year trend are clearly unusual in the context of the raw CGCM historical trend series and they continue to be unusual in the context of the future NAO projections unless the FD reddening method is applied (Figs. 2d and 3b).

The distribution of maximum NAO trends is sensitive to the total block length, i.e. the period over which the trend series is computed. The future period is here defined as 2024–2098 (75-years). To make direct comparisons to the distribution of maximum trends in the historical period the most recent 75-year block available from the observed Had-SLP2r dataset is chosen: 1947–2021. Figure 7a shows the

Fig. 7 Effect of reddening on distribution of NAO trends. The distribution of **(a)** *maximum* and **(b)** *minimum* 31-year trends in the 75-year period 2024–2098 for four future SSP experiments, and for the historical (HIST) period 1947–2021 (using SSP585 beyond 2014). **c** The distribution of *near-term* 31-year trends 2024–2054 (SSP), and the historical 31-year trends 1963–1993 (HIST). **d** The distribution of *long-term* 71-year trends 2024–2094 (SSP), and the historical 71-year trends 1950–2020 (HIST). CGCM distributions are shown using raw model output (gray) and after applying the fractional difference reddening method with $d_R = 0.24$ (red), with the median line in the centre of the 25 to 75 percentile box and total range whiskers. The observed maximum and minimum 31-year (71-year in **(d)**) trends are shown (dashed horizontal lines) with the percentage of model members above/below these thresholds displayed above/below



raw CGCM distributions of maximum 31-year NAO trends for the historical period (historical experiments combined with SSP585 beyond 2014) and for the future period (four different future SSP experiments). After FD reddening, the historical and future CGCM distributions of maximum NAO trends are wider and shifted in location towards more positive values. The CGCM future distributions of maximum trends are generally similar to the recent historical distribution, except for the lowest forcing scenario SSP126 which is shifted towards slightly lower magnitude trends. Using the same method for minimum 31-year NAO trends, FD reddening shifts the CGCM distribution of minimum trends to more negative values compared to the raw distributions (Fig. 7b). After reddening, the distributions of minimum 31-year trends for the lower radiative forcing future scenarios (SSP126 and SSP245) are similar to the recent historical distribution and are centred close to the minimum observed 31-year NAO trend (-0.437 sd/decade) with negative exceedance probability $q'(-0.437)$ being close to 50%. For the higher radiative forcing future scenarios (SSP370 and SSP585) the distribution is shifted towards less extreme minimum trends and $q'(-0.437)$ is reduced to around 30%.

In summary, the CMIP6 multi-model ensemble of NAO simulations suggests that the distribution of *maximum* 31-year NAO trends will remain relatively unchanged in the next 75 years compared to the previous period, except for the lower radiative forcing scenario which forecasts a shift towards weaker positive trends. The distribution of *minimum* 31-year NAO trends remains relatively unchanged in the future under lower radiative forcing scenarios, but the higher radiative forcing scenarios lead to a shift of the distribution towards weaker negative trends.

It has been shown that a strong positive NAO trend such as the maximum observed 31-year trend in 1963–1993 (0.737 sd/decade) can account for at least half the winter warming in the Northern Hemisphere extra-tropics, while a similarly strong negative trend from the 1920s to 1970s reduced the winter warming by half (Iles and Hegerl 2017; Scaife et al. 2005). It therefore seems plausible that a minimum negative trend with absolute magnitude equal to that of the maximum observed (-0.737 sd/decade) could potentially offset around half of the winter warming over the three decades. These magnitudes of trends are within the spread of the FD reddened CGCMs (Fig. 7a and b). The associated exceedance probability estimates $q(0.737)$ range from 3 to 7% across the four future scenario experiments, with SSP126 having the lowest probability (Fig. 7a; Table 2). The negative-exceedance probability estimates $q'(-0.737)$ range from 3 to 9%, with SSP126 having the highest probability (Table 2). The total range of possible minimum and maximum 31-year trends across all four future experiments goes from -1.12 to 1.05 sd/decade. In pressure units, this range is -8.86 hPa/decade to 8.60 hPa/decade (using the

Table 2 Influence of the future scenario on the CGCM distribution of 31-year extreme NAO trends

	SSP126	SSP245	SSP370	SSP585
$q(0.737)$	3.10	5.43	6.98	5.43
$q'(-0.737)$	8.53	5.04	2.71	3.10
$q'(-0.435)$	52.7	50.8	34.5	26.0

Reddened CMIP6 CGCM multi-model empirical rank probabilities (shown as percentages) for exceedance probability $q(z) = \Pr(\max \text{ trend} \geq z)$ and negative-exceedance probability $q'(z) = \Pr(\min \text{ trend} \leq z)$ for block maximum and minimum 31-year NAO trends in the period 2024–2098 (z displayed in column 1). CGCM statistics are shown to three significant figures. The CMIP6 NAO index series have first been reddened using the FD filter with difference parameter $d_R = 0.24$

non-standardised version of the NAO index). Without FD reddening, none of the CGCM future simulations have extreme 31-year trends exceeding the thresholds ± 0.737 sd/decade. In general, the likelihood of future extreme maximum (minimum) trends increases (decreases) as the level of radiative forcing increases, though the rarity of the chosen threshold (± 0.737 sd/decade) makes the result less clear than the relationship with the longer-term trends described in Section 4.1 (and later in Section 4.4). The likelihood of these high magnitude future trends is fairly low (less than 10%) but these probabilities would be greatly underestimated if just raw CGCM output was used which could lead to an underestimation of the related impacts discussed in the introduction.

4.3 Distribution of NAO near-term future climate projections

For projections of regional climate change, a key timescale is out to the middle of the 21st century, for example the UK Climate Projections (UKCP, Lowe et al. 2018) issue projections for the UK for 2041–2060 as this is a planning-based timescale that can be more important for their users than longer climate projections out to the end of the century. Figure 7c uses box and whisker plots to show the CMIP6 CGCM distribution of 31-year NAO trends for the specific future period 2024–2054 for four future SSP scenarios. None of the raw future experiments simulate a trend as large as the maximum observed NAO trend. Similar results have been noted by Schurer et al. (2023) who found that none of the CMIP6 near-term projections of 28-year NAO trends (2023–2050) are as large as the maximum observed.

As discussed in Section 3, FD reddening with the upper estimate of $d = 0.24$ is needed to widen the CGCM distribution of 1963–1993 NAO trends sufficiently to encompass the observed trend at this time-step (Fig. 7c), though still only one member out of 258 has a trend exceeding this observed value ($p(0.737) = 0.4\%$). The FD reddened CGCM

distributions of near-term future 31-year NAO trends stay centred about zero, except for the lowest forcing scenario SSP126 which estimates a slightly negative mean trend. The reddened CGCM future NAO trends have a large ensemble range due to internal variability, spanning -1.12 to 1.00 sd/decade across the four scenarios (or -6.62 to 7.71 hPa/decade using the *non-standardised* NAO index), which is 74% wider than the range for the raw CGCMs (-0.59 to 0.63 sd/decade). Even so, the likelihood of a future trend comparable to the maximum or minimum observed is fairly low with $p(0.737)$ less than 2% and $p'(-0.435)$ less than 7%. These future NAO trend distributions are comparable to those from the covariance approach of Eade et al. (2022) which assumes a stationary climate and leads to probabilities $p(0.737)=0.6\%$ and $p'(-0.435)=6.9\%$ using Eq. (13) with $\sigma_Y^2 = 1$ and difference parameter $d = 0.24$ (see Eade et al. 2022 for details). In summary, these near-term projections of NAO trends for 2024–2054 are not overly sensitive to the choice of radiative forcing scenario, showing weak-to-no forced trend, but with a large ensemble range due to internal variability which is greatly underestimated when using raw climate model output.

4.4 Distribution of NAO long-term future climate projections

Figure 7d shows the distribution of future NAO trends out to the end of the century for the specific 71-year period 2024–2094 for the four CMIP6 future SSP scenarios. As discussed in Section 3, FD reddening with the upper estimate of $d=0.24$ is needed to widen the CGCM 1950–2020 distribution of trends sufficiently to encompass the observed trend at this time-step (Fig. 7d). Virtually none of the raw future experiments simulate a trend as large in absolute magnitude as the maximum or minimum observed 71-year NAO trends (just one SSP585 ensemble member exceeds the observed maximum such that $p(0.186)=0.4\%$). After FD reddening the exceedance (p) and negative-exceedance (p') probabilities relative to the observed maximum (0.186 sd/decade) and minimum (-0.215 sd/decade) 71-year trends respectively are increased but are still small (less than 7%).

The differences between the future scenarios are clearer after the FD reddening method is applied, with increased radiative forcing across the scenarios leading to a positive shift in the distribution of long-term trends in the NAO. This impact of radiative forcing scenarios is not apparent for near-term trends (Section 4.3) as the radiative forcing scenarios are initially fairly similar and don't significantly diverge until further into the future (O'Neill et al. 2016). The reddened CGCM exceedance probability $p(0.186)$ increases with the strength of the radiative forcing (while $p'(-0.215)$ decreases) (Fig. 7d) and the ensemble mean trends increase from -0.03 sd/decade to 0.05 sd/decade (-0.24 to 0.35 hPa/

decade using the *non-standardised* NAO index). This range in ensemble mean trends is over double that obtained from the raw CGCMs: -0.01 to 0.02 sd/decade (-0.09 to 0.16 hPa/decade using the *non-standardised* NAO index). For these long-term future projections, use of raw climate model output leads to an underestimation of both the externally forced NAO trend signal and the range of plausible trends due to internal variability (Fig. 7d).

5 Discussion and conclusions

GCM simulations greatly underestimate the plausible range of NAO trends and extreme multi-decadal NAO trends on timescales of three decades and beyond. This study has shown that a simple reddening method can be used to improve the autocorrelation structure and the multi-decadal variability of the winter NAO in CGCM historical simulations and future projections in terms of the externally forced signal and natural internal variability. The method has been tested on the large-scale seasonal mean NAO index, but could equally be applied to indices on smaller temporal and spatial scales. The FD reddening method is more effective than the AR(1) method for achieving distributions of NAO autocorrelation parameters and NAO multi-decadal trends comparable to those observed. These methods could also be applied to shorter term prediction systems such as seasonal-to-decadal forecasts and may help to improve the representation of clusters of positive or negative NAO years.

There has been considerable research showing how extreme multi-decadal trends in the NAO can have serious implications for regional climate change and society relevant impacts in the northern hemisphere (Scaife et al. 2005; Deser et al. 2017; O'Reilly et al. 2021; Smith et al. 2022). The relationship of multi-decadal NAO trends to northern hemisphere regional winter temperature and European precipitation means that the range of plausible multi-decadal trends for these variables are also likely to be greatly underestimated, thus having a concerning impact on the fidelity of near-term climate projections in these regions. For example, the 1963–1993 NAO trend was the maximum 31-year trend in the historical record and in this period the NAO accounted for around half of the winter warming in the Northern Hemisphere extra-tropics (Scaife et al. 2005; Iles and Hegerl 2017; Deser et al. 2017) but without reddening the CGCMs underestimate the likelihood of this NAO trend by at least a factor of ten. O'Reilly et al. (2021) show that constraining the North Atlantic multi-decadal variability in CGCMs to match that observed leads to a considerable increase in the uncertainty in future projections of temperature and precipitation in the Northern extra-tropics. Our reddening method is only applied to the NAO part of this variability,

but it has the advantage that the climate models also retain the externally forced variability enabling a comparison of the different future forcing scenarios.

CMIP6 future projections show a small systematic increase in long-term (2024–2094) NAO ensemble mean trends relative to the magnitude of the radiative forcing scenario going from -0.09 hPa/decade (low radiative forcing) to 0.16 hPa/decade (high radiative forcing). This range more than doubles after reddening: -0.24 to 0.35 hPa/decade. There is also a related shift in the distribution of extreme 31-year NAO trends, which is clearer after reddening: The likelihood of future maximum 31-year trends comparable to 1963–1993 ranges from 3 to 7% across the radiative forcing scenarios after reddening, whereas none of the raw CMIP6 projections simulate such a positive trend. Near-term projections of the next 31 years (2024–2054) are less sensitive than long term trends to the future scenario. After reddening they still show weak-to-no forced trend in the models but have a 74% larger ensemble range (around ± 1 standard deviation per decade). This level of internal variability could increase or decrease regional climate change signals in the Northern Hemisphere by amounts that are greatly underestimated when using raw climate models.

The lack of autocorrelation or persistence of the NAO in CGCMs has been shown to be closely linked to the signal-to-noise paradox (Zhang et al. 2021; Strommen and Palmer 2018; Strommen 2020). This is a phenomenon originally noticed in seasonal forecasts of the winter NAO which exhibit a high correlation between the ensemble mean and observations (a measure of the predictable component in the observations) but a relatively low signal-to-total-variance ratio (a measure of the predictable component in the CGCMs defined as the ratio of the standard deviation of the ensemble mean to the total ensemble standard deviation). This leads to the ratio of predictable components being greater than the expected value of one (Scaife et al. 2014; Eade et al. 2014; Scaife and Smith 2018; Smith et al. 2020; Klavans et al. 2021). The AR(1) and FD reddening methods lead to an increase in the year-to-year persistence and inflate both the strength of the ensemble mean trend series and the total ensemble standard deviation. It would be interesting to see how reddening methods might influence the ratio of predictable components for seasonal-to-decadal hindcasts of the winter NAO. There is still a lack of understanding as to the cause of the signal-to-noise paradox but seeking to explain why the autocorrelation of the NAO index is underestimated in CGCMs may help with resolving this problem.

Appendix 1: Table of CMIP6 Models

Table 3

Appendix 2: Accounting for CGCMs with non-zero autocorrelation function

AR(1) reddening

If the original index series Y is an AR(1) process with non-zero autocorrelation parameter ρ_Y , then Y can be written as $Y = (1 - \rho_Y B)^{-1} \epsilon$. Substituting this into Eq. (3) and expanding the brackets shows that the reddened series is now an AR(2) process:

$$Y^R = \gamma(1 - (\varphi + \rho_Y)B + \varphi\rho_Y B^2)^{-1} \epsilon. \quad (17)$$

Using the Yule-Walker equations for auto-regressive processes (Hamilton 1994), the lag-1 year autocorrelation for this new AR(2) process is given by

$$\rho_R = \frac{\varphi + \rho_Y}{1 + \varphi\rho_Y}. \quad (18)$$

Thus, for given values of ρ_R estimated from the observed index series and ρ_Y estimated from a CGCM index series, φ is found using:

$$\varphi = \frac{\rho_R - \rho_Y}{1 - \rho_R\rho_Y} \quad (19)$$

It can be shown that the variance of the AR(1) process Y and the AR(2) process Y^R are given by

$$\text{Var}(Y) = \frac{\text{Var}(\epsilon)}{1 - \rho_Y^2} \quad (20)$$

$$\text{Var}(Y^R) = \frac{\gamma^2 \text{Var}(\epsilon)}{1 - (\varphi + \rho_Y)^2 \frac{(1 - \varphi\rho_Y)}{(1 + \varphi\rho_Y)} - \varphi^2 \rho_Y^2} \quad (21)$$

(Hamilton 1994). To make these variances equal, the scaling parameter γ is

$$\gamma^2 = \frac{1 - (\varphi + \rho_Y)^2 \frac{(1 - \varphi\rho_Y)}{(1 + \varphi\rho_Y)} - \varphi^2 \rho_Y^2}{1 - \rho_Y^2}. \quad (22)$$

FD reddening

If the original index series Y is a FD process with non-zero difference parameter d_Y , then Y can be written as $Y = (1 - B)^{-d_Y} \epsilon$. Substituting this into Eq. (7) and expanding the brackets shows that the reddened series is also a FD process:

$$Y^R = \gamma(1 - B)^{-(d + d_Y)} \epsilon. \quad (23)$$

Table 3 CMIP6 CGCM historical and future SSP experiments used to model the NAO with columns showing the institution, model names and the number of realisations (ensemble members)

Institution	Model name	Realisations
AS-RCEC	TaiESM1	1
AWI	AWI-CM-1-1-MR	1
BCC	BCC-CSM2-MR	1
CAMS	CAMS-CSM1-0 ripf code f1	2
CAS	CAS-ESM2-0	2
CCCma	CanESM5 ripf code p1; p2; CanESM5-1 p1; p2; CanESM5-CanOE	25; 25; 1; 1; 3
CCCR-IITM	IITM-ESM	1
CMCC	CMCC-CM2-SR5; CMCC-ESM2	1; 1
CNRM-CERFACS	CNRM-CM6-1; CNRM-CM6-1-HR; CNRM-ESM2-1	6; 1; 5
CSIRO	ACCESS-ESM1-5	40
CSIRO-ARCCSS	ACCESS-CM2	5
EC-Earth-Consortium	EC-Earth3; EC-Earth3-Veg	5; 2
INM	INM-CM4-8; INM-CM5-0	1; 1
MIROC	MIROC6; MIROC-ES2H; MIROC-ES2L	50; 1; 10
MOHC	UKESM1-0-LL ripf code f2	5
MPI-M	MPI-ESM1-2-LR	30
MRI	MRI-ESM2-0	5
NASA-GISS	GISS-E2-1-G ripf code p3f1; p5f1; GISS-E2-1-H ripf code p1f2; p3f1; GISS-E2-2-G	4; 4; 5; 1; 5
NCAR	CESM2; CESM2-WACCM	3; 1
NIMS-KMA	KACE-1-0-G	3

Where institutions have submitted realisations for multiple models, the individual model names and ensemble sizes are separated by a semi-colon. Where institutions have submitted realisations from different model configurations, the individual configurations (identified by the “p” and/or “f” parts of the “ripf” codes which identify the precise model setup) and ensemble sizes are separated by a semi-colon. All experiments are from the CMIP6 database <https://esgf-index1.ceda.ac.uk/search/cmip6-ceda> (Eyring et al. 2016)

For given values of d_R estimated from the observed index series and d_Y estimated from a CGCM index series, d is found using

$$d = d_R - d_Y. \quad (24)$$

It can be shown that the variance of the FD processes Y and Y^R are given by:

$$\text{Var}(Y) = \frac{\text{Var}(\epsilon)\Gamma(1 - 2d_Y)}{(\Gamma(1 - d_Y))^2} \quad (25)$$

$$\text{Var}(Y^R) = \frac{\gamma^2 \text{Var}(\epsilon)\Gamma(1 - 2d_R)}{(\Gamma(1 - d_R))^2} \quad (26)$$

where $\Gamma(\cdot)$ is the Gamma function (Hosking 1981). To make these variances equal, the scaling parameter γ is given by:

$$\gamma^2 = \frac{\Gamma(1 - 2d_Y)(\Gamma(1 - d_R))^2}{\Gamma(1 - 2d_R)(\Gamma(1 - d_Y))^2}. \quad (27)$$

Acknowledgements R. Eade, A. Scaife and D. Smith were supported by the Met Office Hadley Centre Climate Programme funded by BEIS

and Defra. This work and its contributor (R. Eade) 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.

Authors' contributions Not applicable.

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

Data availability The climate model datasets analysed in this study are available from the CMIP data archive: <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).

Code availability Not applicable.

Declarations

Competing interests 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 R, Ansell T (2006) A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J Clim* 19(2):1850–2004. <https://doi.org/10.1175/jcli3937.1>
- Ambaum MHP, Hoskins BJ, Stephenson DB (2001) Arctic Oscillation or North Atlantic Oscillation? *J Clim* 14:3495–3507. [https://doi.org/10.1175/1520-0442\(2001\)014<3495:aoonao>2.0.co;2](https://doi.org/10.1175/1520-0442(2001)014<3495:aoonao>2.0.co;2)
- Arias PA, et al. (2021) Technical summary. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (eds). *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp 33–144. <https://doi.org/10.1017/9781009157896.002>
- Bartlett MS (1946) On the theoretical specification and sampling properties of autocorrelated time-series. *Suppl J R Stat Soc* 8:27. <https://doi.org/10.2307/2983611>
- Blackport R, Fyfe JC (2022) Climate models fail to capture strengthening Wintertime North Atlantic jet and impacts on Europe. *Sci Adv* 8. <https://doi.org/10.1126/sciadv.abn3112>
- Bracegirdle TJ (2022) Early-to-late winter 20th Century North Atlantic Multidecadal Atmospheric variability in observations, CMIP5 and CMIP6. *Geophys Res Lett* 49. <https://doi.org/10.1029/2022gl098212>
- Bracegirdle TJ, Lu H, Eade R, Woollings T (2018) Do CMIP5 models reproduce observed low-frequency North Atlantic Jet Variability? *Geophys Res Lett* 45:7204–7212. <https://doi.org/10.1029/2018gl078965>
- Buchan J, Hirschi JJ-M, Blaker AT, Sinha B (2014) North Atlantic SST anomalies and the Cold North European Weather events of Winter 2009/10 and December 2010. *Mon Weather Rev* 142:922–932. <https://doi.org/10.1175/mwr-d-13-00104.1>
- Chen S, Wu R, Chen W, Li K (2020) Why does a colder (warmer) Winter tend to be followed by a warmer (cooler) summer over Northeast Eurasia? *J Clim* 33:7255–7274. <https://doi.org/10.1175/jcli-d-20-0036.1>
- Davini P, Cagnazzo C (2013) On the misinterpretation of the North Atlantic Oscillation in CMIP5 models. *Clim Dyn* 43:1497–1511. <https://doi.org/10.1007/s00382-013-1970-y>
- 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 et al (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys Res Lett* 41:5620–5628. <https://doi.org/10.1002/2014gl061146>
- Eade R, Stephenson DB, Scaife AA, Smith DM (2022) Quantifying the rarity of extreme multi-decadal trends: how unusual was the late twentieth century trend in the North Atlantic. *Oscillation? Clim Dynamics* 58:1555–1568. <https://doi.org/10.1007/s00382-021-05978-4>
- Eyring V, Bony S, Meehl GA et al (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* 13:4430–4440. [https://doi.org/10.1175/1520-0442\(2000\)013<4430:tpsac>2.0.co;2](https://doi.org/10.1175/1520-0442(2000)013<4430:tpsac>2.0.co;2)
- Gastineau G, Frankignoul C (2015) Influence of the North Atlantic SST variability on the atmospheric circulation during the Twentieth Century. *J Clim* 28:1396–1416. <https://doi.org/10.1175/jcli-d-14-00424.1>
- Gray LJ, Scaife AA, Mitchell DM et al (2013) A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *J Geophys Res: Atmos* 118:13:405–413. <https://doi.org/10.1002/2013jd020062>
- Hamilton JD (1994) *Time series analysis*. Princeton University Press, Princeton
- Hardiman SC, Dunstone NJ, Scaife AA et al (2022) Missing eddy feedback may explain weak signal-to-noise ratios in climate predictions. *Npj Clim Atmos Sci* 5:1–8. <https://doi.org/10.1038/s41612-022-00280-4>
- Hosking JRM (1981) Fractional differencing. *Biometrika* 68:165–176. <https://doi.org/10.1093/biomet/68.1.165>
- Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: Regional temperatures and Precipitation. *Science* 269:676–679. <https://doi.org/10.1126/science.269.5224.676>
- Hurrell JW (2001) The North Atlantic Oscillation. *Science* 291:603–605. <https://doi.org/10.1126/science.1058761>
- Hurrell JW, Van Loon H (1997) Decadal variations in climate associated with the north atlantic oscillation. *Clim Change* 36:301–326. <https://doi.org/10.1023/a:1005314315270>
- Iles C, Hegerl G (2017) Role of the North Atlantic Oscillation in decadal temperature trends. *Environ Res Lett* 12:114010. <https://doi.org/10.1088/1748-9326/aa9152>
- Jerez S, Trigo RM, Vicente-Serrano SM et al (2013) The impact of the North Atlantic Oscillation on renewable energy resources in Southwestern Europe. *J Appl Meteorol Climatol* 52:2204–2225. <https://doi.org/10.1175/JAMC-D-12-0257.1>
- Kim M-K, McCarl BA (2005) The Agricultural Value of Information on the North Atlantic Oscillation: yield and economic effects. *Clim Change* 71:117–139. <https://doi.org/10.1007/s10584-005-5928-x>
- Klavans JM, Cane MA, Clement AC, Murphy LN (2021) NAO predictability from external forcing in the late 20th century. *NPJ Clim Atmos Sci* 4. <https://doi.org/10.1038/s41612-021-00177-8>
- 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>
- Lee J-Y, Marotzke J, Bala G, Cao L, Corti S, Dunne J et al (2021) Future Global Climate: scenario-based projections and Near-Term Information. In: Masson-Delmotte V, Zhai P, Pirani A, Connors SL, Péan C, Berger S et al (eds) *Climate Change 2021 – the physical science basis: Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate*

- Change. Cambridge University Press, Cambridge, pp 553–672. <https://doi.org/10.1017/9781009157896.006>
- 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
- Mastrandrea MD, Field CB, Stocker TF et al (2010) Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC)
- McKenna CM, Maycock AC (2022) The role of the North Atlantic Oscillation for projections of Winter Mean Precipitation in Europe. *Geophys Res Lett* 49. <https://doi.org/10.1029/2022gl099083>
- Mosedale TJ, Stephenson DB, Collins M, Mills TC (2006) Granger causality of coupled climate processes: Ocean Feedback on the North Atlantic Oscillation. *J Clim* 19:1182–1194. <https://doi.org/10.1175/jcli3653.1>
- O'Neill BC, Tebaldi C, van Vuuren DP et al (2016) The scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci Model Dev* 9:3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- O'Reilly CH, Befort DJ, Weisheimer A et al (2021) Projections of northern hemisphere extratropical climate underestimate internal variability and associated uncertainty. *Commun Earth Environ* 2. <https://doi.org/10.1038/s43247-021-00268-7>
- Palin EJ, Scaife AA, Wallace E et al (2016) Skillful seasonal forecasts of Winter disruption to the U.K. transport system. *J Appl Meteorol Climatol* 55:325–344. <https://doi.org/10.1175/jamc-d-15-0102.1>
- Scaife AA, Smith D (2018) A signal-to-noise paradox in climate science. *NPJ Clim Atmos Sci* 1. <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. <https://doi.org/10.1029/2005gl023226>
- Scaife AA, Ineson S, Knight JR et al (2013) A mechanism for lagged North Atlantic climate response to solar variability. *Geophys Res Lett* 40:434–439. <https://doi.org/10.1002/grl.50099>
- Scaife AA, Arribas A, Blockley E et al (2014) Skillful long-range prediction of European and north American winters. *Geophys Res Lett* 41:2514–2519. <https://doi.org/10.1002/2014gl059637>
- Scaife AA, Kucharski F, Folland CK, Kinter J, Brönnimann S, Fereday D, Fischer AM, Grainger S, Jin EK, Kang IS, Knight JR, Kusunoki S, Lau NC, Nath MJ, Nakaegawa T, Pegion P, Schubert S, Sporyshev P, Syktus J, Yoon JH, Zeng N, Zhou T (2009) The CLIVAR C20C project: selected twentieth century climate events. *Climate Dynamics* 33(5):603–614. <https://doi.org/10.1007/s00382-008-0451-1>
- Schurer A, Hegerl GC, Goosse H et al (2023) Role of multi-decadal variability of the winter North Atlantic oscillation on northern hemisphere climate. *Environ Res Lett* 18:044046–044046. <https://doi.org/10.1088/1748-9326/acc477>
- 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* 31:8313–8338. <https://doi.org/10.1175/jcli-d-18-0168.1>
- Smith DM, Scaife AA, Eade R et al (2020) North Atlantic climate far more predictable than models imply. *Nature* 583:796–800. <https://doi.org/10.1038/s41586-020-2525-0>
- Smith D, Gillett NP, Simpson IR et al (2022) Attribution of multi-annual to decadal changes in the climate system: the large ensemble single forcing model Intercomparison Project (LES-FMIP). *Front Clim* 4. <https://doi.org/10.3389/fclim.2022.955414>
- Stephenson DB, Pavan V, Bojariu R (2000) Is the North Atlantic Oscillation a random walk? *Int J Climatol* 20:1–18. [https://doi.org/10.1002/\(sici\)1097-0088\(200001\)20:1<1::aid-joc456>3.0.co;2-p](https://doi.org/10.1002/(sici)1097-0088(200001)20:1<1::aid-joc456>3.0.co;2-p)
- Stephenson DB, Pavan V, Collins M et al (2006) North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment. *Clim Dyn* 27:401–420. <https://doi.org/10.1007/s00382-006-0140-x>
- Strommen K (2020) Jet latitude regimes and the predictability of the North Atlantic Oscillation. *Q J R Meteorol Soc* 146:2368–2391. <https://doi.org/10.1002/qj.3796>
- Strommen K, Palmer T (2018) Signal and noise in regime systems: a hypothesis on the predictability of the North Atlantic Oscillation. *Q J R Meteorol Soc* 145:147–163. <https://doi.org/10.1002/qj.3414>
- Thompson DWJ, Barnes EA, Deser C et al (2015) Quantifying the role of internal climate variability in future climate trends. *J Clim* 28:6443–6456. <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* 12:064017. <https://doi.org/10.1088/1748-9326/aa69c6>
- Timlin MS, Alexander MA, Deser C (2002) On the reemergence of North Atlantic SST anomalies. *J Clim* 15:2707–2712. [https://doi.org/10.1175/1520-0442\(2002\)015<2707:OTRONA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2707:OTRONA>2.0.CO;2)
- Uvo CB, Berndtsson R (2002) North Atlantic Oscillation; a climatic Indicator to predict hydropower availability in Scandinavia. *Hydrol Res* 33:415–424. <https://doi.org/10.2166/nh.2002.0016>
- Watanabe M, Tatebe H (2019) Reconciling roles of sulphate aerosol forcing and internal variability in Atlantic multidecadal climate changes. *Clim Dyn* 53:4651–4665. <https://doi.org/10.1007/s00382-019-04811-3>
- Wilks DS (2006) Statistical methods in the atmospheric sciences, second. Elsevier, Amsterdam
- Wu R, Chen S (2020) What leads to persisting surface air temperature anomalies from winter to following spring over mid- to high-latitude Eurasia? *J Clim* 33:5861–5883. <https://doi.org/10.1175/JCLI-D-19-0819.1>
- Wunsch C (1999) The interpretation of short climate records, with comments on the North Atlantic and Southern oscillations. *Bull Am Meteorol Soc* 80:245–255. [https://doi.org/10.1175/1520-0477\(1999\)080<0245:tioscr>2.0.co;2](https://doi.org/10.1175/1520-0477(1999)080<0245:tioscr>2.0.co;2)
- Zanardo S, Nicotina L, Hilberts AGJ, Jewson SP (2019) Modulation of economic losses from European floods by the North Atlantic Oscillation. *Geophys Res Lett* 46:2563–2572. <https://doi.org/10.1029/2019gl081956>
- Zhang W, Kirtman BP, Siqueira L et al (2021) Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled model. *Clim Dyn* 56:2895–2913. <https://doi.org/10.1007/s00382-020-05621-8>

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