University of Exeter

Department of Mathematics and Statistics

# Extreme multi-decadal trends in the North Atlantic Oscillation

Rosemary Eade

September 2023

Supervised by

David B. Stephenson

Adam A. Scaife

Doug M. Smith

Submitted by Rosemary Eade to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, September 2023.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

(signature) R. Eade .....

# Abstract

Stochastic processes are shown to be useful tools for quantifying extreme trends in climate indices. The variance of the trend distribution is shown to generally increase with autocorrelation, with an increase in extreme trend exceedance probabilities. The winter North Atlantic Oscillation (NAO) index has weak autocorrelation which is underestimated in historical climate models and helps to explain the underestimation of extreme trends. The maximum observed 31-year NAO trend occurred in 1963-1993 and is estimated to have a 1 in 20 chance of being exceeded in the 144-year historical record using fitted stochastic models. Climate models and stochastic models without autocorrelation underestimate this probability as a 1 in 200 chance. The NAO trend in the 1963-1993 window was identified due to its unusual nature. If this window is wrongly treated as a randomly chosen single window, the exceedance probability is further reduced (a 1 in 1000 chance). Post-processing methods are proposed to increase the low autocorrelation in climate models and are shown to improve the simulation of extreme trends and also increase the variance of ensemble mean trends. Future projections show a small systematic increase in end-of-century NAO ensemble mean trends relative to the magnitude of the radiative forcing. The probability of a maximum 31-year trend greater than that observed is 3-7% in the next 75-years (under the higher “business as usual” radiative forcing scenario), which is similar to the historical model probability for the last 75-years. Near-term projections of the next 31-years (2024-2054) are relatively insensitive to the scenario, showing no forced trend in the models but a large ensemble range due to internal variability (-7.41 to 7.68 hPa/decade) which could increase or decrease regional climate change signals in the Northern Hemisphere by magnitudes that are underestimated when using raw climate model output.

# Acknowledgements

Thank you to my supervisors, David Stephenson, Adam Scaife and Doug Smith, for their guidance and support and thank you to Mat Collins, James Screen and Mark Baldwin for their help at the start of this PhD.

This work and its contributor (Rosemary Eade) were funded by the UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund.

# Contents

# List of Tables

**Table 5.1. Trend exceedance probabilities for maximum 31-year trends in the period 1862-2005.** Probabilities (shown as percentages) of trend exceedance $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ in the historical period 1862-2005, where $z$ is the maximum observed NAO trend (0.737 sd/decade from Table 4.3). Stochastic model probabilities are included from Table 4.5 where stochastic processes were fitted to the HadSLP2r NAO index. For the AR(1) and FD model fits, average $q$ = 4.76% ~ 1/20 (intervals in brackets represent the uncertainty from the stochastic model fits, see Section 4.5.1).

**Table 5.2. Sensitivity of maximum trend exceedance probabilities to NAO definitions.** CGCM empirical probabilities (shown as percentages) for $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ in the historical period 1862-2005, where $z$ is the maximum observed 31-year NAO trend using the NAO index definition displayed in column 1 and the observed NAO trend series from (a) HadSLP2r and (b) C20C. Probabilities for definitions (4) and (5) are not included for (a) as HadSLP2r underestimates MSLP variability.

**Table 6.1. Effect of the reddening methods on the CGCM ensemble variability of moving window trends.** The standard deviation of 31-year moving window NAO trend series from CMIP5+6 simulations for the total CGCM ensemble and for the CGCM ensemble mean over the historical period 1862-2005 (in units of sd/decade), and also the ensemble range for the linear trend on the specific time window centred on 1978 ($Z_{1978}$, 1963-1993). CGCM standard deviations 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 $\rho_R$ and $d_R$ in the first column. The percentage increase in standard deviation and range values for the reddened CGCMs relative to the raw CGCM values are shown in brackets. The standard deviation of the observed trend series is shown using HadSLP2r.

**Table 6.2. Trend exceedance probabilities for maximum 31-year trends in the historical period.** Probabilities (shown as percentages) of trend exceedance $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ in the historical period 1862-2005, where $z$ is the maximum observed trend (0.737 sd/decade). CGCM probabilities are shown using the raw NAO index series and for the reddened

NAO index series where the original CGCM lag-1 year autocorrelation is either assumed to be zero ("$\rho_Y$ = 0") or implied by the sample estimate for each model separately ("$\rho_Y$ <= model"). Probabilities for both AR(1) and FD reddening methods are shown, with autocorrelation and difference parameters as indicated by $\rho_R$ and $d_R$ in the first column. Probabilities for matching stochastic processes are also shown ("Stochastic Process") and are comparable to the values in Table 4.5.

**Table 7.1 Long-term trends in future CMIP6 NAO trend series.** The CMIP6 ensemble mean and median of time mean trend $\bar{Z}$ for the future block 2024-2098 for the SSP126 and SSP585 experiments using raw CMIP6 output (column 2 and 3), with * indicating the ensemble mean is significantly different to zero (see Section 7.2.3). The median for FD reddened CMIP6 output using $d_R$ = 0.15 and $d_R$ = 0.24 (columns 4 and 5). $\bar{Z}$ has units of sd/decade. The percentage of ensemble members with $\bar{Z}$ greater than zero is shown in brackets. Values are also shown for the historical experiments for the block 1947-2021, using SSP585 projections for years beyond 2014.

**Table 7.2 CMIP6 Extreme NAO trend exceedance probabilities.** Exceedance probabilities (as percentages) for CMIP6 maximum and minimum 31-year NAO trends in the 75-year periods 1947-2021 (historical) and 2024-2098 (future SSP), relative to +/-0.737 sd/decade respectively (the magnitude of the maximum observed 31-year NAO trend). Probabilities are shown using the FD reddened model output with difference parameter $d_R$ = 0.15 or 0.24 as specified in brackets.

**Table 7.3. Near-term NAO trend projections.** The ensemble mean, 95% prediction interval (P.I.) and range of near-term projections for the 2024-2054 NAO trend in CMIP6 SSP585 future experiments for raw model output and for FD reddened output with $d_R$ = 0.24 (in units of sd/decade). The percentage increase (or decrease) in the mean and the length of the P.I. and range are shown for the reddened CGCMs relative to the raw CGCM values.

**Table A.1**: CMIP6 CGCM historical experiments used to model the NAO with columns showing the institution, model names and the number of realisations

(ensemble members). 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).

**Table A.2**: CMIP5 CGCM historical experiments used to model the NAO with columns showing the institution, model name and the number of realisations (ensemble members) as in Table A.1. All experiments are from the CMIP5 database https://esgf-node.llnl.gov/search/cmip5 (Taylor et al., 2012).

**Table A.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) as in Table A.1. This is a subset of the models from Table A.1 that are available for the historical experiments and all four SSP scenario experiments assessed in Chapter 7 (SSP126, SSP245, SSP370, SSP585). All experiments are from the CMIP6 database https://esgf-index1.ceda.ac.uk/search/cmip6-ceda (Eyring et al., 2016).

# List of Figures

**Figure 2.7:** Lag-1 year autocorrelation estimated within moving 60-year windows for the winter mean NAO index (December to March). Source: Hurrell and van Loon (1997) Figure 7.

**Figure 2.8:** Aggregated variance plot for multi-year means of observed winter NAO vs a white noise model. Source: Stephenson et al. (2000) Figure 2.

**Figure 2.9:** Aggregated variance plots of winter (December to February) NAO multi-year mean and linear trend. Observed NAO index (black dashed line) compared to two different GCMs (red low resolution and blue medium resolution), and the empirical white noise values in (a) (dotted grey diagonal line). Source: Menary et al. (2018) Figure 4.

**Figure 3.1. The relationship of the variance of moving window trends with autocorrelation.** The trend variance is shown for the 3 approaches discussed in Section 3.3 with trend window length 31 (K = 15): The covariance approach for AR(1) (black dashed line) and FD (black dotted line) processes; the simulation approach using a selection of AR(1) (diamond) and FD (circle) simulations of length 130 years; the pre-whitening method from Cochrane & Orcutt (1949) (grey solid line). Trends are calculated for a standardised index, so the variance of trends is in units $year^{-2}$, which can be thought of as $(sd/year)^2$ where sd refers to the standard deviation of the index pre-standardisation. The simulation approach for extended FD simulations with trend series length $10^4$ ("+") and $10^5$ ("*") are included for the case $\rho = 0.9$.

**Figure 3.2. The relationship of the variance of moving window trends with window length.** The trend variance is shown for the covariance approach from Section 3.3.1 for AR(1) (black dashed line) and FD (black dotted line) processes with lag-1 autocorrelation of 0.2 and for a white noise process (grey solid line).

**Figure 3.3. The relationship of the variance of moving window trends with autocorrelation for a longer window length.** The trend variance is shown for the covariance approach from Section 3.3.1 for AR(1) (black dashed line) and FD (black dotted line) processes for a trend window of length 81 ($K = 40$).

**Figure 3.4. The uncertainty in parameter estimates for the distribution of moving window trends.** The trend variance is shown for the covariance approach from Section 3.3.1 for AR(1) (black solid line) processes for a trend window of length 31 ($K$ = 15) with methods to show the uncertainty in parameter estimates due to the finite length of trend series (Section 3.3.4). The uncertainty due to the sample trend variance estimate (black dotted line) and the uncertainty due to the autocorrelation parameter $\rho$ estimate using simulations (grey dashed line) and the Bartlett formula (black dashed line).

**Figure 3.5. Autocorrelation in moving window trend series.** The average lag-1 autocorrelation for moving window trends is shown for the AR(1) and FD simulations (window length 31), plotted against the lag-1 autocorrelation of the underlying index time series.

**Figure 3.6. Example exceedance probabilities for multi-decadal trends.** The exceedance probability $p = \Pr(Z_i \geq z)$ is shown for relatively extreme trend thresholds $z$ for a white noise process (white solid line) with grey shading representing the uncertainty due to stochastic model fitting (Section 3.3.5). Probability curves are also shown for an AR(1) (black solid line) and FD (grey solid line) index time series with lag-1 autocorrelation of $\rho = 0.2$.

**Figure 3.7. The relationship of maximum trend distribution parameters with the autocorrelation of the original index for independent trend values.** (a) Scale and (b) location parameters of the Gumbel distribution of extreme trends, calculated from Equation 3.16 and 3.18 for an independent trend series, vs the lag-1 autocorrelation parameter of the original stochastic process. The window length is 31 and the trend series length is 100. Black dashed lines are for AR(1) processes with the corresponding autocorrelation parameter. Grey dashed lines are for FD processes with difference parameters matched to the displayed autocorrelation values.

**Figure 3.8. The relationship of maximum trend distribution parameters with the autocorrelation of the original AR(1) process.** (a) Shape, (b) scale and (c) location parameters of the GEV distributions fitted to simulated maximum trends vs the lag-1 autocorrelation of the original AR(1) time series

(black solid line) for window length 31 and trend series length $m$ = 100 (Section 3.4.2). Scale and location parameters for a fitted Gumbel distribution are shown (grey solid line), and the location parameter for a fitted Gumbel distribution using the scale parameter for the independent case (grey dotted line). Black dashed lines are Gumbel parameters for the fully independent case (Section 3.4.1).

**Figure 3.9. The relationship of maximum trend distribution parameters with the autocorrelation of the original FD process.** (a) Shape, (b) scale and (c) location parameters of the GEV distributions fitted to simulated maximum trends vs the lag-1 autocorrelation of the original FD time series (black solid line) for window length 31 and trend series length $m$ = 100 (Section 3.4.2). Scale and location parameters for a fitted Gumbel distribution are shown (grey solid line), and the location parameter for a fitted Gumbel distribution using the scale parameter for the independent case (grey dotted line). Black dashed lines are Gumbel parameters for the fully independent case (Section 3.4.1).

**Figure 3.10. Example exceedance probabilities for extreme trends.** The exceedance probability $q$ is shown relative to the trend threshold $z$ for GEV distribution fits to stochastic simulations (window length 31): white noise (white solid line in grey shading), AR(1) ($\rho$ = 0.2, black solid line) and FD ($d$ = 0.167, black dashed line) processes. For each process, the empirical probabilities are shown as dark-grey dot-dash lines close to the GEV distribution fit. The 95% prediction interval (grey shading) represents the uncertainty due to finite sample sizes of (a) 100 and (b) 1000 simulations (Section 3.4.4).

**Figure 4.1. First EOF pattern for MSLP DJF seasonal mean.** The full available period of (a) HadSLP2r and (b) C20C datasets are used, with the standardised NAO index nodes marked as black crosses (Iceland and Azores).

**Figure 4.2. First EOF pattern for MSLP DJF seasonal and daily mean data.** The full period of the EMSLP dataset is used to calculate EOFs on (a) seasonal and (b) daily mean data, with the standardised NAO index nodes marked as black crosses (Iceland and Azores).

**Figure 4.3. Interannual variability of the NAO and its trend in observations.** (a) The standardised DJF mean NAO index for HadSLP2r (black solid line), C20C (black dashed line) and L99 (black dotted line). (b) The 31-winter moving window linear trend estimate is shown for the same datasets with the maximum trend shown as a horizontal black line centred on 1978.

**Figure 4.4. Autocorrelation function for the DJF NAO index.** The lagged autocorrelations are shown for lag 0 to 10 years for winter mean NAO using HadSLP2r. The 95% prediction interval for the autocorrelation assuming a white noise process is shown by black dashed lines, with the upper value marginally lower than the NAO lag-1 year autocorrelation.

**Figure 4.5. Lag-1 year autocorrelation patterns for MSLP DJF seasonal mean.** The full available period of (a) HadSLP2r and (b) C20C datasets are used, with the standardised NAO index nodes marked as black crosses (Iceland and Azores).

**Figure 4.6 Return Plot for NAO Trend with empirical probability-based uncertainty estimates.** Exceedance probabilities (top axis) and return period (bottom axis) from fitted (a) AR(1) and (b) FD processes (thick black dashed curves) with shaded 95% prediction interval based on empirical probabilities from trend series simulations (Section 4.4.3). Empirical probabilities for HadSLP2r standardised NAO index shown as open circles (Section 4.4.1). Probability curves are included for the AR(1) and FD processes with lag-1 autocorrelation $\rho$ in the set {0.0, 0.1, 0.2, 0.3, 0.4} (thin black dashed curves with labelled $\rho$ values).

**Figure 4.7 Return Plot for NAO Trend with variance based uncertainty estimates.** Exceedance probabilities (top axis) and return period (bottom axis) from fitted (a) AR(1) and (b) FD processes (thick black dashed curves) with shaded 95% prediction interval based on variance of trend series simulations (Section 4.4.3). Empirical probabilities for HadSLP2r standardised NAO index shown as open circles (Section 4.4.1). Probability curves are included for the AR(1) and FD processes with lag-1 autocorrelation $\rho$ in the set {0.0, 0.1, 0.2, 0.3, 0.4} (thin black dashed curves with labelled $\rho$ values).

**Figure 5.1. First EOF for MSLP DJF seasonal mean.** HadSLP2r observations (a) and the multi-model mean of the EOF patterns from CGCMs (b) using the full period available for each dataset, with the Iceland and Azores NAO node locations marked as black crosses.

**Figure 5.2. Multi-decadal NAO trends in CGCMs.** The 31-winter moving linear trend estimate is shown for the set of CMIP5 and CMIP6 CGCM simulations of the NAO (grey solid lines) and for the multi-model ensemble mean (grey dot-dash line). The HadSLP2r (black solid line) and C20C (black dash line) observed trends are overlaid, with the maximum trend shown as a horizontal black line, centred on 1978.

**Figure 5.3 Empirical distribution of block maxima NAO trends in CGCMs.** The probability density function (kernel density estimate, see Appendix B.6 for details) for the CGCM maximum 31-year NAO trend values with empirical 95% prediction interval (thick black horizontal lines). The observed HadSLP2r maximum trend is shown as a dashed vertical line.

**Figure 5.4. Return period plot for NAO moving window trend.** Empirical exceedance probabilities (top axis) and return period (bottom axis) from CGCM simulations (grey crosses) and observations (black open circles) of 31-year moving window NAO trends. Black dot-dash curve shows the Gaussian distribution fitted to the CGCM trends with the 95% prediction interval for a single time-series shaded in grey (Section 5.3.2). Black dashed curves show the exceedance probabilities using the covariance approach (Section 4.4.2) for the AR(1) ($\rho$ = 0.169) and white noise ($\rho$ = 0.0) processes fitted to the observed NAO index time series.

**Figure 5.5 Empirical distribution of Moving Window NAO trends in CGCMs.** The probability density function (kernel density estimate, see Appendix B.6 for details) for the CGCM 31-year moving window NAO trend values (black curve) with empirical 95% prediction interval (black horizontal line). The probability density function of the Gaussian fit is overlaid (grey curve). The observed HadSLP2r maximum trend is shown as a dashed vertical line.

**Figure 5.6. The relationship of maximum NAO trends to NAO autocorrelation in CGCMs.** The maximum NAO trend for each CGCM simulations is shown relative to the model estimate of the lag-1 year autocorrelation $\rho$ (grey "x" with 95% prediction ellipse) with the correlation (r) and p-value (p) displayed in the bottom right. Observed values are shown for HadSLP2r (black circle, and horizontal line) and C20C (black square).

**Figure 5.7 Empirical distribution of lag-1 year autocorrelation for NAO index.** Box plots of the empirical distribution of lag-1 year autocorrelation ($\rho$) estimates from the 538 individual CGCM simulations and the expected distribution of sample $\rho$ estimates from a white noise model using the Bartlett formula (Bartlett, 1946) (Section 5.3.3; Equation 3.13). The box plots show the median line in the centre of the 25- to 75- percentile box with thick black whiskers showing the 95% prediction interval (2.5- to 97.5- percentiles). For the CGCM distribution thin black total range whiskers are included. Observed values are shown for HadSLP2r (black circle) and C20C (black square).

**Figure 5.8: Empirical distribution of time mean trends for CGCM historical NAO.** The probability density function (kernel density, see Appendix B.6) for the CGCM time means of 31-year moving window NAO trend series for the period 1862-2005 (black curve). The probability density function for the Gaussian white noise model overlaid (grey curve). Empirical 95% prediction intervals are shown by black (CGCM) and grey (white noise) horizontal lines. The observed time mean trend for 1862-2005 (HadSLP2r) is shown (dashed black vertical line).

**Figure 5.9 Empirical distribution of block maxima NAO trends and pseudo maximum NAO trends in CGCMs.** The probability density function (kernel density, see Appendix B.6) for the CGCM pseudo maximum NAO trend values calculated from the maximum and minimum trends at the Azores and Iceland NAO nodes respectively regardless of timing (grey curve) as described in Section 5.3.5 with the empirical 95% prediction interval (thick grey horizontal line). The distribution for the original CGCM block maxima trends is shown (black solid line) as in Figure 5.3, with the observed (HadSLP2r) maximum trend shown by the dashed vertical line.

**Figure 5.10. Empirical distributions of MSLP extreme trends at NAO nodes in GCMs.** The probability density function (kernel density, see Appendix B.6) for (a) the maximum MSLP trend at the Azores NAO node and (b) the minimum MSLP trend at the Iceland NAO node using the CGCM (black curve) simulations for MSLP over the period 1862-2005. Horizontal black lines show the empirical 95% prediction intervals for the CGCM simulations. Observed (HadSLP2r) estimates are shown as black dashed vertical lines.

**Figure 5.11 Empirical distribution of NAO interannual variability in CGCMs.** Box plots showing the empirical distribution of standard deviation (sd) estimates for the period 1862-2005 from the 538 individual CGCM simulations for (4) “Anomalised point index” and (5) “Anomalised regional average index” (labelled as in Table 5.2). The box plots show the median line in the centre of the 25 to 75 percentile box with thick black whiskers showing the 95% prediction interval (2.5 to 97.5 percentiles) and thin black total range whiskers. Observed standard deviation estimates are shown for the same period using HadSLP2r (black circle) and C20C (black square).

**Figure 5.12 Likelihood of maximum trends versus window length.** Trend exceedance probabilities for block maxima NAO trends for a range of window lengths. Probabilities of maximum 31-year NAO trend exceedance $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ for window length 11 to 81 years ($K$ in range 5 to 40) in the historical period 1862-2005 using the empirical probabilities from CGCM experiments (black solid line) relative to the observed (HadSLP2r) maximum trend thresholds $z$. Probability thresholds 10% and 1% are shown as labelled grey horizontal lines, representing *very unlikely* and *exceptionally unlikely* categories of the IPCC likelihood scale.

**Figure 5.13 Likelihood of minimum trends versus window length.** Trend exceedance probabilities for block minima NAO trends for a range of window lengths. Probabilities of minimum 31-year NAO trend exceedance $q' = \Pr(\min\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \leq z)$ for window length 11 to 81 years ($K$ in range 5 to 40) in the historical period 1862-2005 using the empirical probabilities from CGCM experiments (black solid line) as relative to the observed (HadSLP2r)

minimum trend thresholds *z*. Probability thresholds 10% and 1% are shown as labelled grey horizontal lines, representing *very unlikely* and *exceptionally unlikely* categories of the IPCC likelihood scale.

**Figure 5.14 Distribution of block maxima NAO trends for winter months.** The probability kernel density functions for the CGCM maximum 31-year NAO trend values with empirical 95% prediction interval (thick black horizontal lines) for the winter season DJF, extended winter season DJFM and individual months December to March. Observed (HadSLP2r) maximum trends are shown as dashed vertical lines and the percentage of CGCM values above these thresholds displayed as *q*.

**Figure 6.1: Reddening effect on white noise time series.** (a) A length 144 time series from a stochastic white noise process with variance one (black solid line). Dotted lines show the reddened version of this time series using the AR(1) (red) and FD (blue) method with lag-1 autocorrelation parameter $\rho_R$ = 0.3 and fractional difference parameter $d_R$ = 0.23. (b) Moving window trend series for the time series in (a).

**Figure 6.2: 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 (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.

**Figure 6.3: Reddening effect on a single CGCM NAO simulation.** (a) Raw NAO Index time series (black) from a single CGCM simulation (CMIP6 HadGEM3-GC3.1-MM). Dotted lines show the reddened version of this time series using the AR(1) (red) and FD (blue) method with lag-1 year autocorrelation parameter $\rho_R$ = 0.3 and fractional difference parameter $d_R$ =

0.23. (b) The 31-year moving window trend series of the CGCM NAO index series in (a), relative to the central year of the window.

**Figure 6.4 Reddening effect on CGCM ensemble distribution of NAO trend series.** Ensemble mean NAO 31-year trend series for CMIP5+6 simulations using (a) raw model output, or FD reddened output with difference parameter (b) $d_R$ = 0.15 and (c) $d_R$ = 0.24 (black dashed curves), relative to the central year of the window. The light and dark grey shading shows the ensemble range and empirical 95% prediction interval respectively. The observed 31-year trend series is shown for HadSLP2r (black solid curve) with the maximum value identified (horizontal thin dotted black line).

**Figure 6.5: Return Period Plots for NAO moving window trends for reddened CGCMs.** Empirical exceedance probabilities (top axis) and return period (bottom axis) for CGCM 31-year moving window NAO trend series after applying (a) the AR(1) reddening method ($\rho_R$ = 0.17 and $\rho_R$ = 0.32) and (b) the FD reddening method ($d_R$ = 0.15 and $d_R$ = 0.24) shown as red curves identified in the key by solid and dashed lines respectively. Probabilities are also shown for the raw CGCM simulations (black solid curve) and observations (black open circles, using HadSLP2r). The maximum observed trend is identified by the horizontal black line.

**Figure 6.6: Return Period Plots for NAO maximum trends for reddened CGCMs.** Empirical exceedance probabilities (top axis) and return period (bottom axis) for the maximum 31-year NAO trend in the 144 year historical record (1862-2005) computed from the reddened CGCM trend series after applying (a) the AR(1) reddening method ($\rho_R$ = 0.17 and $\rho_R$ = 0.32) and (b) the FD reddening method ($d_R$ = 0.15 and $d_R$ = 0.24), shown as red curves identified in the key by solid and dashed lines respectively. Probabilities are also shown for the raw CGCM simulations (black solid curve) and the maximum observed trend is identified by the horizontal black line.

**Figure 7.1: CMIP6 Effective Radiative Forcing.** Time series showing the total anthropogenic part of the effective radiative forcing ($Wm^{-2}$) for historical CMIP6 experiments (black solid curve) and for future scenarios SSP126 (blue),

SSP245 (green), SSP370 (orange) and SSP585 (red). The natural part of the effective radiative forcing is shown separately for the historical (black dashed curve) and future (black dotted curve) experiments. This figure has been adapted from the IPCC AR6 Technical Summary to just show the scenarios discussed in this chapter and to have more contrasting colours that match later figures. Source: Arias et al. (2012) Figure TS.4.

**Figure 7.2: Interannual variability of the NAO.** Time series of observed NAO winter mean index using HadSLP2r (grey line, relative to *left axis*) in units of standard deviation (sd). CMIP6 multi-model ensemble mean series (relative to *right axis with a smaller range*) for historical (black line) combined with four different future SSP scenarios (coloured lines) from 2015 onwards.

**Figure 7.3: Multi-decadal variability of the NAO.** Time series of 31-year moving window average of HadSLP2r observed winter NAO index (black solid line). CMIP6 multi-model ensemble mean series for historical (black dashed line) combined with four different future SSP scenarios (coloured dashed lines). The multi-model range and 95% prediction interval are shown by light and dark grey shading for the historical experiments and future scenario SSP585. The vertical dashed line marks where the CGCM values begin to diverge due to inclusion of future projections and the vertical dotted line marks from where the values purely use future projections.

**Figure 7.4: Multi-decadal trends in the NAO.** Time series of observed 31-year moving window linear trends of winter NAO index using HadSLP2r (black solid line). CMIP6 multi-model ensemble mean series for historical (black dashed line) combined with four different future SSP scenarios (coloured dashed lines). The multi-model range and 95% prediction interval are shown by light and dark grey shading for the historical experiments and future scenario SSP585. The vertical dashed line marks where the CGCM values begin to diverge due to inclusion of future projections and the vertical dotted line marks from where the values purely use future projections.

**Figure 7.5 Distribution of time-mean 31-year trends in future NAO series.** The probability density function (kernel density, see Appendix B.6) of time

means of 31-year trends ($\bar{Z}$) computed for each of the 165 ensemble members over the 75-year period 2024-2098 is shown for the four different CMIP6 SSP future experiments (coloured curves identified in key) using (a) raw model output and (b) FD reddened model output with $d_R$ = 0.24.

**Figure 7.6: Longer-term trends in the NAO.** Time series of HadSLP2r 71-year moving window winter NAO trends (black solid line). CMIP6 ensemble mean series for historical (black dashed line) combined with four future SSP scenarios (coloured lines), with shading for the multi-model range (light grey) and 95% prediction interval (dark grey) (using SSP585 beyond 2014). The vertical dashed line marks where the CGCM values begin to diverge due to inclusion of future projections and the vertical dotted line marks from where the values purely use future projections. (a) uses raw CMIP6 output, (b) and (c) use FD reddened CMIP6 output with $d_R$ = 0.17 and 0.24 respectively.

**Figure 7.7 Historical and future distributions of maximum 31-year NAO trends.** The probability density function (kernel density, see Appendix B.6) of maximum 31-year NAO trends in the 75-year period 2024-2098 for the four different CMIP6 SSP future experiments using raw model output (coloured curves). The distribution of historical CMIP6 maximum trends is shown for the 75-year period 1947-2021 (black curve), using SSP585 beyond 2014 and the maximum observed (HadSLP2r) trend for this period is shown in grey (vertical line).

**Figure 7.8 Effect of reddening on distribution of future maximum NAO trends.** Box and whisker plots of the distribution of maximum 31-year NAO trends in the 75-year period 2024-2098 for the SSP126 and SSP585 future experiments using raw model output (grey box plots) and after applying the fractional difference ("+FD") reddening method with $d_R$ = 0.24 (red box plots), displaying the median line in the centre of the 25 to 75 percentile box and total range whiskers. Box and whisker plots for the historical (HIST) CMIP6 maximum trends are shown for the 75-year period 1947-2021, using SSP585 beyond 2014. The observed maximum 31-year trend is shown as the dashed horizontal line, and the percentage of model members exceeding this threshold is displayed above each box plot.

**Figure 7.9 Near-term NAO trends.** Box and whisker plots for CMIP6 historical ("HIST") 1963-1993 NAO trends and for CMIP6 future 2024-2054 NAO trends using scenarios SSP126 and SSP585. Grey box plots use the raw CMIP6 model output (median line in centre of 25 to 75 percentile box; total range whiskers). Red box plots use CMIP6 model output after applying the fractional difference reddening method ("+FD") with $d_R$ = 0.24. The observed 1963-1993 NAO trend is shown as the dashed horizontal line.

**Figure 7.10 Anomaly index near-term future NAO trends.** As Figure 7.9 but using the *anomaly* based NAO index (without standardisation) so that units are in hPa/decade, and the observed maximum trend 1963-1993 is based on the C20C MSLP data (5.59 hPa/decade). Box and whisker plots for CMIP6 historical ("HIST") 1963-1993 NAO trends and for CMIP6 future 2024-2054 NAO trends using scenarios SSP126 and SSP585. Grey box plots use the raw CMIP6 model output (median line in centre of 25 to 75 percentile box; total range whiskers). Red box plots use CMIP6 model output after applying the fractional difference reddening method ("+FD") with $d_R$ = 0.24. The observed 1963-1993 NAO trend is shown as the dashed horizontal line.

# Publications

Material from Chapters 3, 4 and 5 has been published in:

Eade, R., Stephenson, D. B., Scaife, A. A., and Smith, D. M. (2021) Quantifying the rarity of extreme multi-decadal trends: how unusual was the late twentieth century trend in the North Atlantic Oscillation? *Climate Dynamics* 58(5-6): 1555–1568.

This paper was led by Rosemary Eade based on Chapters 3, 4 and 5 of this thesis and it is a condensed version of the full chapters. The paper co-authors are the supervisors for this PhD who helped guide the research throughout the PhD.

# 1. Introduction

Changes in regional atmospheric climate variables such as temperature and precipitation over the coming few decades (multi-decadal timescales) are of interest to society, but it is the extreme changes that are most important to predict as these can lead to increased likelihood of extreme events that can have severe impacts on society (Scaife et al., 2005; Deser et al., 2017; O'Reilly et al., 2021; Smith et al., 2022a). A common method to quantify the strength of multi-decadal climate change is to use moving window linear trend analysis, for example trends in large scale climate modes of variability such as the North Atlantic Oscillation (NAO) or the North Atlantic jet (Raible et al., 2005; Semenov et al., 2008; Scaife et al., 2008, 2009; Deser et al., 2017; Bracegirdle et al., 2018; Bracegirdle, 2022). Rapid changes in climate modes such as the NAO can have considerable impacts on regional atmospheric variability. Multi-decadal trends in the winter NAO have been shown to have a strong influence on Eurasian trends in temperature and snow cover (Ye et al., 2022). The unusually large positive NAO trend from the 1960s to 1990s has been shown to account for at least half of the winter warming in the northern hemisphere extra-tropics (Scaife et al., 2005), while the unusually large negative trend from 1920 to 1971 more than halved the winter warming (Iles and Hegerl, 2017).

Modelling methods are required to quantify the distribution of maximum trends in the observed record, using statistical and/or dynamical models. Extreme NAO trends in the observed record are not well simulated by current state-of-the-art General Circulation Models (GCMs) (Bracegirdle et al., 2018, 2022) which is concerning as it is likely that the influence of the NAO on future projections of regional climate change will also be underestimated. It is therefore important to understand and quantify whether the GCMs may be underestimating extreme NAO trends and find ways to improve the realism of the models.

## 1.1 Aims

The aim of this thesis is to outline methods to quantify the distribution of moving window trends and extreme trends in climate indices and make predictions of how these distributions may change in the future. The winter NAO is chosen as

a case study with a moving window length of 31-years. Key questions that will be addressed are:

1. How can we better diagnose and quantify extreme trends in climate indices?
2. How realistically do General Circulation Models simulate extreme trends in the NAO?
3. Can post-processing of GCM simulations improve the representation of extreme trends in the NAO?
4. How might extreme trends in the NAO change in the future?

## 1.2 Structure of this thesis

Chapter 2 reviews the current state of knowledge about winter NAO multi-decadal variability and the motivation behind choosing it as a case study, including methods to distinguish between externally forced and unforced components of the NAO variability.

Chapter 3 outlines the mathematical definition of a multi-decadal linear trend in a climate index series. From this definition, the distribution of moving window trends is mathematically derived by fitting stochastic models to the original index series. The relationship of the distribution of extreme trends to that of the moving window trend series is discussed and a simulation approach is proposed to estimate the distribution of block maxima trends and the effect of uncertainties in distribution parameter estimates.

Chapter 4 applies the trend analysis methods from Chapter 3 to the winter NAO case study by first fitting simple stochastic processes to the observed NAO index series and then estimating the related distributions of moving window trends and extreme trends. These distributions are used to quantify the likelihood of the maximum observed 31-year NAO trend (1963-1993) in the historical record, which can be used as a simple estimate of the probability of such a trend occurring again in the future under the assumption of stationarity.

Chapter 5 uses GCM simulations from the Coupled Model Intercomparison Project (CMIP) phases 5 (CMIP5) and 6 (CMIP6) (Taylor et al., 2012; Eyring et al., 2016) large ensembles of historical experiments to estimate the distribution

of extreme NAO trends. Comparisons are made to the fitted stochastic model distributions from Chapter 4 to investigate whether GCMs may be underestimating multi-decadal NAO trends and show how this may relate to a lack of autocorrelation in the NAO index series.

Chapter 6 proposes new post-processing reddening methods to increase the autocorrelation or long-range dependence of the GCM NAO index series and improve the representation of extreme trends in the NAO. This is the first known application of such a reddening method to postprocess climate model output and increase the multi-decadal variability of the NAO index.

Chapter 7 uses CMIP6 climate projections to quantify how the behaviour of extreme trends in the NAO may change in the future, making use of the reddening methods from Chapter 6 to better simulate the range of plausible future trends.

Conclusions and ideas for future work are given in Chapter 8.

# 2. Background

## 2.1 Outline

The motivation to assess the multi-decadal variability of the climate system is to improve understanding of observed periods of rapid change which have a significant impact on society, such as occurred in the North Atlantic in the second half of the 20th century. This chapter reviews what is currently known about the multi-decadal variability of the North Atlantic Oscillation (NAO) in winter, highlighting the importance of multi-decadal trends in the NAO for northern hemisphere winter climate variability and how General Circulation Models (GCMs) tend to underestimate the magnitude of NAO trends. Methods are discussed for evaluating multi-decadal variability in terms of moving window filters, and for quantifying the role of externally forced components relative to internal variability components. In general it is found that for the full historical period back to the mid-1800s, the NAO interannual and multi-decadal variability is consistent with internal variability, however when restricting the period to just the late 20th century, there is still a lack of consensus as to whether there is evidence of a forced signal. Finally, a brief description is given of the various drivers and mechanisms that may influence NAO multi-decadal variability.

## 2.2 The North Atlantic Oscillation

The North Atlantic Oscillation is the dominant large-scale mode of atmospheric variability in the North Atlantic region related to changes in the North-South pressure gradient (Walker and Bliss, 1932) and can be quantified by an index representing the difference in pressure between the Azores and Iceland regions (Jones et al., 1997; Hurrell, 1995; Hurrell et al., 2003). Statistically it is the dominant mode of interannual variability in the North Atlantic as it relates to the first Empirical Orthogonal Function (EOF) of monthly mean sea level pressure (MSLP) variability, with consistent North Atlantic teleconnection patterns in other atmospheric variables (Ambaum et al., 2001). The NAO also manifests as changes to the mean wind flow (Wanner et al., 2001), with an alternative definition based on the variability of westerly wind strength in the mid-latitude North Atlantic (40-60N) (Greatbatch, 2000), and is strongly linked to the Atlantic storm track and jet stream (e.g. Woollings et al., 2015).

### 2.2.1 Observed NAO multi-decadal variability

The NAO exhibits decades with a persistent positive or negative phase and also decades of rapid change (Hurrell et al., 2003). In the latter half of the 20th Century the observed winter NAO experienced a large positive trend, with “unprecedented strongly positive NAO index values” (Hurrell, 1995) (Figure 2.1), while decades before and after displayed large negative trends seen in Figure 2.2 (Hanna et al., 2015). This change of behaviour between different decades has been termed *multi-decadal variability*. Going beyond direct observations to use proxy records, low-frequency variability can be identified in the North Atlantic over several hundreds of years, shown in Figure 2.3 (Luterbacher et al., 1999). The North Atlantic Jet has multi-decadal variability that is closely related to the NAO (e.g. Woollings et al., 2010) and also exhibits significant long-term positive trends for the winter jet speed and latitude over the historical observation record (e.g. Hallam et al., 2022). In addition to multi-decadal variability, the observed interannual variance of the NAO has increased over the last century, shown in Figure 2.4 (Hanna et al., 2015).

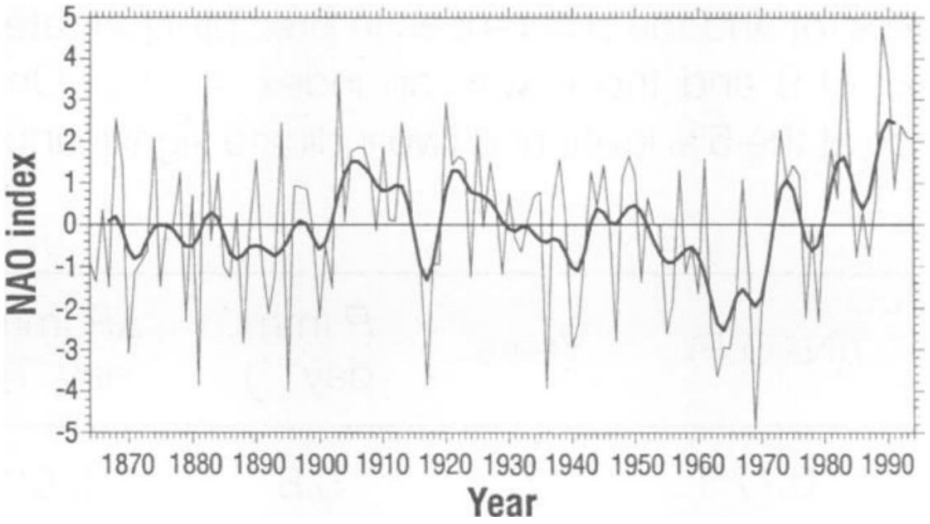


**Figure 2.1:** Normalised time series of winter mean NAO observations (thin solid lines) and a low-pass filtered version (thick solid lines). The NAO index is defined as the difference of normalised MSLP at locations in Stykkisholmur and Lisbon, for December to March seasonal averages. Source: Hurrell (1995) Figure 1A.

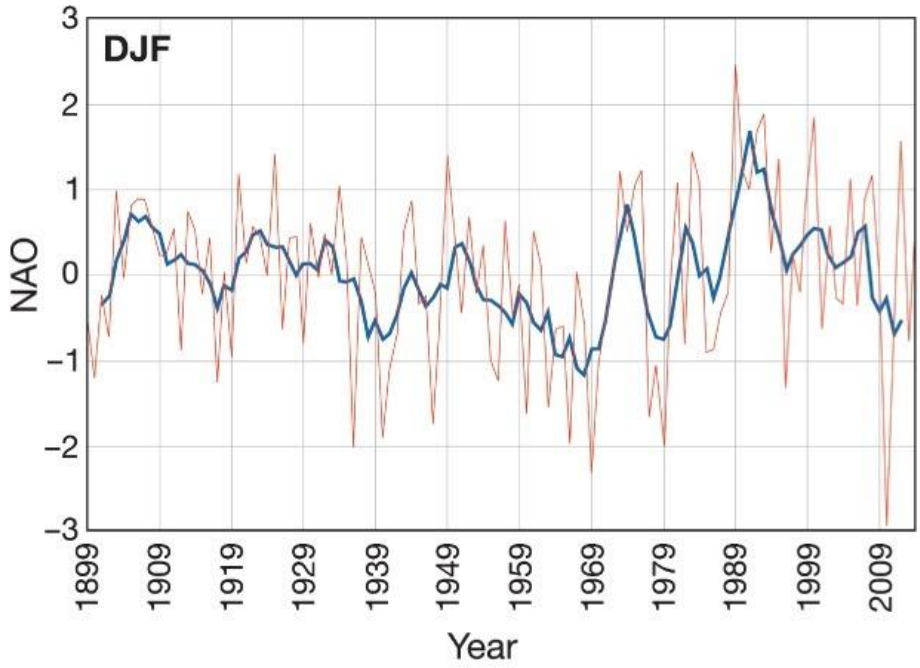


**Figure 2.2:** Winter NAO values plotted annually (thin lines) and 5-year running means (thick lines). The NAO index is defined as the first EOF of observed December-to-February seasonal mean MSLP anomalies. Source: Hanna et al. (2015) Figure 1.

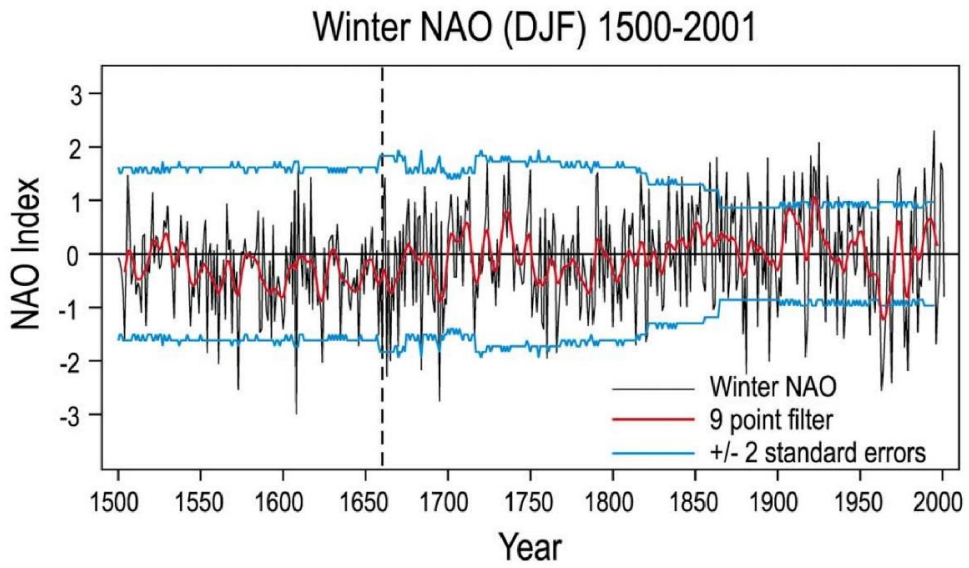


**Figure 2.3:** Normalised time series of winter mean NAO index (black solid line) and a low-pass filtered version (red solid lines). This is a reconstruction using instrumental measurements of surface temperature, MSLP and precipitation as predictors for the NAO index based on observed differences of MSLP between Iceland and the Azores, for December to February mean. Source: Luterbacher et al. (2001) Figure 3.

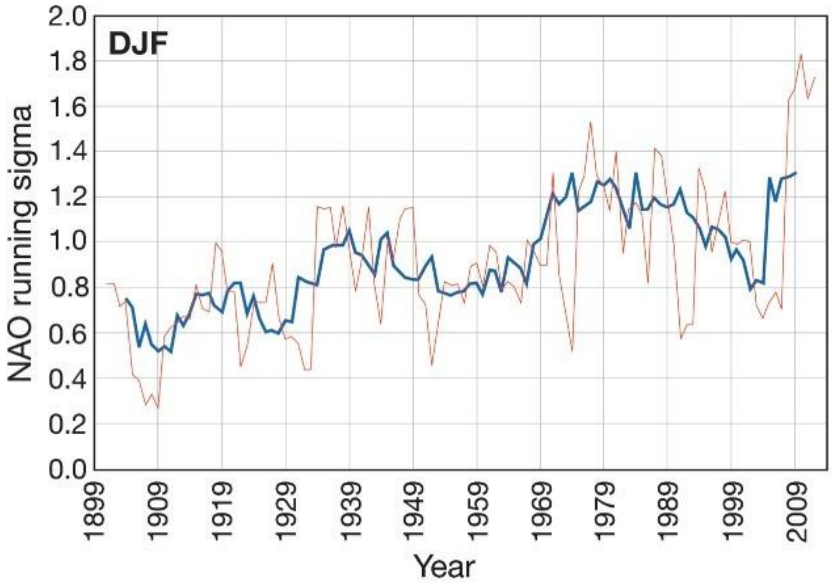


**Figure 2.4:** Running standard deviation of NAO index (defined as in Figure 2.2) for 5- and 11-year windows (thin and thick lines). Source: Hanna et al. (2015) Figure 3.

### 2.2.2 Impacts of NAO multi-decadal variability

The impact of NAO variability on European and North American regional climate has been widely discussed in the literature in terms of seasonal forecasting (e.g. Scaife et al., 2014). The winter NAO 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 et al., 2001) which in turn effects 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). Understanding multi-decadal trends in the NAO is therefore important for all these sectors when planning for future adaptation.

Increased interest in multi-decadal NAO trends occurred in the late 1990s, when links were identified between the observed Atlantic variability and Northern hemisphere winter temperature variability. 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). More generally, the large positive NAO trend from the 1960s to 1990s has been shown to account for at least half of the winter warming in the northern hemisphere extra-tropics in this period (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 the *minimum* 52-year trend in the NAO historical record. It is therefore important that climate models are capable of simulating these *extreme* multi-decadal trends in the NAO as such dynamic effects may enhance or counteract some of the thermodynamic regional climate warming signal over the coming decades (e.g. Scaife et al., 2008; Shepherd, 2014; Deser et al., 2017; Fereday, 2018).

NAO multi-decadal variability has been shown to influence regions beyond nearby Europe and North America, for example surface temperature in South-

Central China and Eurasia (Zuo et al., 2016; Ye et al., 2022). There is also a growing body of research showing that the NAO multi-decadal trends can have a strong influence on the variability of other variables, such as European rainfall (Deser et al., 2017) and Eurasian snow cover (Ye et al., 2022).

### 2.2.3 Anthropogenic forcing of the NAO

The 1960s-1990s NAO trend inspired the key question as to whether this large positive trend could be attributed to climate change. In the early 2000s, just after the peak of a positive phase of the NAO, a number of studies showed evidence that the dominant driver of the then recent positive NAO trend might be anthropogenic forcing (Gillett et al., 2003; Shindell et al., 1999) (proposed mechanisms are summarised in Section 2.5.2.7). Gillett et al. (2003) showed that most GCM experiments with increasing greenhouse gas forcing analogous to the 1960s to 1990s lead to a positive NAO, however this response was not uniform across all models. Indeed, in the decades that followed, the NAO has exhibited a negative trend which cannot be simply explained by anthropogenic forcing in most GCMs (Pinto and Raible, 2012; Hanna et al., 2015).

Osborn et al. (1999) showed that a coupled GCM (HadCM2) forced by greenhouse gas and aerosol concentrations partially simulated the positive trend in the NAO over the second half of the 20$^{th}$ century, and predicted a downturn in the first half of the 21st century which is weaker when the aerosol forcing is removed (Figure 2.5). This doesn't explain all the variability observed, but with comparison to a related control run (no external forcings just natural internal variability) they found evidence to reject the null hypothesis that the observed NAO variability is just due to internal variability and weather noise.

Multi-model studies show considerable variation in the response of the NAO to greenhouse gas forcings. CMIP2 greenhouse gas forced multi-model simulations (1% increase per year over 80 years) exhibit a slight increase in the NAO trend compared to matching control simulations (without the greenhouse gas forcing), but with large variation between models (Kuzmina et al., 2005; Stephenson et al., 2006). The Coupled Model Intercomparison Project Phase 5 (CMIP5) (Taylor et al., 2012) multi-model simulations with historical forcings and projections for the future (RCP8.5 scenario), including greenhouse gas and

aerosol emission, show an overall increase in the NAO index over the 20th to 21st century for the ensemble mean, and some multi-decadal variability, but this is relatively weak compared to the large variability between models as shown in Figure. 2.6 (Hanna et al., 2015). More recently, the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) simulations have been used to show that the models simulate the NAO interannual variability well, but still underestimate the multi-decadal variability (Lee et al., 2021; Bracegirdle, 2022). There is still a lack of consensus as to whether the 1960s to 1990s NAO trend can be explained by internal variability alone or whether it was externally forced. Much of this analysis relies on trusting that climate models can simulate both the internal and externally forced components of NAO multi-decadal variability. The performance of current generation GCMs is discussed in Section 2.4.4.

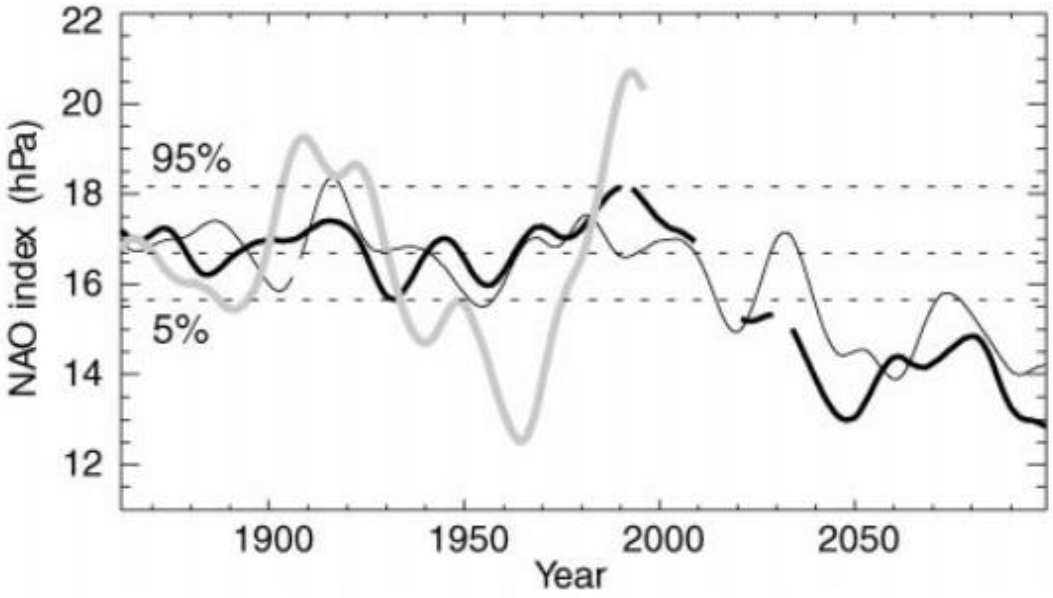


**Figure 2.5:** 30-year low pass filtered winter mean NAO index (December to March) for observations (grey thick line), the ensemble mean from a GCM with greenhouse gas forcings (think black line) and from a GCM with both greenhouse gas and aerosol forcings (thick black line). Source: Osborn et al. (1999) Figure 9.

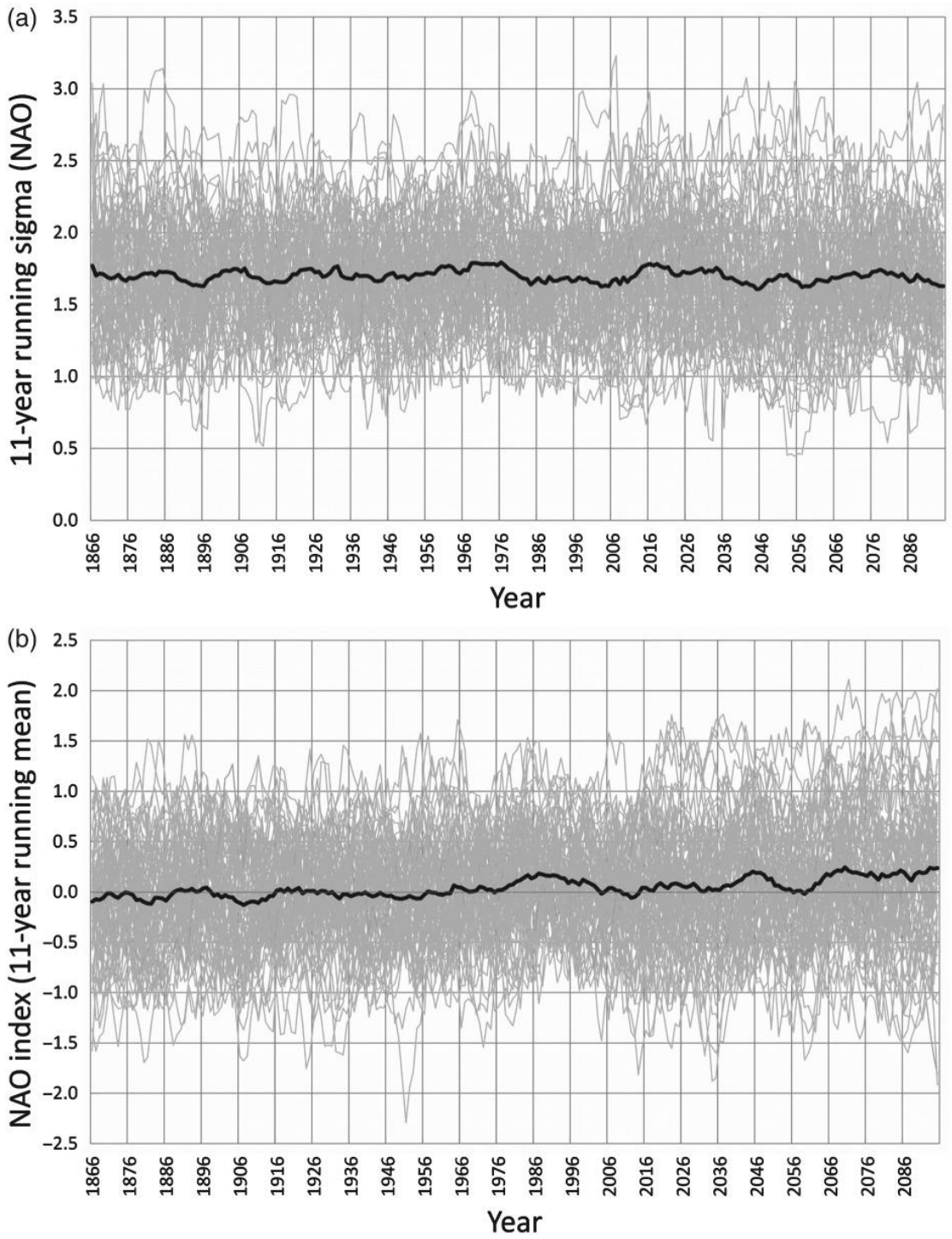


**Figure 2.6:** The standard deviation (top) and mean (bottom) on moving 11-year windows for the NAO index from CMIP5 historical GCMs (grey thin lines). The version for the multi-model ensemble mean is shown in thick black. Source: Hanna et al. (2015) Figure 10.

## 2.3 Evaluation of multi-decadal variability

### 2.3.1 Moving window averages

A common way to assess the multi-decadal variability part of a time series is to apply a smoothing technique to filter away the higher frequency variability. A simple method is to calculate multi-year averages of a time series on M-year moving windows (some positive integer M) which helps to identify the periods where a certain phase persists. For example, Scaife et al. (2005) calculate the observed 10-year mean winter NAO index for non-overlapping windows, while Stephenson et al. (2000) calculate 10-year running medians of winter NAO on moving windows, both noting the change in phase from negative in the 1960s to positive in the 1990s. This technique attempts to remove the short-term variability component (assumed to be unpredictable) from a longer-term variability component that has potential predictability and may be externally forced (Zwiers et al., 1987).

### 2.3.2 Moving window trends

An alternative method when considering multi-decadal variability is to compute the moving window linear trend series, which helps to highlight periods of rapid change. This alternative filtering method is frequently used to assess climate indices, for example, trends in the North Atlantic Oscillation (Deser et al., 2017; Scaife et al., 2005, 2009; Semenov et al., 2008; Raible et al., 2005) and in the North Atlantic jet characteristics (Bracegirdle et al., 2018; Bracegirdle, 2022; Blackport and Fyfe 2022; Schurer, 2023). 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). Unusual or extreme trends are of particular interest as they represent rapid periods of change that society may not be expecting and is not well adapted to. Moving window trend analysis has also been used 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). Trends in these studies are typically estimated using Ordinary Least Squares (OLS) regression, and this method is presented in Chapter 3.

## 2.4 Modelling multi-decadal internal variability

There is some debate as to how much of the multi-decadal variability of the NAO is an aggregation of short-term weather variations, which are unpredictable beyond a few weeks, and how much may be forced by internal ocean-atmosphere feedbacks or by external drivers, which are potentially predictable (Zwiers et al., 1987). Common methods to quantify these components make use of stochastic models or general circulation models (GCMs).

Stochastic processes are a set of random variables with some pre-defined mathematical properties, and for atmospheric variables such as the NAO, these will need to be continuous stochastic processes (e.g., Wilks, 2006). They are a useful tool for emulating climate index variability over time as they can be generated very quickly and can replicate some of the variability seen in observations. GCMs are more computationally expensive to run but they include known physics and chemistry elements of the climate system so can lead to a better understanding of the behaviour of the real world.

### 2.4.1 Stochastic models of daily mean NAO index

The observed *daily* variability of the NAO is frequently modelled as a stochastic red noise process using a first order auto-regressive model, AR(1) (Feldstein, 2000, 2002; Franzke, 2009; Keeley et al., 2009). This AR(1) model can then be used to estimate the distribution of longer-term variability that would be expected without any additional forcing or non-stationarity, and test the null hypothesis that this variability can instead be explained by the simple aggregation of weather noise.

Feldstein (2000) applies this potential predictability approach to show that the winter mean NAO variability for the period 1864-1997 can be explained without any external forcing, i.e. the null hypothesis is accepted when simulating the daily weather noise element as an AR(1) process with parameters estimated from the observed NAO time series. In contrast, for the late 20$^{th}$ century period 1958-1997 the null hypothesis is rejected as it explains just 61% of the interannual variability.

Keeley et al. (2009) estimate how the within season daily variability is influenced by the interannual variability, by estimating the AR(1) lag-5 day autocorrelation parameter from daily NAO values after the removal of the interannual variability (using the 5-day lag rather than 1-day lag autocorrelation as they say this leads to a better fit of the data in this case). For the period 1957-2002 they estimate that 70% of the winter NAO index interannual variability is actually externally forced, so only 30% is explained by the null hypothesis of weather noise. Their method leads to a lower estimate of the AR(1) autocorrelation parameter than Feldstein (2000), which can be shown to reduce the model variability estimate in most cases (see Chapter 3 on how increased autocorrelation can lead to an increase in the variance of the moving window trend series).

Feldstein (2002) extends the method of Feldstein (2000) to test the significance of the Arctic Oscillation (AO) and NAO trends over the period 1967-1997. They generate a large sample of 90-day simulations based on a red noise model with lag-1 day autocorrelation parameter estimated from the observations, and use randomly generated initial values. Distinct 30-year trends are then calculated leading to an estimate of the trend distribution due simply to aggregation of weather noise. The observed 1967-1997 NAO trend compared to this distribution is shown to be statistically significant at the 99% confidence level, suggesting that there must be an additional factor, other than weather noise, that leads to the magnitude of the observed trend, such as atmospheric coupling with the ocean or some external forcing.

In summary, these studies suggest that for the full historical period back to the mid-1800s, the NAO interannual variability is consistent with internal variability without any external forcing, however when restricting the period to just the late 20$^{th}$ century, there is evidence of a forced signal. However, a warning is given against the validity of results for the late 20$^{th}$ century due to the short sample size which happens to coincide with a period of large positive trend that may itself alter the stochastic model parameter estimates (Feldstein, 2000). Also, these studies assume independence of individual winters - Feldstein (2002) uses a randomly generated initial value for each season and Keeley et al.

(2009) argue that daily autocorrelation becomes insignificant beyond 30 days – which is shown not to be a valid assumption in Section 2.4.2.

### 2.4.2 Stochastic models of seasonal mean NAO index

The variability of the winter *seasonal mean* NAO can also be modelled as a stochastic process to help interpret the longer-term variability of the NAO. Wunsch (1999) sets up the null hypothesis that the longer-term variability of the extended-winter NAO index can be explained by the aggregation of seasonal mean variability modelled as a stochastic process. Wunsch (1999) shows that the winter mean NAO exhibits weak red noise, and simulations using this fitted AR(1) process can lead to multi-decadal trends of similar magnitude to that observed from the 1960s to 1990s. They conclude that the NAO multi-decadal variability can be explained as an aggregation of seasonal mean variability and find no evidence for the requirement of an externally forced component. Similar results were found by Franzke (2009) and Stephenson et al. (2000), though the latter show that winter mean NAO variability requires a long-range fractional difference model to fully account for the long-term persistence of the NAO index.

Wunsch (1999) also show that simulations from AR(1) processes can produce large multi-decadal fluctuations in the autocorrelation estimate when calculated over relatively short moving windows, similar to the observed fluctuations shown in Figure 2.7 (Hurrell and van Loon, 1997). This leads to the warning that short samples of observed time series exhibiting red noise should be treated carefully to avoid underestimating the level of autocorrelation and wrongly rejecting the null hypothesis.

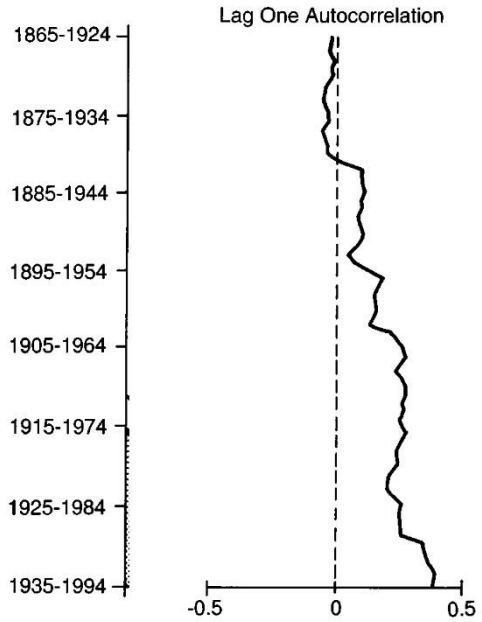


**Figure 2.7:** Lag-1 year autocorrelation estimated within moving 60-year windows for the winter mean NAO index (December to March). Source: Hurrell and van Loon (1997) Figure 7.

Aggregated variance analysis is a simple tool to assess long-term variability in time series for a range of time-scales to see how quickly this variability dissipates. This method computes the variance of time means on moving windows of increasing length M years, as seen in Figure 2.8 (Stephenson et al., 2000). Stochastic models lead to an analytic way of calculating the expected behaviour under the null hypothesis that the longer-term NAO variability is explained by aggregation of the shorter-term variability and weather noise. Stephenson et al. (2000) show that the observed multi-year mean of the winter NAO index has rather higher aggregated variance than expected for a simple white noise process which should show a decay in variance proportional to 1/M (M = length of moving window) (Figure 2.8). Franzke (2009) use an empirical method to calculate the expected behaviour for a red noise process by calculating the mean and 95% prediction intervals from a large ensemble of AR(1) simulations. I extended this aggregated variance method to assess the variance of the M-year trend (for increasing integers M) in Figure 2.9 (Menary et al., 2018, Figure 4), but this is missing a comparable analytic solution which will be derived later in Chapter 3.

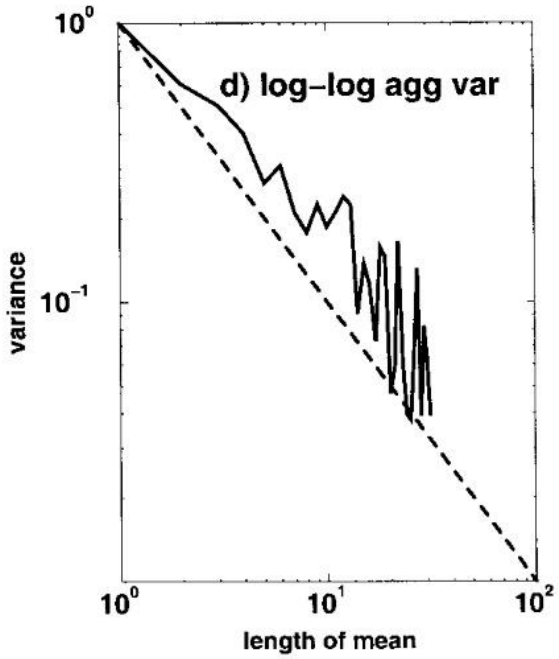


**Figure 2.8:** Aggregated variance for multi-year means of observed winter NAO vs a white noise model. Source: Stephenson et al. (2000) Figure 2.

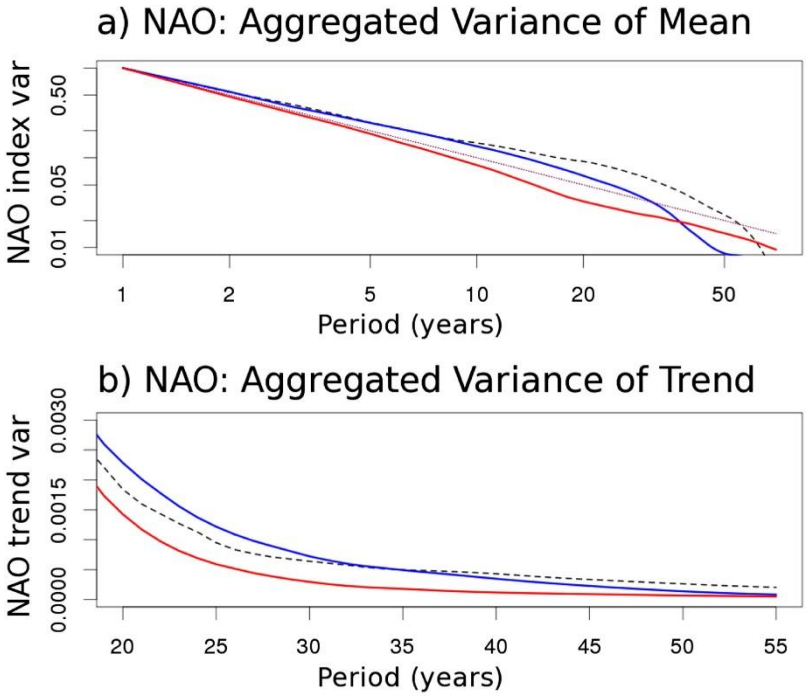


**Figure 2.9:** Aggregated variance plots of winter (December to February) NAO multi-year mean and linear trend. Observed NAO index (black dashed line) compared to two different GCMs (red low resolution and blue medium resolution), and the empirical white noise values in (a) (dotted grey diagonal line). Source: Menary et al. (2018) Figure 4.

### 2.4.3 Climate models and internal variability

Another frequently used method to estimate the natural internal variability of the climate system is to use coupled GCM control simulations, i.e. models with a stationary level of external forcing such as pre-industrial climate conditions. GCM control runs can be used as a robust null hypothesis against which to compare the observed variability or the variability of GCMs which include additional external forcing components that reflect past or future changes. Examples of these types of studies were summarised in Section 2.2.3, generally showing a weak response of the NAO to the external forcing (Osborn et al., 1999; Kuzmina et al., 2005; Stephenson et al., 2006).

Advantages of GCMs over stochastic models are that they do not rely on assumptions about the underlying statistical distribution of the climate variable, and they can incorporate the current state of knowledge on climate system dynamics, thermodynamics and feedbacks between the ocean and atmosphere (and land and sea ice) which may occur on a range of timescales. These simulations are more computationally expensive than stochastic models, but can still be run for many hundreds of years, i.e. for much longer timescales than the available observations datasets. This type of analysis assumes that climate models are able to realistically simulate both the internal and externally forced components of climate variability, though this is not always the case as discussed in Section 2.4.4.

### 2.4.4 Climate model performance

Many recent studies have shown that GCMs underestimate the multi-decadal variability of the NAO. Scaife et al. (2009) found that none of the GCMs they tested could reproduce the strong positive trend observed in 1965-1995, likely in part due to an inadequate stratosphere in the models (Scaife et al., 2005). More recently, the Coupled Model Intercomparison Project Phase 5 and 6 (CMIP5 and 6) (Taylor et al., 2012; Eyring et al., 2016) multi-model ensembles of historically forced experiments were assessed to compare extreme trends in the winter NAO and jet indices (Bracegirdle et al., 2018; Bracegirdle, 2022; Blackport and Fyfe, 2022; Schurer et al., 2023). They found that these GCMs still generally underestimate the extreme multi-decadal trends in the NAO (and

jet strength) compared to observations, despite successfully representing the interannual variability (Kravtsov, 2017; Wang et al., 2017). 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. Possible causes of the GCM underestimation of multi-decadal trends include:

- An underestimation of the natural internal variability of the atmosphere
- A failure to realistically respond to external forcing
- An underestimation of the ocean-atmosphere interactions

(Gastineau and Frankignoul, 2015; Blackport and Fyfe, 2022; Bracegirdle, 2022).

## 2.5 Causes of NAO multi-decadal variability

### 2.5.1 Null hypothesis – natural internal variability

There are many hypotheses about what may be driving the year-to-year variability of the NAO. These drivers may also be important for multi-decadal variability if they have a long response time, or if they are themselves forced by another low-frequency driver. To test the relative influence of single or multiple drivers, their impact can be compared against the null hypothesis that the variability is just due to a build-up of aggregated weather noise (as discussed in Section 2.4.1 and 2.4.2 using stochastic models) or the null hypothesis that the variability can be explained by natural variations within a GCM control simulation that does not contain these drivers (as discussed in Section 2.4.3).

### 2.5.2 Drivers and mechanisms of NAO multi-decadal variability

A variety of drivers of the NAO multi-decadal variability have been proposed in the literature in terms of observation and modelling studies, many of which highlight the importance of ocean-atmosphere coupling and a well resolved stratosphere as outlined in the following sections. Some of these drivers can be thought of as internal to the coupled climate system as a whole, such as the variability of the stratosphere, oceans and Arctic sea ice, and may involve coupled feedback mechanisms with the NAO. Other drivers can be thought of as external to the climate system such as solar variability, volcanic eruptions

and anthropogenic emissions, without any feedback from the NAO. These may force the NAO directly or indirectly via the stratosphere, oceans and sea ice.

*2.5.2.1 The stratosphere*

Baldwin and Dunkerton (2001) demonstrated the importance of the stratosphere in leading changes in the NAO by about 60 days, showing that observed changes in the stratospheric circulation and wind fields, in particular the strong northern hemisphere stratospheric polar vortex, descend down and influence the troposphere. Periods with a strong polar vortex show a shift in the daily NAO index towards more positive values, and more negative values when the vortex is weaker. Kidston et al. (2015) explain that a stronger polar vortex leads to reduced polar cap pressure causing air to rise and cool at the pole, balanced by a movement of air equatorward where it descends. This reduces surface pressure at the poles and increases pressure in the tropics, leading to a poleward shift of the tropospheric jet and a positive NAO response. Likewise, a weaker polar vortex leads to a negative NAO response.

The troposphere can influence the stratosphere via planetary Rossby waves which propagate upwards and weaken the polar vortex (Kidston et al., 2015). This can lead to a sudden stratospheric warming (SSW), for example in January 2009 the temperature of the polar stratosphere jumped from -70C to -10C in 4 days, which leads to a negative NAO response in the month following the warming (Kidston et al., 2015).

GCM experiments with observed stratospheric forcings lead to improved simulations of North Atlantic winter variability that can better replicate the observed NAO increase from the 1960s to 1990s (Scaife et al., 2005). In terms of winter seasonal variability, the simulation of SSW events is a crucial part of the prediction skill. The correlation of winter mean NAO forecasts with observations over past decades drops from a significant value of 0.62 to an insignificant value of 0.09 if the members with SSW events are removed from the ensemble (Scaife et al., 2016). Periods with a SSW event lead to a significant mean shift towards negative NAO values. The initialisation of the polar high-level stratosphere also improves seasonal predictability, shown to explain around 22% of the winter NAO variability (Nie et al., 2019).

A variety of other mechanisms that lead to changes in the polar vortex may thus effect surface weather, e.g. tropical SSTs, solar variability, volcanic eruptions, greenhouse gas concentrations, and these may evolve on a range of timescales (Baldwin and Dunkerton, 2001; Kidston et al., 2015; Smith et al., 2014).

*2.5.2.2 Atlantic multi-decadal oscillation*

The Atlantic Multi-decadal Oscillation (AMO) relates to persistent phases of the winter NAO with positive AMO years showing a tendency for negative NAO patterns (Smith et al., 2014). This is in part due to the influence of the NAO on the North Atlantic ocean. The warm phase of the AMO in the 1990s links to an increase in the Atlantic Meridional Overturning Circulation (AMOC), which ocean models show to be caused by the predominantly positive winter NAO in the 1980s (Robson et al., 2012). The observed multi-decadal variability of the AMOC is strongly related to the multi-decadal variability seen in occurrences of SSWs (Reichler et al., 2012). They use GCMs to show that the positive NAO response to a strengthening of the polar vortex leads to surface cooling in the northern North Atlantic which persists over a number of years, strengthens, and descends within the ocean column, leading to low frequency variability of the ocean circulation.

Ocean-atmosphere coupling in GCMs is important for capturing the multi-decadal variability of the NAO. Omrani et al. (2014) show that large scale Atlantic warming can drive changes in the stratosphere leading to SSWs which then follow the stratospheric pathway for influencing the NAO. Atmosphere-only GCMs forced by observed Atlantic SSTs can capture the NAO like winter response, but only if the model has an adequate stratosphere. Delayed oscillator models show that coupling of the NAO with Atlantic Ocean SSTs is key for realistic multi-decadal NAO variability with instantaneous and delayed feedbacks increasing the persistence of the NAO response (Sun, C., et al., 2015; Omrani et al., 2022). This highlights the importance of both a fully resolved stratosphere and ocean-atmosphere coupling for GCMs to realistically simulate the multi-decadal variability of the NAO.

*2.5.2.3 Tropical ocean influence*

Hurrell et al (2004) use multi model atmosphere-only GCMs forced with observed SSTs to show that SSTs are able to explain much of the interannual variability of the winter NAO (winter season JFM) and some of the long-term positive trend from the 1950s to 2000s. Regionally forced experiments (global tropical ocean forcing versus Atlantic tropical ocean forcing) show that the leading influence comes from the tropics, and mainly the Pacific and Indian oceans rather than the Atlantic itself, where the surface temperature warming has a related increase in local rainfall (Hoerling et al., 2001; Hurrell et al., 2004). Hurrell et al. (2004) find that the positive NAO trend in this period using the atmosphere-only GCM is significantly different to the sampling distribution calculated from such 50-year trends in a control experiment forced with just climatological SSTs (which is centred around zero), but it is still only half the magnitude of the observed 50-year trend.

Seabrook et al. (2023) show that the positive phase of tropical Pacific decadal variability leads to a strengthening of the stratospheric polar vortex (SPV) on decadal time scales which is consistent with a positive NAO response. In contrast, on interannual time scales the Pacific variability is dominated by the El Nino Southern Oscillation and in this case the positive phase is known to cause a weakening of the SPV and an associated negative NAO response. They show that the decadal response is related to a build up of water vapour in the stratosphere which cools the poles and strengthens the extratropical jet (related to an increase in the meridional temperature gradient).

Peterson et al. (2002) use a primitive dry atmospheric model forced with realistic diabatic heating to show that, while the tropical forcing is most important for the NAO multi-decadal trend in the second half of the 20$^{th}$ century, the extra-tropical forcing is important for the interannual variability. They also manage to simulate an eastward shift in the NAO spatial pattern in this period, similar to that observed, but only if fully global diabatic forcings are included.

The Indian Ocean warming since the 1950s is at least partially due to increased greenhouse gas concentrations, and is well simulated in coupled GCMs with greenhouse gas forcings as well as in atmosphere-only GCMS forced with

observed Indian Ocean SST (Hoerling et al., 2004). However, while the Indian Ocean has continued to warm since the 1990s, the NAO has experienced a downward trend (Hoerling et al., 2004; Hurrell et al., 2004) so there isn't a simple consistent relationship between these variables.

Yu and Lin (2016) show that the observed NAO is significantly correlated with rainfall in the tropical Indian ocean, but not SSTs. They propose that this rainfall variability leads to a shift in the local jet, influencing the northern hemisphere jet stream and leading to a response in the NAO. Tropical rainfall, especially in the Indian and West Pacific oceans, is identified as a key driver in the atmosphere-only GCMs of Hurrell et al. (2004). Forcing the Indian ocean with a warming trend in the atmosphere-only GCMs of the same magnitude as observed (1950-2000) leads to enhanced Indian Ocean rainfall and a positive NAO response (Hoerling et al., 2004), while a sudden step change in temperature based on the change from 1950 to 2000 shows that this NAO response takes three weeks to emerge, likely through changes in the storm track. Molteni et al. (2015) note that a deficiency of atmosphere-only GCMs is that they can't simulate the feedback of convection onto SSTs causing the tropical rainfall response to be too uniform with longitude. Tropical rainfall variability also accounts for a large part of the skill (40% of the variability) in seasonal forecasts of winter NAO, by generating Rossby waves that travel poleward in the upper troposphere (peaking at 200hPa height) from key source regions in the tropical oceans (Scaife et al., 2017).

*2.5.2.4 Arctic sea ice*

It has been shown in modelling studies that reducing Arctic sea ice generally leads to a negative NAO response, however large ensembles are needed for this response to be apparent as the strength of the response is generally underestimated and there is large variation between models in terms of timing, strength and sign of this signal (Mori et al., 2014; Cohen et al., 2014; Screen et al., 2018; Smith et al., 2022b). Results differ across modelling studies due to differences in the experimental setup, such as the time scale of the sea ice change or the location of the sea ice change, and differences between the models used.

There are a variety of possible mechanism by which reduced Arctic sea ice may influence the NAO. GCMs are fairly consistent in their simulation of local thermodynamic responses with a warming at the surface where the sea ice is reduced, but the dynamical responses vary considerably (Screen et al., 2018; Smith et al., 2022b). Reduced Arctic sea ice leads to local surface warming and a reduction in the surface temperature gradient from the equator to the pole which weakens the mid-latitude westerly jet, in line with a negative NAO response (Screen, 2017). Reduced Arctic sea ice may also weaken the AMOC (Sevellec et al., 2017; Suo et al., 2017) which could influence the NAO via ocean heat transport. The regional pattern of Arctic sea ice loss causes local changes in pressure and asymmetries in the flow which can enhance wave activity and weaken the stratospheric polar vortex leading to a negative NAO response through the stratospheric pathway described in Section 2.5.2.1 (Cohen et al., 2014; Francis and Vavrus, 2012). Reduced Arctic sea ice leads to increased snow fall over Siberia (as it causes an increase in precipitation) (Ghatak et al., 2010). This strengthens the Siberian high which expands into northern Europe, the north pole and eastern North America, shifting the Icelandic low south-westwards and leading to a negative NAO response (Cohen et al., 1999).

Experiments with large changes in sea ice, such as future projections versus present day, tend to show a stronger negative NAO response (e.g. Deser et al., 2010) whereas gradually changing sea ice reflecting recent observed year-to-year variability tend to show weaker signals (e.g. Screen et al., 2013). Future projections tend to reduce sea ice everywhere, whereas historical to present day sea ice variability has a dipole pattern of change in the Atlantic (Labrador sea versus the Greenland, Iceland and Norwegian (GIN) seas) and the Pacific (Sea of Okhotsk versus Bering Sea) (Deser et al., 2000). Atmosphere-only models show that the atmospheric response to regional sea ice loss depends on the location and the size of the loss as reducing sea ice in the Atlantic sector leads to a weaker winter polar vortex whereas the Pacific sector leads to a strengthening (Sun, L., et al., 2015; McKenna et al., 2018). For moderate changes in Atlantic or Pacific sea-ice, this leads to a negative or positive Arctic Oscillation (AO) response respectively, as expected via the stratospheric pathway, however for large sea ice reduction in the Pacific the response is

actually a negative AO with similar magnitude to that from the analogous Atlantic forcing (McKenna et al., 2018). They explain these differences as a result of two competing pathways, one via the stratospheric polar vortex, and one via the tropospheric mechanism relating to eddy feedbacks. Screen (2017) use atmosphere-only GCM experiments to show that differences between the responses to regional sea ice loss are related to constructive or destructive interference between the Rossby waves forced by the sea ice loss and the background wave pattern of the model climatology.

A recent coordinated large multi-model study has shown that very large ensembles are needed to see a significant response in the NAO and North Atlantic jet to Arctic sea ice loss, and the underestimation of the climate model response is related to an underestimation of the eddy feedback, the strength of which also partially explains the differences between models (Smith et al., 2022b).

The NAO has also been shown to influence Arctic sea ice. Deser et al. (2000) show that in observations of winter (JFM) mean NAO for the period 1958-1997, periods of positive NAO seem to force Arctic sea ice variability with higher sea ice concentration in the Labrador sea and lower in the GIN seas which is consistent with the main dipole pattern of sea ice variability in this region using EOF analysis. Strong et al. (2010) use observations of weekly mean NAO for the December to April months (1978-2008) and a vector auto-regressive model to show that NAO forced changes in Arctic sea ice can then have a lagged negative feedback effect on the NAO on timescales of several weeks.

*2.5.2.5 Solar forcing*

Solar radiation variability has been shown to impact atmospheric circulation on decadal (e.g. Ineson et al., 2011) and multi-decadal (e,g, Shindell et al., 2001) timescales. The transition from a maximum to a minimum period leads to a cooling of the upper stratosphere due to reduced ozone heating from incoming ultra-violet radiation (UV). The strongest stratospheric cooling occurs in the tropics leading to a reduced temperature gradient between the equator and the poles, balanced by an easterly wind anomaly which weakens the polar vortex. This then follows the stratospheric pathway of Section 2.5.2.1 leading to a

negative NAO response at the surface. Further analysis shows that for a solar maximum period, the observed NAO has a lagged response which builds to a positive NAO pattern which peaks after about 3-4 years and later reverses after about seven years, though this is mainly due to a significant change in the southern node rather than the full pattern (Gray et al., 2013). They suggest that the lagged response is due to ocean-atmosphere feedbacks with the ocean response to the stratosphere then feeding back onto the NAO. There is also evidence of solar activity forcing changes in the tropical Pacific which can influence the NAO (see Section 2.5.2.3) e.g. by generating Rossby waves (Swingedouw et al., 2011; Chiodo et al., 2016).

Long GCM control simulations forced with solar maximum versus solar minimum UV conditions demonstrate the same mechanisms observed that lead to a negative NAO response (Ineson et al., 2011). Gray et al. (2013) also find that long GCM control runs with a realistically varying solar cycle capture the NAO response to solar minimum and maximum periods, but they get the maximum response at lag zero rather than the lagged response seen in observations that increases over a few years. In contrast, multiple short GCM experiments perturbed with a sudden increase in UV do simulate a lagged NAO response to the forcing which starts weakly and increases out to four years, along with a related SST tripole pattern, though it is still weaker than reality (Scaife et al., 2013).

*2.5.2.6 Volcanic forcing*

Large tropical volcanic eruptions such as Pinatubo (1991) that inject aerosols into the stratosphere lead to a global surface cooling response, but also a positive AO or NAO response in the following winters and warming in Eurasia and North America, with longer lags for higher latitude eruptions (Robock and Mao, 1992). Volcanic aerosol in the lower stratosphere causes ozone depletion leading to cooling at high latitudes and a strengthening of the polar vortex (Stenchikov, 2002) which then follows the stratospheric pathway in Section 2.5.2.1 to a positive NAO response.

A positive NAO response is also found in atmosphere-only GCM experiments forced with Pinatubo like aerosols over five years, but the response is much

weaker (Ottera et al., 2008). GCM experiments can show a more faithful response if they have a realistic QBO, for example when also forced by observed stratospheric winds (Stenchikov, 2004). They show that the NAO response to aerosols is enhanced when there is a westerly QBO which further strengthens the polar vortex, and also allows stratospheric planetary waves to travel from the northern hemisphere south across the equator. Historical simulations used in the Intergovernmental Panel on Climate Change (IPCC) 4$^{th}$ Assessment Report (AR4) capture the tropospheric cooling, but the positive AO/NAO response and winter surface warming are rather weaker than observed (Stenchikov, 2006). More recently, CMIP5 models still struggle to capture the strength of this NAO/AO response, linked to issues with QBO or perhaps an overly stable polar vortex that is hard to shift (Driscoll et al., 2012). Decadal prediction systems (using coupled GCMs) simulate a positive NAO response in the first winter after an eruption, but the amplitude is much weaker than expected from observations (Hermanson et al., 2020). Seasonal forecasts can capture a more realistic NAO/AO response, but this is thought to be due to persistence of observed initial conditions rather than the influence of volcanic forcing and the stratosphere (Marshall et al., 2009).

*2.5.2.7 Anthropogenic forcing*

Increased greenhouse gas forcing in GCMs leads to an overall warming of the troposphere, but in contrast a cooling of the lower stratosphere (Cubasch et al., 2001). This leads to an increase in the meridional temperature gradient around the tropopause, as the tropics warm while the high-latitudes cool, enhancing zonal winds and strengthening the polar vortex (Shindell et al., 1999; Gillett et al., 2003). This can result in a positive NAO response as per Section 2.5.2.1, however there are considerable differences across models with some showing opposite or no response (Section 2.2.3). Shindell et al. (1999) highlight the importance of including the stratosphere in their GCMs (atmosphere- mixed layer ocean model) for simulating a positive trend comparable to the observed NAO in the second half of the 20$^{th}$ century when forced by the observed increase in greenhouse gas concentrations for this period (and projections thereafter). Without this they find that GCMs simulate only a weak decrease in MSLP over the Arctic related to the surface warming more strongly at higher latitudes than the tropics. Gillett et al. (2002) find that increasing the height of

their model (atmosphere- slab ocean model) to include the stratosphere does not increase the response of the AO to increased greenhouse gas forcing. This could be due to differences between the model physics, or differences in the experiment setup as Gillett et al. (2002) apply a doubling of carbon-dioxide (compared to a control simulation with pre-industrial concentration levels) which relates to the projected mid-21st century conditions (van Vuuren et al., 2011), whereas Shindell et al. (1999) apply the observed increase in greenhouse gas concentrations over time.

Greenhouse gas forcings may influence the NAO variability indirectly through changes to the oceans or the stratosphere. Section 2.5.2.3 noted how anthropogenic warming of the tropical oceans may in turn influence the increasing trend of the NAO (Hoerling et al., 2001; Hurrell et al., 2004), especially the Indian Ocean (Hoerling et al., 2004), however a different balance of mechanisms may be needed to explain the decreasing NAO trend since the 1990s when the Indian ocean temperature has continued to rise (e.g. Roxy et al., 2014).

# 3. Distribution of extreme multi-decadal trends

## 3.1. Introduction

This chapter defines the concept of multi-decadal trends in climate indices (Section 3.2) and describes how to model the distribution of moving window trends (Section 3.3) and extreme trends (Section 3.4). The goal is to find a methodology to estimate exceedance probabilities for trends such that in later chapters this approach can be applied to thresholds of interest based on climate index time series. Gaussian stochastic processes are used to represent a time series with some year-to-year memory and the related distributions of trends in these time series are explored. Where observed records of climate indices are relatively short, it can be informative to fit a model to the original index series and then quantify the associated distribution of trends for that model, rather than try to fit a distribution to the moving window trend series directly. For this chapter the focus is on extreme positive trends, but the approaches described can easily be inverted to assess extreme negative trends.

## 3.2. Definition of a multi-decadal trend

A *multi-decadal trend* is here defined as an Ordinary Least Squares (OLS) trend in time for a window of length 31 years, i.e. the slope parameter estimate obtained from linear regression of the index on time. Other window lengths will also be considered. Shpakova et al. (2020) suggest that the maximum 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 the 11 year solar cycle. A standard time scale considered to be decadal variability is defined as 10-30 years by Meehl et al. (2009). For observation based time series which normally have a length of order one century, a convenient window length to satisfy all of these criteria is of order three decades, hence the choice of 31-years.

Moving window trends are 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 $\{Y_1, Y_2, \ldots, Y_n\}$ of length $n$ years, the OLS slope parameter estimate of the trend ($Z_i$) in a window of length $2K+1$ years centred at

time $i$ = 1+$K$, 2+$K$, …, $n$–$K$ is given by

$$Z_i = \frac{\sum_{j=-K}^{K}\left(i+j-\overline{(i+j)}\right)\left(Y_{i+j}-\bar{Y}\right)}{\sum_{j=-K}^{K}\left(i+j-\overline{(i+j)}\right)^2} = \frac{\sum_{j=-K}^{K} jY_{i+j} - \sum_{j=-K}^{K} j\bar{Y}}{\sum_{j=-K}^{K} j^2}$$

where the over-bars indicate a mean over time. $Z_i$ is calculated over years $i$-$K$, $i$-$K$+1, …, $i$+$K$+1 such that the time mean of $j$ over these years is $\bar{j}$ = 0. For this single i'th window the time mean $\bar{Y}$ is independent of $j$, so $\sum_{j=-K}^{K} j\bar{Y} = \bar{Y}\sum_{j=-K}^{K} j = 0$ and

$$Z_i = \frac{1}{h^2}\sum_{j=-K}^{K} jY_{i+j} \qquad (3.1)$$

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

The window length 2$K$+1 is chosen to be an odd number to avoid the central point being half-way between two years, e.g. $K$ = 15 for a moving window length of 31 years (~3 decades). From the filtered time series $Z_i$, the *extreme trend* can be identified as max {$Z_{1+K}$, $Z_{2+K}$, …, $Z_{n-K}$}.

From Equation 3.1, the resulting trend values are a linear combination of the index values, so the trend time series $Z$ is a moving average filtered version of index series $Y$ with the 2$K$+1 filter weights:

$$\left\{\frac{-\mathrm{K}}{h^2}, \frac{-(\mathrm{K}-1)}{h^2}, \ldots, \frac{\mathrm{K}-1}{h^2}, \frac{K}{h^2}\right\}.$$

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($p$, 2$K$+1) (e.g. Wilks, 2006, Chapter 8).

## 3.3. Distribution of multi-decadal trends

To quantify how unusual an observed trend is for a single predefined window within a climate index series, it is necessary to understand the distribution of such Z values in a series and then estimate the exceedance probability for that magnitude of trend. A common approach in the climate literature is to model internal atmospheric variability as a stationary stochastic process, such as a white noise or known process, and use knowledge of this process to quantify the associated distribution of linear trends (e.g. Santer et al., 2000; Feldstein 2002). Three approaches are discussed below:

- A covariance approach that defines the trend variance directly from Equation 3.1 assuming that *Y* is a pre-defined stochastic process (Section 3.3.1);
- A simulation approach using a large set of simulations of *Y* generated from a pre-defined stochastic process (Section 3.3.2);
- The pre-whitening method from Cochrane & Orcutt (1949) (Section 3.3.3).

The time series in this thesis will be thought of as representing climate indices for atmospheric variables which typically have fairly low levels of dependence between years. For example, the winter mean North Atlantic Oscillation (NAO) has been shown to exhibit some weak dependence between winters with lag-1 year autocorrelation of around 0.1 to 0.2 (e.g. Stephenson et al., 2000; Pozo-Vázquez et al., 2001; Scaife et al., 2014).

### 3.3.1. Covariance approach

The probability distribution of the moving window trend *in any predefined year* is determined by the moving window filter weights and the distribution and dependency 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 aggregate climate indices because of the central limit theorem (e.g. Wilks, 2006).

For index series that have a deterministic part that is a function of time $F(k)$, such as a global warming trend or climate change signal, and a non-deterministic part that can be represented by a first order stationary process such as an ARMA, termed here as random variable $V_k$, then the index series $Y_k$ can be represented as

$$Y_k = V_k + F(k) + \alpha$$

where α is a constant. The trend series, using Equation 3.1 is then

$$Z_i = \frac{1}{h^2}\sum_{j=-K}^{K} jV_{i+j} + \frac{1}{h^2}\sum_{j=-K}^{K} jF(i+j).$$

The expectation of $Z_i$ is given by

$$E[Z_i] = \frac{1}{h^2}\sum_{j=-K}^{K} jE[V_{i+j}] + \frac{1}{h^2}\sum_{j=-K}^{K} jE[F(i+j)].$$

In this case, the random variable $V_k$ is 1st order stationary so the expectation of $V_k$ does not depend on the year $k$ ($E[V_k] = E[V]$), and function $F(k)$ is deterministic so $E[F(k)] = F(k)$, leading to

$$E[Z_i] = \frac{1}{h^2}E[V]\sum_{j=-K}^{K} j + \frac{1}{h^2}\sum_{j=-K}^{K} jF(i+j) = \frac{1}{h^2}\sum_{j=-K}^{K} jF(i+j). \qquad (3.2)$$

By substitution of Equation 3.1, the variance of $Z_i$ is given by

$$\boldsymbol{Var}[Z_i] = \frac{1}{h^4}\sum_{j=-K}^{K}\sum_{k=-K}^{K} jk\boldsymbol{Cov}[Y_{i+j}Y_{i+k}].$$

The covariance part simplifies to

$$\boldsymbol{Cov}[Y_{i+j},Y_{i+k}] = \boldsymbol{Cov}[V_{i+j} + F(i+j),\ V_{i+k} + F(i+k)] = \boldsymbol{Cov}[V_{i+j},\ V_{i+k}]$$

as $F$ is a deterministic function ($F(k)$ is just a fixed value dependent on $k$), so the covariance for the index $Y_k$ is only dependent on the stochastic part $V_k$. The variance $\sigma_z^2$ of the moving window trends (hereafter referred to as the *trend variance*) is then given by

$$\sigma_z^2 = \frac{\sigma_Y^2}{h^4}\sum_{j=-K}^{K}\sum_{k=-K}^{K} jk\rho_{k-j} \qquad (3.3)$$

where $\rho_{k-j}$ is the autocorrelation function and $\sigma_Y^2$ is the variance of the stationary part of the original index process. Woodward and Gray (1993) used sample moment estimates of lagged autocorrelation parameters $\rho_1$, $\rho_2$, … and a similar equation to Equation 3.3. to estimate the trend variance. However, for a stochastic process that has a known parametric form, a precise estimate of the trend variance can be calculated from Equation 3.3.

From Equations 3.2 and 3.3, the trend $Z_i$ in any year $i$ will be Gaussian distributed

$$Z_i \sim N(\bar{Z}_i, \sigma_z^2) \qquad (3.4)$$

and the probability of the trend in year $i$ exceeding a threshold value of $z$ is then simply given by

$$p = Pr(Z_i > z) = 1 - Pr(Z_i \le z) = 1 - \Phi\left(\frac{z - \bar{Z}_i}{\sigma_z}\right) \qquad (3.5)$$

where Φ(.) is the cumulative distribution function of the standard normal. For a given threshold *z*, the probability of exceedance is determined by the trend mean $\bar{Z}_i$ and the trend variance $\sigma_z^2$.

For index series where the deterministic part is a simple linear trend

$$F(k) = \beta k$$

then, from Equation 3.2, the mean of the trend is independent of *k* so

$$E[Z_i] = \frac{1}{h^2}\sum_{j=-K}^{K} j\beta(i+j) = \frac{1}{h^2}\beta i \sum_{j=-K}^{K} j + \frac{1}{h^2}\beta \sum_{j=-K}^{K} j^2 = \beta \tag{3.6}$$

which means Equation 3.4 can be written as

$$Z_i \sim N(\beta, \sigma_z^2) \tag{3.7}$$

The rest of this chapter will consider the specific case where the index series are 1st order stationary, i.e. $F(k) = 0$. Using Equation 3.6 with $\beta = 0$, it can be seen that for this stationary case $\bar{Z} = 0$ and

$$Z_i \sim N(0, \sigma_z^2). \tag{3.8}$$

For a white noise process, the lag autocorrelation $\rho_{k-j} = 1$ if k = j and $\rho_{k-j} = 0$ otherwise, so from Equation 3.3 the trend variance is simply

$$\sigma_z^2 = \frac{\sigma_Y^2}{h^2}. \tag{3.9}$$

For the short-range dependence AR(1) process $Y_{i+1} = \rho Y_i + \varepsilon_{i+1}$ (where $\rho$ is the autocorrelation parameter such that at lag-1 $\rho_1 = \rho$, and $\varepsilon_{i+1}$ are independent Gaussian variables), the lag autocorrelation function at time lag $k - j$ is

$$\rho_{k-j} = \rho^{|k-j|}. \tag{3.10}$$

For a long-range dependence fractional difference (FD) process $(1 - B)^d Y_{i+1} = \varepsilon_{i+1}$ (where *d* is the difference parameter, and the backward shift operator *B* is such that $BY_t = Y_{t-1}$), the lag autocorrelations decay slower than exponential and have the form

$$\rho_{k+1} = \frac{\rho_k(k+d)}{k+1-d} \tag{3.11}$$

for $k = 0, 1, 2, \ldots$ (Hosking, 1981).

The black dashed curve in Figure 3.1 shows the variance of moving window trend series (trend variance) calculated numerically using Equations 3.3 and 3.10 for AR(1) processes over a range of values of the lag-1 autocorrelation parameter $\rho$, all with time series variance $\sigma_Y^2 = 1$. Equation 3.3 captures the expected trend variance dependence on $\rho$, increasing slowly with $\rho$ for small values, but then more rapidly as $\rho$ increases. 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, for a stationary process, consecutive values in the index time series become very similar to one another.

The black dotted curve in Figure 3.1 shows the trend variance for FD processes calculated numerically using Equations 3.3 and 3.11. In this case the difference parameters ($d$) have been matched to the lag-1 year autocorrelation parameters by noting that $\rho_1(d) = d/(1 - d)$ (Hosking, 1981). The trend variance is slightly higher for a FD process than for an AR(1) process for moderate positive values of lag-1 autocorrelation, but for larger lag-1 autocorrelation the FD trend variance increases more slowly, peaking earlier and at a considerably lower trend variance than for the AR(1) process. The moving window trend variance is not unduly sensitive to the choice of stochastic model when lag-1 autocorrelation is low-to-moderate ($\rho < 0.4$).

Figure 3.2 shows how the variance of moving window trends decreases monotonically with respect to the increasing window length for a white noise process. The AR(1) and FD processes initially have larger variance than the white noise process, but tend towards the variance of the white noise process as the window length gets larger. This is a consequence of values at the end of the window having less dependence on values at the start of the window as the window length increases. For larger window lengths, the qualitative relationship of trend variance to lag-1 year autocorrelation $\rho$ remains similar, but the peak in variance shifts towards higher $\rho$ values, seen in Figure 3.3 for window length 81.

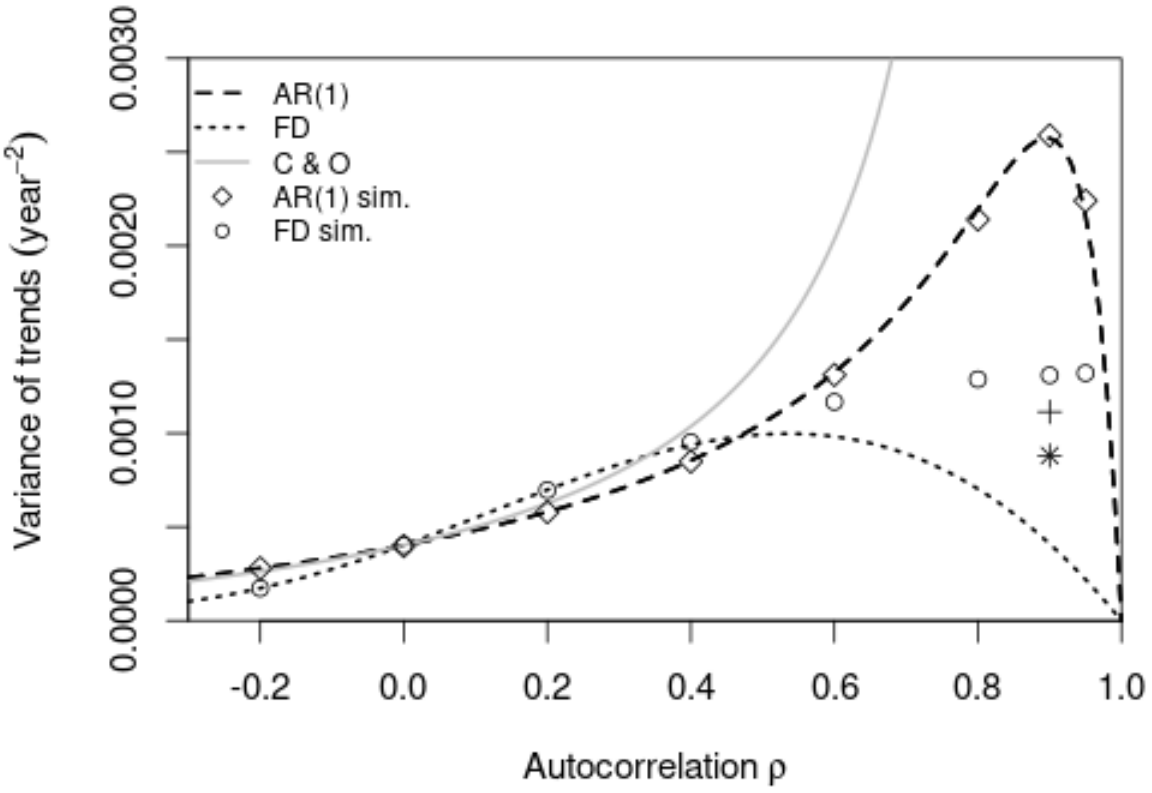


**Figure 3.1. The relationship of the variance of moving window trends with autocorrelation.** The trend variance is shown for the 3 approaches discussed in Section 3.3 with trend window length 31 (K = 15): The covariance approach for AR(1) (black dashed line) and FD (black dotted line) processes; the simulation approach using a selection of AR(1) (diamond) and FD (circle) simulations of length 130 years; the pre-whitening method from Cochrane & Orcutt (1949) (grey solid line). Trends are calculated for a standardised index, so the variance of trends is in units year$^{-2}$, which can be thought of as (sd/year)$^2$ where sd refers to the standard deviation of the index pre-standardisation. The simulation approach for extended FD simulations with trend series length $10^4$ ("+") and $10^5$ ("*") are included for the case $\rho = 0.9$.

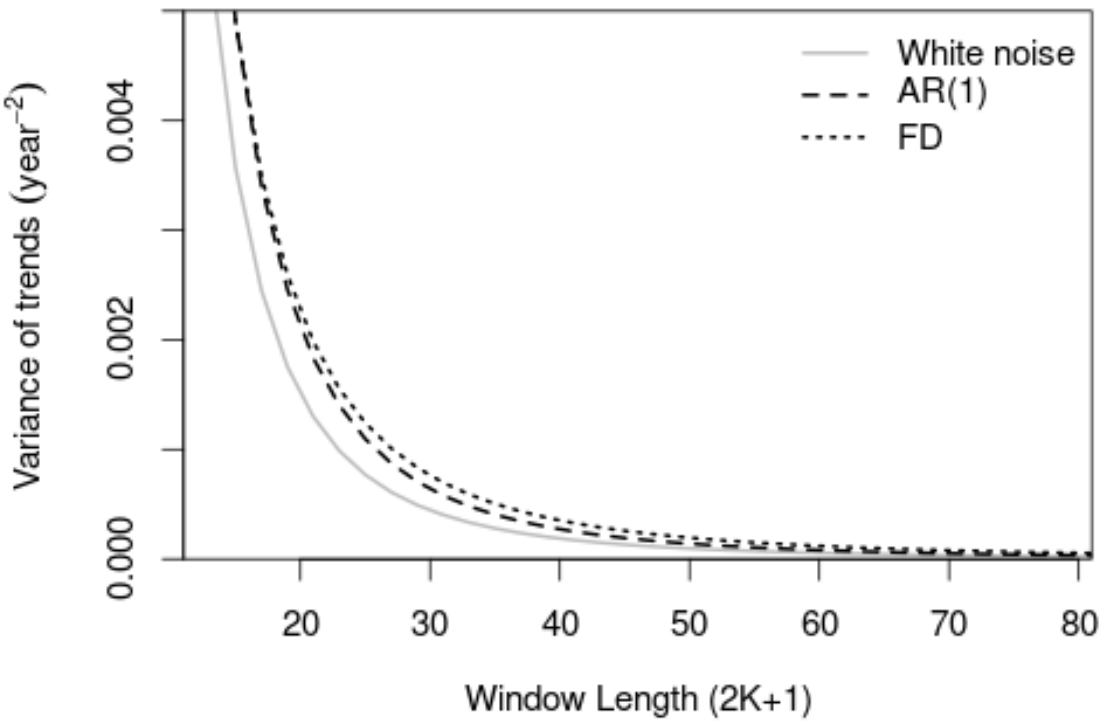


**Figure 3.2. The relationship of the variance of moving window trends with window length.** The trend variance is shown for the covariance approach from Section 3.3.1 for AR(1) (black dashed line) and FD (black dotted line) processes with lag-1 autocorrelation of 0.2 and for a white noise process (grey solid line).

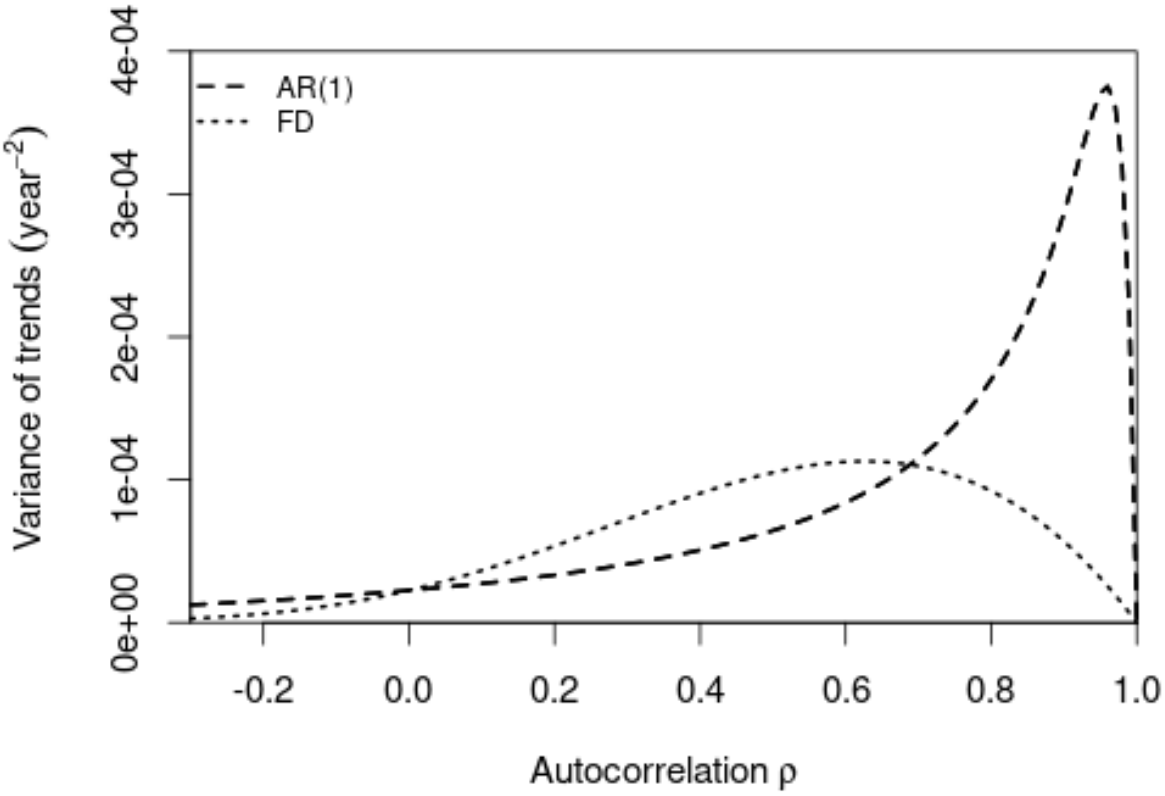


**Figure 3.3. The relationship of the variance of moving window trends with autocorrelation for a longer window length.** The trend variance is shown for the covariance approach from Section 3.3.1 for AR(1) (black dashed line) and FD (black dotted line) processes for a trend window of length 81 ($K$ = 40).

### 3.3.2. Simulation approach

An alternative method to quantify the distribution of moving window trends is to use a stochastic simulation approach (e.g. Feldstein 2002). 5000 random stochastic simulations of length $n$ = 130 have been generated from both short-range dependent AR(1) and long-range dependent FD Gaussian stochastic processes (see Appendix B.1 for code details). From these, the moving window trend series have been computed for window length 31, leading to 5000 trend series of length $m$ = 100, each with a sample trend variance estimated directly from the $m$ values. The length $m$ = 100 was chosen as most observation based climate index time series have length of order 100.

The mean value of the trend variance estimates is shown in Figure 3.1 for stochastic simulations with lag-1 autocorrelation parameter $\rho$ in the set {-0.2,

0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95}. For the AR(1) processes the trend variance estimates from the simulations closely match the numerical solution Equation 3.3, even for high autocorrelation values. For the FD processes (with $d$ parameters matched to the same set of $\rho$ parameters), the simulations agree well with the numerical solution Equation 3.3 for the low to moderate levels of autocorrelation found for most annual atmospheric climate index time series, which generally have a year-to-year autocorrelation of less than 0.4. For time series such as the winter North Atlantic Oscillation, the numerical covariance approach and the simulation approach are fairly interchangeable with the former having the benefit of requiring less computations.

Beyond $\rho \approx 0.4$ (i.e. $d \approx 0.3$) the simulation approach overestimates the variance of moving window trends compared to the covariance approach for the FD process. This is due to the relatively short simulations (trend series length 100) as, for processes with a high level of long-range dependence, the asymptotic behaviour needs a much longer series to be applicable. Extending 10 of the FD simulations to have trend series length $10^4$ and then $10^5$ for $\rho = 0.9$ ($d = 0.474$) brings the simulated variance estimate closer to the numerical solution but there is still a large bias (Figure 3.1 “+” and “*”). However, this high level of long-range dependence is not apparent in atmospheric climate indices such as the NAO so both the simulation and covariance approaches are adequate.

### 3.3.3. Cochrane and Orcutt approach

A common approach used in the climate literature is to calculate the standard error for a trend using linear regression which assumes that the residuals about the trend are independent identically distributed Gaussian random variables (e.g. Santer et al., 2000; Thompson et al., 2015). This standard error estimate is then scaled by an additional factor to account for autocorrelation in the residuals if required, following Cochrane & Orcutt (1949). For an AR(1) process with known autocorrelation function, as the expected trend is zero (Equation 3.6 with $\beta = 0$), the residuals about the expected trend will have the same autocorrelation function as the original process and the standard error for the

trend (referred to here as $\sigma^2_{z\_co}$) is a scaled version of the trend variance for a white noise process (Equation 3.9):

$$\sigma^2_{z\_co} \approx \frac{\sigma^2_Y}{h^2}\frac{(1+\rho)}{(1-\rho)} \tag{3.12}$$

(Santer et al., 2000; Cochrane & Orcutt, 1949; Wilks, 2006; Chandler and Scott, 2011). The $\sigma^2_{z\_co}$ estimate of the trend variance tends to infinity as $\rho$ approaches 1 (Figure 3.1). This method does not behave well for large values of autocorrelation $\rho$, compared to the covariance based approach (Equation 3.3) and the simulation based approach, however for moderate $\rho$ typical of annual atmospheric climate indices ($\rho < 0.4$) there is little difference in these methods.

### 3.3.4. Uncertainty in trend variance estimates

The distribution of trends can be estimated directly from an observation based climate index trend series, but there will be considerable sampling uncertainty in the variance estimate due to the fairly short periods of data available. Alternatively, a stochastic process can be fitted to the observed climate index time series by estimating the autocorrelation function parameters. This section compares the uncertainty arising from estimating the trend variance directly with the uncertainty arising from estimating the autocorrelation function and then computing the expected trend variance using the covariance method.

To investigate the uncertainty in the sample trend variance, the 5000 simulated trend series described in Section 3.3.2 are used to calculate 5000 individual trend variance estimates and form an empirical distribution. Figure 3.4 shows the 95% prediction interval for direct sample estimates of the trend variance (dotted lines) from the AR(1) simulations using lag-1 autocorrelation values $\rho$ in the set {0.0, 0.1, 0.2, 0.3, 0.4} and trend series length $m$ = 100 (from a 130-year index series), using the empirical 2.5th and 97.5th percentiles. The prediction interval demonstrates considerable uncertainty. This is unsurprising given that there are very few independent 31-year windows in a 130-year time series (just four non-overlapping windows). This suggests that it is difficult to obtain a robust estimate of the trend variance directly from a single observed climate index time series.

To investigate the uncertainty arising from the estimate of the autocorrelation function, the distribution of the lag-1 year autocorrelation parameter can be assumed to be Gaussian such that the 95% prediction interval for the lag-1 year autocorrelation parameter $\rho$ for an index series of length $n$ is given by the Bartlett formula

$$\rho \pm 1.96 \sqrt{\frac{(1-\rho^2)}{n}} \tag{3.13}$$

(Bartlett, 1946) where 1.96 is the threshold such that the probability $P(|X| < 1.96) = 0.95$ for the standard Gaussian variable $X$. Alternatively, the distribution can be approximated by fitting an AR(1) distribution to each of the 5000 simulated AR(1) trend series (Section 3.3.2) and using the individual sample $\rho$ estimates to form an empirical distribution.

Figure 3.4 shows how the uncertainty in the autocorrelation parameter effects the estimate of the trend variance by using Equation 3.3 with $\rho$ equal to the 2.5th and 97.5th percentiles of the $\rho$ distribution using the Bartlett formula (Equation 3.13) (black dashed lines). This uncertainty estimate is very similar if the empirical 2.5th and 97.5th percentiles of sample $\rho$ estimates are used (Figure 3.4, grey dashed lines). These methods give estimates of the 95% prediction interval for the trend variance based on the uncertainty in the autocorrelation parameter estimates, which are narrower than the prediction interval based on the sample trend variance estimates. This is likely due to the large amount of autocorrelation in the *trend* series which decreases the effective sample size of the series (only four independent 31-year windows in the series) and increases the level of uncertainty in the variance parameter estimate. Using the AR(1) and FD simulated trend series (Section 3.3.2) to make estimates of the lag-1 autocorrelation for each individual *trend* series, Figure 3.5 shows that the average lag-1 autocorrelation ($\rho_z$) for the moving window trend series Z is large (quickly reaching above 0.9) and not overly sensitive to the lag-1 autocorrelation $\rho$ of the original index Y.

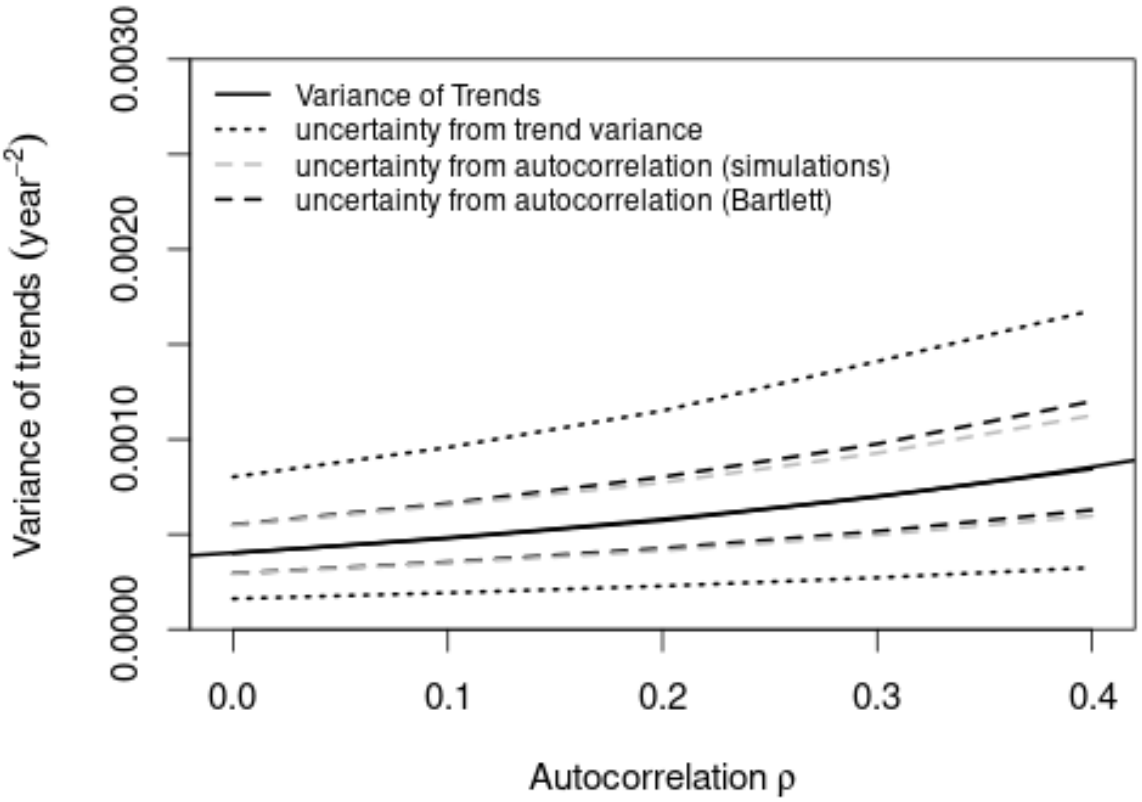


**Figure 3.4. The uncertainty in parameter estimates for the distribution of moving window trends.** The trend variance is shown for the covariance approach from Section 3.3.1 for AR(1) (black solid line) processes for a trend window of length 31 ($K$ = 15) with methods to show the uncertainty in parameter estimates due to the finite length of trend series (Section 3.3.4). The uncertainty due to the sample trend variance estimate (black dotted line) and the uncertainty due to the autocorrelation parameter $\rho$ estimate using simulations (grey dashed line) and the Bartlett formula (black dashed line).

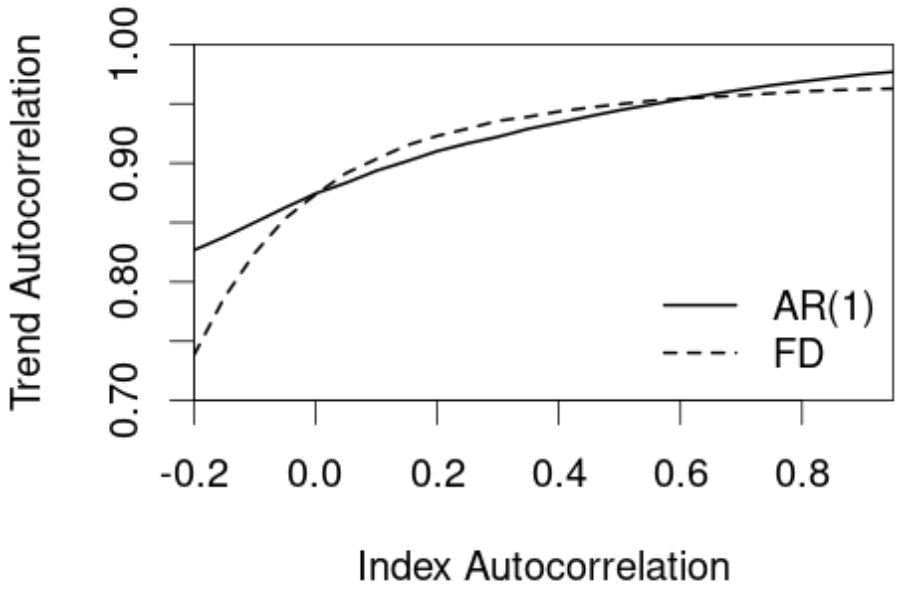


**Figure 3.5. Autocorrelation in moving window trend series.** The average lag-1 autocorrelation for moving window trends is shown for the AR(1) and FD simulations (window length 31), plotted against the lag-1 autocorrelation of the underlying index time series.

### 3.3.5. Sensitivity of exceedance probabilities to trend distribution

To quantify how unusual an observed trend is in a climate index series, an estimate of $\Pr(Z_i \geq z)$ is needed from Equation 3.4, i.e. the probability that the trend centred on a specific year $i$ exceeds a threshold $z$. Section 3.3.4 has shown that it is difficult to estimate the distribution of trends in a series directly from a short trend series due to the high level of autocorrelation in moving trend series and noting that for high exceedance thresholds $z$ such trends may have rarely been observed. Instead, it is useful to consider the distribution and autocorrelation structure of the original climate index series, and then use the covariance approach (Section 3.3.1) to estimate the trend distribution.

Exceedance probabilities $\Pr(Z_i > z) = 1 - \Phi\left(\frac{z}{\sigma_z}\right)$ for a white noise trend series ($\sigma_Y$ = 1, window length 31, $\bar{Z}_i$= 0) are shown in Figure 3.6 for a range of thresholds $z$ which might be considered relatively extreme, given that the Intergovernmental Panel on Climate Change (IPCC) definition of *very unlikely* is

an event that occurs with probability less than 0.1 (IPCC likelihood scale, Appendix B.2). Equation 3.3 is used to calculate $\sigma_z$, and the uncertainties due to stochastic model fitting to finite time series (Section 3.3.4) are shown as grey shading using the 2.5th and 97.5th percentiles from the Bartlett formula with $\rho$ = 0.0 (Equation 3.13). For an AR(1) process with a typical weak red noise autocorrelation parameter of 0.2, the exceedance probabilities are higher than for the white noise process and outside of the 95% uncertainty range, while the exceedance probabilities for a matching FD process ($d$ = 0.167) are higher still.

Although the trend variance shows little sensitivity to the choice of stochastic model when weak year-to-year dependence is assumed (Figure 3.1), the effect of wrongly assuming no year-to-year dependence, as for the white noise process, could substantially alter conclusions based on exceedance probabilities. For example, for a white noise process the trend threshold $z$ = 0.0330 sd/year equates to a 95-percentile event, i.e. the exceedance probability is $1 - \Phi\left(\frac{0.0330}{\sigma_z}\right)$ = 0.05. However, the exceedance probability for this threshold using the AR(1) model with $\rho$ = 0.2 is 0.0852, almost double that of the white noise process, and using the FD model ($d$ = 0.167) the probability is 0.106, which is over double that of the white noise process and changes the IPCC description from *very unlikely* to *unlikely*.

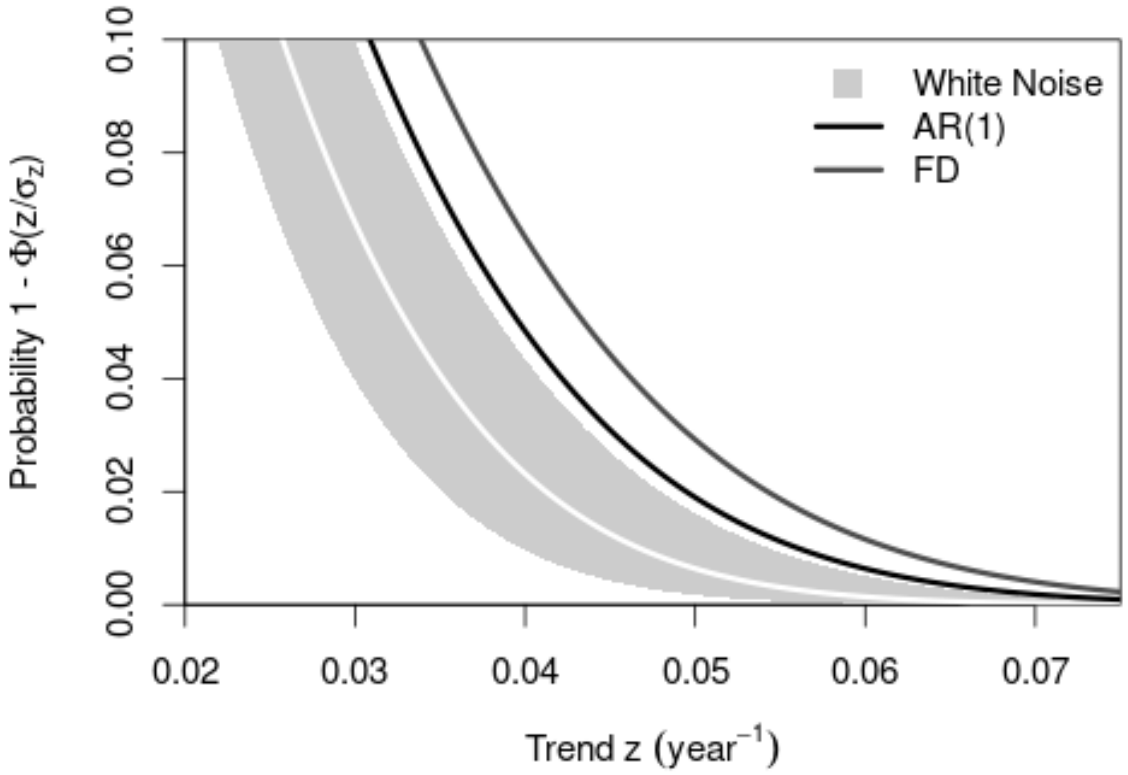


**Figure 3.6. Example exceedance probabilities for multi-decadal trends.** The exceedance probability $p = \Pr(Z_i \geq z)$ is shown for relatively extreme trend thresholds $z$ for a white noise process (white solid line) with grey shading representing the uncertainty due to stochastic model fitting (Section 3.3.5). Probability curves are also shown for an AR(1) (black solid line) and FD (grey solid line) index time series with lag-1 autocorrelation of $\rho = 0.2$.

## 3.4. Distribution of 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 $q(z)$:

$$q(z) = Pr(max\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \geq z) \quad (3.14)$$

where $n$ is the length of the original index series and the window length is 2K+1.

### 3.4.1. Independent trend values

Extreme Value Theory for block maxima can be used to explain how the asymptotic distribution of the maximum trend relates to properties of an index series if the series consists of *independent* random variables. This is not the case for moving window trend series, as the application of the moving trend

filter leads to a high level of autocorrelation in the trend series (Figure 3.5) which increases further if there is autocorrelation in the underlying series.

If the trend series consisted of independent Gaussian variables $Z_i^* \sim N(0, \sigma_z^2)$, the maximum value distribution would asymptotically converge to a Gumbel distribution, a special case of the Generalised Extreme Value (GEV) distribution with a shape parameter of zero, location $\mu^*$ and scale $\sigma^*$ such that

$$q = 1 - exp\left\{-exp\left(-\frac{(z-\mu^*)}{\sigma^*}\right)\right\} \tag{3.15}$$

(Kinnison, 1985 chapter 7; Coles et al., 2001 chapter 3; Wilks, 2006, chapter 4). It is also known that Gaussian processes with short-range dependence such as ARMA processes have extremes that are asymptotically independent (Davis et al., 2013), and so will also lead to GEV distributions with a shape parameter of zero.

For independent $Z_i$ values, the Gumbel parameters are linear functions of $\sigma_z$ and vary with the block length, which in this case is the length of the trend series $m = n - 2K$. It has been shown using an approximation of the Gaussian probability density function that as $m$ tends to infinity

$$\mu_m^* = \sigma_z\left\{\sqrt{2\ln(m)} - \frac{ln(ln(m)) + ln(4\pi)}{2\sqrt{2\,ln(m)}}\right\} \tag{3.16}$$

$$\sigma_m^* = \sigma_z \frac{1}{\sqrt{2\,ln(m)}} \tag{3.17}$$

(Kinnison, 1985, chapter 7). These location and scale parameters are just scaled versions of $\sigma_z$, as shown in Figure 3.7 for maximum trends in a trend series of length $m = 100$ where $\sigma_z$ has been calculated for AR(1) and FD processes using Equation 3.3. For the weak autocorrelation values typical of yearly climate indices, e.g. around 0.0 to 0.2 for the NAO, these parameters are not overly sensitive to the choice of stochastic process as $\sigma_z$ is similar for AR(1) and FD models.

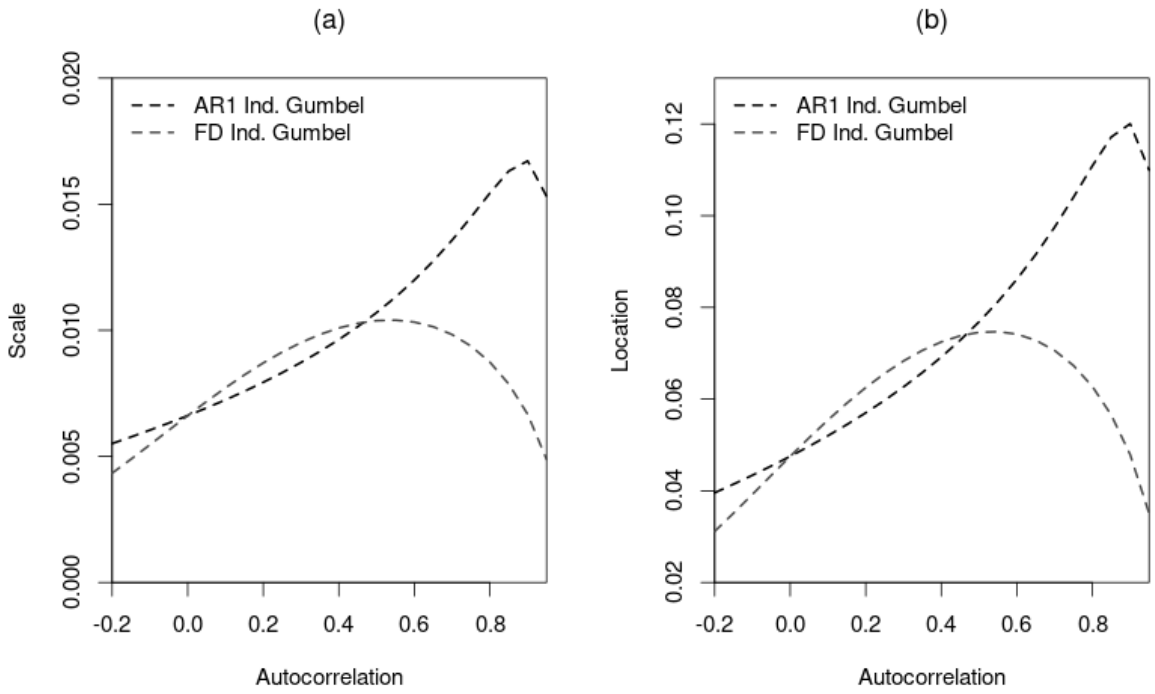


**Figure 3.7. The relationship of maximum trend distribution parameters with the autocorrelation of the original index for independent trend values.** (a) Scale and (b) location parameters of the Gumbel distribution of extreme trends, calculated from Equation 3.16 and 3.18 for an independent trend series, vs the lag-1 autocorrelation parameter of the original stochastic process. The window length is 31 and the trend series length is 100. Black dashed lines are for AR(1) processes with the corresponding autocorrelation parameter. Grey dashed lines are for FD processes with difference parameters matched to the displayed autocorrelation values.

## 3.4.2. Dependent trend values

For moving window trend series, there is a high level of autocorrelation due to the moving trend filter (Figure 3.5), so the sample of trend values cannot be treated as independent. For dependent trend values the Gumbel distribution of the block maxima has the same scale parameter as the independent case but a reduced location parameter:

$$\sigma_m = \sigma_m^* \tag{3.18}$$

$$\mu_m = \mu_m^* + \sigma_m^* \ln(\theta) \tag{3.19}$$

where $\theta$ is the extremal index in the range (0, 1] quantifying the level of dependence with $\theta = 1$ when $Z_i$ are independent (Coles et al., 2001).

To investigate the behaviour in this dependent case, stochastic simulations are needed to go beyond the theory for the independent case as it is not possible to derive simple analytic solutions for $\theta$. Simulations of maximum moving window trends can be obtained from the stochastic simulations described in Section 3.3.2. For each of the 5000 simulations from a specified stochastic process, the maximum moving window trend is identified and a GEV distribution is then fitted to these 5000 values (see Appendix B.4 for code details).

Figure 3.8 shows the GEV parameter estimates (see Appendix B.4 for code details) for maximum trends from the set of AR(1) processes. The best GEV distribution fit has a negative shape parameter, i.e. the distribution is a Weibull rather than a Gumbel. The relationship of the fitted location and scale parameters to the lag-1 autocorrelation $\rho$ are similar to the independent case, as expected from the asymptotic theory, with the peak location and scale values occurring for the same $\rho$ values. The location parameters for the dependent case are reduced compared to the independent case as expected from Equation 3.19, however, the fitted scale parameters are considerably larger than for the independent case so are not consistent with Equation 3.18. Fitting a Gumbel distribution to the data (i.e. setting the shape parameter to be zero) leads to very similar scale and location parameters compared to those from the best GEV distribution fit. If the fitted Gumbel distribution is additionally set to have the scale parameter from Equation 3.17 the location parameter is reduced marginally further. The asymptotic theory is useful for understanding the relationship of the GEV parameters to the distribution of moving window trends, but trend series of length $m$ = 100 are too short for this theory to be applicable.

Similar results are found for simulated extreme trends using FD processes (Figure 3.9) - the fitted GEV has a negative shape parameter with reduced location parameter and increased scale parameter compared to the independent case. However, for larger levels of autocorrelation, the fitted location and scale parameters continue to increase with autocorrelation, contrary to the independent case which peaks for moderate levels of autocorrelation (around $\rho$ = 0.4 in this case) and then decreases. This relates to Section 3.3.2 which found that moving window trends simulated from FD

processes lead to a trend variance estimate that diverges from the numerical solution for $\rho > 0.4$. If the fitted Gumbel distributions are set to have the independent scale parameters from Equation 3.17 the fitted location parameters more closely resemble the behaviour of the independent case but with much lower values (Figure 3.9c, dotted line).

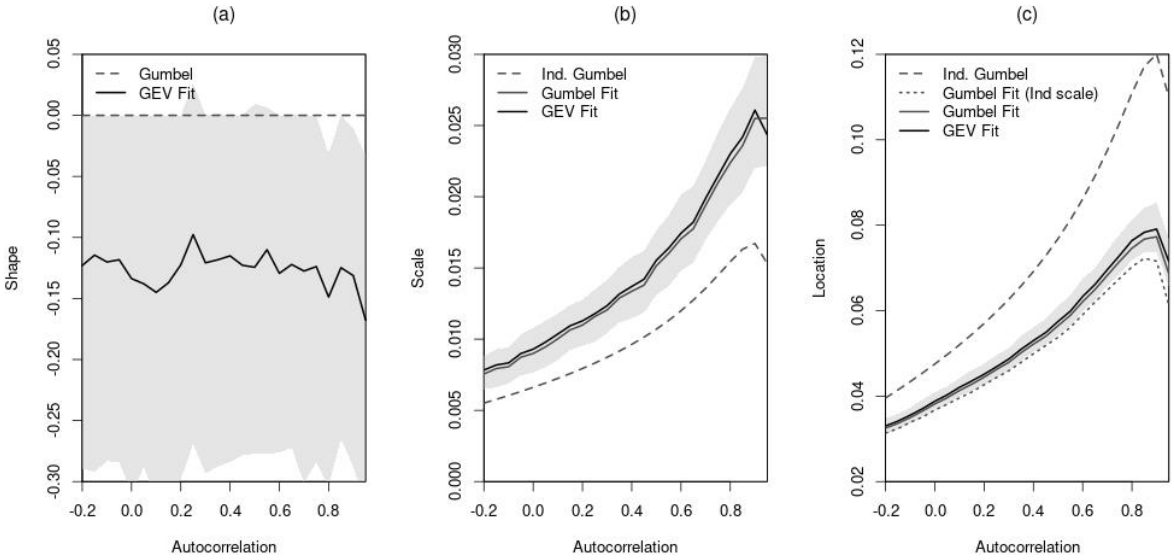


**Figure 3.8. The relationship of maximum trend distribution parameters with the autocorrelation of the original AR(1) process.** (a) Shape, (b) scale and (c) location parameters of the GEV distributions fitted to simulated maximum trends vs the lag-1 autocorrelation of the original AR(1) time series (black solid line) for window length 31 and trend series length $m$ = 100 (Section 3.4.2). Scale and location parameters for a fitted Gumbel distribution are shown (grey solid line), and the location parameter for a fitted Gumbel distribution using the scale parameter for the independent case (grey dotted line). Black dashed lines are Gumbel parameters for the fully independent case (Section 3.4.1).

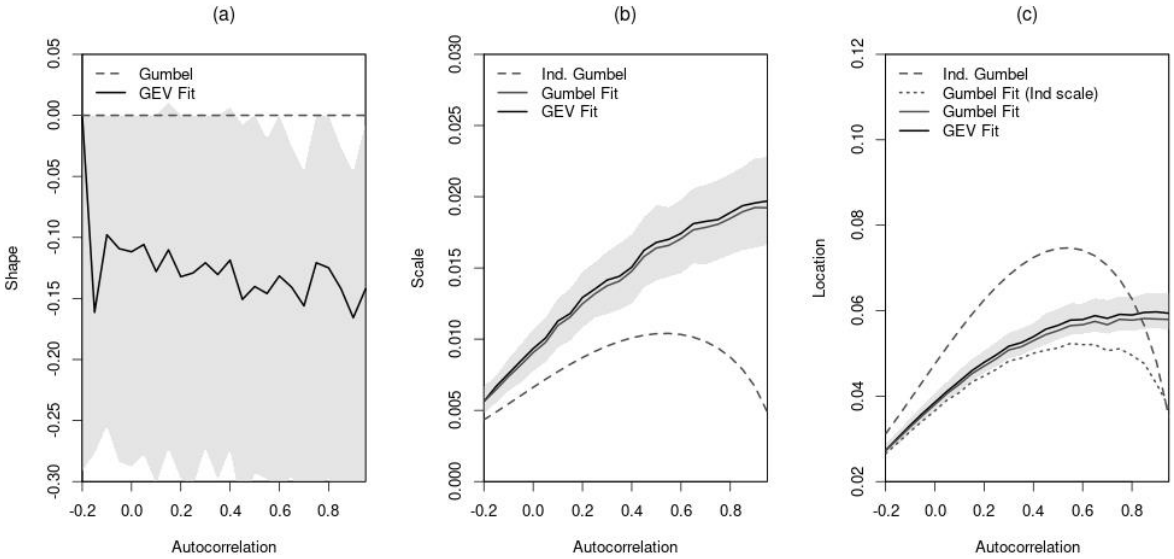


**Figure 3.9. The relationship of maximum trend distribution parameters with the autocorrelation of the original FD process.** (a) Shape, (b) scale and (c) location parameters of the GEV distributions fitted to simulated maximum trends vs the lag-1 autocorrelation of the original FD time series (black solid line) for window length 31 and trend series length *m* = 100 (Section 3.4.2). Scale and location parameters for a fitted Gumbel distribution are shown (grey solid line), and the location parameter for a fitted Gumbel distribution using the scale parameter for the independent case (grey dotted line). Black dashed lines are Gumbel parameters for the fully independent case (Section 3.4.1).

### 3.4.3. Uncertainty in extreme value distribution parameter estimates

In most cases it will not be possible to estimate the distribution of extreme multi-decadal trends directly from an observation based climate index time series as the time series are too short. More often a modelling approach will be needed, for example using the simulated maximum trend values from the stochastic simulations described in Section 3.4.2 or from General Circulation Model (GCM) simulations of the climate index of interest. For the stochastic simulation approach, a massive set of simulations can easily be generated such that the uncertainty in GEV parameter estimates should be fairly low. For the GCM approach, such model runs are prohibitively computer intensive such that the largest sets of simulations are generally of order 100 and come from

coordinated modelling projects such as the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016).

To test the uncertainty in GEV parameter estimates given a sample size of 100 simulations, a bootstrap resampling method was applied to the stochastic simulations described in Section 3.4.2. For a specified stochastic process, the 5000 simulated maximum trends were used to repeatedly pick random subsets of 100 values (without replacement) and fit a GEV distribution in each case (1000 random subsets in total). The 95% prediction interval was then determined using the empirical 2.5th and 97.5th percentiles for each of the GEV parameters. This interval is shown in Figures 3.8 and 3.9 as grey shading for AR(1) and FD simulations for each of the GEV parameters. There is a large amount of uncertainty in the shape parameter estimate, but it is generally negative implying a Weibull distribution. The 95% prediction interval is fairly narrow for the scale and location parameters compared to the level of variation related to the autocorrelation of the underlying index time series. The location and scale parameters for the independent trend window case (Equation 3.16 and 3.17) are far outside of the prediction intervals for the dependent case, which means that the independent case is too different to the dependent case to be helpful in estimating exceedance probabilities. This is related to the high level of autocorrelation in the moving window trend series (Figure 3.5), even when there is no autocorrelation in the underlying index.

### 3.4.4. Sensitivity of exceedance probabilities to extreme trend distribution

To quantify how unusual the maximum observed trend is in a climate index series, an estimate of the exceedance probability $q(z) = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ is needed (Equation 3.14), i.e. the probability that the maximum trend in a series exceeds a threshold $z$. Figure 3.10a shows exceedance probabilities $q(z)$ for a range of thresholds $z$ where the underlying index series $Y$ has been generated from a white noise process. Fitting a GEV distribution to the sample of maximum trends (Figure 3.10a white solid line) gives very similar probabilities to the empirical probabilities based on sample rank values (black dot-dash line), as expected.

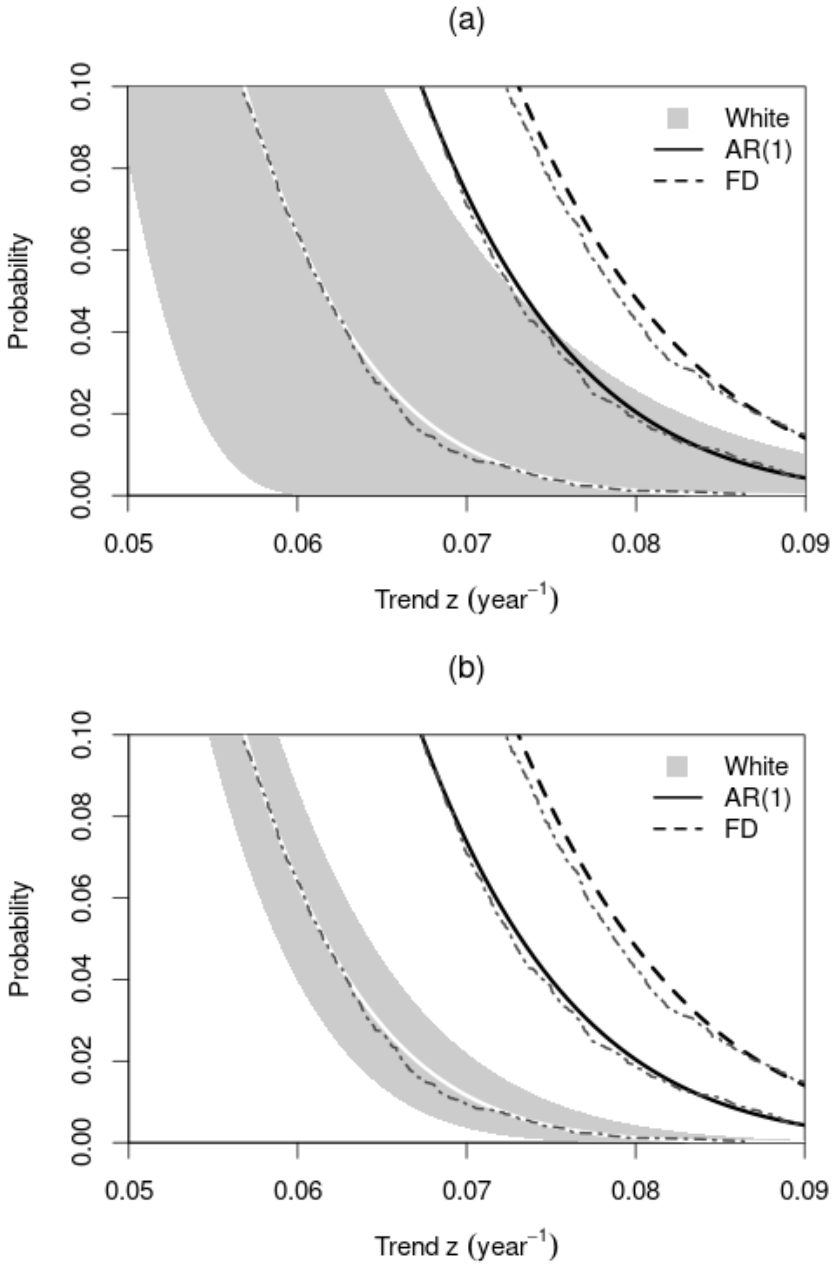


**Figure 3.10. Example exceedance probabilities for extreme trends.** The exceedance probability $q$ is shown relative to the trend threshold $z$ for GEV distribution fits to stochastic simulations (window length 31): white noise (white solid line in grey shading), AR(1) ($\rho$ = 0.2, black solid line) and FD ($d$ = 0.167, black dashed line) processes. For each process, the empirical probabilities are shown as dark-grey dot-dash lines close to the GEV distribution fit. The 95% prediction interval (grey shading) represents the uncertainty due to finite sample sizes of (a) 100 and (b) 1000 simulations (Section 3.4.4).

To explore the sensitivity of exceedance probabilities to the uncertainty in the GEV parameter estimates due to relatively small sample size (100 maximum trend values), the bootstrap resampling method from Section 3.4.3 is used to get a probability estimate from the fitted GEV distributions for each of the 1000 subsamples of the maximum trend data. The 95% prediction interval for these probabilities is defined using the empirical 2.5th and 97.5th percentiles from these sample GEV distribution fits (grey shading in Figure 3.10a), and shows that this estimate of the sampling uncertainty for fitted GEV probabilities is almost as large as the uncertainty due to the choice of stochastic process for the original time series (white noise, AR(1) or FD). Thus, even for seemingly large GCM ensembles such as from CMIP6, there will be considerable uncertainty in the estimates of the distribution of extreme trends.

Despite the apparent larger uncertainty in the shape parameter than the other GEV parameters (Figures 3.8 and 3.9), the uncertainty in the exceedance probabilities is not all coming from the shape parameter. If the bootstrap resampling method is repeated with the shape parameter fixed to be that from the full 5000 simulations, the 95% prediction interval is almost unchanged.

To explore the sensitivity of exceedance probabilities to the autocorrelation in the original index time series, Figure 3.10a shows exceedance probabilities for AR(1) and FD index time series with lag-1 autocorrelation of $\rho = 0.2$ and difference parameter $d = 0.167$, with similar results for both the GEV fit (solid black and grey lines) and empirical probabilities based on sample rank values (black dot-dash lines). For these cases with weak autocorrelation such as is seen in the atmosphere, exceedances probabilities are substantially increased and outside of the 95% prediction interval for white noise processes except for very large trend values.

Repeating this bootstrap analysis for a larger sample size of 1000 maximum trend values greatly reduces the levels of uncertainty (Figure 3.10b), showing that fairly robust estimates of the extreme trend distribution for stochastic model time series should be possible provided a large number of trend simulations are generated. In this case the uncertainty due to the fitting of a stochastic process to the original time series becomes more important.

## 3.5. Summary

In the context of trends and extreme trends in stationary stochastic Gaussian time series, this chapter has proposed approaches to:

- Define the distribution of multi-decadal trends and extreme trends.
- Assess the uncertainty in distribution parameter estimates.
- Assess the influence of these uncertainties on trend exceedance probabilities.

These approaches are also applicable to time series that consist of a stationary part and a deterministic signal such as a long-term linear trend.

For the distribution of moving window trends, a numerical method is proposed that calculates the trend variance based on properties of the underlying index time series (autocorrelation function, variance, and moving window length). For the distribution of block maxima trends, a numerical method to describe the relevant GEV distribution (Gumbel) might be achievable if the trend series was long enough and the trend values were independent. Unfortunately, such asymptotic behaviour is not achieved for typical observed climate variable trend series due to the shortness of the series (block length) and the high level of serial dependence created by the moving window trend filter. Instead, stochastic simulations are used here to estimate the distribution of maximum trends, either empirically or by fitting a GEV distribution which in this case is shown to be a Weibull distribution. By considering stationary stochastic models, it has been shown that the chance of extreme trends is determined by the variance of the moving window trend process, which generally increases when there is more autocorrelation in the index series. If the underlying index series has a deterministic part such as a positive long-term linear trend, then the distribution of moving window trends is shifted away from zero to instead be centred around this long-term trend and the probability of extreme positive trends will also increase.

Later chapters will apply these approaches to climate index time series from observations and multi-model ensembles of GCM simulations. For these applications, it has been shown that the length of time series and the number of

GCM simulations available are important considerations when estimating the distribution parameters for trend and extreme trend distributions. Given the relatively short length of most observation based climate index time series, there will be large uncertainty in an estimate of the variance of multi-decadal trends obtained directly from the moving window trend series, shown in Section 3.3.4 using stochastic simulations. However, if the time series can be well modelled as a stationary Gaussian stochastic process, then more robust estimates of the trend and extreme trend distributions may be obtained based on the behaviour of the fitted stochastic process. In this case one uncertainty to consider is that of the stochastic process distribution parameters when estimated from observed time series. This uncertainty has been addressed (in Sections 3.3.4 and 3.4.3) by estimating a plausible range of autocorrelation parameters that could describe the observed NAO series (using the 95% prediction interval assuming a Gaussian distribution) and computing exceedance probabilities based on the end points of the interval. An alternative more Bayesian like method would be to produce a large sample of plausible parameter estimates drawn from the Gaussian distribution for autocorrelation, compute exceedance probabilities based on each and then compute the 95% prediction interval. These two methods should give similar results, so the former simpler method was chosen. Finally, there is also uncertainty related to the choice of stochastic model. For this reason, both the AR(1) and FD models will be assessed in later chapters to see how sensitive the analysis is to the stochastic model choice. If the observed NAO index series is not stationary, then this methodology will not lead to accurate probabilities, thus the validity of this assumption is tested in Chapter 4.

# 4. Extreme multi-decadal historical trends in the North Atlantic Oscillation

## 4.1. Introduction

This chapter applies methods from Chapter 3 to the winter North Atlantic Oscillation (NAO) index. The NAO definition and observation/reanalysis data sets are outlined in Section 4.2. Stochastic models are fitted to the NAO index in Section 4.3, and are used to model the distribution of multi-decadal trends and extreme trends in Sections 4.4 and 4.5. The goal is to identify block maxima trends in the observed NAO index time series and quantify the probability of seeing such trends under present day climate conditions. This analysis assumes that the NAO multi-decadal variability can be reasonably represented by a stationary stochastic process that represents the aggregation of atmospheric noise (e.g. Wunsch, 1999; Feldstein, 2000) with some year-to-year memory. Stationary Gaussian stochastic processes are fitted to the NAO using a first order auto-regressive AR(1) process and a fractional difference (FD) process to capture the short- and long- term memory of the NAO.

## 4.2. The North Atlantic Oscillation

### 4.2.1 Definition of the NAO

The winter NAO index is defined using Mean Sea Level Pressure (MSLP) averaged over the winter season December to February (DJF), where January is used to define the year. Analysis is repeated for three observation datasets:

- HadSLP2r gridded observation dataset, 1851-2020, 171 years (Allan and Ansell, 2006);
- 20th century reanalysis data V2 (“C20C”), 1872-2012, 141 years (Compo et al., 2011);
- NAO reconstructed index (“L99”), 1659-2001, 343 years, based on a mix of station data (pressure, temperature and precipitation) and proxy data fitted to the standardised NAO index outlined below (Luterbacher et al., 1999, 2001).

Using the gridded MSLP datasets, a *standardised* NAO index is defined to match the definition of the L99 NAO reconstructed index which is based on Hurrell et al. (1995):

$$N(t) = \frac{(P(t) - \bar{P})}{s_P} - \frac{(R(t) - \bar{R})}{s_R}$$

$$NAO(t) = \frac{(N(t) - \bar{N})}{s_N} \qquad (4.1)$$

*NAO*(*t*) at time *t* is the standardised difference in DJF MSLP at the two main nodes of NAO variability, that is Ponta Delgada, Azores at 37.7N, 25.7W (*P*) minus Reykjavík, Iceland at 65.0N, 22.8W (*R*), using the nearest grid boxes. $\bar{P}$ and $\bar{R}$ are the sample means and $s_P$ and $s_R$ are the sample standard deviations of *P* and *R* MSLP index series respectively, calculated over the full length of the MSLP datasets i.e. different periods for different datasets. $\bar{N}$ and $s_N$ are the sample mean and standard deviation of *N* calculated over climate periods specified in the analysis sections below, generally 1862-2005.

The time series for each MSLP node used in the NAO definition (Equation 4.1) are first standardised individually so that neither node dominates the variability. This definition is based on the index used in Hurrell (1995), except that the Azores is used as the southern node instead of Lisbon, and means that the index is comparable to the leading principal component of winter MSLP variability in the North Atlantic region. The standardisation of the NAO index series makes it easier to compare all observation time series simultaneously with the same stochastic processes (variance equal to one), without having to consider differences in the interannual variance of the individual MSLP datasets. In particular, it enables the longer gridded observation dataset HadSLP2r to be used, despite its known underestimation of MSLP variability (e.g. Semenov et al., 2008). It is also useful in Chapter 5 when comparing NAO index time series from observation datasets to those from Coupled General Circulation Model (CGCM) experiments as the sensitivity of trend distributions to the difference in stochastic behaviour (e.g. year-to-year memory) can be assessed separately to the sensitivity to differences in overall MSLP variability. For these reasons, the standardised NAO index (Equation 4.1) is used for the majority of this thesis, and the sensitivity of results to the NAO index definition are discussed later in Section 5.4.1.

Figure 4.1 shows the Empirical Orthogonal Function (EOF) patterns for the North Atlantic region using HadSLP2r and C20C MSLP datasets (see Appendix B.5 for code details). The North Atlantic region is defined here as 20-90N and 90W-40E, as in Hurrell (1995) but extended to 90N as in Deser et al. (2000) and Ambaum et al. (2001). The full available period of data is used for each dataset to get the most robust estimate of the pattern. The EOF patterns are consistent with one another, with just small differences as would be expected due to the different horizontal resolutions and different time periods of the two datasets. In both cases, the main nodes of variability are slightly shifted compared to those defining the *standardised* NAO index (Azores 37.7N, 25.7W and Iceland 65.0N, 22.8W), particularly the southern node. Differences may just be due to the station based NAO index definition being constrained by the availability of land based station data with long historical records (Portugal, Azores and Iceland). However, the EOF pattern is also sensitive to the temporal resolution of the MSLP data used, shown in Figure 4.2 using an alternative MSLP observation dataset, EMSLP (Ansell et al., 2006), which has daily data for 1850-2003 over the European-North Atlantic region. Calculating DJF means from this MSLP data and then calculating the first EOF pattern (Figure 4.2a) leads to a similar pattern to those in Figure 4.1, whereas using the daily MSLP data directly (days within season DJF) leads to an EOF pattern (Figure 4.2b) with the southern node closer to the Azores and the northern node shifted east of Iceland. Despite the sensitivity of the EOF pattern to temporal resolution, the Azores and Iceland node positions used in Equation 4.1 remain close to the main nodes of variability in the NAO patterns so should be adequate for capturing the NAO variability.

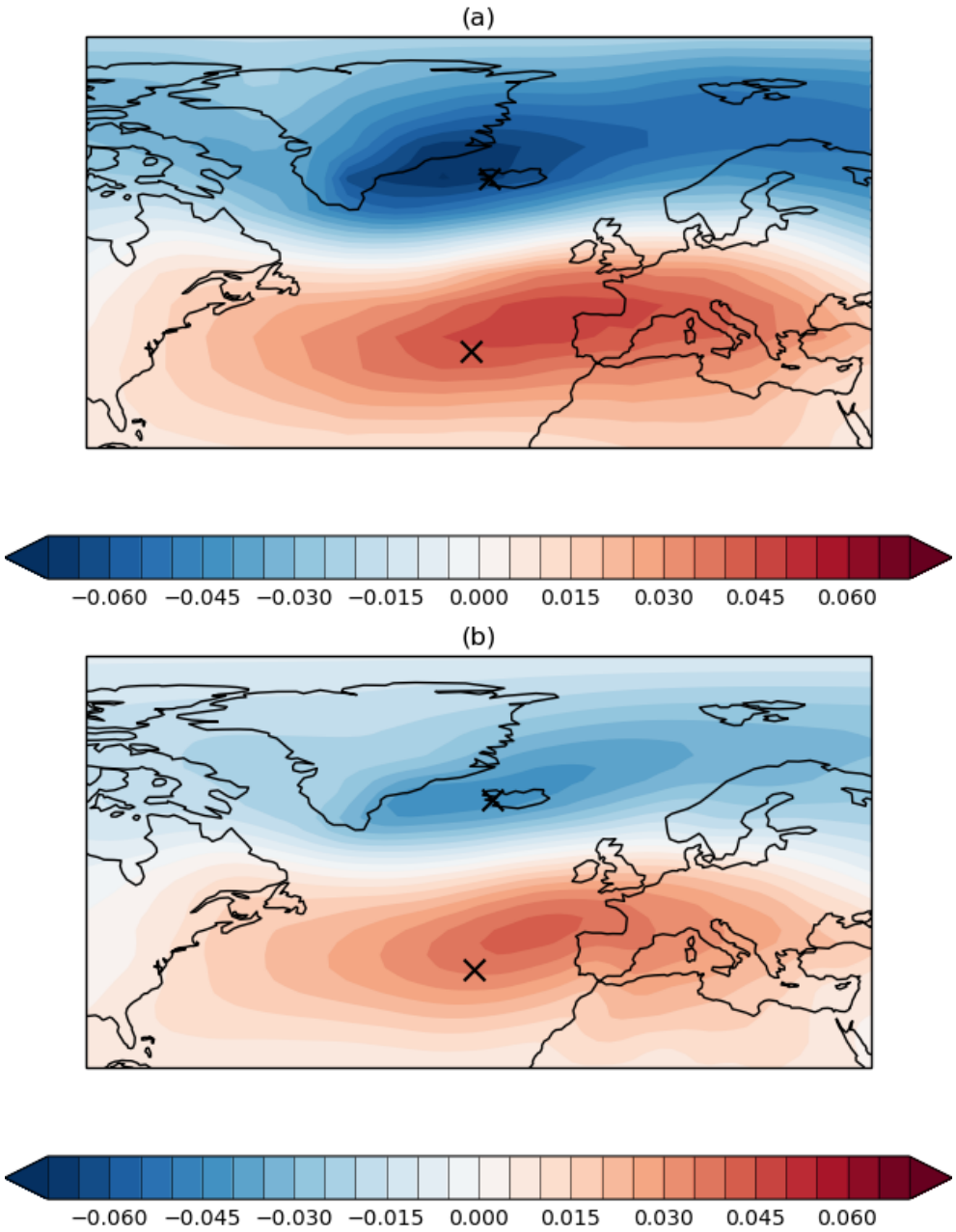


**Figure 4.1. First EOF pattern for MSLP DJF seasonal mean.** The full available period of (a) HadSLP2r and (b) C20C datasets are used, with the standardised NAO index nodes marked as black crosses (Iceland and Azores).

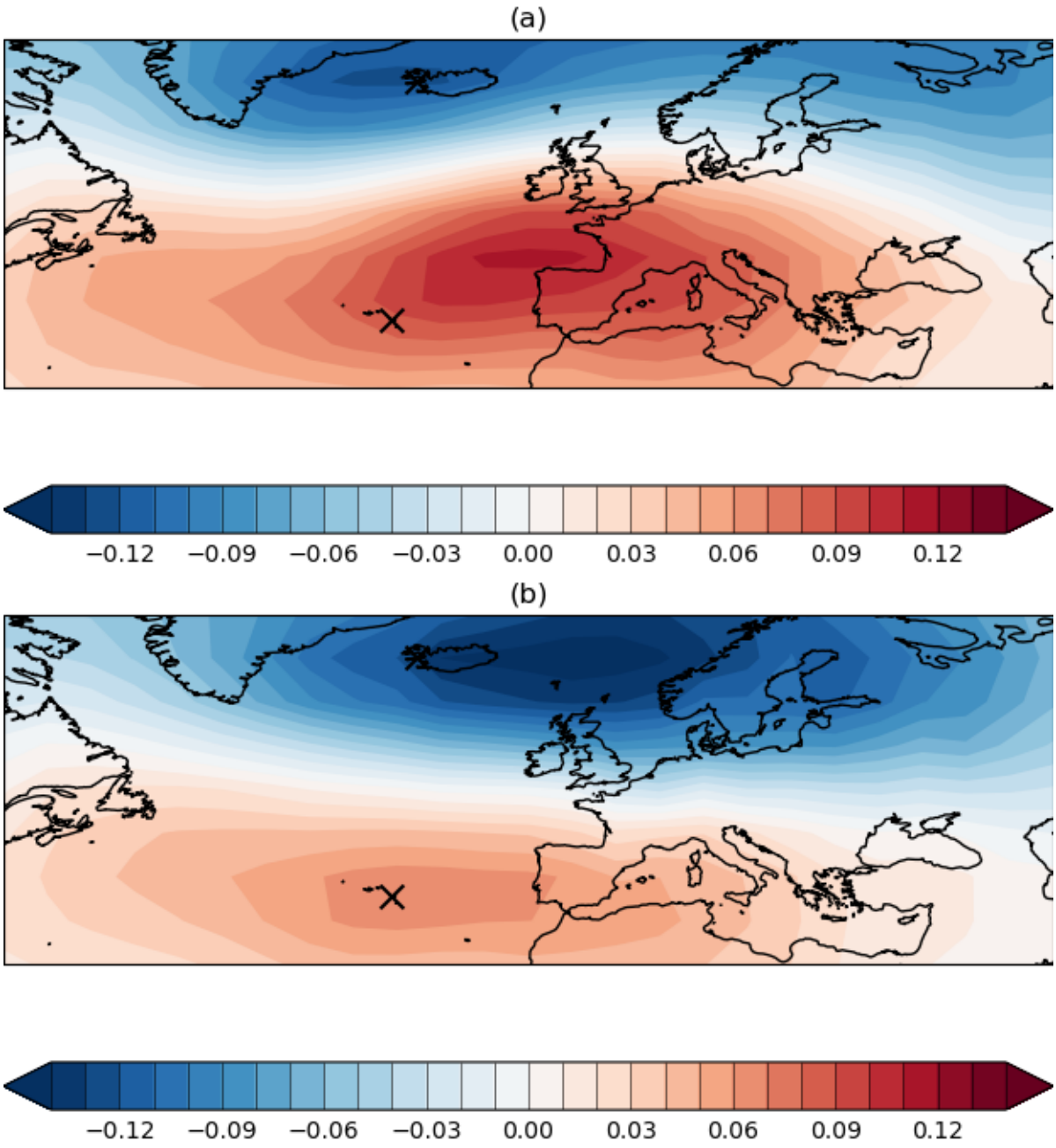


**Figure 4.2. First EOF pattern for MSLP DJF seasonal and daily mean data.** The full period of the EMSLP dataset is used to calculate EOFs on (a) seasonal and (b) daily mean data, with the standardised NAO index nodes marked as black crosses (Iceland and Azores).

### 4.2.2 NAO variability and multi-decadal trends

Figure 4.3a shows the NAO index for all three NAO datasets in units of standard deviation (sd) as the index has been standardised over the common period of 1872-2001. All three datasets show consistent timing of peaks and troughs in the NAO index over the common period, with a prominent shift from large negative values in the 1960s to large positive values in the 1990s.

Figure 4.3b shows the 31-winter moving window trend time-series calculated as in Section 3.2 with units sd/decade (i.e. 10 times the analogous value in units of sd/year). As discussed in Chapter 3, the window length of 31 winters is chosen as it is within the time frame generally considered as multi-decadal variability, with a large enough sample of data points to calculate a fairly robust estimate of the linear trend over time, whilst being short enough to have multiple non-overlapping samples within the observed time series. The maximum 31-winter trend occurs for the winters 1963 to 1993, i.e. the window centred on 1978 with years referring to the January in DJF, with magnitude close to 0.7 sd/decade for all three observation datasets (Table 4.1). This maximum NAO trend is noticeably greater than any of the previous peak 31-year trends, including those in the whole 343 years of L99 reconstruction, i.e. this is an *extreme multi-decadal trend*. Over the 3 decades this is equivalent to a total shift of ~2 standard deviations of the winter mean NAO interannual variability.

| Max Trend (sd/decade) | (a) | (b) |
|---|---|---|
| HadSLP2r | 0.725 | 0.737 |
| C20C | 0.701 | 0.708 |
| L99 | 0.654 | 0.659 |

**Table 4.1. Maximum 31-year NAO trends.** Values are shown using the entire NAO time series available from HadSLP2r, C20C and L99 datasets, standardised over (a) the common period 1872-2001 and (b) the CGCM period 1862-2005 such that units are in sd/decade (i.e. 10 times the analogous value in units of sd/year).

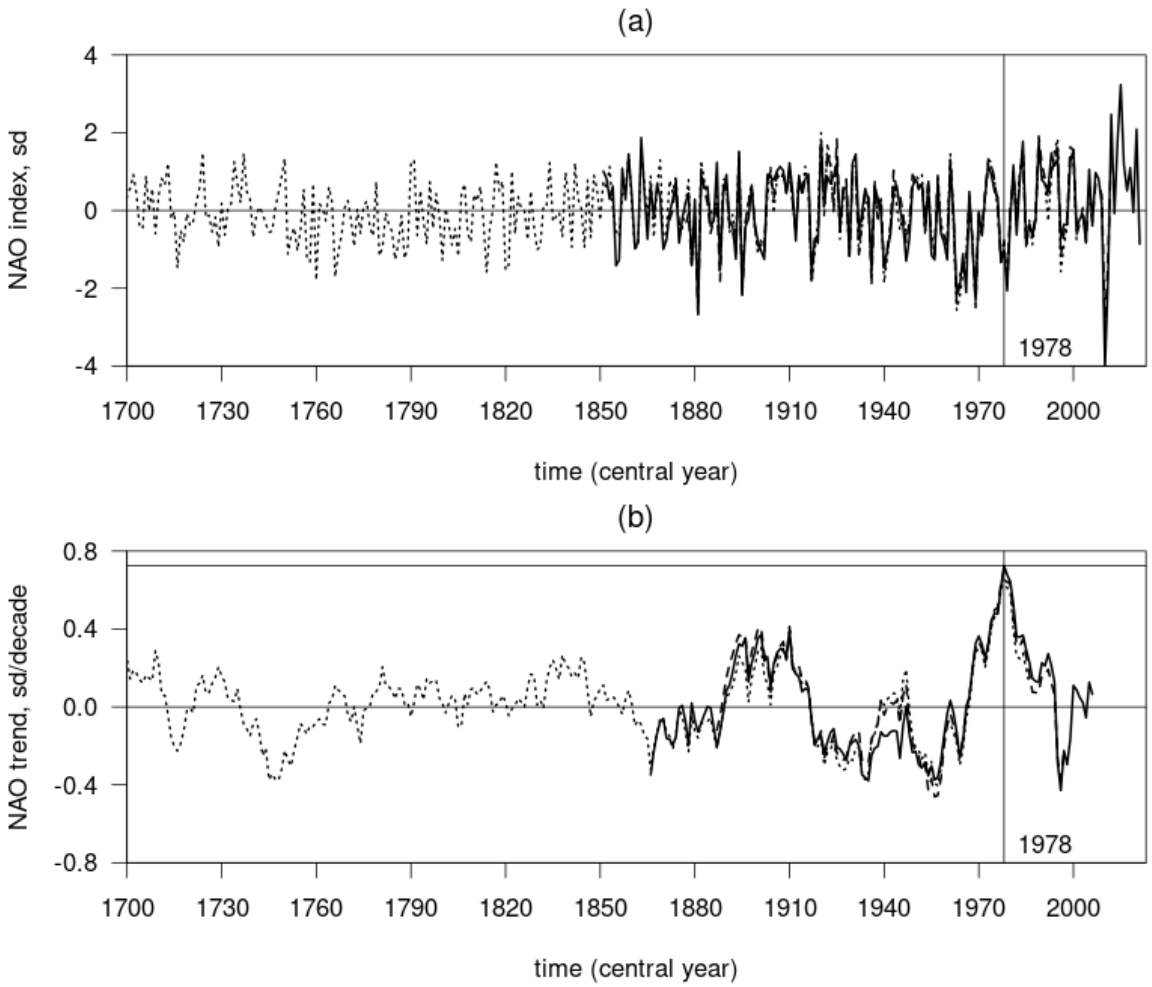


**Figure 4.3. Interannual variability of the NAO and its trend in observations.** (a) The standardised DJF mean NAO index for HadSLP2r (black solid line), C20C (black dashed line) and L99 (black dotted line). (b) The 31-winter moving window linear trend estimate is shown for the same datasets with the maximum trend shown as a horizontal black line centred on 1978.

The precise values of the maximum 31-year NAO trends depend on the climate period used for standardisation as well as the observation dataset. This can be seen by comparing Table 4.1 column (a) to column (b) where the latter uses 1862-2005 as the climate period for standardisation, chosen as this is the period available from the CGCM experiments analysed in Chapter 5. For C20C and L99 a subset of 1862-2005 period is used as these datasets do not cover the whole period. The maximum trends are only slightly sensitive to these definitions, and are the same to an accuracy of one decimal place (0.7 sd/decade), so these details should not overly effect exceedance probabilities.

### 4.2.3 Co-variability of MSLP trends at NAO nodes

The NAO index using MSLP at the two nodes Ponta Delgada (Azores) and Reykjavík (Iceland) was originally defined as such due to the high level of North Atlantic variability captured by these points, as seen in the EOF pattern (Figure 4.1) and the long-term availability of historical observations at these locations. On interannual timescales, the winter mean MSLP at these two nodes is strongly anti-correlated with correlation -0.520 using HadSLP2r (1862-2005) and -0.669 using C20C (1872-2005). Analysing the 31-year MSLP trend series at these two nodes, the timing of the maximum observed 31-year NAO trend coincides with both the positive peak in 31-year trends for the Azores and the negative peak in 31-year trends for Iceland (extreme trends centred on 1978 except for HadSLP2r maximum Azores trend which is centred on 1979). This consistency in multi-decadal variability is also shown by a high level of anti-correlation between the two moving window trend series (-0.849 for HadSLP2r NAO and -0.939 for C20C for their respective time periods 1862-2005 and 1872-2005). These correlation estimates are fairly insensitive to excluding the period of interest, with values -0.749 (HadSLP2r) and -0.907 (C20C) when winters 1960 onwards are excluded. Thus, it seems that these locations are strongly related on multi-decadal time-scales as well as shorter time-scales.

## 4.3. Fitting stochastic processes to the NAO index

To estimate the distribution of trends in the NAO index, stochastic models are fitted to the observed NAO index time series and then the covariance approach from Section 3.3.1 is applied. The most common stochastic models used to represent the NAO are short-range dependence red noise (AR(1)) processes (e.g. Wunsch, 1999; Feldstein, 2000; Thompson et al., 2015) and long-range dependence (FD) processes (Stephenson et al., 2000). These models assume that the NAO index series has stationary variance. The NAO index in Figure 4.3 looks like the variance has increased over recent years, but this is mainly because of the extreme negative and positive NAO winters in 2010 and 2015 respectively while differences in the pre-1850 NAO series are likely due to it being based on proxy data rather than gridded observations. The recent decade with extreme winter NAO events is too short to make a robust assessment as to

whether the variance has significantly changed, but if the variance has increased then this would lead to an increase in the magnitude of multi-decadal trends as $\sigma_Y^2$ would increase in Equation (3.3). The extreme NAO years and the pre-1850 proxy data period are outside of the time period used to standardise and assess the distribution of NAO trends (1862-2005) thus the assumption of stationary variance is deemed valid.

Figure 4.4 shows the autocorrelation function for the NAO index using HadSLP2r. The autocorrelation function does not simply decay with lag, which has led some studies to consider long range dependence processes, however for the datasets assessed here the autocorrelation is very low beyond lag-1 year so higher order auto-regressive models are not considered. The lag-1 year autocorrelation estimates are significantly different to that of a white noise process at the 5% significance level, falling just outside the 95% prediction interval expected for white noise (Figure 4.4, black dashed lines) estimated using Bartlett's formula (Bartlett, 1946; Equation 3.13) with $n$ = 170 and $\rho$ = 0.0.

Similar autocorrelation and fractional difference parameter estimates are found when fitting AR(1) and FD processes to the three observation datasets (Table 4.2), though with some sensitivity to the time period used. Using the Bartlett formula (Bartlett, 1946; Equation 3.13) for the uncertainty in the HadSLP2r lag-1 year autocorrelation estimate of $\rho$ = 0.169 with $n$ = 170 (time series length) gives a 95% prediction interval for $\rho$ of (0.0212, 0.317). The $\rho$ estimates in Table 4.2 for different datasets and time periods are therefore consistent as they all fall within this range.

Figure 4.5 shows the lag-1 year autocorrelation patterns for DJF seasonal mean MSLP from HadSLP2r and C20C. The Azores and Iceland nodes (marked as black crosses) are both positioned over regions of relatively high lag-1 year autocorrelation. The observed winter mean NAO index consistently exhibits moderate levels of year-to-year dependence, coming from both the Azores and Iceland nodes, and this should be considered when investigating the distribution of multi-decadal NAO trends.

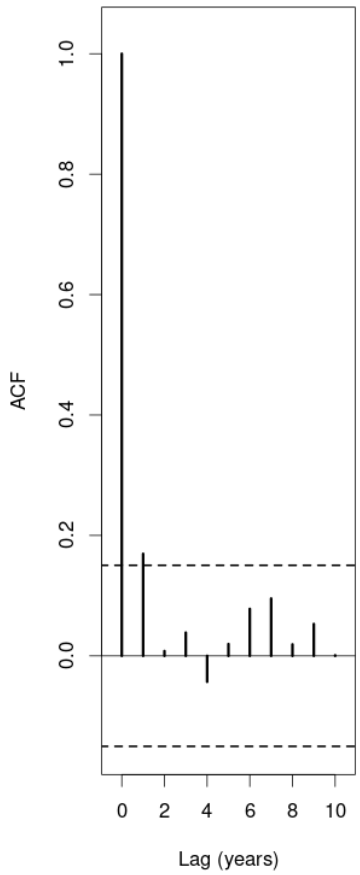


**Figure 4.4. Autocorrelation function for the DJF NAO index.** The lagged autocorrelations are shown for lag 0 to 10 years for winter mean NAO using HadSLP2r. The 95% prediction interval for the autocorrelation assuming a white noise process is shown by black dashed lines, with the upper value marginally lower than the NAO lag-1 year autocorrelation.

| | AR(1) $\rho$ parameter | FD $d$ parameter |
|---|---|---|
| HadSLP2r | 0.169 | 0.123 |
| C20C | 0.142 | 0.0919 |
| L99 | 0.0975 | 0.0614 |
| HadS 1872-2001 | 0.158 | 0.126 |
| C20C 1872-2001 | 0.156 | 0.116 |
| L99 1872-2001 | 0.171 | 0.0978 |

**Table 4.2. AR(1) and FD model fits to the NAO index.** Lag-1 year autocorrelation $\rho$ and difference parameter $d$ estimates are shown for the entire NAO index time series available from HadSLP2r, C20C and L99 datasets, and for the period 1872-2001 that overlaps with all the datasets.

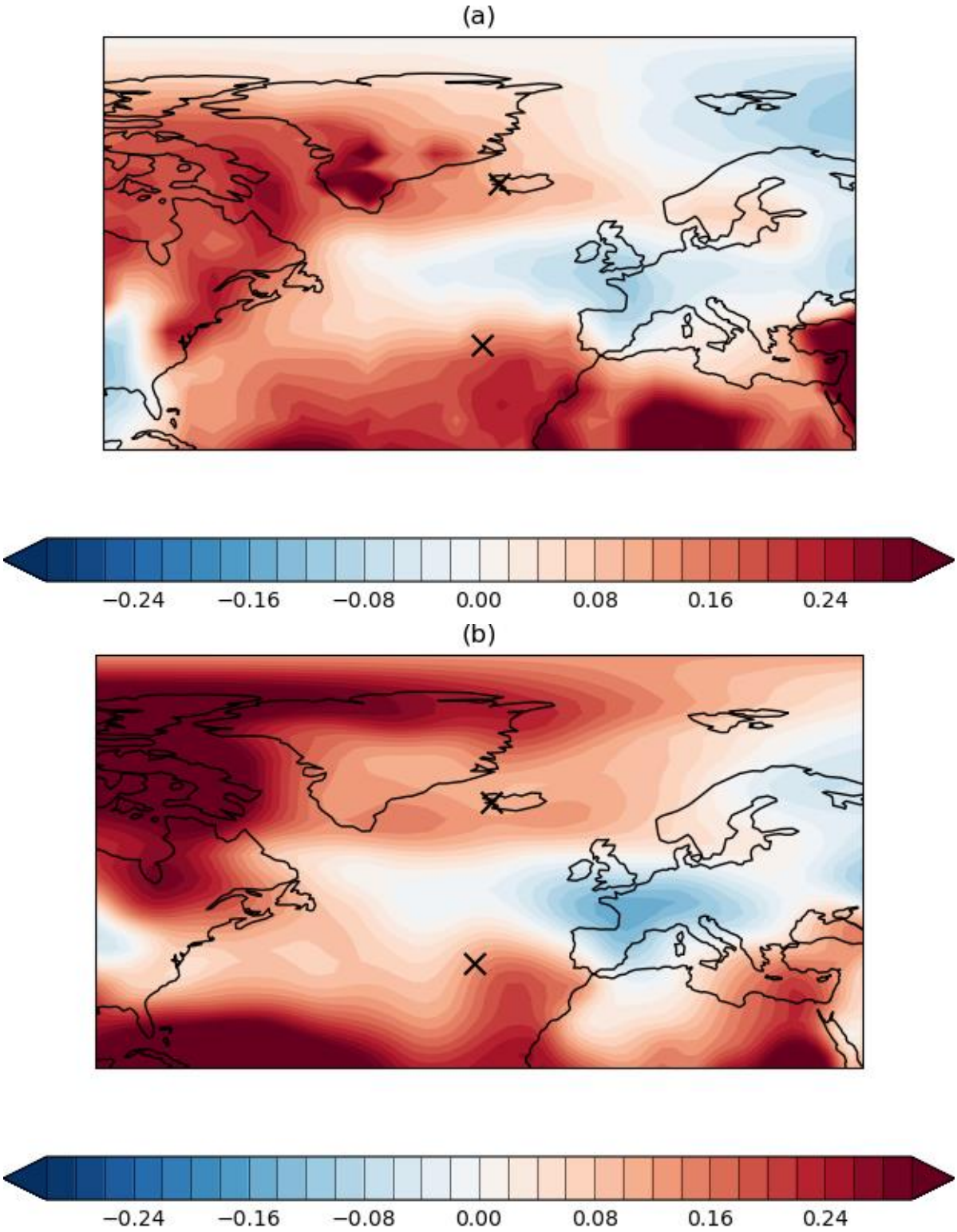


**Figure 4.5. Lag-1 year autocorrelation patterns for MSLP DJF seasonal mean.** The full available period of (a) HadSLP2r and (b) C20C datasets are used, with the standardised NAO index nodes marked as black crosses (Iceland and Azores).

## 4.4. Distribution of multi-decadal trends

### 4.4.1. Empirical probabilities

To quantify the unusualness of the 1963-1993 NAO trend within the available observed record it is insufficient to use the observed NAO index time series alone as this trend has been specially chosen as it is the *maximum* in the time series. This means that the simple method of estimating the empirical exceedance probability based on sample rank values for the filtered trend series ($Z$) will be governed by the length of the available observed trend series ($m$) rather than the relative number of occurrences as there has only been one such value in the trend series. This leads to the empirical rank probability

$$p = \Pr(Z_i \geq z) = 1/m$$

for the threshold $z = \max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\}$. For example, using the full NAO time series from each of the three observation datasets, the empirical rank probability $p$ is estimated to be 1/140 = 0.714% (HadSLP2r, Table 4.3), 1/111 = 0.901% (C20C) and 1/313 = 0.319% (L99), noting that for a time series of length $n$, the length of the 31-year moving window trend series $m = n - 30$.

| Model | $p$ | Uncertainty from $\sigma_z^2$ | Uncertainty from $\rho$ |
|---|---|---|---|
| Empirical | 0.714 % | n/a | n/a |
| White Noise | 0.0121 % | ($6.31 \times 10^{-7}$, 0.413) % | (0.00129, 0.0681) % |
| AR(1) $\rho = 0.169$ | 0.0822 % | ($2.44 \times 10^{-5}$, 1.27) % | (0.0159, 0.309) % |
| FD $d = 0.123$ | 0.141 % | ($6.38 \times 10^{-5}$, 1.77) % | (0.0199, 0.584) % |

**Table 4.3. Trend exceedance probabilities for 31-year trends.** Probabilities of trend exceedance (shown as percentages) for a single 31-winter trend $p = \Pr(Z_i \geq 0.0737$ sd/year) using HadSLP2r empirical estimates (Section 4.4.1) and stochastic model fits to HadSLP2r (Section 4.4.2). Uncertainty levels are shown due to the trend variance estimates ($\sigma_z^2$) and autocorrelation estimates ($\rho$) using 95% prediction intervals (Section 4.4.3).

### 4.4.2. Covariance approach

The maximum 31-year NAO trend using the HadSLP2r observations is $z$ = 0.737 sd/decade (Table 4.1). Using the covariance approach from Section 3.3.1, the probability of the trend in year $i$ exceeding this threshold $z$ = 0.737 sd/decade is given by Equation 3.3

$$p = Pr(Z_i > 0.737) = 1 - \Phi\left(\frac{0.737}{\sigma_z}\right)$$

The trend variance $\sigma_z^2$ is found from Equation 3.4 with $\sigma_y$ = 1.0, K = 15 and $h^2 = 15 \times (15+1)(2\times 15+1)/3$ = 2480 (Section 3.3.1) combined with the fitted stochastic model parameters ($\rho$ and $d$) from Table 4.2.

For the fitted AR(1) process with $\rho$ = 0.169,

$$\sigma_z^2 = \frac{1}{2480^2}\sum_{j=-15}^{15}\sum_{k=-15}^{15}\left(jk \times 0.169^{|k-j|}\right) \qquad (4.2)$$

such that $p$ = 0.0822%.

For the fitted FD process with $d$ = 0.123,

$$\sigma_z^2 = \frac{1}{2480^2}\sum_{j=-15}^{15}\sum_{k=-15}^{15}\left(jk \times \rho_{k-j}\right) \qquad (4.3)$$

where $\rho_{k-j}$ is determined iteratively using Equation 3.11, leading to $p$ = 0.141%.

If instead it is assumed that the NAO index has no short- or long- range dependence, a white noise stochastic process can be used instead such that ${\sigma_z}^2$ = 1/2480 (Equation 3.9) and $p$ = 0.0121%, i.e. $p$ is an order of magnitude smaller than for the stochastic models which include year-to-year memory.

The fitted stochastic model exceedance probability estimates average around a 1 in 1000 chance ($p$ = 0.1%). Ignoring the year-to-year dependence when modelling the NAO timeseries underestimates the exceedance probabilities by a factor of 10 ($p$ = 0.01%). The HadSLP2r empirical rank probability estimate $p$ = 0.714% (Table 4.3) is considerably larger than the fitted AR(1) and FD model estimates, but this empirical estimate is not very robust as it is totally dependent on the length of the time series. Using the whole L99 trend series leads to a smaller probability estimate $p$ = 0.319%. These observed empirical

rank probabilities are within the uncertainty range for the AR(1) and FD processes related to the sampling uncertainty of the variance estimate $\sigma_z^2$ (Table 4.3) so are consistent with the fitted AR(1) and FD processes (the method to calculate these intervals is outlined in Section 4.4.3). However, these observed empirical rank probabilities are in the upper end of the uncertainty range which suggests that extreme NAO trends may be more likely in the observations than one might expect from these simple stochastic models.

### 4.4.3. Sampling uncertainty

Figure 4.6 shows the distribution of moving window trends for the standardised NAO index in terms of the exceedance probability $p = \Pr(Z_i \geq z)$ and return period ($1/p$) for moderate to extreme $z$ thresholds. These probability curves are plotted for the Gaussian distribution with mean zero and variance $\sigma_z^2$ from Equation 4.2 for the fitted AR(1) process (Figure 4.6a) and from Equation 4.3 for the fitted FD process (Figure 4.6b). A set of probability curves are also shown for the AR(1) and FD processes with lag-1 autocorrelation $\rho$ in the set {0.0, 0.1, 0.2, 0.3, 0.4} (thin black dashed lines with labelled $\rho$ values). The empirical probabilities for HadSLP2r from Section 4.4.1 are shown as open circles, using the period 1862-2005 to match the length of CGCM experiments assessed later in Section 5. These empirical probabilities are fairly close to the fitted probability curves for low trend thresholds but are clearly larger for higher thresholds.

To quantify the uncertainty on the exceedance probability $p$ when estimated from a finite trend series, a set of 5000 stochastic simulations are generated for the fitted AR(1) process ($\rho = 0.169$) and fitted FD process ($d = 0.123$), both with standard deviation $\sigma_Y = 1.0$ and length $n = 144$ years (representing 1862-2005), leading to a set of simulated trend series of length $m = 144 - 30 = 114$ years calculated as in Section 3.2. For each simulated trend series, the empirical probabilities of trend exceedance are stored for a set of thresholds ($z$) in the range 0.01 to 0.09. The 95% prediction interval from these 5000 sets of probabilities is used to estimate the uncertainty on the empirical probability estimate of $p$ due to finite sample size, shaded dark grey in Figure 4.6. Another set of simulations is generated for a white noise process, with the related prediction interval shaded light grey (Figure 4.6). The region where these

intervals overlap is shaded medium grey to enable the plotting of both intervals on the same graph. These prediction intervals are a fairly good indication of the expected uncertainty in empirical probabilities calculated from a single filtered trend series (*Z*) derived from a relatively short observation dataset. The level of sampling uncertainty is very large, meaning that the empirical distribution from a relatively short series of observed moving window trends should be treated with caution as it may not be very representative of the true distribution. The prediction interval curves are step-like due to the discrete set of possible empirical exceedance probability values {0, 1/114, 2/114, …} from a trend series of length 114. Note that as the trend thresholds get more extreme, the prediction interval lower bound quickly tends to zero and the return period tends to infinity. Trend thresholds of magnitude greater than or equal to the maximum observed NAO trend ($z$ = 0.737 sd/decade) are so rarely sampled, even within 5000 simulations, that the prediction interval upper bound also quickly tends to zero for the fitted stochastic processes, while no such trends are simulated using the white noise process, at which point this method of estimating sampling uncertainty ceases to be meaningful.

An alternative method to estimate the sampling uncertainty for more extreme trend thresholds is proposed using sample variance estimates. The simulation method can be adapted to make use of the assumption that the 31-year trend *Z* has a Gaussian distribution with variance $\sigma_z^2$. For each simulated trend series, the sample variance of the moving window trends is stored. The empirical 95% prediction interval from these 5000 values is used to estimate the effect of $\sigma_z^2$ sampling uncertainty on the Gaussian probabilities. The prediction interval from the fitted stochastic processes are shaded dark grey in Figure 4.7 and summarised in Table 4.3 for trend threshold $z$ = 0.0737 sd/year ("Uncertainty from $\sigma_z^2$"). The white noise prediction interval is shaded light grey, and medium grey represents where these regions overlap (Figure 4.7). In this case the prediction interval curves are similar to those in Figure 4.6, but they are continuous as they inherit the continuous nature of the Gaussian distribution, so can be used to estimate the sampling uncertainty for any trend threshold.

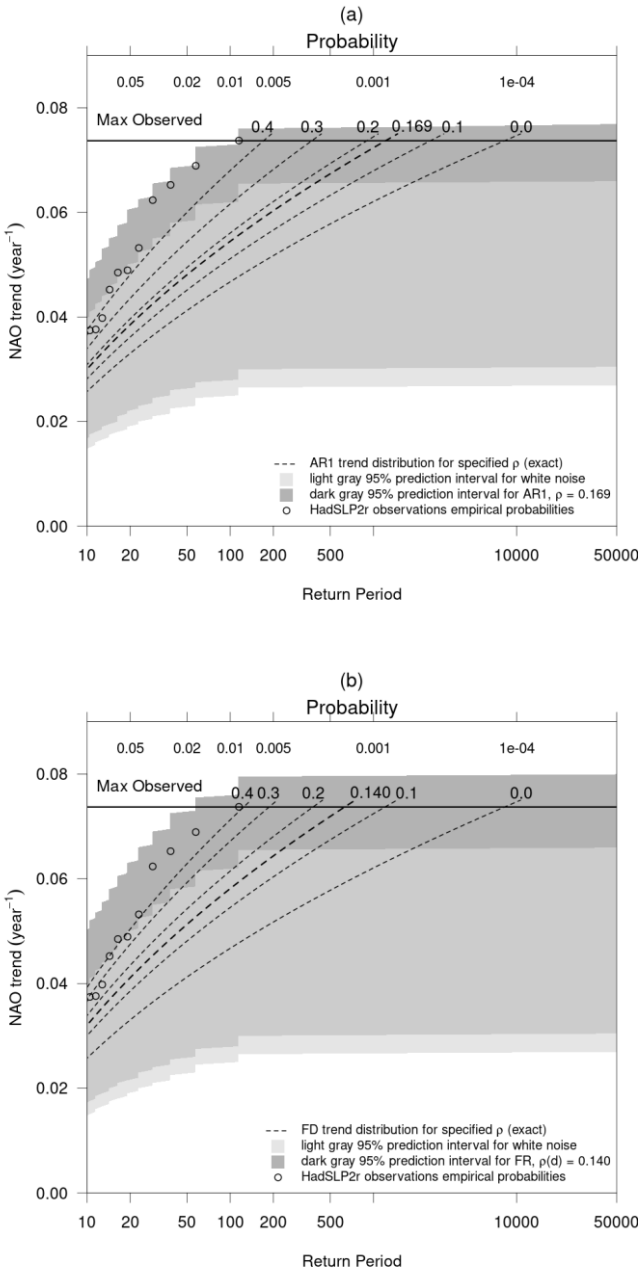


**Figure 4.6 Return Plot for NAO Trend with empirical probability-based uncertainty estimates.** Exceedance probabilities (top axis) and return period (bottom axis) from fitted (a) AR(1) and (b) FD processes (thick black dashed curves) with shaded 95% prediction interval based on empirical probabilities from trend series simulations (Section 4.4.3). Empirical probabilities for HadSLP2r standardised NAO index shown as open circles (Section 4.4.1). Probability curves are included for the AR(1) and FD processes with lag-1 autocorrelation $\rho$ in the set {0.0, 0.1, 0.2, 0.3, 0.4} (thin black dashed curves with labelled $\rho$ values).

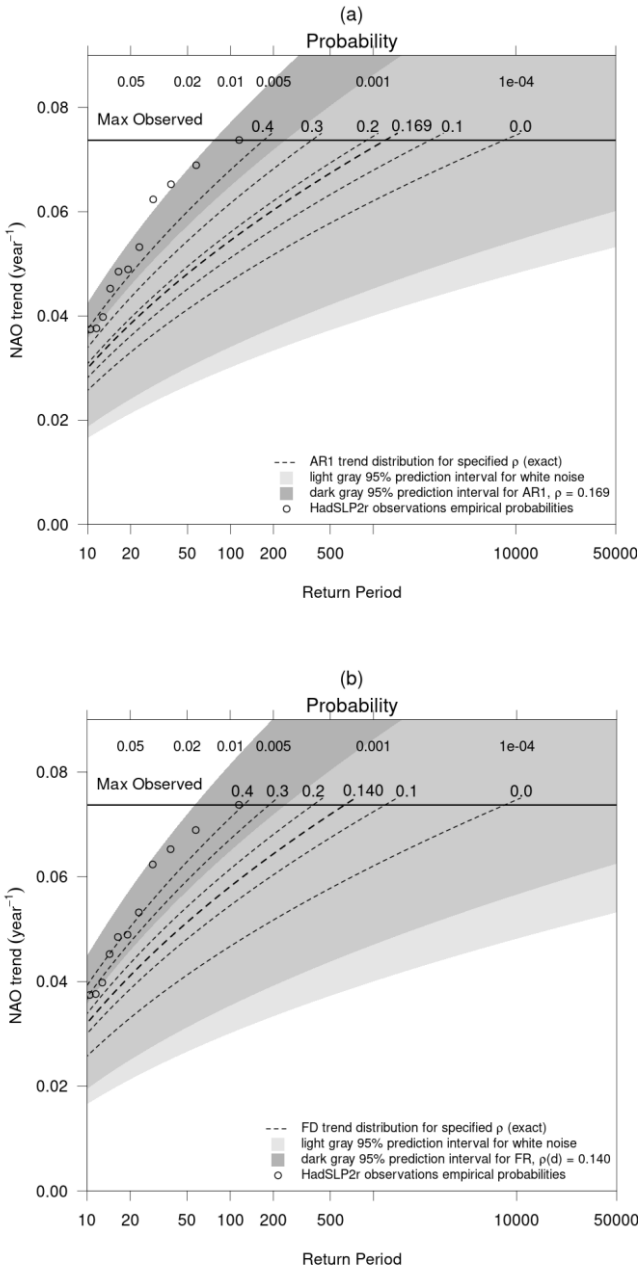


**Figure 4.7 Return Plot for NAO Trend with variance based uncertainty estimates.** Exceedance probabilities (top axis) and return period (bottom axis) from fitted (a) AR(1) and (b) FD processes (thick black dashed curves) with shaded 95% prediction interval based on variance of trend series simulations (Section 4.4.3). Empirical probabilities for HadSLP2r standardised NAO index shown as open circles (Section 4.4.1). Probability curves are included for the AR(1) and FD processes with lag-1 autocorrelation $\rho$ in the set {0.0, 0.1, 0.2, 0.3, 0.4} (thin black dashed curves with labelled $\rho$ values).

The observed empirical probabilities clearly lie outside of the probability prediction intervals for a white noise process (Figure 4.6a and 4.7a, light grey shading), showing that it is inadequate to neglect the year-to-year dependence when modelling the NAO time series. For moderate trend thresholds the observed empirical probabilities lie within the prediction intervals for the fitted AR(1) and FD processes (Figure 4.6 and Figure 4.7 dark grey shading), thus the difference to fitted model probabilities can likely be explained by the considerable sampling uncertainty in empirical probabilities caused by the relatively short observed time series available and the lack of occurrence of the larger trend values. For more extreme trend thresholds, greater than about 0.04 sd/year, the observed empirical probabilities lie close to the upper edge of the prediction intervals. From Figure 4.3b it can be seen that the timing of these exceedances are all concentrated around the maximum trend centred on 1978. This under-sampling of higher magnitude trends in the observations means that the related empirical probabilities lack robustness and are likely consistent with the fitted AR(1) and FD processes.

It is also useful to quantify the uncertainty in the estimate of the probability $p$ given that the stochastic processes have been fitted to a finite time series. Using the Bartlett formula (Bartlett 1946; Equation 3.13) for the uncertainty in the estimate of lag-1 autocorrelation parameter given the finite time series length $n$ = 170 and $\rho$ = 0.169 gives the 95% prediction interval of (0.0212, 0.317). Using this range of parameters for an AR(1) process, and the related difference parameters for a FD process using $d(\rho_1) = \rho_1/(1 + \rho_1)$ (Hosking, 1981), leads to an estimate of how the uncertainty in stochastic model fit may affect the estimate of $p$. These prediction intervals are summarised in Table 4.3 ("Uncertainty from $\rho$") for trend threshold $z$ = 0.0737 sd/year, alongside the prediction intervals quantifying the effect of $\sigma_z^2$ sampling uncertainty ("Uncertainty from $\sigma_z^2$"). The sampling uncertainty in $p$ coming from the autocorrelation estimate $\rho$ is considerably smaller than the uncertainty from trying to model the distribution of trends directly from the filtered time series $Z$ (related to the uncertainty in $\sigma_z^2$). This justifies the approach of fitting a stochastic model to the original NAO index time series, as it should give more robust probability estimates than using the observed trend series directly,

especially for higher trend thresholds which are under-sampled in the observed trend series.

### 4.4.4. Stationarity of observed NAO trends

The stochastic model covariance approach to quantify exceedance probabilities in Section 4.4.2 assumed that the NAO index series is first order stationary such that the expectation of the trend $E[Z_i]$ is equal to zero. To test the validity of this assumption, the time mean trend $\bar{Z}$ is computed for the observed 31-year NAO trend series ($Z_i$) over the historical period 1862-2005 (144 years):

$$\bar{Z} = \frac{1}{m} \sum_{i=1877}^{1990} Z_i \,. \tag{4.4}$$

The 31-year moving trend series $Z_i$ is labelled here by the central year ($i$) and $m$ is the length of the 31-year trend series ($m$ = 144-31+1 = 114 in this case). For the observed NAO index series using HadSLP2r, $\bar{Z}$ = 0.0354 sd/decade.

To test whether the observed $\bar{Z}$ value is consistent with a stationary process, the distribution of time mean trends $\bar{Z}$ is estimated using NAO trend simulations from each of the fitted stochastic models in Section 4.4.3, which are stationary by definition. For each fitted stochastic process, the time mean trend $\bar{Z}$ is computed for each of the 5000 simulations of moving window trend series (length 114 to emulate the observed NAO trend series). Relative to these empirical stochastic distributions (white noise, AR(1) and FD), the probability of $\bar{Z}$ having absolute magnitude greater than the observed estimate of $\bar{Z}$ (0.0354 sd/decade) is in the range 14% to 35% (Table 4.4), so there is no evidence to reject the null hypothesis that $\bar{Z}$ comes from a stationary stochastic model distribution with mean zero (applying standard hypothesis testing at the 5% level, see Appendix B.3). Therefore, it seems that the assumption of stationarity for the observed NAO trend series is valid, given the sampling uncertainty implied by the stochastic model distributions of $\bar{Z}$ due to the relatively short historical period available. This is in agreement with previous studies summarised in Sections 2.4.1 and 2.4.2 that found that the multi-decadal variability of the NAO is consistent with simple stochastic models of internal variability with some short- or long- range dependence when assessed over the full historical record.

| NAO model | % of $\lvert\bar{Z}\rvert$ > 0.0354 sd/decade |
|---|---|
| White noise | 14.2 % |
| AR(1) $\rho$ = 0.169 | 22.7 % |
| FD $d$ = 0.123 | 34.6 % |

**Table 4.4 Likelihood of stationarity for observed historical NAO series.** The percentage of fitted stochastic simulations of 31-year trend series $Z$ (length 114) with absolute magnitude of time mean $\bar{Z}$ greater than the observed value 0.0354 sd/decade for the historical period 1862-2005 (see Section 4.4.4).

## 4.5. Distribution of extreme multi-decadal trends

When considering the chance of exceeding the 1963-1993 trend value, it is necessary to take into account that this period has been pre-selected because of it being the maximum of all the values in the historical trend series. The approach chosen here is to use extreme value distributions for block maxima, building on the distribution of moving window linear trends to estimate the distribution of the maximum value in the trend series (block). Using the notation of Section 3.4 (Equation 3.14) and the threshold value of $z$ = 0.737 sd/decade from Section 4.4, the exceedance probability of interest in this case is

$$q(z) = Pr(max\{Z_{1+K}, Z_{2+K}, \dots, Z_{n-K}\} \geq z). \qquad (4.5)$$

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 by testing end-of-series trends for a range of window lengths, specifically ten to fifty years, that encompass the 32-year window that was chosen by “eyeballing” the data. They use a large sample of simulations from stochastic time series (length ten to fifty years) to find the empirical probability that at least one of those window lengths simulates a trend with magnitude greater than that observed. Adjusting the notation of Section 4.4 such that $Z^{(w)}$ is the variable representing the end-of-series trend for window

length *w* years and $z^{(w)}$ is the observed end-of-series trend for window length *w* years, then Percival and Rothrock (2005) are essentially calculating the probability:

$$p_{PR} = Pr\left(\left|Z^{(10)}\right| \geq \left|z^{(10)}\right| \; or \; \left|Z^{(11)}\right| \geq \left|z^{(11)}\right| \; or \; \ldots or \; \left|Z^{(50)}\right| \geq \left|z^{(50)}\right|\right). \qquad (4.6)$$

The extreme value method seems more appropriate when considering the chance of exceeding the 1963-1993 NAO trend as the window length has been pre-chosen based on definitions of multi-decadal variability (Section 4.2.2) and the temporal position is not constrained to be at the end of the time series. In this case it is the position of the window in time, rather than the length of the window, that has been chosen based on looking at the data and finding the maximum value in the moving trend series. If the window length had been chosen to maximise the observed trend, then it could be interesting to develop a method that combines the two approaches.

### 4.5.1. Stochastic simulations of extreme trends

To estimate the exceedance probability *q*(*z*) (Equation 4.5) for the *maximum* moving window trend in a series, the large set of stochastic simulations from Section 4.4.3 are used to model the distribution of *maximum* trends for the fitted stochastic processes with $\rho$ = 0.169 or *d* = 0.123 (Table 4.2) and standard deviation $\sigma_Y$ = 1.0. To quantify the uncertainty in these estimates, additional sets of simulations are generated for stochastic processes based on the 95% prediction interval for lag-1 year autocorrelation $\rho$ = 0.0212 and 0.317 which translates to the interval *d* = 0.0208 and 0.241 (Hosking, 1981). For this analysis, the stochastic time series length is set as *n* = 144 years to match the length of CGCM experiments assessed later in Section 5 (1862-2005), leading to a set of trend series (blocks) of length *m* = 144 – 30 = 114 years calculated as in Section 3.2. The exceedance probability *q*(*z*) from Equation 4.5 is then computed using the empirical probabilities based on the proportion of simulated trend series which have a maximum trend greater than the threshold *z* = 0.737 sd/decade.

The AR(1) and FD fitted models estimate maximum trend exceedance probabilities *q*(*z*) to be 3.46% and 6.06% respectively for trend threshold

$z$ = 0.737 sd/decade (Table 4.5), which averages at around a 1 in 20 chance. Using the Intergovernmental Panel on Climate Change (IPCC) likelihood scale (Appendix B.2), this would be termed a *very unlikely* chance of occurrence. Using the 95% prediction interval for lag-1 autocorrelation $\rho$ of (0.0212, 0.317) leads to an estimate of the uncertainty in the probability $q$ given that the stochastic processes have been fitted to a finite time series: (0.780%, 12.3%) for the AR(1) model and (1.34%, 19.4%) for the FD model (Table 4.5). The estimate of $q(z)$ is significantly lower if instead it is assumed that the NAO index can be modelled as a white noise stochastic process, i.e. with no short- or long-range dependence, for which $q(z)$ = 0.600% (Table 4.5). This aligns with results from Section 4.4.2 which found that ignoring the year-to-year dependence when modelling the NAO timeseries will lead to critically underestimating the exceedance probabilities for moving window trends.

| Model | $q$ | Uncertainty |
|---|---|---|
| White noise | 0.600 % | NA |
| AR(1) | 3.46 % | (0.780, 12.3) % |
| FD | 6.06 % | (1.34, 19.4) % |

**Table 4.5. Extreme trend exceedance probabilities for maximum 31-year trends in the period 1862-2005.** Probabilities (shown as percentages) of trend exceedance $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ for 31-year maximum NAO trends using stochastic model fits to HadSLP2r with parameters $\rho$ = 0.169 and $d$ = 0.123 (Table 4.2) relative to the observed maximum trend $z$ = 0.737 sd/decade (Table 4.1). For AR(1) and FD model fits, average $q$ = 4.76% ~ 1/20. The intervals in brackets represent the uncertainty from the stochastic model fits (based on empirical 2.5th and 97.5th percentiles for $\rho$, see Section 4.5.1).

## 4.6. Summary

This chapter has confirmed that the NAO trend for the period 1963-1993 is the maximum in the observed record for all three observation datasets assessed. Applying the approaches from Chapter 3 to assess the distribution of extreme

31-year trends in the NAO, the exceedance probability estimates for 31-year trends relative to this threshold (0.737 sd/decade) are:

- A 1 in 20 chance for the block maxima NAO trend in a block of length 144 years (1862-2005), estimated using stochastic simulations from the fitted stochastic processes with weak year-to-year memory.
- A 1 in 200 chance for the block maxima NAO trend estimate if the year-to-year memory is ignored (using a fitted white noise process)
- A 1 in 1000 chance if the 1963-1993 NAO trend is treated as occurring in a randomly chosen single window, i.e. neglecting to account for it being chosen specifically as it is the maximum trend in the historical record.

The NAO multi-decadal variability over the full historical period has been shown to be consistent with the fitted stochastic models, however the 1963 to 1993 extreme NAO trend is described as a "*very unlikely*" occurrence based on the block maxima probabilities above. This result is only valid if the NAO variability in the historical period is stationary, as shown in Section 4.4.4. This is in agreement with the literature summarised in Sections 2.4.1 and 2.4.2, which found that the NAO multi-decadal variability in the full historical period is consistent with simple models of internal variability, but this consistency is weaker when just assessing the late 20$^{th}$ century period. The stochastic models are not sufficient to partition the variability into internal variability, feedback mechanism and externally forced components as some of the short- and long-range dependence detected in the observed NAO index may itself be coming from the feedback mechanisms and external forcing. The stochastic models are a useful tool for estimating the distribution of extreme NAO trends, as it is not possible to do this using observations alone, and will provide a useful framework for testing the ability of GCMs to simulate extreme NAO trends.

# 5. Climate modelling of trends in the North Atlantic Oscillation

## 5.1. Introduction

This chapter uses Coupled General Circulation Model (CGCM) experiments (described in Section 5.2) to simulate the NAO, using the *standardised* NAO index as defined in Section 4 (Equation 4.1) for the observational analysis, based on mean sea level pressure (MSLP). The goal is to quantify the CGCM based probability of seeing such extreme trends as those identified in the observed NAO index trend series (Section 4) under present day climate conditions. CGCM simulations are used to estimate the distribution of extreme 31-year trends in the winter NAO index (Section 5.3.1) and relate this to the distribution of moving window trends (Section 5.3.2). Methods from Chapter 3 are used to fit stochastic models to the CGCM simulations of the NAO index to compare the level of dependence in CGCMs relative to that seen for the observed NAO and assess the effect of this on the trend distributions (Section 5.3.3). The method applied in Chapter 4 assumed that the observed NAO index is a stationary process. The stationarity of the historical CGCMs is assessed in Section 5.3.4. The observed maximum 31-year NAO trend occurred at the time of the maximum Azores and minimum Iceland MSLP trends and the 31-year moving window trend series at these locations have a high magnitude of anti-correlation. This level of co-variability is found to be less apparent in the historical CGCMs (Sections 5.3.5). The sensitivity of results to NAO index definition, window length, winter season definition and extreme trend definition are discussed in Section 5.4.

## 5.2. Climate models

### 5.2.1 Climate model data

Historical simulations from state-of-the-art CGCMs from the Coupled Model Intercomparison Project Phase 6, CMIP6 (Eyring et al., 2016) are assessed alongside those from Phase 5, CMIP5 (Taylor et al., 2012) to form a large ensemble of MSLP output. From CMIP6 there are 435 simulations from 56 models (Appendix A, Table A.1), while from CMIP5 there are 103 simulations from 42 models (Appendix A, Table A.2) making a total of 77472 years (538

simulations x 144 years). These are continuous transient simulations driven by external historical forcings (solar, volcanic, and anthropogenic). They cover the common period of winters 1862-2005, where January is used to define the year.

### 5.2.2 NAO indices

NAO index time series are calculated as in Section 4.2 (Equation 4.1, replicated as Equation 5.1 in Section 5.4.1 below) for each individual CGCM simulation separately with standardisation using the sample standard deviation from the same common period of 1862-2005. Where the model ensemble has size greater than one, the sample standard deviation is calculated across all members combined. These time series are then filtered in the same way as for the observed NAO to calculate the 31-winter moving window linear trends.

Figure 5.1 shows the first EOF pattern for winter mean MSLP over the North Atlantic (20-90N and 90W-40E) using observations (HadSLP2r, Allan and Ansell, 2006) and the multi-model mean of the first EOF patterns from the combined set of CMIP6 and CMIP5 CGCM experiments (see Appendix B.5 for code details). Where the model ensemble has size greater than one, the EOF is calculated across all members combined. The CGCM EOF patterns are similar to those from the MSLP observations. The similarity of the CGCM and Observed EOF patterns, and the closeness of the main nodes of variability to Iceland and the Azores, indicates that the NAO index definition from Equation 4.1 is suitable for a comparison of NAO behaviour in observations and CGCMs.

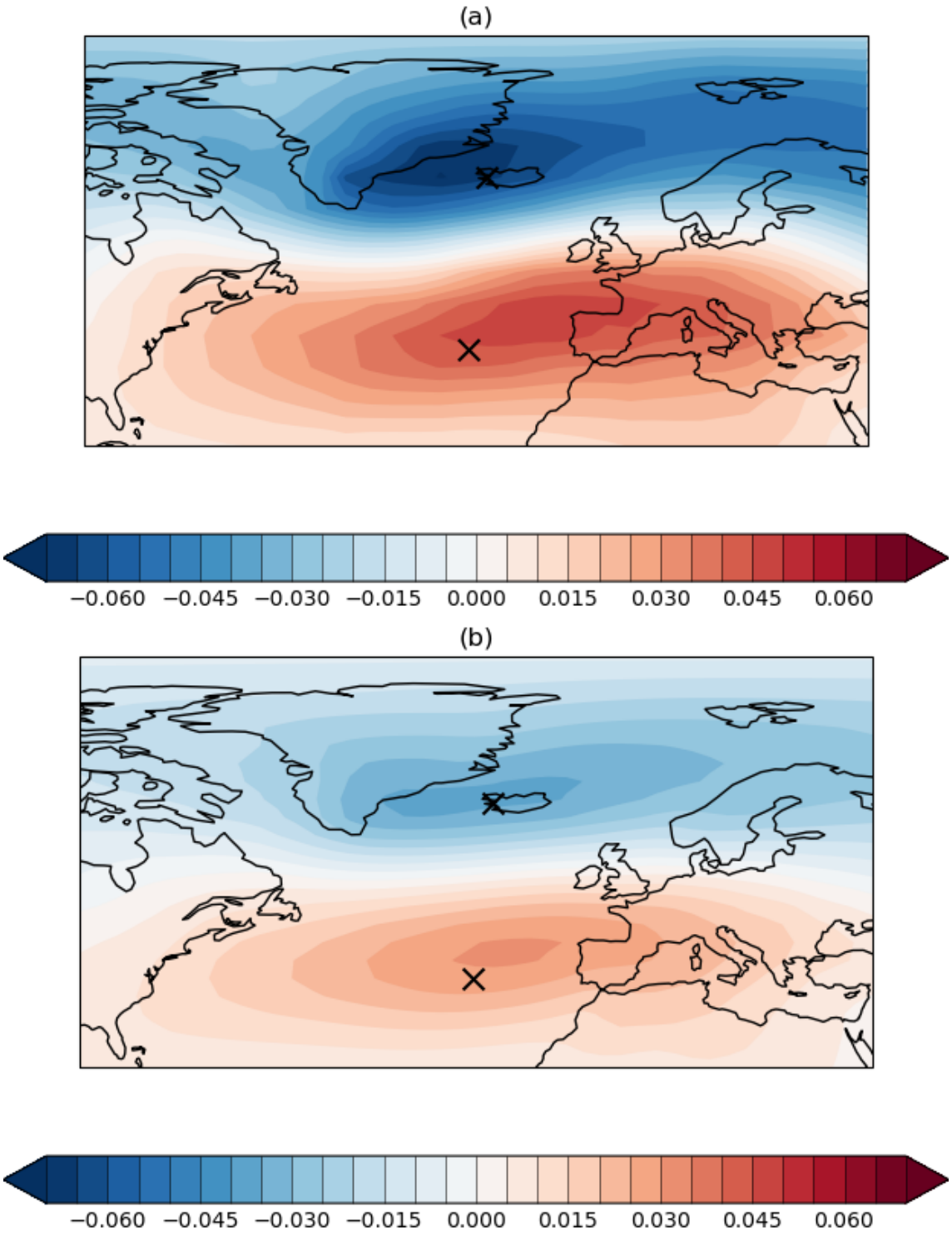


**Figure 5.1. First EOF for MSLP DJF seasonal mean.** HadSLP2r observations (a) and the multi-model mean of the EOF patterns from CGCMs (b) using the full period available for each dataset, with the Iceland and Azores NAO node locations marked as black crosses.

## 5.3. Climate model distribution of NAO extreme trends

### 5.3.1. Empirical distribution of extreme trends

Figure 5.2 shows the moving window trend series for each of the CGCM simulations (ensemble members) for their common period of 1862-2005, and for the observation dataset HadSLP2r and the 20$^{th}$ century reanalysis data V2 (C20C, Compo et al., 2011). For the specific time window 1963-1993, for which the observed NAO trend is at the maximum value, the multi-model ensemble mean across all simulations has only a very weak positive trend and none of the individual simulations predict a trend as large as that observed (Figure 5.2).

The empirical distribution of extreme trends in the CGCMs is calculated by taking the maximum 31-winter trend from each CGCM simulation, over the period 1862-2005 (block maxima for blocks of length 114 trend points). There is no consistency in the timing of maximum linear trends across the models, with maxima spread fairly evenly across the whole time period (Figure 5.2). This suggests that the NAO multi-decadal variability in the CGCMs for this period is primarily caused by internal variability and is not responding strongly to their common boundary condition forcing (solar, volcanic, and anthropogenic).

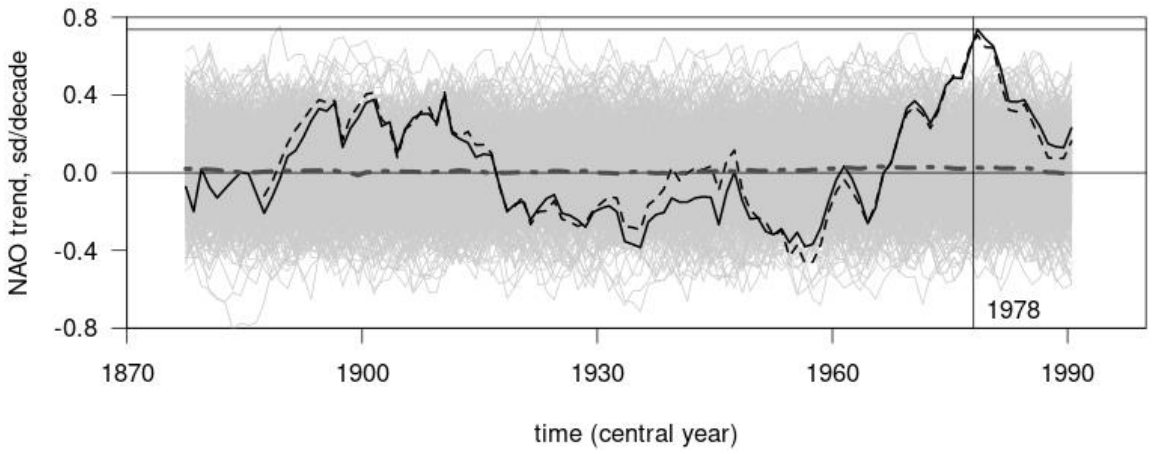


**Figure 5.2. Multi-decadal NAO trends in CGCMs.** The 31-winter moving linear trend estimate is shown for the set of CMIP5 and CMIP6 CGCM simulations of the NAO (grey solid lines) and for the multi-model ensemble mean (grey dot-dash line). The HadSLP2r (black solid line) and C20C (black dash line) observed trends are overlaid, with the maximum trend shown as a horizontal black line, centred on 1978.

Figure 5.3 shows the distribution of maximum NAO trends from the CGCM simulations, applying a kernel density fit (see Appendix B.6 for details) to the 538 CGCM values. The observed maximum NAO (shown for HadSLP2r) is outside of the empirical 95% prediction interval based on the 2.5th and 97.5th percentiles from the CGCM maximum trend distribution. Of the 538 CGCM simulations, only 2 have a maximum trend greater than the maximum observed, so the empirical estimate of the exceedance probability $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ (Equation 4.4) is therefore 2/538 (Table 5.1) which would be termed an *exceptionally unlikely* occurrence (less than 1 in 100 chance) using the Intergovernmental Panel on Climate Change (IPCC) likelihood scale (Appendix B.2). These results are similar to those from Bracegirdle et al. (2018) which found that the CMIP5 historical experiments very rarely simulate a 30-year trend in the winter NAO index or Atlantic jet strength of magnitude comparable to that observed in the 1960s to 1990s anywhere in their 150-year period of simulation.

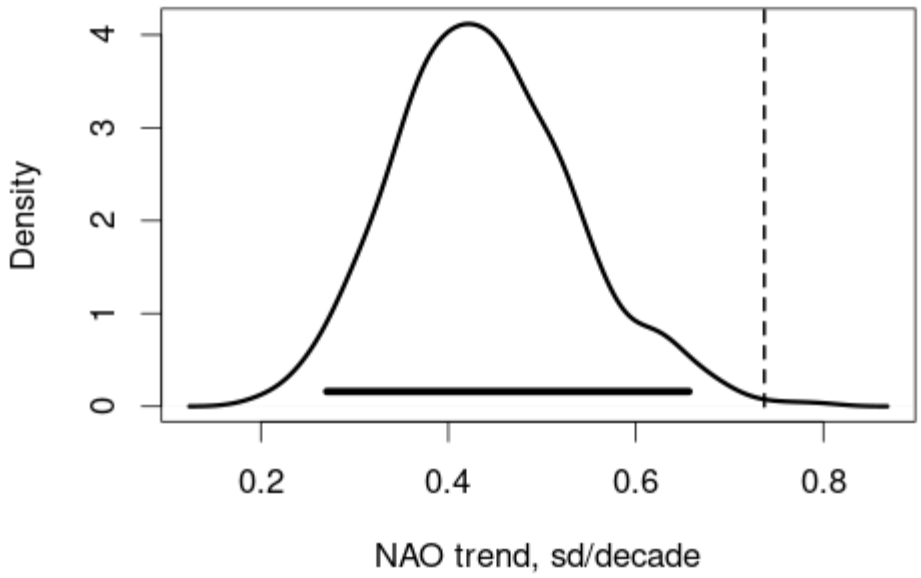


**Figure 5.3 Empirical distribution of block maxima NAO trends in CGCMs.** The probability density function (kernel density estimate, see Appendix B.6 for details) for the CGCM maximum 31-year NAO trend values with empirical 95% prediction interval (thick black horizontal lines). The observed HadSLP2r maximum trend is shown as a dashed vertical line.

| Model | Prob. $q$ as % |
|---|---|
| CGCM | 0.372% (2/538) |
| White Noise | 0.600% |
| AR(1) | 3.46% (0.780%, 12.30%) |
| FD | 6.06% (1.34%, 19.40%) |

**Table 5.1. Trend exceedance probabilities for maximum 31-year trends in the period 1862-2005.** Probabilities (shown as percentages) of trend exceedance $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ in the historical period 1862-2005, where $z$ is the maximum observed NAO trend (0.737 sd/decade from Table 4.3). Stochastic model probabilities are included from Table 4.5 where stochastic processes were fitted to the HadSLP2r NAO index. For the AR(1) and FD model fits, average $q$ = 4.76% ~ 1/20 (intervals in brackets represent the uncertainty from the stochastic model fits, see Section 4.5.1).

The CGCM empirical distribution of extreme NAO trends considerably underestimate exceedance probabilities compared to the stochastic model estimates from Chapter 4. Given there are 538 CGCM simulations in this study, it is expected that around 18 to 32 of them will simulate a maximum NAO trend greater than the observed maximum based on the fitted stochastic model probabilities in Table 5.1 ($538 \times q$). Instead, only two CGCM simulations exceed the observed maximum trend which leads to the empirical probability estimate $q$ = 2/538. This CGCM probability estimate is only around 10% of the probability estimates from Section 4 using stochastic model fits which averaged around a 1 in 20 chance (Table 5.1). The CGCM probability estimate is below the lower bound of the uncertainty range based on the stochastic model fits to observations (defined in Section 4.5.1) and is even lower than the white noise estimate which assumes no short- or long- range dependence between winter seasons (Table 5.1). This suggests that the CGCMs significantly underestimate the exceedance probabilities for extreme NAO trends, and this may in part be due to an underestimation of the year-to-year dependence in the winter NAO index.

Semenov et al. (2008) used two CGCMs to deduce that the distribution of 30-year NAO trends in CGCMs 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 CGCM 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 the CMIP5+6 estimate of $2/(538 \times 114) = 3.26 \times 10^{-5}$. For the large CGCM ensemble used in this thesis, it cannot be argued that the lack of extreme multi-decadal NAO trends in CGCM simulations is due to the CGCM sample size being too small. Using the data 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 CMIP5+6 simulations), which is consistent with the results from Table 5.1, i.e. that the 1960s to 90s trend is a *very unlikely* occurrence in CGCMs.

### 5.3.2. Empirical distribution of moving window trends

The distribution of extreme trends in a series is related to the distribution of moving window trends (Section 3.4). For the maximum trend in the observed NAO trend series, the distribution of such trends is unknown and must be estimated using a model such as the stochastic models in Section 4.5 and the CGCMs in this chapter. However, for the distribution of moving window trends, the empirical distribution can be obtained directly from the observation trend series for comparison with model-based distributions.

Figure 5.4 shows the distribution of moving window trends in terms of the exceedance probability $p = \Pr(Z_i \geq z)$ (top axis) and return period $1/p$ (bottom axis) for the CGCM NAO simulations. The CGCM empirical probabilities are calculated from the total multi-model ensemble of 61332 moving window trend values (538 simulations x 114 trend values), shown as grey crosses (based on sample rank values). The CGCM empirical probability curve leads to exceedance probabilities that are considerably lower than those from the fitted

AR(1) model from Section 4.4.2 (black dash line marked “0.169” in Figure 5.4) and are closer to those of the fitted white noise model from Section 4.4.3 (black dash lines marked “0.0” in Figure 5.4). This suggests that the CGCMs underestimate the magnitude of moving window multi-decadal NAO trends and, as for the distribution of extreme NAO trends (Section 5.3.1), this can be partially explained by a lack of year-to-year memory compared to the observed NAO index.

The grey shading in Figure 5.4 represents the 95% prediction interval for exceedance probabilities estimated from a single trend-series. To quantify this uncertainty on the exceedance probability *p* when estimated from a single finite trend-series (e.g. one CGCM member or one observation dataset), the empirical 97.5th and 2.5th percentiles of the sample of trend variance estimates from the 538 individual CGCM members are chosen here to define upper and lower bounds for the Gaussian distribution (with zero mean) exceedance probabilities. The empirical exceedance probabilities for NAO observations (from Section 4.4.1), shown as open circles in Figure 5.4 (using the period 1862-2005 to match the length of CGCM experiments), are far outside of the CGCM prediction interval suggesting that the CGCM experiments significantly underestimate the observed exceedance probabilities for 31-year moving window NAO trends.

It was argued in Section 4.4.3 that the observed NAO trend series are too short to estimate the trend distribution directly from moving window trend data, however for this large sample of CGCM NAO data (61332 moving window trend), a more robust distribution estimate should be possible. Figure 5.5 shows the empirical distribution of 31-year moving window NAO trends from the CGCM simulations (black curve). A Gaussian distribution can be fitted to the NAO trend values, using the total variance calculated across these 61332 trend values and a mean value of zero. The sample mean for the CGCM trends is very close to zero, so fixing the mean to be zero has little influence on the results. This Gaussian probability density function and the associated exceedance probability curve are displayed in Figure 5.5 (grey curve) and Figure 5.4 (black dot-dash lines marked “CGCM”) respectively. The Gaussian distribution is a fairly good statistical model fit when compared to the empirical

distribution of the CGCM moving window trends for moderate trend thresholds (Figure 5.5 grey and black curves).

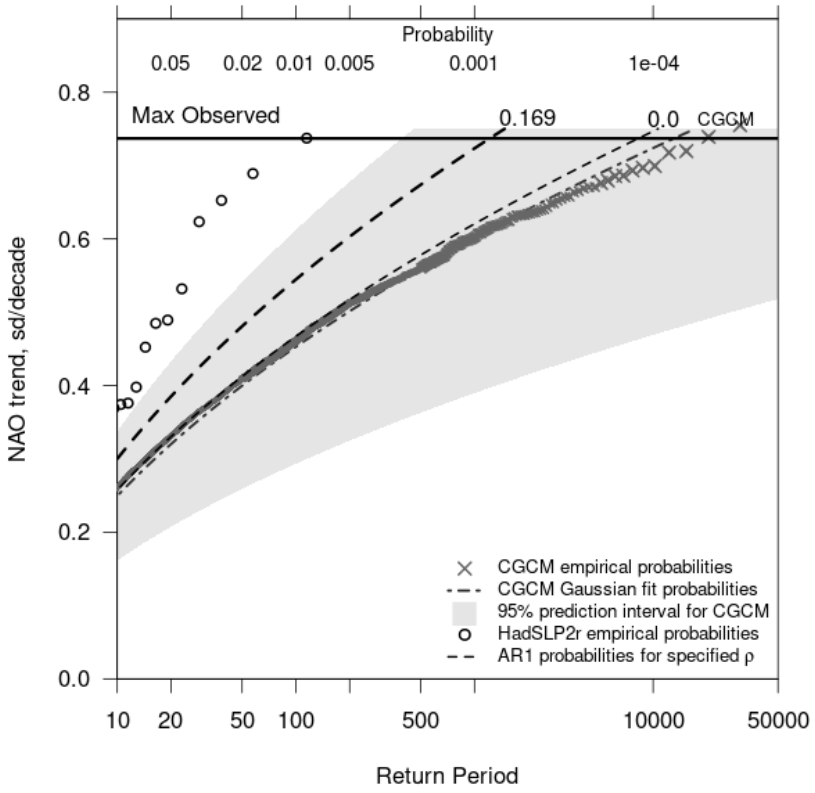


**Figure 5.4. Return period plot for NAO moving window trend.** Empirical exceedance probabilities (top axis) and return period (bottom axis) from CGCM simulations (grey crosses) and observations (black open circles) of 31-year moving window NAO trends. Black dot-dash curve shows the Gaussian distribution fitted to the CGCM trends with the 95% prediction interval for a single time-series shaded in grey (Section 5.3.2). Black dashed curves show the exceedance probabilities using the covariance approach (Section 4.4.2) for the AR(1) ($\rho$ = 0.169) and white noise ($\rho$ = 0.0) processes fitted to the observed NAO index time series.

For extreme trend thresholds the Gaussian fit marginally overestimates exceedance probabilities (Figure 5.4) compared to the empirical probabilities (grey crosses), though this difference is overly accentuated due to the logarithmic scale of the probability axes. These slight differences in the tails of

the CGCM Gaussian and empirical probability curves are likely just due to the increased uncertainty in the rank probability estimates since in these cases there is a lack of occurrence of large trends in the CGCMs comparable to the maximum observed. Overall, the CGCM empirical probabilities for moving window trends are well represented by the Gaussian distribution fit which is good evidence that the covariance approach from Section 4.4.2 is justified in modelling the NAO trend series as Gaussian.

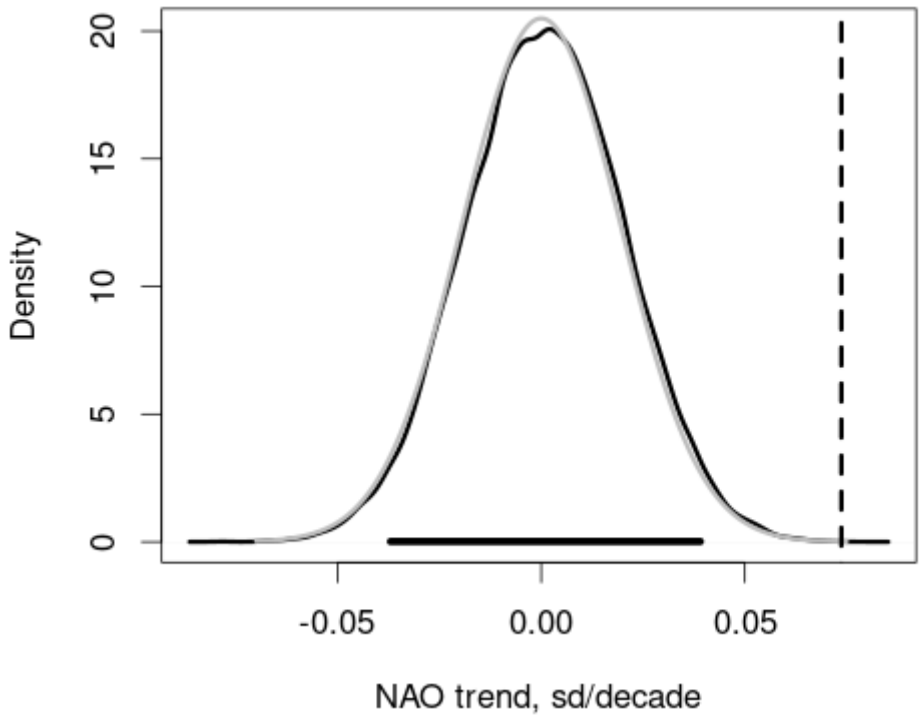


**Figure 5.5 Empirical distribution of Moving Window NAO trends in CGCMs.** The probability density function (kernel density estimate, see Appendix B.6 for details) for the CGCM 31-year moving window NAO trend values (black curve) with empirical 95% prediction interval (black horizontal line). The probability density function of the Gaussian fit is overlaid (grey curve). The observed HadSLP2r maximum trend is shown as a dashed vertical line.

### 5.3.3. Fitting stochastic processes to climate models

The distribution of moving window trends is related to the variability of the underlying index by Equations 3.3 and 3.4 (Section 3). The NAO index has been standardised (Equation 4.1) so it is only the autocorrelation function that is important here. Figure 5.6 shows the maximum NAO trend for each CGCM simulation relative to the model lag-1 year autocorrelation estimate $\rho$ for the winter mean NAO index, with just two simulations having a maximum trend greater than the observed HadSLP2r maximum trend (shown by the horizontal black line). Where the model ensemble has size greater than one, $\rho$ is calculated by first concatenating all the members so that it represents a model characteristic rather than a single simulation characteristic. The scatterplot spread is represented by the 95% prediction interval ellipse (for details see Appendix B.7). This shows no clear relationship between the level of autocorrelation in the CGCMs and the magnitude of extreme NAO trends. The correlation across the scatterplot points is very small ($r = 0.079$) and although the p-value is small ($p = 0.066$) it is still above the standard threshold of 0.05 used to reject the null hypothesis that the correlation is equal to zero, so the correlation value is not significant in this case. The models with the highest autocorrelation estimates in Figure 5.6 just have one member per model. Using the model with the largest ensemble size (CMIP6-MIROC6) to estimate autocorrelation for single members, the spread in these estimates (-0.17 to +0.15) is as large as the spread between model estimates, while the total model estimate is just 0.01, thus single members are not sufficient to identify if one model has autocorrelation greater than another (large ensembles are needed). This means that for this set of CGCMs the autocorrelation metric $\rho$ does not seem able to identify single models that are superior for simulating more extreme NAO trends, despite the expected relationship between the maximum trend and lag-1 year autocorrelation $\rho$ (Equation 3.3). This is likely due to a general lack of year-to-year memory in the CGCM NAO indices, where the spread in $\rho$ estimates about zero is due to noise from sampling short time series, rather than a physical difference between models.

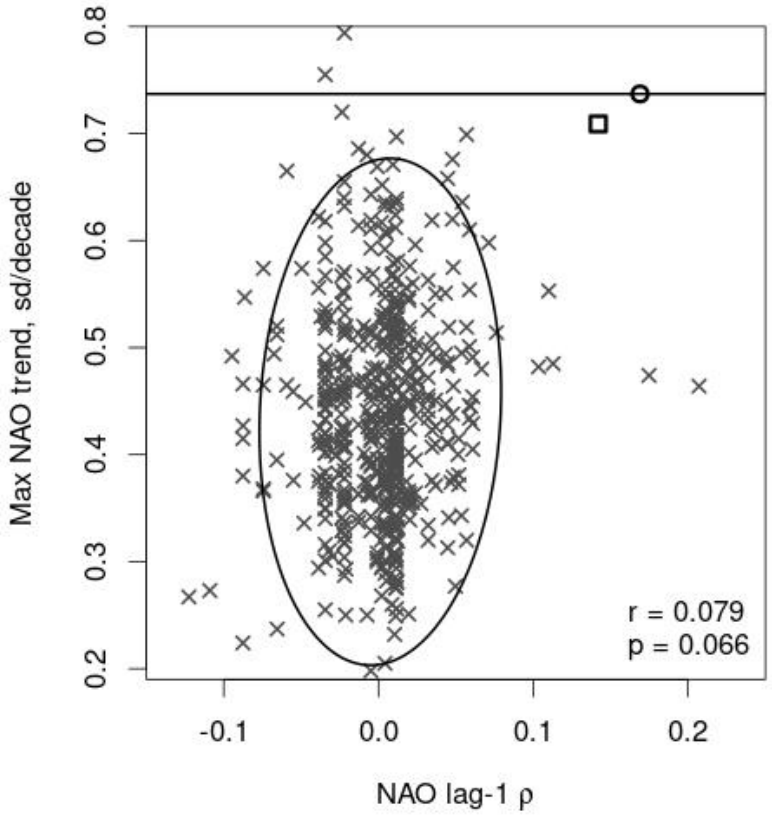


**Figure 5.6. The relationship of maximum NAO trends to NAO autocorrelation in CGCMs.** The maximum NAO trend for each CGCM simulations is shown relative to the model estimate of the lag-1 year autocorrelation $\rho$ (grey "x" with 95% prediction ellipse) with the correlation (r) and p-value (p) displayed in the bottom right. Observed values are shown for HadSLP2r (black circle, and horizontal line) and C20C (black square).

Figure 5.7 uses a boxplot to show the distribution of the lag-1 year autocorrelation $\rho$ estimates from the 538 individual CGCM NAO simulations using empirical estimates of the prediction intervals (2.5th to 97.5th percentiles whiskers, 25th to 75th percentile box, and central median line shown). The expected distribution of $\rho$ estimates from a white noise process of the same length is included using the Bartlett formula (Bartlett 1946; Equation 3.13) for the uncertainty in the estimates of lag-1 autocorrelation parameter given the finite time series length $n$ = 144 and $\rho$ = 0. The observed estimates of lag-1 year autocorrelation $\rho$ are significantly larger than the CGCM estimates (outside of the empirical 95% prediction intervals) and only five of the 538 CGCM estimates are greater than the HadSLP2r estimate (13 greater than the C20C estimate).

The HadSLP2r estimate of $\rho$ is also outside the white noise model 95% prediction interval (though the C20C estimate is slightly smaller). The CGCM distribution of $\rho$ estimates is centred close to zero and has very similar 75% and 95% prediction intervals to the expected intervals from the white noise process. This is further evidence that the CGCMs underestimate the level of autocorrelation in the NAO index series and are more consistent with a white noise process. Similar results were found for the jet stream variability in Simpson et al. (2018).

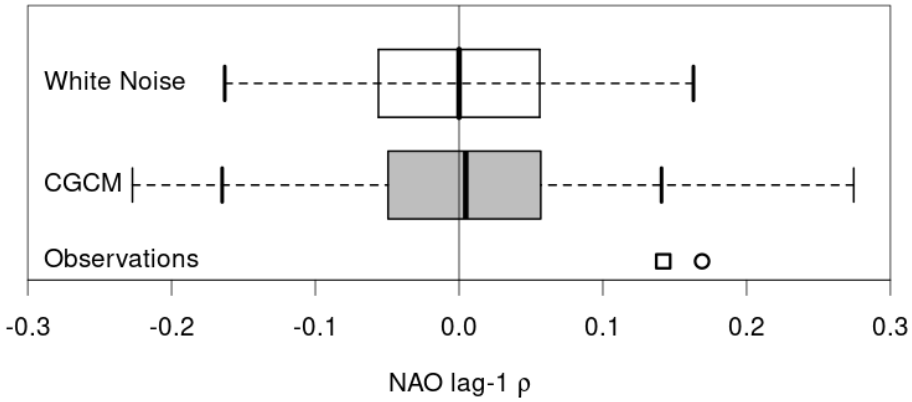


**Figure 5.7 Empirical distribution of lag-1 year autocorrelation for NAO index.** Box plots of the empirical distribution of lag-1 year autocorrelation ($\rho$) estimates from the 538 individual CGCM simulations and the expected distribution of sample $\rho$ estimates from a white noise model using the Bartlett formula (Bartlett, 1946) (Section 5.3.3; Equation 3.13). The box plots show the median line in the centre of the 25- to 75- percentile box with thick black whiskers showing the 95% prediction interval (2.5- to 97.5- percentiles). For the CGCM distribution thin black total range whiskers are included. Observed values are shown for HadSLP2r (black circle) and C20C (black square).

## 5.3.4. Stationarity of the NAO in climate models

The stationarity of the CGCM NAO index series can be examined in the same way as for the observed NAO index (Section 4.3), by assessing the distribution of time mean trends ($\bar{Z}$; Equation 4.4). Figure 5.8 shows the CGCM distribution of time means of 31-year trend series ($\bar{Z}$) for the period 1862-2005 using the

538 estimates of $\bar{Z}$ from the CMIP5+6 historical experiments (black curve). The distribution of $\bar{Z}$ for a white noise process of the same length is shown (grey curve) using the 5000 white noise stochastic model simulations from Chapter 4. The CGCM historical distribution of $\bar{Z}$ is shifted towards slightly more positive values than the stationary white noise process, with 68% of the CGCM $\bar{Z}$ values greater than zero but only 15% greater than the observed value of $\bar{Z}$ for the same period using HadSLP2r observations (Figure 5.8 dashed line). The sample mean of $\bar{Z}$ is small (0.0103 sd/decade) but is significantly different to zero (using a standard hypothesis test at the 5% level, see Appendix B.3) which suggests a weak level of non-stationarity in the CGCMs. Despite this positive shift in $\bar{Z}$, the CGCMs still underestimate the distribution of multi-decadal trends compared to observations and fitted stochastic models (Section 5.3.2).

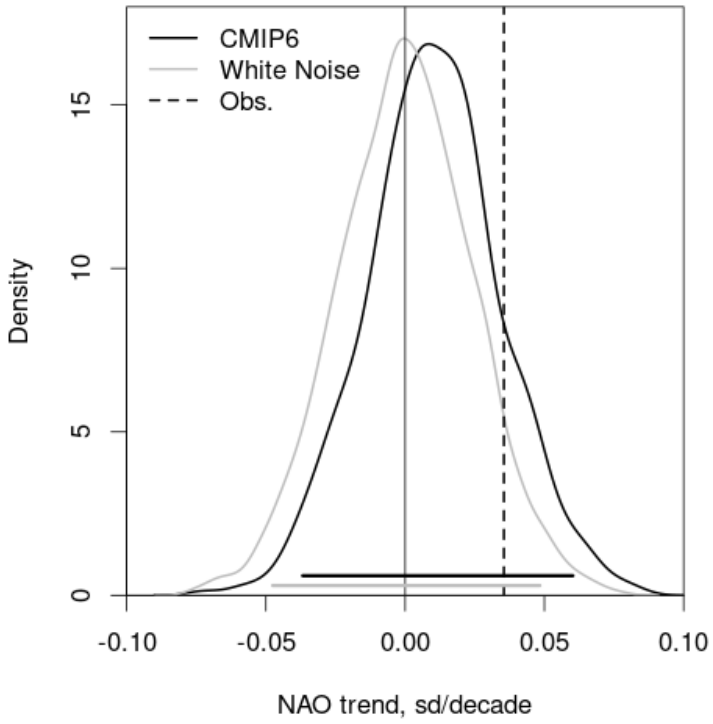


**Figure 5.8: Empirical distribution of time mean trends for CGCM historical NAO.** The probability density function (kernel density, see Appendix B.6) for the CGCM time means of 31-year moving window NAO trend series for the period 1862-2005 (black curve). The probability density function for the Gaussian white noise model overlaid (grey curve). Empirical 95% prediction intervals are shown by black (CGCM) and grey (white noise) horizontal lines. The observed time mean trend for 1862-2005 (HadSLP2r) is shown (dashed black vertical line).

### 5.3.5 Co-variability of pressure at NAO nodes

The NAO index has been defined using winter mean MSLP at the two main nodes of North Atlantic MSLP variability: Ponta Delgada, Azores and Reykjavík, Iceland. It was shown in Chapter 4 that there is a high level of anti-correlation between the moving window trend series at these points (-0.849 for HadSLP2r NAO for the period 1862-2005) and the observed maximum 31-year MSLP trend for the Azores occurs at the same time (within a year) as the minimum 31-year trend for Iceland. The CGCM historical experiments greatly underestimate this measure of co-variability with an average correlation of -0.559 across all 538 simulations (time period 1862-2005). The observed estimate of trend co-variability is on the very edge of the CGCM empirical 95% prediction interval (-0.853, -0.0809) (defined by the 2.5th and 97.5th percentiles), with only 14 out of the 538 simulations (2.6%) having a correlation value greater in absolute magnitude than that observed. This is consistent with the lack of coincident timing of CGCM extreme trends at the two node locations as only 19.7% of the CGCM simulated 31-year trend series have the maximum Azores trend within a year of the minimum Iceland trend. This CGCM deficiency in co-variability on multi-decadal timescales does not seem to be due to a problem on seasonal timescales as the CGCM co-variability for winter seasonal mean MSLP at Iceland and the Azores is consistent with that observed. For winter mean MSLP, the observed Azores and Iceland time series are fairly strongly anti-correlated, with correlation -0.520 (HadSLP2r), as would be expected based on the first EOF pattern for winter mean MSLP over the North Atlantic (Figure 5.1). The CGCM experiments are consistent with observations for this interannual variability, with the average of correlation values across all simulations for the period 1862-2005 being -0.592. The CGCM deficiency in the co-variability of NAO node trend series is of interest as it might relate to the CGCM's underestimation of maximum NAO trends.

To test the effect of coincident timing of maximum and minimum trends at the Azores and Iceland nodes on the NAO maximum trends, pseudo maximum NAO trends are calculated by combining the maximum 31-year MSLP trend at the Azores southern node and the minimum MSLP trend at the Iceland northern node regardless of their timing. Figure 5.9 shows the distribution of these

pseudo maximum NAO trends (grey curve) relative to the original distribution of CGCM maximum NAO trends (black curve) as in Figure 5.3. The pseudo maximum trend distribution is clearly shifted towards larger trend values, but the observed maximum NAO trend is still beyond the new empirical 95% prediction interval (grey horizontal line). The new empirical estimate of exceedance probability *q* is 8/538 = 1.49%, which is four times the original CGCM probability (2/538, Table 5.1), but this is still termed a *very unlikely* occurrence (less than 1 in 10 chance, IPCC likelihood scale, Appendix B.2) and is still smaller than the AR(1) and FD fitted model estimates (3.46% and 6.06% from Table 5.1). Thus, for the CGCM simulations, the lack of coincident timing of extreme trends for the individual NAO nodes likely contributes to the underestimation of extreme NAO trends relative to those observed, but this may itself be related to the lack of autocorrelation in the winter mean NAO index series which has a greater overall effect on the estimation of extreme trends.

Figure 5.10 shows the CGCM distribution of maximum MSLP trends at the Azores NAO node and minimum MSLP trends at the Iceland NAO node. The empirical 95% prediction intervals are shown by horizontal lines, using the empirical 2.5th and 97.5th percentiles from the CGCM extreme trend output. The observed extreme trend values (HadSLP2r; dashed vertical lines in Figure 5.10) are on the extreme edge of these prediction intervals. Of the 538 CGCM simulations, only 19 have maximum trends at the Azores NAO node greater than the maximum observed (block maxima empirical exceedance probability 0.0353 for trends above observed maximum) and only 15 have minimum trends at the Iceland NAO node less than the minimum observed (block minima empirical probability 0.0279 for trends below observed minimum). None of the individual simulations appear in both of these subsets. This analysis of MSLP trends at the individual NAO nodes shows that the CGCM underestimation of maximum NAO trends is related to the underestimation of the magnitude of both Azores maximum trends and Iceland minimum trends.

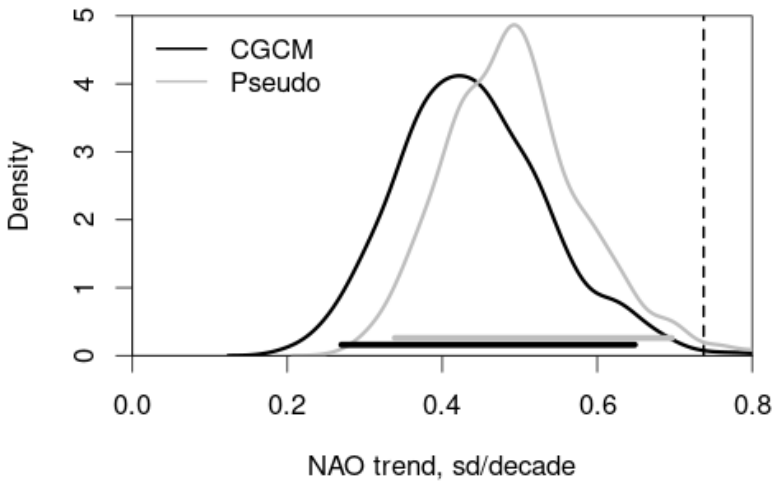


**Figure 5.9 Empirical distribution of block maxima NAO trends and pseudo maximum NAO trends in CGCMs.** The probability density function (kernel density, see Appendix B.6) for the CGCM pseudo maximum NAO trend values calculated from the maximum and minimum trends at the Azores and Iceland NAO nodes respectively regardless of timing (grey curve) as described in Section 5.3.5 with the empirical 95% prediction interval (thick grey horizontal line). The distribution for the original CGCM block maxima trends is shown (black solid line) as in Figure 5.3, with the observed (HadSLP2r) maximum trend shown by the dashed vertical line.

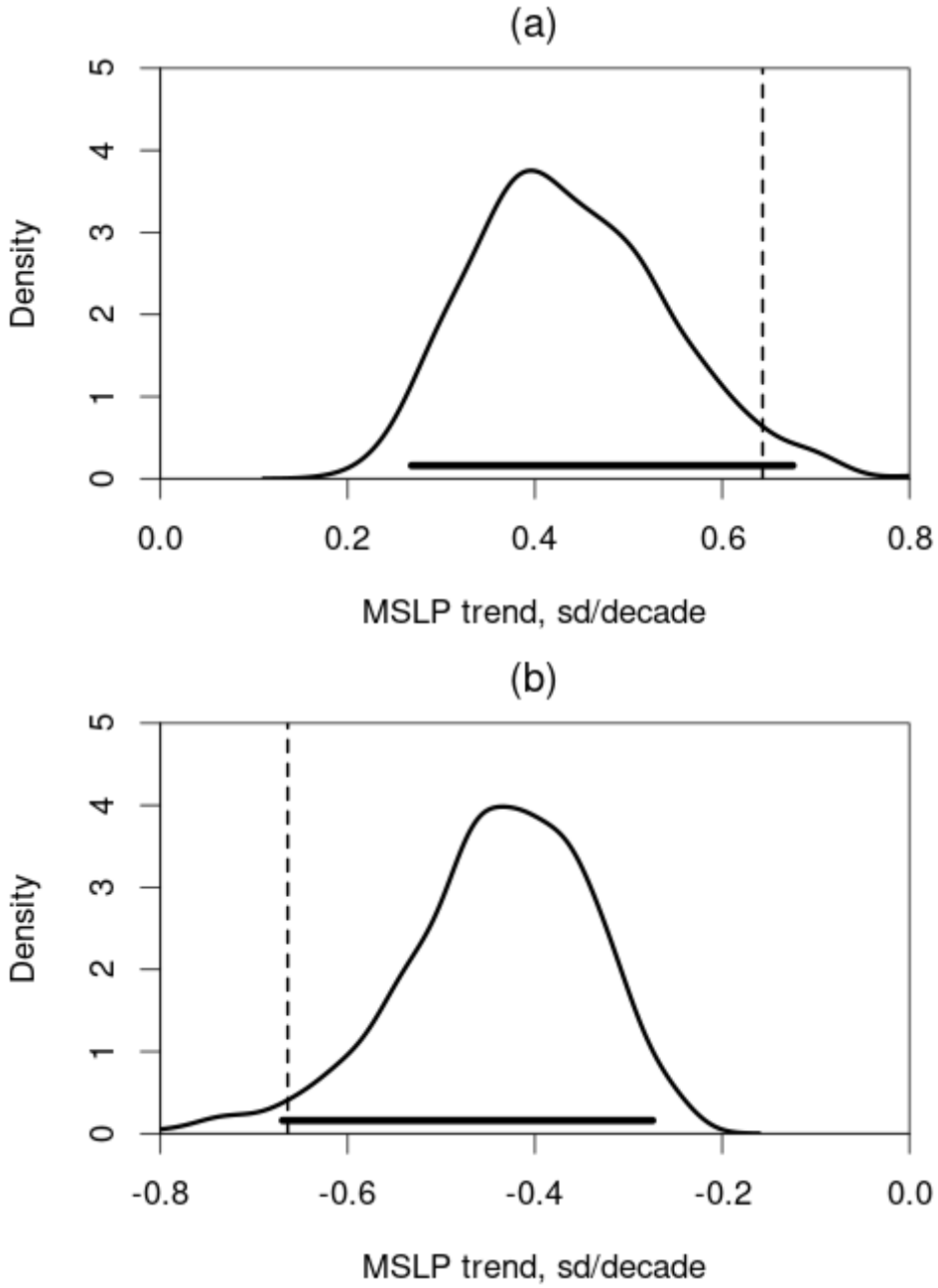


**Figure 5.10. Empirical distributions of MSLP extreme trends at NAO nodes in GCMs.** The probability density function (kernel density, see Appendix B.6) for (a) the maximum MSLP trend at the Azores NAO node and (b) the minimum MSLP trend at the Iceland NAO node using the CGCM (black curve) simulations for MSLP over the period 1862-2005. Horizontal black lines show the empirical 95% prediction intervals for the CGCM simulations. Observed (HadSLP2r) estimates are shown as black dashed vertical lines.

## 5.4. Sensitivity to NAO trend definitions

### 5.4.1. NAO index definition

The analysis of NAO trends has so far focussed on a single NAO index definition from Hurrell et al. (1995) which is the standardised difference of standardised winter mean MSLP series at Ponta Delgada (*P*) in the Azores and Reykjavík (*R*) in Iceland (using the nearest grid boxes) with the notation from Section 4.2.1 (Equation 4.1) replicated below:

**1. NAO Standardised point index**:

$$N(t) = \frac{(P(t) - \bar{P})}{s_P} - \frac{(R(t) - \bar{R})}{s_R}$$

$$NAO(t) = \frac{(N(t) - \bar{N})}{s_N} \qquad (5.1)$$

$\bar{P}$ and $\bar{R}$ are the sample means and $s_P$ and $s_R$ are the sample standard deviations of the P and R MSLP index series respectively, calculated over the full period available i.e. different periods for different datasets. $\bar{N}$ and $s_N$ are the sample mean and standard deviation of *N*, calculated over the CGCM common period 1862-2005.

To test the sensitivity of CGCM results to the NAO index definition, alternative definitions are considered as defined below:

**2. NAO Standardised regional average index:** Stephenson et al. (2006) define an NAO index using large regions rather than single grid boxes, noting that this should allow for shifts in the precise NAO position with time. They use the difference in regionally averaged DJF MSLP between the subtropical Mid-Atlantic to Southern Europe region 90W-60E, 20-55N and the North Atlantic-Northern Europe region 90W-60E, 55-90N. This index is defined as in Equation 5.1 but using the northern and southern regional means in place of *R* and *P* point values.

**3. NAO Empirical Orthogonal Function (EOF) index**: The leading principal component of winter MSLP variability in the North Atlantic region, defined here as 20-90N and 90W-40E as in Hurrell (1995) but extended to 90N as in Deser

et al. (2000) and Ambaum et al (2001). The full length of the MSLP dataset is used to calculate the EOF pattern (see Appendix B.5 for code details), i.e. different periods for different datasets. By definition, this index is in a standardised form, and for fair comparison between observation and model values the index time series is then re-standardised over the CGCM common period 1862-2005.

**4. NAO Anomalised point index**: The difference in DJF MSLP for Ponta Delgada (*P*) minus Reykjavík (*R*) but without standardisation so the actual pressure difference in hPa can be considered (e.g. Scaife et al., 2005).

$$NAO(t) = P(t) - R(t) - \overline{(P - R)}. \qquad (5.2)$$

This index time series is calculated as an anomaly with respect to the sample mean $\overline{(P - R)}$ calculated over the CGCM common period 1862-2005.

**5. NAO Anomalised regional average index:** The difference in large regional averages (as described in definition 2) but based on anomaly MSLP output using Equation 5.2 (with the northern and southern regional means in place of *R* and *P* point values).

Table 5.2 shows the CGCM maximum trend exceedance probability $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ for CGCM NAO trend series, computed as in Section 5.3.1 for the period 1862-2005, using each of the alternative NAO index definitions as displayed in column 1. In each case, the threshold *z* is defined as the maximum observed 31-year NAO trend for the observed trend series which has also been computed using the NAO index definition displayed in column 1. The empirical CGCM exceedance probabilities are shown in columns 2 and 3 as percentages, relative to the *z* thresholds calculated from the two observation datasets HadSLP2r and C20C. All of the exceedance probabilities are less than those from the fitted stochastic models in Table 5.1, except for the NAO anomalised regional average index (definition 5) which is marginally higher than the AR(1) model estimate but lower than the FD model estimate. However, NAO definitions 4 and 5 in Table 5.2 are in units of hPa (anomalised NAO point index and anomalised NAO regional average index), so these CGCM exceedance probabilities are not directly comparable to the stochastic model estimates in Table 5.1 as the interannual variance of the NAO index needs to

be considered as well as the autocorrelation (Equation 3.3). In general, the result that CGCMs underestimate extreme NAO trends is insensitive to the NAO definitions and observation datasets used.

| NAO index definition | (a) | (b) |
|---|---|---|
| (1) Standardised point index | 0.372 % | 0.558 % |
| (2) Standardised regional average index | 1.86 % | 0.929 % |
| (3) EOF index | 1.30 % | 0.186 % |
| (4) Anomalised point index | n/a | 1.86 % |
| (5) Anomalised regional average index | n/a | 3.72 % |

**Table 5.2. Sensitivity of maximum trend exceedance probabilities to NAO definitions.** CGCM empirical probabilities (shown as percentages) for $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ in the historical period 1862-2005, where $z$ is the maximum observed 31-year NAO trend using the NAO index definition displayed in column 1 and the observed NAO trend series from (a) HadSLP2r and (b) C20C. Probabilities for definitions (4) and (5) are not included for (a) as HadSLP2r underestimates MSLP variability.

Figure 5.11 shows the CGCM distribution of interannual standard deviation estimates for NAO index definitions 4 and 5. HadSLP2r is known to underestimate MSLP interannual variability (e.g. Semenov et al. 2008), and this can be seen in Figure 5.11 comparing the NAO standard deviation from HadSLP2r to the CGCM distribution of sample standard deviations. For this reason, the probability estimates for definitions 4 and 5 in Table 5.2 are only included for C20C and not HadSLP2r as CGCM exceedance probabilities could be greatly over-estimated. For NAO index definition 5 the CGCMs over-estimate the NAO index standard deviation relative to the C20C estimate, whereas this is not an issue for definition 4. The higher exceedance probability for definition 5 compared to the other NAO index definitions is likely due to the CGCMs overestimating the NAO index standard deviation, leading to an overestimate compared to the stochastic models which are fitted to the observed NAO standard deviation.

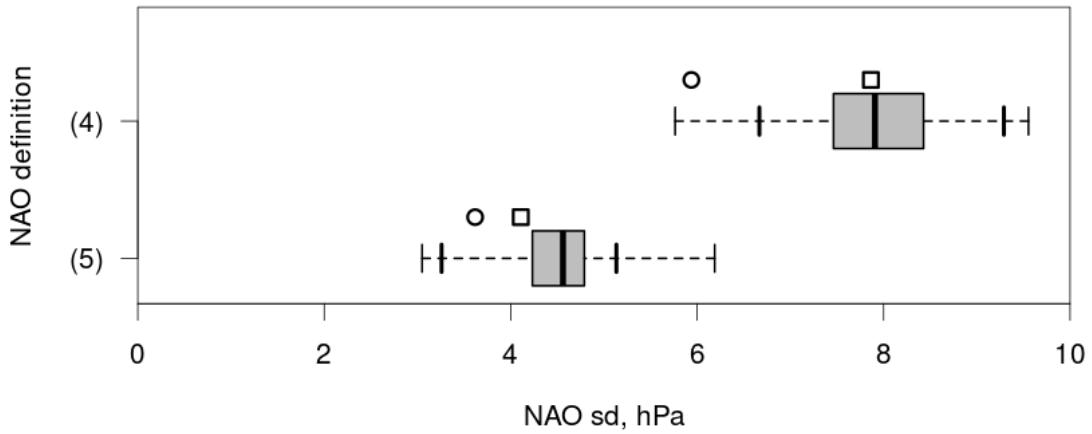


**Figure 5.11 Empirical distribution of NAO interannual variability in CGCMs.** Box plots showing the empirical distribution of standard deviation (sd) estimates for the period 1862-2005 from the 538 individual CGCM simulations for (4) “Anomalised point index” and (5) “Anomalised regional average index” (labelled as in Table 5.2). The box plots show the median line in the centre of the 25 to 75 percentile box with thick black whiskers showing the 95% prediction interval (2.5 to 97.5 percentiles) and thin black total range whiskers. Observed standard deviation estimates are shown for the same period using HadSLP2r (black circle) and C20C (black square).

## 5.4.2. Window length

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. Figure 5.12 shows the CGCM empirical exceedance probabilities $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ for maximum NAO trends in the historical period 1862-2005 relative to the observed (HadSLP2r) maximum trend thresholds $z$ for window lengths in the range 11 to 81 years. The observed maximum trends are *very unlikely* in the CGCM models for windows of length 29 to 53 years (probabilities less than 10%) and *exceptionally unlikely* for windows of length 31 to 45 (probabilities less than 1%), using categories from the IPCC likelihood scale (Appendix B.2). Exceedance probabilities are also shown for the fitted white noise process and

the fitted AR(1) and FD processes with autocorrelation $\rho$ = 0.169 and $d$ = $\rho/(1+\rho)$ = 0.145 (from Chapter 4). The FD model gives the highest exceedance probabilities but for trend windows of length 31 to 47 the observed maximum trends are still *very unlikely* with probabilities less than 10%, while CGCM probabilities are less than 1% and close to the fitted white noise probabilities. For window lengths of around three to four decades, the observed maximum NAO trends are termed *very unusual*, even when using the fitted AR(1) and FD models which should capture some of the observed short- and long- range dependence characteristics. The CGCMs underestimate maximum NAO trend exceedance probabilities for window lengths up to at least 71 years compared to the fitted FD and AR(1) models, and are generally more comparable to the white noise probability estimates. This underestimation of long-term NAO trends is also noted in Blackport and Fyfe (2022) for the specific single window trend 1951-2020. 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. In general, the result that CGCMs underestimate extreme NAO trends is insensitive to the window length and the observed maximum trends for window lengths of three to four decades are the most unusual.

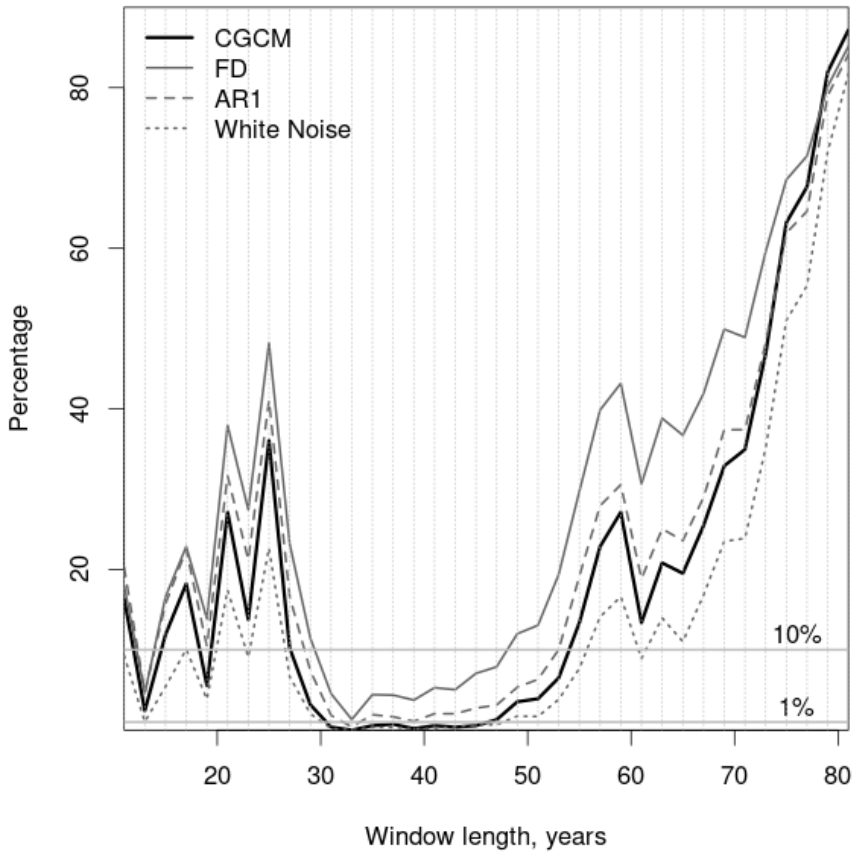


**Figure 5.12 Likelihood of maximum trends versus window length.** Trend exceedance probabilities for block maxima NAO trends for a range of window lengths. Probabilities of maximum NAO trend exceedance $q = \Pr(\max\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \geq z)$ for window length 11 to 81 years ($K$ in range 5 to 40) in the historical period 1862-2005 using the empirical probabilities from CGCM experiments (black solid line) relative to the observed (HadSLP2r) maximum trend thresholds $z$. Probability thresholds 10% and 1% are shown as labelled grey horizontal lines, representing *very unlikely* and *exceptionally unlikely* categories of the IPCC likelihood scale.

### 5.4.3. Minimum multi-decadal NAO trends

Figure 5.13 is like Figure 5.12 but for minimum multi-decadal NAO trends, showing the CGCM empirical exceedance probability $q' = \Pr(\min\{Z_{1+K}, Z_{2+K}, \ldots Z_{n-K}\} \leq z)$ for minimum 31-year NAO trends in the historical period 1862-2005 relative to the observed (HadSLP2r) minimum trend thresholds $z$. The observed minimum trends are *very unlikely* in the CGCM models for windows of length 43 to 81 years and *exceptionally unlikely* for most of the windows lengths in range 47 to 81 years. The FD model again gives the

highest exceedance probabilities, but for trend windows of length 47 to 81 the observed maximum trends are still mainly *very unlikely* with probabilities less than 10%. For these minimum NAO trends, the CGCMs again underestimate exceedance probabilities for all window lengths, but in this case it is the longer-term observed trends that are most unusual.

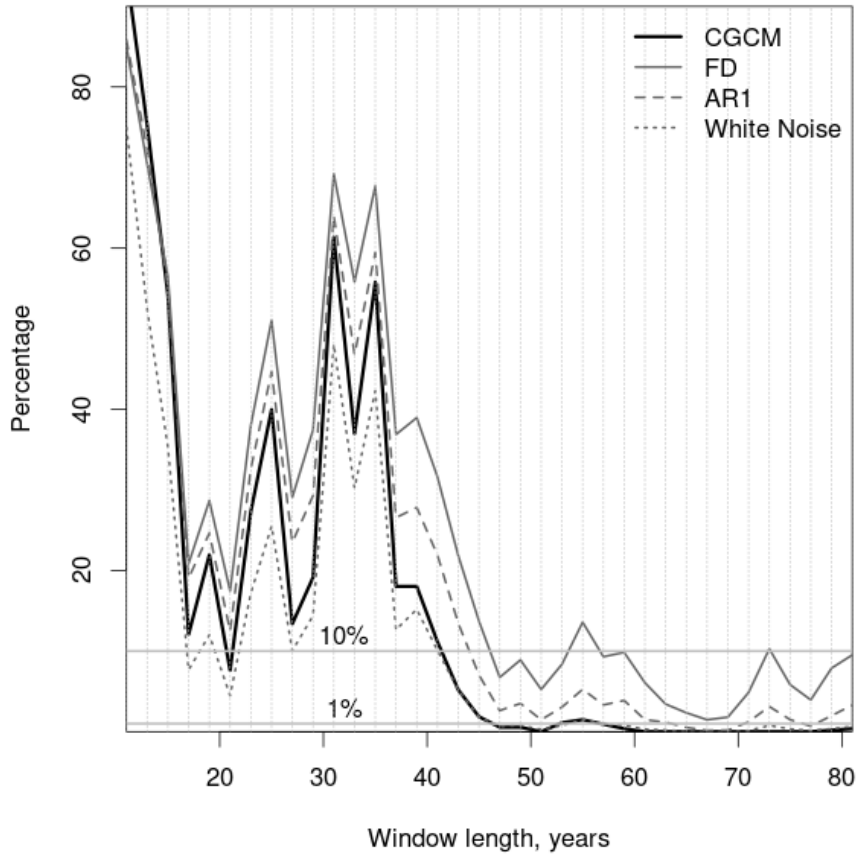


**Figure 5.13 Likelihood of minimum trends versus window length.** Trend exceedance probabilities for block minima NAO trends for a range of window lengths. Probabilities of minimum NAO trend exceedance $q'$ = Pr(min{$Z_{1+K}$, $Z_{2+K}$, … $Z_{n-K}$} ≤ $z$) for window length 11 to 81 years ($K$ in range 5 to 40) in the historical period 1862-2005 using the empirical probabilities from CGCM experiments (black solid line) as relative to the observed (HadSLP2r) minimum trend thresholds $z$. Probability thresholds 10% and 1% are shown as labelled grey horizontal lines, representing *very unlikely* and *exceptionally unlikely* categories of the IPCC likelihood scale.

### 5.4.4 Winter season definition

To test the sensitivity of the CGCM results to the precise definition of the winter season, the analysis has been repeated using HadSLP2r and CGCMs for the extended winter season December-to-March (DJFM) and the individual months December to March for the standardised NAO index. For the DJFM season, the results are consistent with those for DJF, with none of the CGCM simulations reproducing the magnitude of the DJFM observed maximum 31-year NAO trend (period 1862-2005). Breaking down the trend analysis to look at individual winter months, it seems that the inconsistency between the CGCM and observed block maxima winter NAO trends is concentrated in January, the only single winter month for which the observed maximum trend is outside of the CGCM empirical 95% prediction interval (Figure 5.14). January has a considerably higher maximum observed 31-year NAO trend than for the other months, whereas the CGCM maximum trend distributions are unchanged for the different months. This is consistent with findings from Simpson et al. (2018) who showed that there is greater multi-decadal variability for the observed North Atlantic zonal mean wind in late winter than early winter, but that this seasonality is not apparent in the large ensemble of CMIP5 CGCM pre-industrial control experiments they analysed. This issue is also apparent in more recent models as Bracegirdle (2022) showed that in reanalysis datasets the multi-decadal atmospheric variability in the North Atlantic peaks in late winter for NAO and jet indices, while CMIP6 and CMIP5 historical experiments underestimate this variability. This lack of seasonality is also apparent in the CGCM distributions of extreme NAO trends (Figure 5.14) so may relate to the underestimation of extreme winter NAO trends in the CGCMs. This could be related to the ocean-atmosphere interactions being stronger in the real world than in CMIP models, especially in late winter (Gastineau and Frankignoul, 2015; Bracegirdle, 2022).

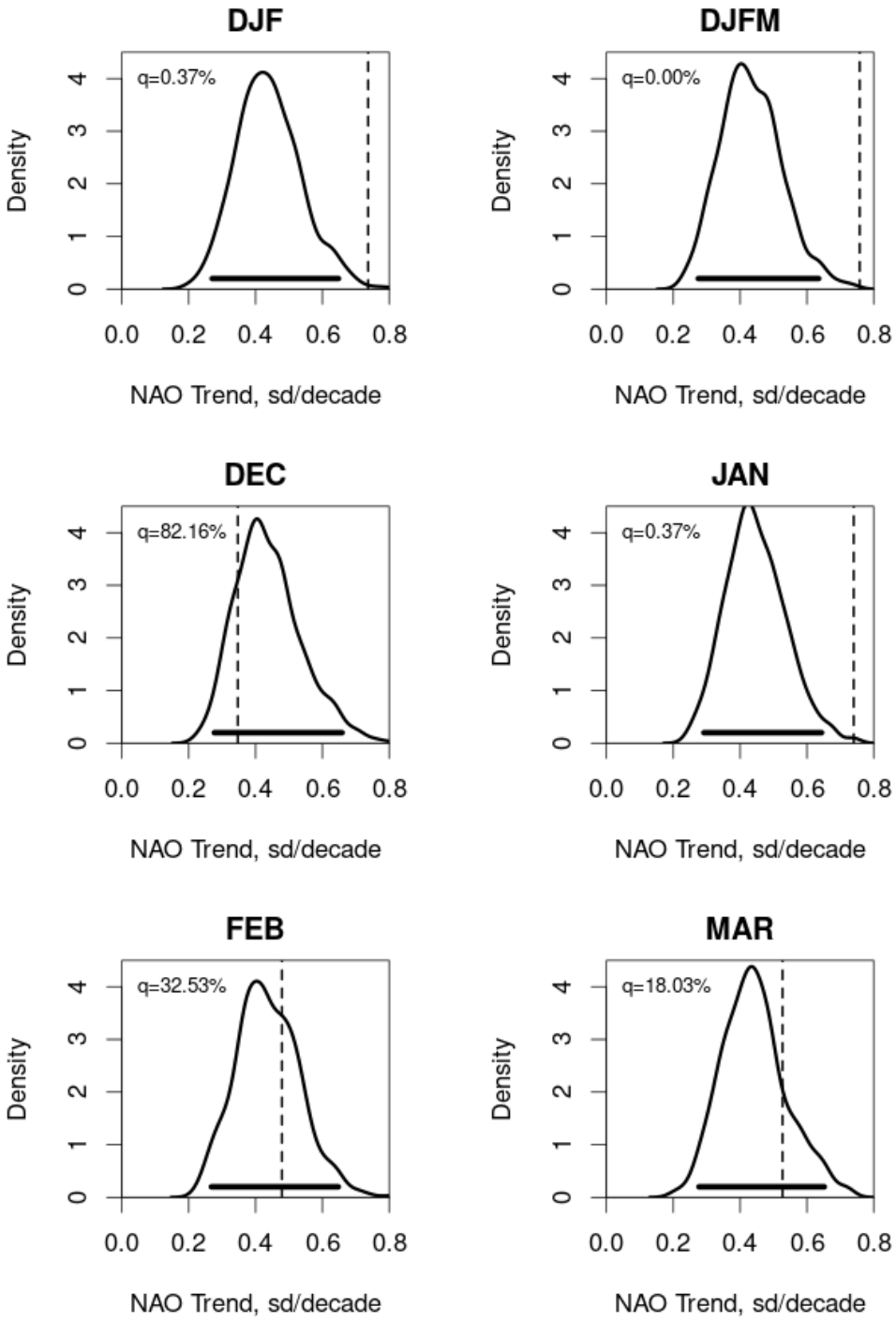


**Figure 5.14 Distribution of block maxima NAO trends for winter months.** The probability kernel density functions for the CGCM maximum 31-year NAO trend values with empirical 95% prediction interval (thick black horizontal lines) for the winter season DJF, extended winter season DJFM and individual months December to March. Observed (HadSLP2r) maximum trends are shown as dashed vertical lines and the percentage of CGCM values above these thresholds displayed as *q*.

## 5.5. Summary

CGCMs underestimate the magnitude of extreme multi-decadal NAO trends, characterising the maximum observed trend as having around a 1 in 200 chance of occurrence (block maxima trend exceedance probability in period 1862-2005), which is only about 10% of the magnitude of the estimates using the stochastic AR(1) and FD model fits from Chapter 4. This suggests that the observed extreme NAO trends are far more likely than is simulated by CGCMs.

The result that CGCMs underestimate extreme multi-decadal NAO trends is fairly insensitive to the choice of window length, as CGCMs underestimate maximum NAO trend exceedance probabilities for window lengths up to at least 71 years compared to the fitted FD and AR(1) models. This result is also insensitive to the precise NAO index definition and to the observation data sets used. This result is insensitive to the winter definition, with similar results for DJF and DJFM winter seasons. When analysing individual months, it seems that the CGCMs are especially deficient in January. This agrees with results from Simpson et al (2018) and Bracegirdle (2022) who identified differences in observed North Atlantic atmospheric variability in late winter compared to early winter, but this element of seasonality is not apparent in CGCMs. Likewise, the distribution of maximum 31-year NAO trends in CGCMs does not demonstrate any seasonality when comparing individual months whereas the observed January maximum trend appears more extreme than for other winter months.

The underestimation of multi-decadal NAO trends in the CGCMs appears as a lack of autocorrelation in the NAO index series compared to the observed NAO. There is also a lack of co-variability between the MSLP trend series at the NAO nodes (Azores and Iceland) for the CGCMs, but this plays a secondary role in the underestimation of extreme NAO trends and likely relates to the underestimation of autocorrelation.

As summarised in Section 2.4.4, likely causes of the deficiencies in CGCM simulations of the NAO are: An underestimation of the natural internal variability of the atmosphere; a failure to realistically respond to external forcing; an underestimation of the ocean-atmosphere interactions. The lack of

autocorrelation in the CGCM NAO series could be related to any or all of these hypotheses. For the CGCMs, there is no consistency in the timing of extreme NAO trends, suggesting that the CGCM NAO response to external forcing is weak. This is an issue that has been identified in recent studies relating to a signal-to-noise paradox 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 CGCM lack of autocorrelation for the NAO is likely related to the signal-to-noise paradox (Zhang et al., 2019; Strommen and Palmer, 2018; Strommen, 2020). Thus, it may be that CGCMs are underestimating a forced component of NAO multi-decadal variability as well as the internal multi-decadal variability component as represented by stochastic processes. Chapter 6 proposes a method to increase the autocorrelation of the CGCM NAO index series and thereby improve the simulation of extreme multi-decadal NAO trends.

# 6. Reddening post-processing of the North Atlantic Oscillation Index in climate models

## 6.1. Introduction

From Equation 3.3 in Section 3.3.1, it can be seen that an increase in the variance of multi-decadal trends in Coupled General Circulation Model (CGCM) simulations of climate indices can be achieved by increasing the variance of the original index series and/or increasing the autocorrelation. It was shown in Section 5.4.1 and Figure 5.11 that the variance of the NAO Anomalised point index series is fairly well captured by the CGCMs, and for the NAO Standardised point index the time series have already been post-processed such that the variance equals one for the CGCMs and observations. However, it was shown in Section 5.3.3 and Figure 5.7 that the lag-1 year autocorrelation of the NAO index series is underestimated by virtually all of the CGCMs. For other climate indices it may be useful to adjust the variance of the original index, which is as simple as the standardisation method mentioned above, but for the NAO it is the lack of autocorrelation that is of interest to correct. This chapter proposes a post-processing technique to increase the autocorrelation in CGCM simulations of the NAO index, so that the CGCMs better match the year-to-year dependence of the observations. Two methods are proposed here, one to increase the short-term dependence of a climate index using the first order auto-regressive (AR(1)) model and one to increase the long-range dependence using the fractional difference (FD) model. Key questions in this chapter are:

- Can the reddening methods enable CGCMs to simulate the observed NAO multi-decadal variability?
- What is the effect of reddening on the CGCM distribution of maximum multi-decadal NAO trends?
- Which method is more effective, the AR(1) or FD reddening?

The reddening methods are outlined in Section 6.2 and tested on random stochastic model processes. In Section 6.3 the reddening methods are applied to CGCM NAO index series from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) and Phase 5 (CMIP5) (Taylor et al., 2012) historical experiments, as used in Chapter 5.

## 6.2. Reddening method

It was shown in Chapter 5 that the CGCM NAO index series underestimated the distribution of extreme multi-decadal trends, which is related to an underestimation of the low frequency variability due to the lack of autocorrelation. To correct for this, a low-pass linear filter approach is proposed to increase the autocorrelation of the NAO index series using filters 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

$$Y^R = \gamma(1 - \varphi B)^{-d} Y \tag{6.1}$$

(Hosking, 1981) where B is the backward shift operator such that $BY(t) = Y(t - 1)$ and $\gamma$ is a constant chosen to retain the variance of the original series. The parameters for this reddening filter are 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. Higher order stochastic models could be considered, but in general, NAO index series are too short to robustly estimate multiple parameters simultaneously. Two special cases are chosen here using $d = 1$ such that the filter is an AR(1) process or $\varphi = 1$ such that the filter is a FD process as these types of process have already been shown to be fairly good at modelling the NAO (Section 4.3) and they only have one parameter.

### 6.2.1 First order auto-regressive process

If it is assumed that the observed and CGCM NAO index series can be approximated as AR(1) processes, but with the CGCM index series underestimating the level of autocorrelation found in the observations, then it is appropriate to apply an AR(1) filter to increase the level of autocorrelation in the CGCM index series. In this case, the CGCM NAO index series can be written as:

$$Y = (1 - \rho_Y B)^{-1} \varepsilon \tag{6.2}$$

where $\rho_Y$ is the lag-1 year autocorrelation parameter for the raw CGCM index series $Y$ and $\varepsilon$ are independent Gaussian variables. The new reddened index $Y^R$ is defined using Equation (6.1) with $d = 1$:

$$Y^R = \gamma(1 - \varphi B)^{-1} Y = \gamma(1 - \varphi B)^{-1}(1 - \rho_Y B)^{-1} \varepsilon. \tag{6.3}$$

Expanding out these brackets shows that for non-zero $\rho_Y$, $Y^R$ is a second order auto-regressive process, AR(2):

$$Y^R = \gamma(1 - (\varphi + \rho_Y)B + \varphi\rho_Y B^2)^{-1}\varepsilon. \quad (6.4)$$

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

$$\rho_R = \frac{\varphi + \rho_Y}{1 + \varphi\rho_Y}\,. \quad (6.5)$$

Thus, for given values of $\rho_R$ estimated from the observed index series and $\rho_Y$ estimated from a CGCM index series, $\varphi$ is found using:

$$\varphi = \frac{\rho_R - \rho_Y}{1 - \rho_R\rho_Y}\,. \quad (6.6)$$

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 $Y$ and the AR(2) process $Y^R$ are given by:

$$Var(Y) = \frac{Var(\epsilon)}{1 - {\rho_Y}^2} \quad (6.7)$$

$$Var(Y^R) = \frac{\gamma^2 Var(\epsilon)}{1 - (\varphi + \rho_Y)^2\frac{(1 - \varphi\rho_Y)}{(1 + \varphi\rho_Y)} - \varphi^2{\rho_Y}^2} \quad (6.8)$$

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

$$\gamma^2 = \frac{1 - (\varphi + \rho_Y)^2\frac{(1 - \varphi\rho_Y)}{(1 + \varphi\rho_Y)} - \varphi^2{\rho_Y}^2}{1 - {\rho_Y}^2}\,. \quad (6.9)$$

It was shown in Chapter 5 that the NAO index series in CMIP5+6 historical experiments have lag-1 year autocorrelation close to zero. For this special case where $\rho_Y \approx 0$, $Y$ is a white noise process and $Y^R$ is an AR(1) process, which means that Equations 6.6 and 6.9 can be simplified to:

$$\varphi \approx \rho_R \quad (6.10)$$

$$\gamma^2 \approx 1 - {\rho_R}^2. \quad (6.11)$$

### 6.2.2 Fractional difference process

If it is assumed that the observed and CGCM NAO index series can be approximated as FD processes, but with the CGCM index series

underestimating the level of long-range dependence found in the observations, then it is appropriate to apply a FD filter to increase the level of autocorrelation in the CGCM index series. In this case, the CGCM NAO index series can be written as:

$$Y = (1 - B)^{-d_Y}\varepsilon \tag{6.12}$$

where $d_Y$ is the difference parameter for raw CGCM index series *Y* (Hosking, 1981). The new reddened index $Y^R$ is defined using Equation (6.1) with $\varphi = 1$:

$$Y^R = \gamma(1 - B)^{-d}Y = \gamma(1 - B)^{-d}(1 - B)^{-d_Y}\varepsilon\,. \tag{6.13}$$

Expanding out the brackets shows that the difference parameters are additive and $Y^R$ is also a FD process:

$$Y^R = \gamma(1 - B)^{-(d+d_Y)}\varepsilon\,. \tag{6.14}$$

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:

$$d = d_R - d_Y\,. \tag{6.15}$$

The scaling parameter γ is again 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 processes *Y* and $Y^R$ are given by:

$$Var(Y) = \frac{Var(\epsilon)\Gamma(1 - 2d_Y)}{\left(\Gamma(1 - d_Y)\right)^2} \tag{6.16}$$

$$Var(Y^R) = \frac{\gamma^2 Var(\epsilon)\Gamma(1 - 2d_R)}{\left(\Gamma(1 - d_R)\right)^2} \tag{6.17}$$

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

$$\gamma^2 = \frac{\Gamma(1 - 2d_Y)\left(\Gamma(1 - d_R)\right)^2}{\Gamma(1 - 2d_R)\left(\Gamma(1 - d_Y)\right)^2}\,. \tag{6.18}$$

For the special case where $d_Y \approx 0$, Equations 6.15 and 6.18 can be simplified to

$$d = d_R \tag{6.19}$$

$$\gamma^2 = \frac{\left(\Gamma(1 - d_R)\right)^2}{\Gamma(1 - 2d_R)}\,. \tag{6.20}$$

### 6.2.3 Example reddening of white noise simulations

For the simple case where the original index series *Y* is a stochastic white noise process, application of the AR(1) or FD reddening methods will convert *Y* into an AR(1) or FD process respectively. Figure 6.1 (a) shows a sample white noise time series (black) of length 144 years with variance equal to one. The variance has been constrained to be identically equal to one by dividing the series by the sample standard deviation of that series. The AR(1) reddening method is applied to this time series using Equations 6.3, 6.10 and 6.11 with lag-1 autocorrelation parameter $\rho_R$ = 0.3. The value 0.3 is chosen as it is large enough to clearly see differences between the white noise and reddened time series, while still being a plausible estimate for the observed NAO index series (Chapter 4). The FD reddening method is applied using Equations 6.13, 6.19 and 6.20 with difference parameter $d_R$ = 0.23 which is related to $\rho_R$ using $d = \rho/(1+\rho)$ (Hosking, 1981). Both reddening methods retain the original timing of peaks and troughs but lead to accentuated clusters of positive or negative index values (Figure 6.1a). Both reddening methods successfully lead to a substantial increase in the absolute magnitude of 31-year moving window trends for both positive and negative trend periods (Figure 6.1b), and so widen the moving window trend distribution in both directions.

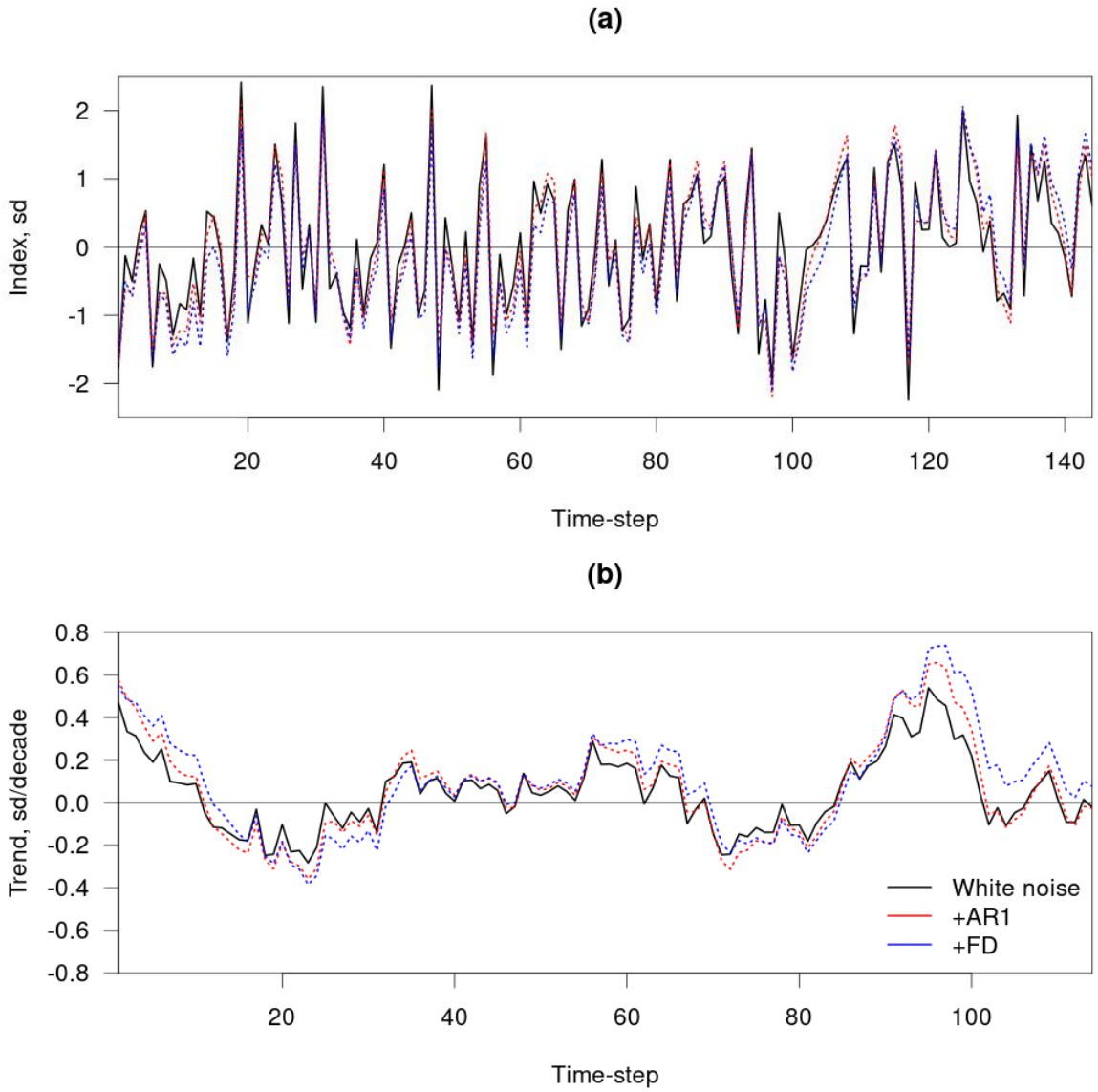


**Figure 6.1: Reddening effect on white noise time series.** (a) A length 144 time series from a stochastic white noise process with variance one (black solid line). Dotted lines show the reddened version of this time series using the AR(1) (red) and FD (blue) method with lag-1 autocorrelation parameter $\rho_R$ = 0.3 and fractional difference parameter $d_R$ = 0.23. (b) Moving window trend series for the time series in (a).

## 6.3. Distribution of NAO trends after reddening climate models

### 6.3.1 Application of reddening methods to climate models

The reddening post-processing methods are applied here to the NAO index series simulated by the CMIP5+6 historical experiments to increase the autocorrelation. The set of CGCM historical experiments assessed is the same as in Chapter 5 (a total of 538 simulations), using the same definition of the NAO index (the standardised difference of the standardised winter mean MSLP series at grid points closest to the Azores and Iceland using the climate period 1862-2005, Equation 4.1). It is assumed that the raw CGCM NAO index series have a lag-1 year autocorrelation of zero (as shown in Chapter 5) such that the reddening process is the same as that applied to the white noise process time series in Section 6.2.3.

Figure 6.2a shows that the distribution of lag-1 year autocorrelation parameter estimates for the 538 raw NAO simulations is centred around zero. 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 4.5.1), thus this interval is also considered when reddening the climate models ($\rho$ 0.02 to 0.32; $d$ 0.02 to 0.24). 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 figures. Figure 6.2b shows the multi-model ensemble mean of lagged autocorrelation estimates for lags up to ten years. 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.

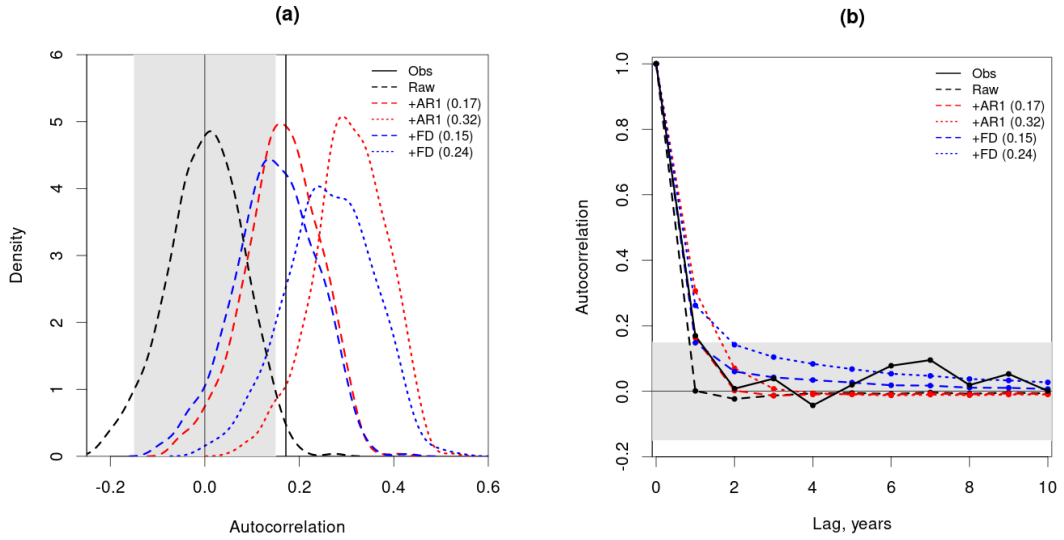


**Figure 6.2: Reddening effect on autocorrelation function of CGCM NAO index**. (a) The distribution of lag-1 year autocorrelation parameter estimates for all the individual raw CGCM historical simulations (black dashed line). The distributions are also shown for the reddened CGCM output using the AR(1) 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.

Figure 6.3 shows the effect of the reddening process on a single CGCM simulation of the NAO index (HadGEM3-GC3.1-MM) using the AR(1) and FD methods with $\rho_R$ = 0.3 and $d_R$ = 0.23 (as in Section 6.2.3). As expected from Section 6.2.3, both reddening methods lead to accentuated clusters of positive or negative phases of CGCM NAO index values while retaining the original timing (Figure 6.3a). This reddening leads to a widening of the CGCM moving window trend distribution in both directions and a substantial increase in the absolute magnitude of both maximum and minimum 31-year NAO trends (Figure 6.3b).

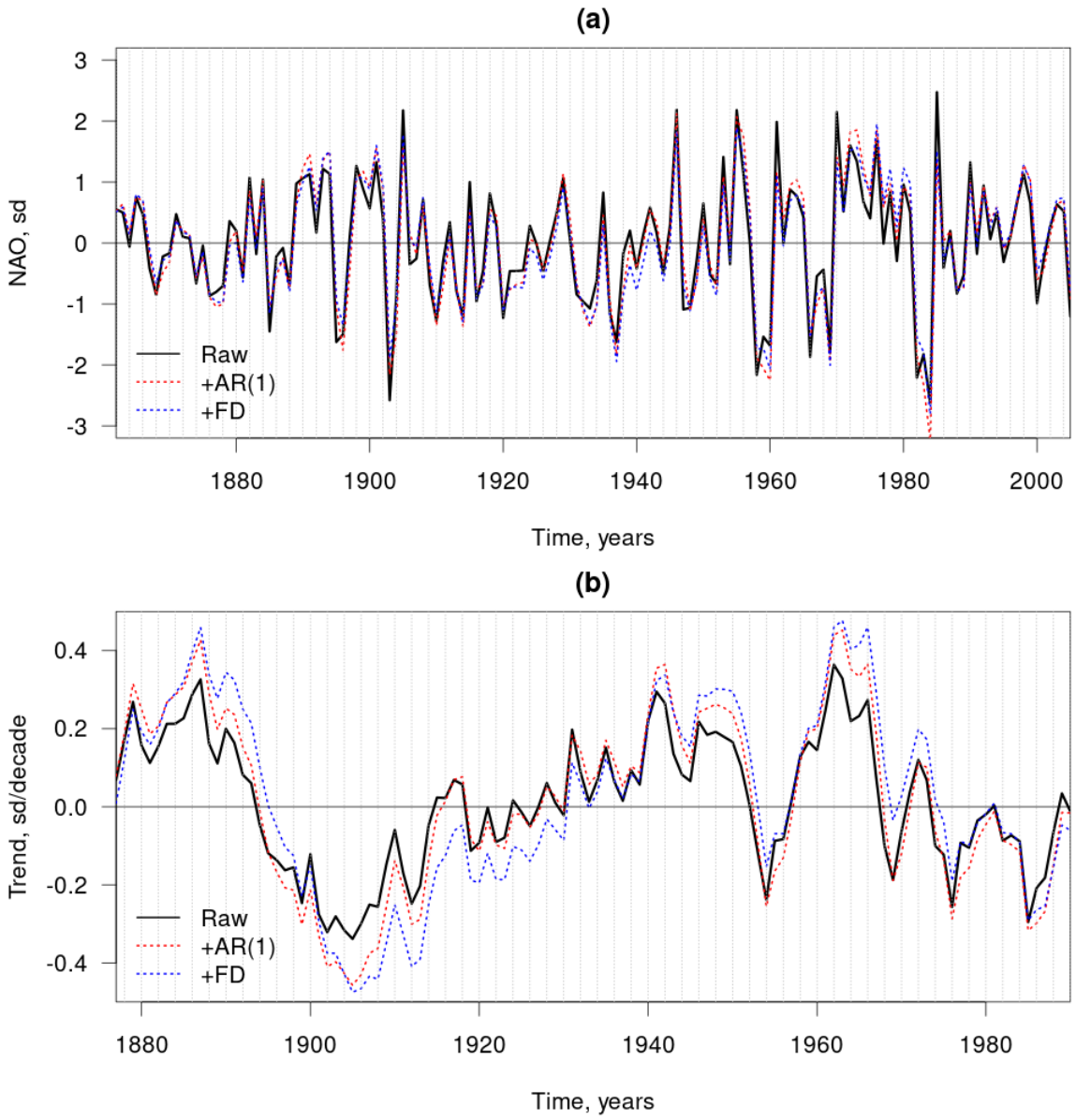


**Figure 6.3: Reddening effect on a single CGCM NAO simulation.** (a) Raw NAO Index time series (black) from a single CGCM simulation (CMIP6 HadGEM3-GC3.1-MM). Dotted lines show the reddened version of this time series using the AR(1) (red) and FD (blue) method with lag-1 year autocorrelation parameter $\rho_R$ = 0.3 and fractional difference parameter $d_R$ = 0.23. (b) The 31-year moving window trend series of the CGCM NAO index series in (a), relative to the central year of the window.

Figure 6.4 shows how the FD reddening method can improve the ability of the CGCM multi-model ensemble to simulate 31-year moving window NAO trends. Figure 6.4a shows the time evolution of the raw CGCM 31-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). In Figure 6.4b and c, two levels of NAO long-range dependence are considered for the FD reddening method: $d_R$ = 0.15 computed from $\rho_R$ = 0.17 which is the lag-1 autocorrelation estimate from Chapter 4 for the observed NAO index series, and $d_R$ = 0.24 computed from $\rho_R$ = 0.32 which is an upper estimate of plausible autocorrelation values from Chapter 4 (97.5th percentile). Over the historical period, the observed 31-year moving window NAO trend series is frequently on the boundary of the raw CGCM ensemble 95% prediction interval. After applying the FD reddening method, the 95% prediction interval and range are inflated to better encompass the observed multi-decadal variability, with higher strengths of reddening leading to a larger increase in the total ensemble standard deviation of moving window trends (Table 6.1). The standard deviation of the observed trend series is about 40% larger than the total standard deviation of the raw CGCM ensemble (Table 6.1). The upper level of autocorrelation is needed for the FD reddened CGCMs to simulate the observed level of trend variability and this also leads to a doubling of the standard deviation of the ensemble mean (Table 6.1). Similar results are found for the AR(1) reddening method but with a smaller increase in standard deviation such that the total ensemble standard deviation of NAO trend series is still a little weak compared to that observed (Table 6.1). The FD reddening method is a useful post-processing tool to enable the CGCM multi-model ensemble to simulate 31-year moving window NAO trend series comparable to those observed, provided the magnitude of reddening is large enough.

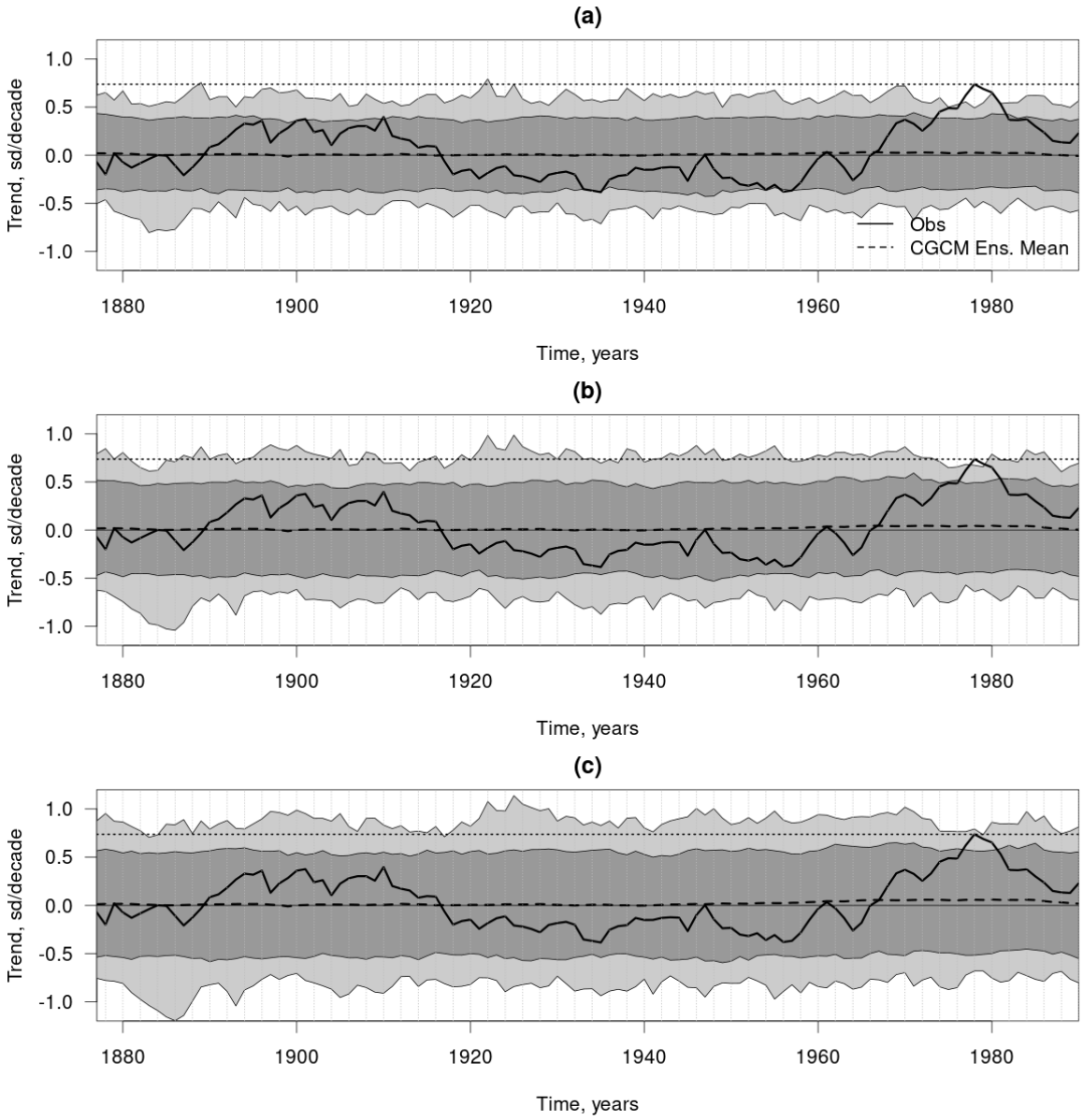


**Figure 6.4 Reddening effect on CGCM ensemble distribution of NAO trend series.** Ensemble mean NAO 31-year trend series for CMIP5+6 simulations using (a) raw model output, or FD reddened output with difference parameter (b) $d_R$ = 0.15 and (c) $d_R$ = 0.24 (black dashed curves), relative to the central year of the window. The light and dark grey shading shows the ensemble range and empirical 95% prediction interval respectively. The observed 31-year trend series is shown for HadSLP2r (black solid curve) with the maximum value identified (horizontal thin dotted black line).

| Data | Total S.D. | S.D. of Ens. Mean | Ens. Range $Z_{1978}$ |
|---|---|---|---|
| Observations | 0.269 | n/a | n/a |
| Raw CMIP | 0.195 | 0.0100 | 1.04 |
| Reddened CMIP: | | | |
| AR(1) $\rho_R$ = 0.17 | 0.227 (16%) | 0.0117 (17%) | 1.20 (16%) |
| FD $d_R$ = 0.15 | 0.248 (27%) | 0.0150 (50%) | 1.31 (27%) |
| AR(1) $\rho_R$ = 0.32 | 0.261 (34%) | 0.0135 (35%) | 1.38 (33%) |
| FD $d_R$ = 0.24 | 0.283 (45%) | 0.0198 (98%) | 1.47 (42%) |

**Table 6.1. Effect of the reddening methods on the CGCM ensemble variability of moving window trends.** The standard deviation of 31-year moving window NAO trend series from CMIP5+6 simulations for the total CGCM ensemble and for the CGCM ensemble mean over the historical period 1862-2005 (in units of sd/decade), and also the ensemble range for the linear trend on the specific time window centred on 1978 ($Z_{1978}$, 1963-1993). CGCM standard deviations 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 $\rho_R$ and $d_R$ in the first column. The percentage increase in standard deviation and range values for the reddened CGCMs relative to the raw CGCM values are shown in brackets. The standard deviation of the observed trend series is shown using HadSLP2r.

For the specific time window 1963-1993 (centred on 1978), the observed 31-year NAO trend is the maximum in the historical record (0.737 sd/decade) and is outside of the raw CGCM spread for this specific 31-year time window, while the CGCM ensemble mean is relatively small (0.0262 sd/decade). After applying the FD reddening method using the upper estimate of $d_R$ = 0.24 the CGCM ensemble range is inflated by 42% (Table 6.1) to encompass the observed 1963-1993 trend (two members out of 538 have a trend greater than that observed), and the multi-model ensemble mean is inflated by 121% (0.0579 sd/decade) though it is still less than 10% of the observed maximum trend. The FD reddening method using the fitted estimate of $d_R$ = 0.15 does not sufficiently widen the distribution of CGCM trends to encompass the observed value, and neither does the AR(1) reddening method (using $\rho_R$ = 0.32 or

$\rho_R$ = 0.17). The FD reddening method improves the ability of the CGCM ensemble to simulate NAO trends comparable to the maximum observed in that specific time window, but only if the magnitude of reddening is large enough.

The effect of the reddening process on CGCM moving window trend exceedance probabilities is shown in Figure 6.5 for 31-year trends. The black solid curve shows the empirical exceedance probabilities and the return period (i.e. the reciprocal of these exceedance probabilities) using the raw CGCM NAO trend series. The observation based empirical exceedance probabilities (using HadSLP2r) are shown as black circles, and these are considerably underestimated by the raw CGCM probability curve (as shown in Section 5.3.2). The pink shading represents an uncertainty range for exceedance probabilities after applying the AR(1) (Figure 6.5a) and FD (Figure 6.5b) reddening methods to the CGCM NAO series using the fitted and upper level estimates of autocorrelation and difference parameters ($\rho_R$ = 0.17 to 0.32; $d_R$ = 0.15 to 0.24). For moderate trend thresholds, the observed exceedance probabilities are fairly well represented by the reddened CGCMs with the observed points situated mainly within the pink shaded interval of reddened CGCM probabilities. For more extreme trend thresholds, the observed empirical probabilities are beyond the pink shaded interval of reddened CGCM probabilities but are closest to the FD reddened curve using $d_R$ = 0.24. Both the AR(1) and FD reddening methods lead to a substantial improvement over using the raw CGCM NAO series which considerably underestimate the exceedance probabilities, but the FD method leads to a better match with the observed distribution.

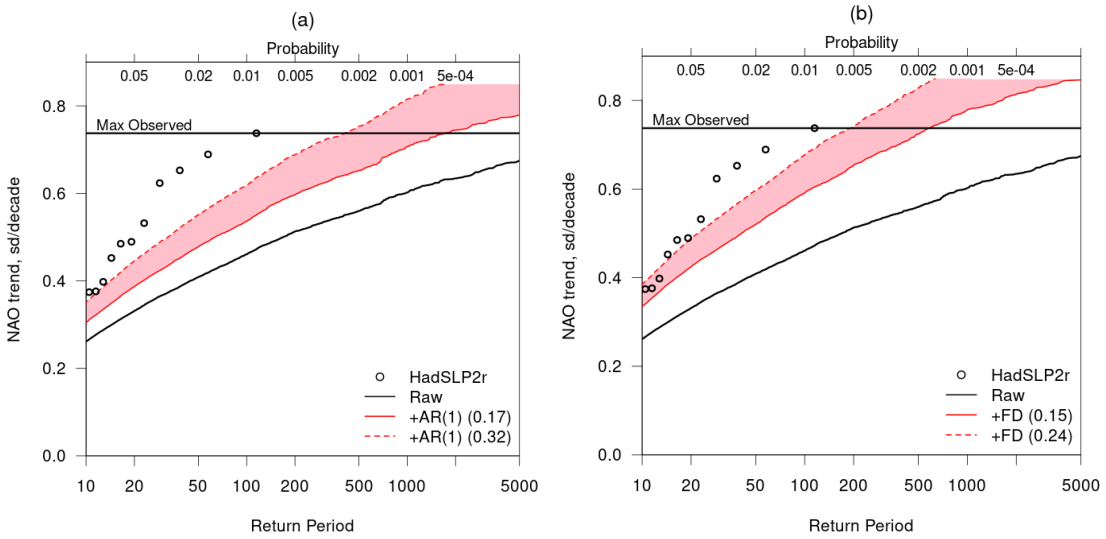


**Figure 6.5: Return Period Plots for NAO moving window trends for reddened CGCMs.** Empirical exceedance probabilities (top axis) and return period (bottom axis) for CGCM 31-year moving window NAO trend series after applying (a) the AR(1) reddening method ($\rho_R$ = 0.17 and $\rho_R$ = 0.32) and (b) the FD reddening method ($d_R$ = 0.15 and $d_R$ = 0.24) shown as red curves identified in the key by solid and dashed lines respectively. Probabilities are also shown for the raw CGCM simulations (black solid curve) and observations (black open circles, using HadSLP2r). The maximum observed trend is identified by the horizontal black line.

## 6.3.2 Distribution of extreme NAO trends after reddening

The effect of the reddening methods on CGCM extreme trend exceedance probabilities are shown in Figure 6.6 for the block maxima 31-year trends in the period 1862-2005. The AR(1) and FD reddening methods increase the CGCM exceedance probabilities in accordance with the level of reddening (Figure 6.6), as would be expected from the stochastic model analysis in Chapter 4.

The CGCM block maxima exceedance probabilities relative to the observed maximum NAO trend threshold (0.737 sd/decade) are shown in Table 6.2 (second column) for the raw CGCM NAO series and for the CGCM NAO series after application of the reddening methods listed in the first column. Note that probabilities in Table 6.2 are presented as percentages to overcome the issue of displaying very small values. Exceedance probabilities based directly on

stochastic processes (from Chapter 4) are shown for comparison (fourth column) using the same levels of lag-1 year autocorrelation and difference parameters displayed in the first column. As shown in Chapter 5, raw CGCM NAO output estimates the observed maximum NAO trend to have around a 1 in 200 chance of occurrence (block maxima in period 1862-2005), which is only about 10% of the probabilities derived from the fitted stochastic AR(1) and FD processes (using $\rho_R$ = 0.17 and $d_R$ = 0.15 respectively). After applying the AR(1) and FD reddening methods to the CGCM NAO series with these same $\rho_R$ and $d_R$ parameters, the exceedance probabilities (relative to 0.737 sd/decade) are increased to around a 1 in 20 chance, matching the stochastic model probabilities from Chapter 4 (Table 6.2 second vs fourth column). If the upper estimates of AR(1) and FD reddening parameters are applied ($\rho_R$ = 0.32 and $d_R$ = 0.24) these exceedance probabilities increase to around a 1 in 10 and a 1 in 6 chance respectively. Application of the reddening methods to the CGCM NAO index series are effective post-processing tools for increasing the CGCM maximum trend exceedance probabilities towards those estimated in Chapter 4 using fitted stochastic processes.

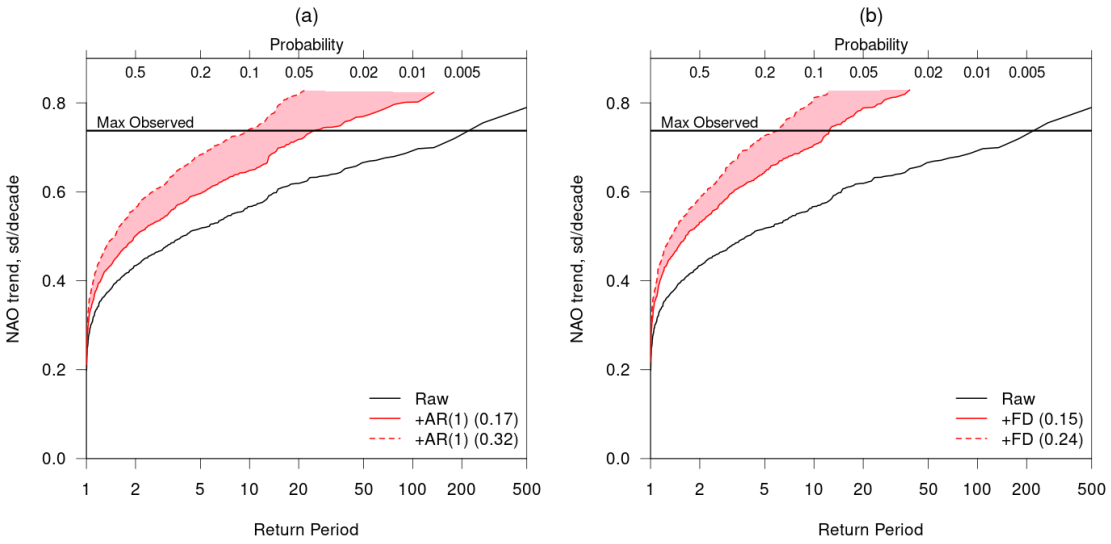


**Figure 6.6: Return Period Plots for NAO maximum trends for reddened CGCMs.** Empirical exceedance probabilities (top axis) and return period (bottom axis) for the maximum 31-year NAO trend in the 144 year historical record (1862-2005) computed from the reddened CGCM trend series after applying (a) the AR(1) reddening method ($\rho_R$ = 0.17 and $\rho_R$ = 0.32) and (b) the FD reddening method ($d_R$ = 0.15 and $d_R$ = 0.24), shown as red curves identified in the key by solid and dashed lines respectively. Probabilities are also shown for the raw CGCM simulations (black solid curve) and the maximum observed trend is identified by the horizontal black line.

| Reddening Method | CGCM, $\rho_Y$ = 0 | | CGCM, $\rho_Y$ <= model | | Stochastic |
|---|---|---|---|---|---|
| Raw/White Noise | 0.372% | (2/538) | 0.372% | (2/538) | 0.600% |
| AR(1) $\rho_R$ = 0.17 | 3.90% | (21/538) | 1.85% | (10/538) | 3.46% |
| FD $d_R$ = 0.15 | 7.99% | (43/538) | 4.27% | (23/538) | 6.06% |
| AR(1) $\rho_R$= 0.32 | 10.2% | (55/538) | 7.43% | (40/538) | 12.30% |
| FD $d_R$ = 0.24 | 16.5% | (89/538) | 10.0% | (54/538) | 19.40% |

**Table 6.2. Trend exceedance probabilities for maximum 31-year trends in the historical period.** Probabilities (shown as percentages) of trend exceedance $q$ = Pr(max{$Z_{1+K}$, $Z_{2+K}$, … $Z_{n-K}$} ≥ $z$) in the historical period 1862-2005, where $z$ is the maximum observed trend (0.737 sd/decade). CGCM probabilities are shown using the raw NAO index series and for the reddened NAO index series where the original CGCM lag-1 year autocorrelation is either assumed to be zero ("$\rho_Y$ = 0") or implied by the sample estimate for each model separately ("$\rho_Y$ <= model"). Probabilities for both AR(1) and FD reddening methods are shown, with autocorrelation and difference parameters as indicated by $\rho_R$ and $d_R$ in the first column. Probabilities for matching stochastic processes are also shown ("Stochastic") and are comparable to the values in Table 4.5.

### 6.3.3 Sensitivity of results to parameter estimates

In Section 6.3.1, the reddening method assumed that the raw CGCM NAO index series have a lag-1 year autocorrelation of $\rho_Y$ = 0. Repeating the analysis to use the individual (non-zero) model sample estimates of $\rho_Y$ does not substantially change the results but leads to slightly lower exceedance probabilities (Table 6.2, third column "$\rho_Y$ <= model" vs. second column). This is because the distribution of CMIP5+6 model $\rho_Y$ estimates is centred around zero but with a fairly large range (Chapter 5). Under the assumption that $\rho_Y$ = 0, the new range of autocorrelation estimates from the reddened simulations will have a similar width to the original range of $\rho_Y$ sample estimates as all simulations are reddened by the same magnitude. In contrast, if the sample estimate of $\rho_Y$ is used for the reddening of each individual simulation, the range of autocorrelation estimates from the reddened simulations will be much narrower than the original range of $\rho_Y$ sample estimates as each simulation is reddened

just enough to achieve an autocorrelation of $\rho_R$. This means that the distribution of maximum trends is likely to be narrower using the $\rho_Y$ sample estimates than when assuming that $\rho_Y$ = 0. Despite these subtle differences, the assumption that $\rho_Y$ = 0 is deemed valid for subsequent analysis 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 $\rho_R$ or $d_R$).

For the convenience of making direct comparisons between the AR(1) and FD reddening methods, the difference parameter $d_R$ has so far been defined based on the observed estimate of $\rho_R$. Estimating $d_R$ directly by fitting a FD process to the observations leads to a slightly lower value (fitted $d_R$ = 0.12 compared to $d_R$ = 0.15 computed from the fitted estimate of $\rho_R$ = 0.17). Applying the FD reddening method with this slightly lower estimate of $d_R$ = 0.12 will lead to maximum trend exceedance probabilities that are slightly lower than when using $d_R$ = 0.15, but the two estimates of $d_R$ are so close that these differences in probabilities will be very small, based on the sensitivity of the variance of trends relative to the magnitude of autocorrelation (Figure 3.1). The direct estimate of $d_R$ could itself be an underestimation of the observed level of long-range dependence as the observed time series is fairly short and the long-range dependence of FD processes mean that a longer time series is needed to robustly estimate the difference parameter. Overall, the sensitivity of maximum trend exceedance probabilities to these methods of parameter estimation is small and does not change the results in this chapter, so it seems reasonable to use the originally proposed reddening method with $d_R$ estimated from $\rho_R$ and the assumption that $\rho_Y$ = 0.

## 6.4. Summary

This chapter has outlined post-processing techniques to increase the autocorrelation in CGCM simulations of climate indices using an AR(1) reddening method and a FD reddening method. Results after application to the NAO index time series show that:

- Application of the AR(1) and FD reddening methods to the CGCM NAO index series lead to a widening of the distribution of moving window

multi-decadal NAO trends that better match the observed multi-decadal variability. After application of the FD reddening method with $d_R$ = 0.24, the total standard deviation of the historical ensemble of 31-year moving window trends is 0.283 sd/decade which is 45% greater than for the raw ensemble and closer to the observed standard deviation (0.269 sd/decade) (Table 6.1). The standard deviation of the ensemble mean is increased by 98%.

- Both of the reddening methods transform the CGCM distributions of maximum multi-decadal NAO trends to match those of the fitted stochastic models from Chapter 4 when applying the same levels of autocorrelation, as would be expected.
- The FD reddening method is more effective than the AR(1) reddening method for enabling the CGCMs to simulate multi-decadal NAO trends as large as those seen in the observations, but a relatively large difference parameter is required ($d_R$ = 0.24) for the CGCMs to simulate moving window trend exceedance probabilities close to those from observations.
- The FD reddening method with $d_R$ = 0.24 widens the CGCM ensemble sufficiently to encompass the maximum observed NAO trend in that specific time window 1963-1993. In this window, the ensemble mean is increased by 121% and the ensemble range is increased by 42% relative to the raw model output (Table 6.1). Use of lower levels of $d_R$ or the AR(1) reddening method with plausible $\rho_R$ values are not sufficient for the CGCM ensemble to encompass the magnitude of the observed trend here.

Relative to the observed maximum 31-year NAO trend threshold (0.737 sd/decade), the fitted stochastic models and the reddened CGCM simulations (applying these fitted levels of autocorrelation) estimate the block maxima 31-year trend exceedance probability (for the period 1862-2005) to be around a 1 in 20 chance which would be termed a *very unlikely* occurrence (less than 1 in 10 chance) using the IPCC likelihood scale (Appendix B.2). This is 10 times the probability based on raw CGCM simulations, which were shown to underestimate trend exceedance probabilities relative to the fitted stochastic model probabilities (Chapter 5). If upper estimates of autocorrelation are used in

the FD reddening method ($d_R$ = 0.24) the exceedance probability is increased to around a 1 in 6 chance, which would be termed an *unlikely* occurrence.

The reddening method has here been assessed for multi-decadal trends in continuous historical simulations for a large-scale seasonal mean index, but it could equally be applied to indices on smaller temporal and spatial scales. This method 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 (Section 6.3.1). The advantage of using CGCMs rather than stochastic processes is that the reddened CGCM NAO index series will still retain dynamical properties of the NAO for example those related to the response to external forcings. This is important for future projections of the NAO as it is of interest to determine how different future forcing scenarios may affect the distribution of extreme multi-decadal NAO trends. However, use of the raw NAO from future projections will likely lead to the same underestimation of multi-decadal NAO variability as was found for the historical experiments. The FD reddening method has been shown to be adequate for adjusting CGCM's to better simulate the observed multi-decadal NAO variability in the past, so it can now be used on climate projections to better simulate multi-decadal NAO variability in the future.

# 7. Future trends in the North Atlantic Oscillation Index

## 7.1. Introduction

This chapter uses Coupled General Circulation Model (CGCM) future climate projections from the Coupled Model Intercomparison Project Phase 6, CMIP6 (Eyring et al. 2016) and the fractional difference (FD) reddening method (Chapter 6) to assess potential changes in the distribution of multi-decadal North Atlantic Oscillation (NAO) trends and extreme trends in the future. Key questions are:

- Does the future radiative forcing scenario effect the distribution of moving window NAO trends in the future?
- Is stationarity a valid assumption for the recent past and future NAO trend series?
- How is the distribution of extreme multi-decadal NAO trends likely to change in the future?
- What are the best near-term projections of NAO multi-decadal trends?

## 7.2 Future climate projections of NAO trends

### 7.2.1 CMIP6 future projection experiments

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 chapter: 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. Figure 7.1 is adapted from the Intergovernmental Panel on Climate Change (IPCC) 6th Assessment Report (AR6), showing how the effective radiative forcing varies according to the different scenarios. 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.

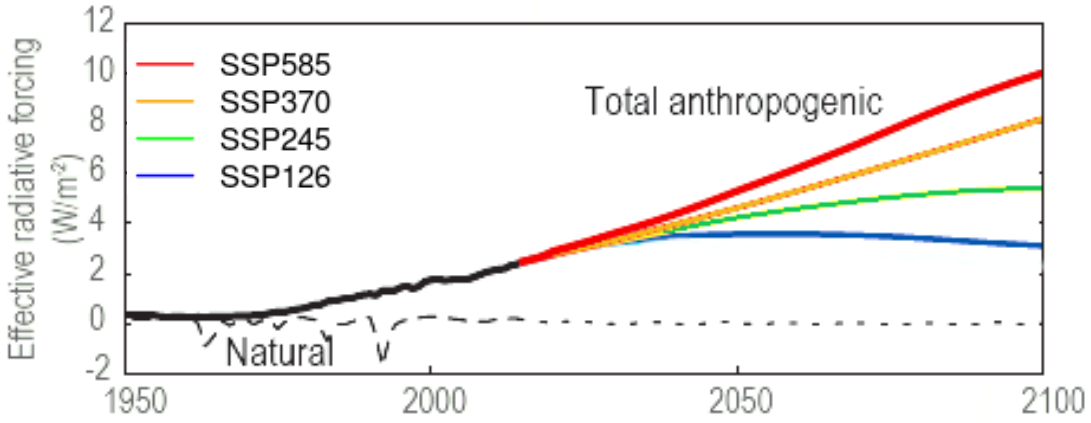


**Figure 7.1: CMIP6 Effective Radiative Forcing.** Time series showing the total anthropogenic part of the effective radiative forcing ($Wm^{-2}$) for historical CMIP6 experiments (black solid curve) and for future scenarios SSP126 (blue), SSP245 (green), SSP370 (orange) and SSP585 (red). The natural part of the effective radiative forcing is shown separately for the historical (black dashed curve) and future (black dotted curve) experiments. This figure has been adapted from the IPCC AR6 Technical Summary to just show the scenarios discussed in this chapter and to have more contrasting colours that match later figures. Source: Arias et al. (2012) Figure TS.4.

For this analysis of future projections, CMIP6 models are only used if they are available for all four future scenarios so that consistent models are used for each scenario and differences in future responses can be attributed to the differences in external forcings rather than model differences. CGCMs from CMIP5 are not included in this analysis so as to avoid having to account for any

differences in the way the future scenarios have been defined. Historical and future CMIP6 simulations are paired by model name and "ripf" code (this identifies the precise model and member setup) so that continuous series of the NAO can be constructed from 1851 to 2100 with winter means and linear trends calculated across the forcing boundary 2014 to 2015. This results in a set of 32 models with a total of 165 members (Appendix A.3).

### 7.2.2 NAO variability in future projections

Figure 7.2 shows the observed winter mean (December to February, DJF) NAO index using HadSLP2r (1851-2021) (Allan and Ansell, 2006) and Figure 7.3 shows the 31-year moving average of this NAO series. The NAO index is based on the standardised difference of the standardised winter mean MSLP series at grid points closest to the Azores and Iceland using the climate period 1862-2005 when calculating the mean and variance (as in previous chapters; Equation 4.1). The same index is calculated from CMIP6 MSLP model output, noting that for years beyond 2014 there is a choice of future scenario.

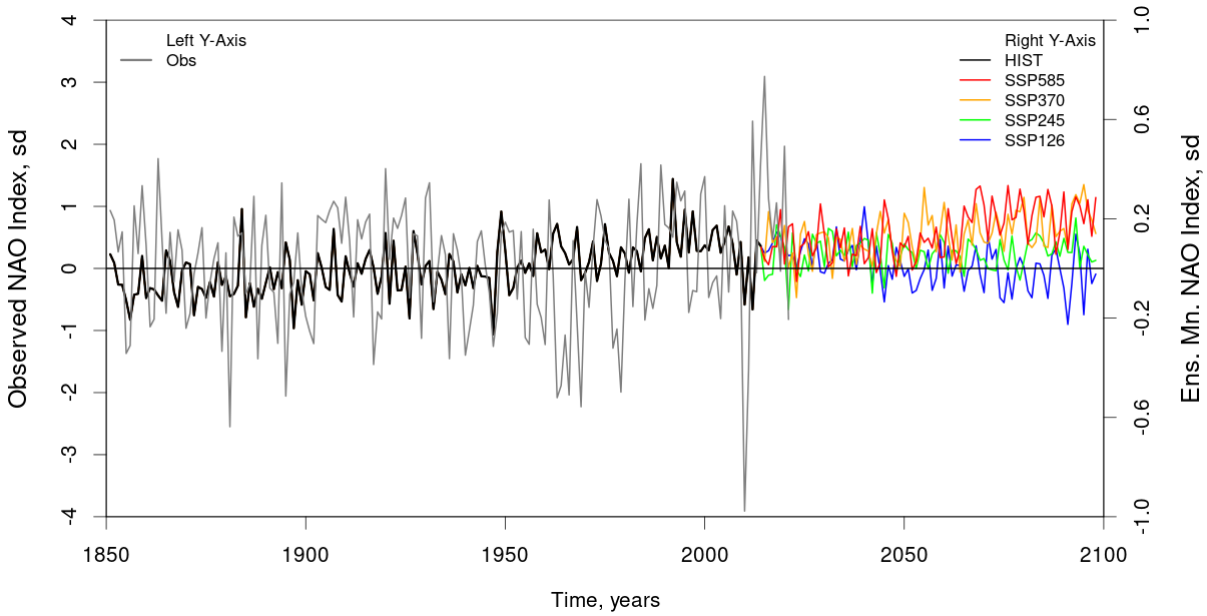


**Figure 7.2: Interannual variability of the NAO.** Time series of observed NAO winter mean index using HadSLP2r (grey line, relative to *left axis*) in units of standard deviation (sd). CMIP6 multi-model ensemble mean series (relative to *right axis with a smaller range*) for historical (black line) combined with four different future SSP scenarios (coloured lines) from 2015 onwards.

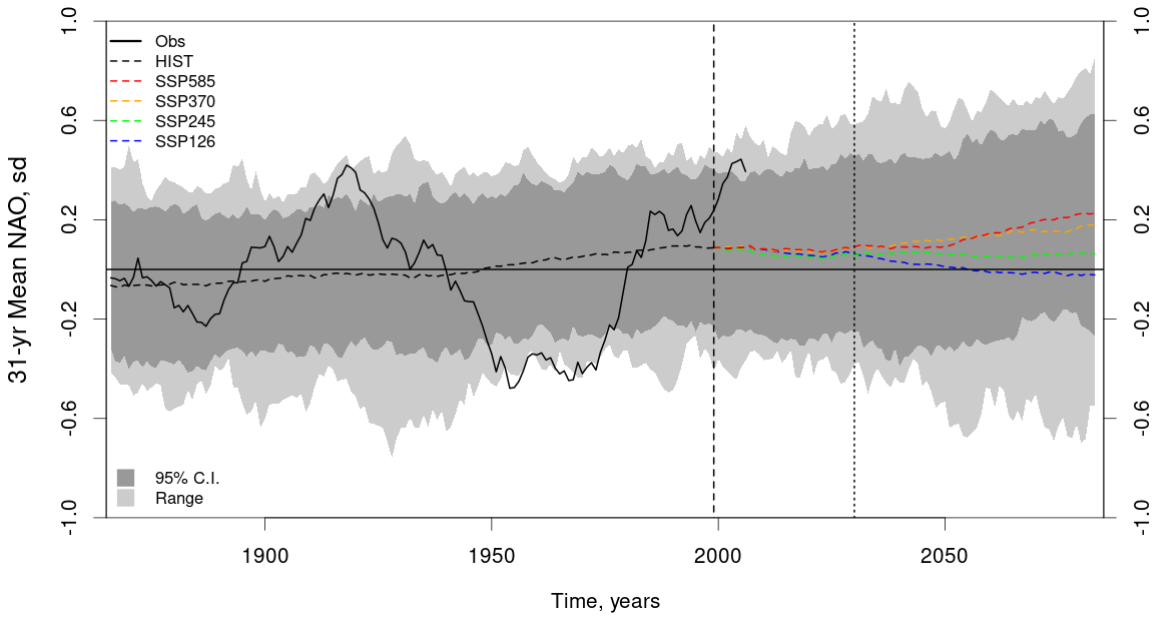


**Figure 7.3: Multi-decadal variability of the NAO.** Time series of 31-year moving window average of HadSLP2r observed winter NAO index (black solid line). CMIP6 multi-model ensemble mean series for historical (black dashed line) combined with four different future SSP scenarios (coloured dashed lines). The multi-model range and 95% prediction interval are shown by light and dark grey shading for the historical experiments and future scenario SSP585. The vertical dashed line marks where the CGCM values begin to diverge due to inclusion of future projections and the vertical dotted line marks from where the values purely use future projections.

The CMIP6 multi-model ensemble mean for the historical NAO series exhibits interannual (Figure 7.2) and multi-decadal variability (Figure 7.3) but, as discussed in Chapter 5, this is much weaker than the observed variability (noting different scales for observations and CGCMs in Figure 7.2). 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 (Figure 7.2 and 7.3). This is especially noticeable for SSP585 which continues with a positive trend similar to that in the historical experiments versus SSP126 which exhibits a negative trend back towards early 20th century values (Figure 7.3), such that by the end of the century the ensemble mean 31-year mean values are 0.223 and -0.0257

respectively (in units of standard deviations) and are significantly different (standard hypothesis test comparing the ensemble mean of paired differences to zero, see Appendix B.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).

Figure 7.3 includes the ensemble spread for the historical and SSP585 experiments, with the total ensemble range shaded light grey and the 95% prediction interval (based on the empirical 2.5th and 97.5th percentiles of the 165 ensemble members) shaded dark grey. The observed 31-year mean NAO index is within the total spread of the CMIP6 ensemble members (independent of time), but for the specific trend windows centred around the 1950s to 1970s, which coincide with the start of the maximum observed NAO trend period 1963-1993, the CMIP6 NAO simulations seem to underestimate the strong negative values observed and the observed index is sometimes outside of the CMIP6 ensemble range.

Figure 7.4 shows the 31-year moving window NAO trend series for the observations and for the CMIP6 multi-model ensemble mean and ensemble spread. The difference in CMIP6 future long term NAO trends influences the multi-decadal moving window trends, with the ensemble mean of the 31-year NAO trend series being generally more positive for the SSP585 experiments than for the SSP126 experiments (Figure 7.4). As discussed in Chapter 5, the observed maximum NAO trend (1963-1993) is clearly unusual in the context of the CMIP6 historical NAO trend series, and it also continues to be unusual in the context of the future NAO projections (Figure 7.4). In Chapters 4 and 5 it was shown that stationarity is a valid assumption for the historical NAO observations and the historical CMIP5+6 NAO simulations. If this stationarity assumption becomes less valid for future NAO variability, this could lead to a shift in the distribution of extreme multi-decadal NAO trends compared to the

historical distribution, and application of the FD-reddening should help to identify this shift.

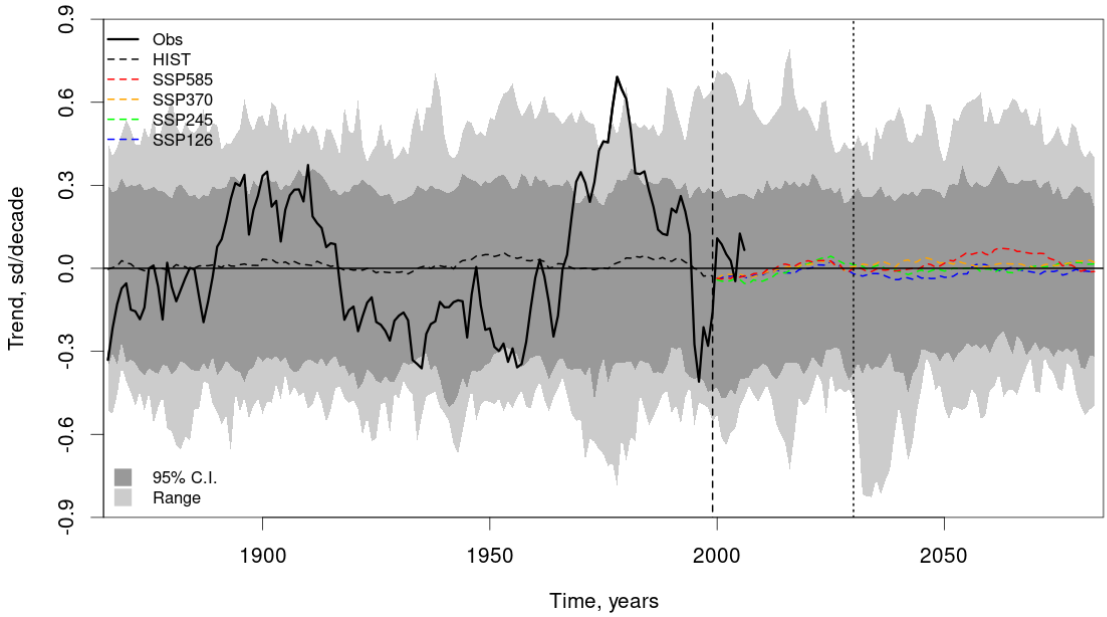


**Figure 7.4: Multi-decadal trends in the NAO.** Time series of observed 31-year moving window linear trends of winter NAO index using HadSLP2r (black solid line). CMIP6 multi-model ensemble mean series for historical (black dashed line) combined with four different future SSP scenarios (coloured dashed lines). The multi-model range and 95% prediction interval are shown by light and dark grey shading for the historical experiments and future scenario SSP585. The vertical dashed line marks where the CGCM values begin to diverge due to inclusion of future projections and the vertical dotted line marks from where the values purely use future projections.

### 7.2.3 Stationarity of NAO trend series in future projections

The stationarity of the CMIP6 NAO trend series can be examined by assessing the ensemble distribution of time mean trends ($\bar{Z}$), as for the analysis in Section 5.3.4. In this chapter, $\bar{Z}$ is calculated over 75-year blocks as the future part of the projections relative to the completion date of this PhD thesis is 2024-2098 (75-years in length). The distribution of $\bar{Z}$ (calculated from the 165 ensemble members) is shown in Figure 7.5a for each of the CMIP6 SSP future experiments. For SSP585 and SSP370 the sample means of $\bar{Z}$ are positive and

significantly different to zero (using a standard hypothesis test at the 5% level, for the sample mean, see Appendix B.3) (Table 7.1). SSP126 has a negative sample mean of $\bar{Z}$ which is also significantly different to zero (Table 7.1). SSP245 has a small sample mean of $\bar{Z}$ that is not significantly different to zero. The greatest shifts are for scenarios SSP126 (negative) and SSP585 (positive). Application of the FD reddening method to these future CMIP6 experiments more than doubles the median of $\bar{Z}$ when using $d_R$ = 0.24 (Table 7.1), and also widens the spread in the distribution of $\bar{Z}$ in both directions (Figure 7.5b) such that the percentage of ensemble members with $\bar{Z}$ greater than zero is left unchanged. These shifts suggest a small non-stationary response to external forcings in the future which is underestimated by the raw CMIP6 NAO index series due to the lack of autocorrelation.

| CMIP6 Experiment | Mean $\bar{Z}$, raw | Median $\bar{Z}$, raw | Median $\bar{Z}$ $d_R$ = 0.15 | Median $\bar{Z}$, $d_R$ = 0.24 |
|---|---|---|---|---|
| SSP126 | -0.0160* | -0.0173 (37%) | -0.0286 (35%) | -0.0364 (32%) |
| SSP585 | 0.0295* | 0.0318 (69%) | 0.0618 (72%) | 0.0805 (74%) |
| Historical | 0.00659 | 0.00726 (58%) | 0.0291 (64%) | 0.0483 (65%) |

**Table 7.1 Long-term trends in future CMIP6 NAO trend series.** The CMIP6 ensemble mean and median of time mean trend $\bar{Z}$ for the future block 2024-2098 for the SSP126 and SSP585 experiments using raw CMIP6 output (column 2 and 3), with * indicating the ensemble mean is significantly different to zero (see Section 7.2.3). The median for FD reddened CMIP6 output using $d_R$ = 0.15 and $d_R$ = 0.24 (columns 4 and 5). $\bar{Z}$ has units of sd/decade. The percentage of ensemble members with $\bar{Z}$ greater than zero is shown in brackets. Values are also shown for the historical experiments for the block 1947-2021, using SSP585 projections for years beyond 2014.

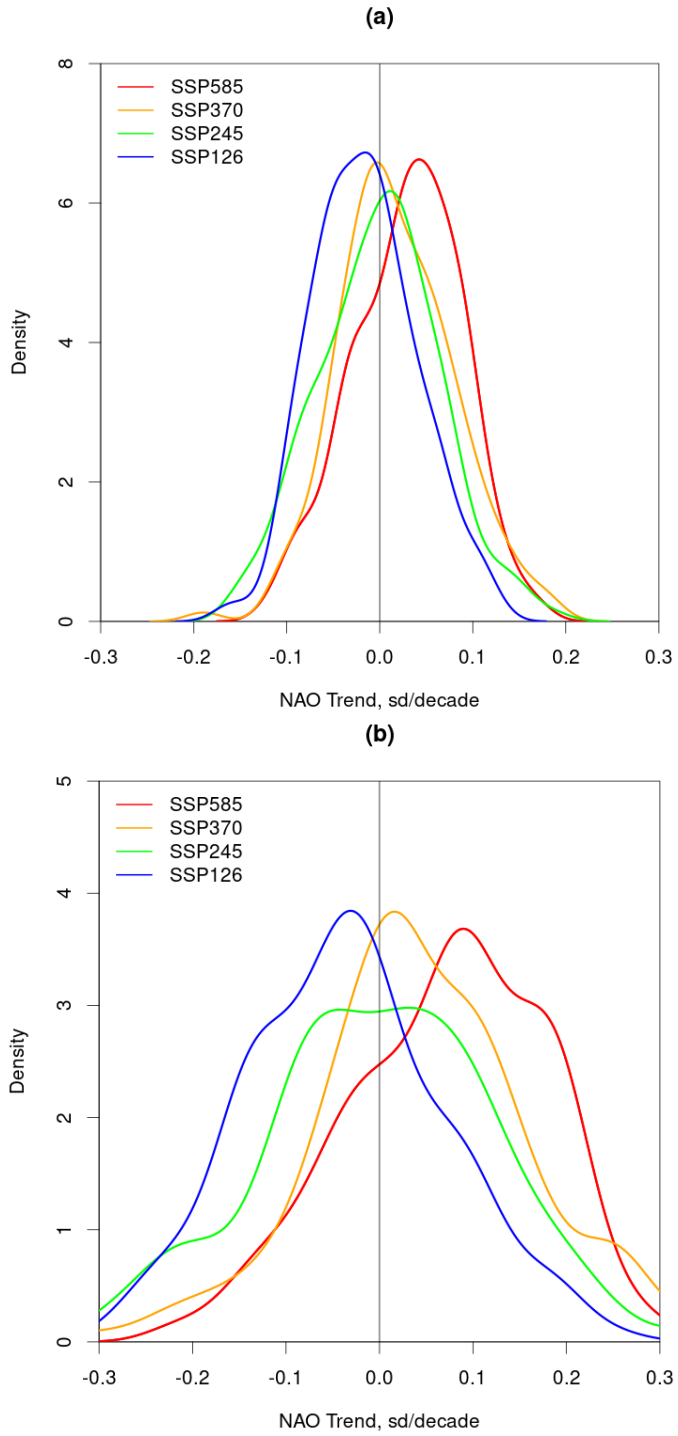


**Figure 7.5 Distribution of time-mean 31-year trends in future NAO series.** The probability density function (kernel density, see Appendix B.6) of time means of 31-year trends ($\bar{Z}$) computed for each of the 165 ensemble members over the 75-year period 2024-2098 is shown for the four different CMIP6 SSP future experiments (coloured curves identified in key) using (a) raw model output and (b) FD reddened model output with $d_R$ = 0.24.

## 7.3 Stationarity of NAO trend series in the recent past

### 7.3.1 Time-mean of NAO trend series

The future period 2024-2098 (75 years) is relatively short compared to the full historical period 1862-2005 (141 years) assessed in Chapter 5. To make direct comparisons of the CMIP6 future experiments to the historical experiments, especially for block maximum trends (discussed further in Section 7.4), it is useful to consider just the most recent 75 years in the historical observation record which for HadSLP2r is the period 1947-2021. A combination of future and historical CMIP experiments are needed for this analysis as the historical CMIP experiments only cover the period up to 2014. Applying the test for stationarity from Section 7.2.3 to CMIP6 historical simulations for the period 1947-2021 (using SSP585 beyond 2014), the ensemble mean of $\bar{Z}$ (0.00659 sd/decade) is not significantly different to zero and only 58% of the raw ensemble members have $\bar{Z}$ greater than zero (Table 7.1). This suggests that the small non-stationary response to external forcings found in Chapter 5 for the period 1862-2005 is not strong enough to be noticeable on the shorter period 1947-2021.

The HadSLP2r observed estimate of $\bar{Z}$ in the period 1947-2021 is 0.178 sd/decade. This is considerably larger than the observed estimate for the historical period 1862-2005 (0.0354 sd/decade) which was computed in Section 4.4.4. To test whether this *observed* $\bar{Z}$ estimate for the more recent period is consistent with a stationary process, it can be compared to the empirical distribution of time mean trends $\bar{Z}$ as estimated from fitted stochastic model simulations using the method from Section 4.4.4, but for the shorter period length of 75 years (trend series of length 45). For the fitted FD process with difference parameter in range (0.15, 0.24) the probability of a time mean trend with absolute magnitude greater than the observed estimate of $\bar{Z}$ = 0.178 sd/decade is in the range 5% to 12%. Applying standard hypothesis testing of this time mean at the 5% level (see Appendix B.3), there is no evidence to reject the assumption of stationarity in the observed NAO index in this more recent period.

The recent observed estimate of $\bar{Z}$ calculated above (0.178 sd/decade) is outside of the range for raw CMIP6 historical experiments (-0.168 to 0.161 sd/decade) for the same time window (combined with SSP585 future experiments beyond 2014). After applying the FD reddening method with $d_R$ in the range (0.15, 0.24), the percentage of reddened ensemble members with $\bar{Z}$ greater than 0.178 sd/decade has the range 2% to 8%. Even after FD reddening of the CMIP6 ensemble, the observed estimate of $\bar{Z}$ is a *very unlikely* occurrence (less than 1 in 10 chance, using the IPCC likelihood scale, Appendix B.2) relative to the CMIP6 distribution of $\bar{Z}$. This is further evidence that the CGCMs underestimate multi-decadal NAO trends, especially in the more recent decades.

### 7.3.2 Longer-term NAO trends

Blackport and Fyfe (2022) identified a related issue when assessing the distribution of 70-year trends in CMIP6 models for the recent historical period. They found that the 1951-2020 linear trend for the NAO is the largest 70-year trend in the historical observation record but only 0.7% of CMIP6 simulations achieve this magnitude of trend for the same period (2 out of 300 members). Similar results are found for the set of 165 CMIP ensemble members used in this chapter: Figure 7.6a shows the 71-year moving window NAO trend series for the HadSLP2r observations and for the CMIP6 ensemble mean for the historical experiments and four future scenario experiments. Shading is used to show the range (light grey) and 95% prediction interval based on the empirical 2.5th and 97.5th percentiles (dark grey) for the historical experiments combined with SSP585 experiments beyond 2014. A 71-year window length has been chosen as the methods described in Chapter 3 onwards are based on odd-number window lengths. The 1950-2020 linear NAO trend is the largest 71-year trend in the historical record (0.186 sd/decade), seen as the second-to-last point on the observed trend series in Figure 7.6a. For this specific time window, the observed 71-year trend is outside of the raw CMIP6 range (Figure 7.6a, light grey shading). Blackport and Fyfe (2022) suggest that this may be due to models underestimating the response to human emissions or underestimating the magnitude of internal multi-decadal variability.

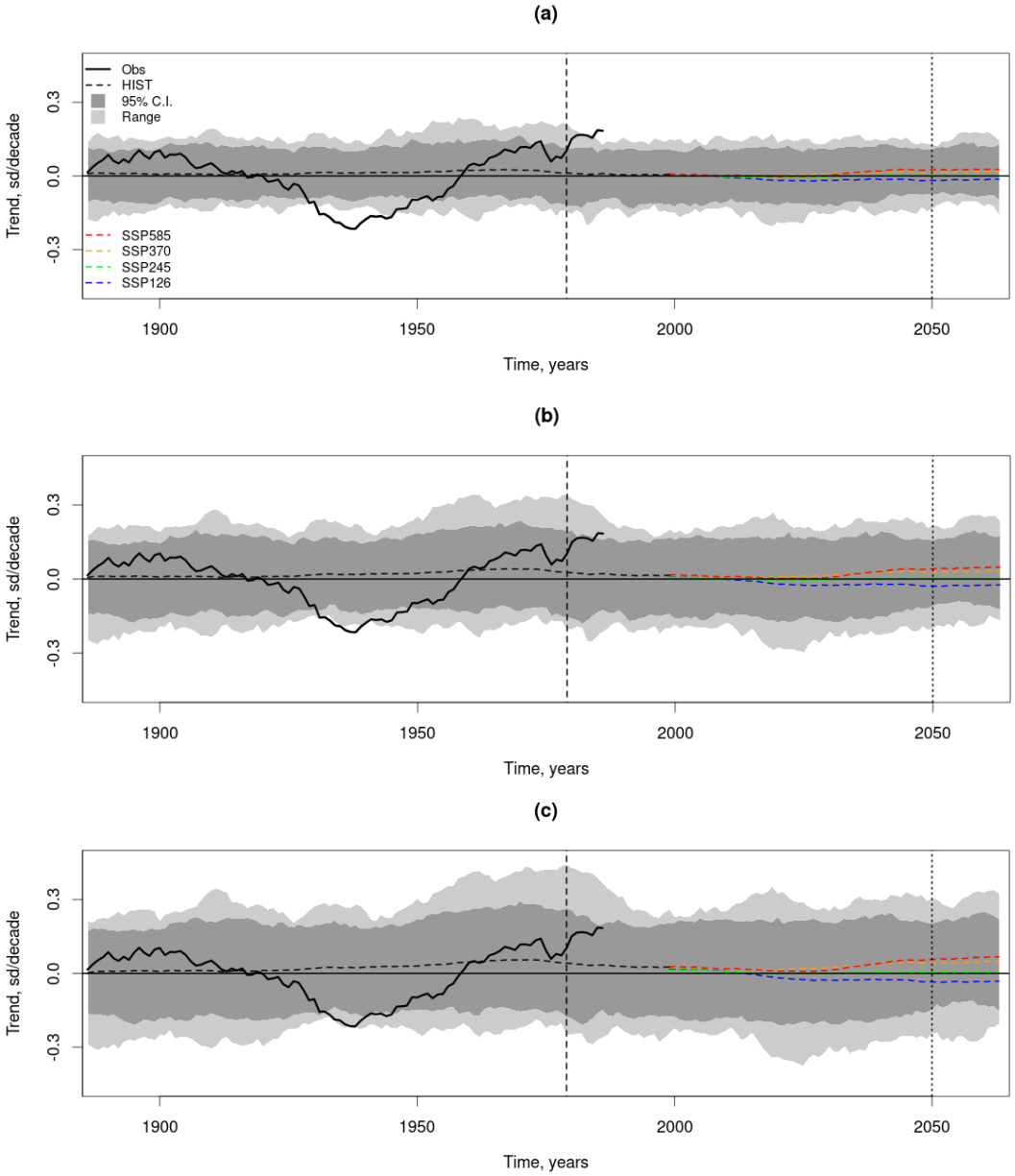


**Figure 7.6: Longer-term trends in the NAO.** Time series of HadSLP2r 71-year moving window winter NAO trends (black solid line). CMIP6 ensemble mean series for historical (black dashed line) combined with four future SSP scenarios (coloured lines), with shading for the multi-model range (light grey) and 95% prediction interval (dark grey) (using SSP585 beyond 2014). The vertical dashed line marks where the CGCM values begin to diverge due to inclusion of future projections and the vertical dotted line marks from where the values purely use future projections. (a) uses raw CMIP6 output, (b) and (c) use FD reddened CMIP6 output with $d_R$ = 0.17 and 0.24 respectively.

Application of the FD reddening method to the CMIP6 NAO series (using $d_R$ = 0.17 and 0.24) leads to a widening of the distribution of 71-year trends that increases with the value of $d_R$ (Figure 7.6b, c). This increases the CMIP6 exceedance probability, for a 1950-2020 NAO trend greater than that observed, to 0.7% with $d_R$ = 0.17 (1 member out of 165) and 3.64% with $d_R$ = 0.24 (6 members out of 165). Even after strong levels of FD reddening, the observed 1950-2020 NAO trend is greatly underestimated by the CMIP6 models and would be termed a *very unlikely* occurrence using the IPCC likelihood scale (Appendix B.2).

The CMIP6 model underestimation of linear NAO trends in the specific 1950-2020 window isn't evidence on its own for non-stationarity in the observed trend series as the window has been pre-selected as an extreme trend window (as discussed in Chapter 3). Also of note is the *minimum* observed 71-year trend (1903-1973) which has a larger absolute magnitude (-0.215 sd/decade) than the *maximum* 71-year trend and is also outside of the raw CMIP6 range for that specific time window (Figure 7.6a). Even after applying the upper estimate of FD reddening ($d_R$ = 0.24) to the CMIP6 NAO series, only 2 members out of 165 simulate a trend more negative than the observed value for the 1903-1973 window (probability 1%) (Figure 7.6c). For these extreme 71-year NAO trends, it is necessary to apply the extreme trend methodology from Chapter 3 and utilise the full historical period of the CMIP6 experiments to better understand whether these extreme trends can be explained by natural internal variability within a stationary climate or whether there is evidence of an externally forced trend in the NAO series.

The observed 1950-2020 and 1903-1973 NAO trends are the block *maximum* and *minimum* 71-year NAO trends respectively in the historical block of length 160 years (1862-2021). The CMIP6 exceedance probability for a maximum 71-year trend above the observed maximum threshold (0.186 sd/decade) in the 160-year block is 4% when using the raw model output, which makes it a *very unlikely* occurrence (IPCC likelihood scale, Appendix B.2). For the reverse case, the observed minimum 71-year NAO trend (-0.215 sd/decade) is outside of the raw CMIP6 range of minimum trends (an *exceptionally unlikely* occurrence). The maximum (minimum) 71-year trend windows start (end)

around the 1950s to the 1970s, which was the period noted in Section 7.2.2 where the observed 31-year average NAO index is below the raw CMIP6 model spread, whereas the observations are well within the model range elsewhere in the historical record. Application of the FD reddening method to these CMIP6 simulations (1862-2021) with $d_R$ in the range (0.15, 0.24) considerably increases the NAO multi-decadal variability increasing the maximum trend exceedance probabilities to be in the range 27% to 49%. For the reverse case, minimum 71-year trends below -0.215 sd/decade, probabilities are increased to be in the range 4% to 12%, which would be termed an *unlikely/very unlikely* occurrence. When considered in the context of extreme 71-year trends, the 1903-1973 minimum NAO trend is more unusual than the 1950-2020 maximum NAO trend relative to the CMIP6 extreme trend distributions.

The FD reddening method with $d_R$ = 0.24 enables CMIP6 models to easily simulate maximum 71-year trends as large as the maximum observed (exceedance probability 49%), but these FD reddened models still underestimate the magnitude of minimum 71-year trends (probability 12% for minimum trends less than the minimum observed). This suggests that the observed maximum 71-year NAO trend may be underestimated in CMIP6 models due to an underestimation of the autocorrelation in the NAO index which leads to an underestimation of the magnitude of internal multi-decadal variability. However, the observed minimum 71-year NAO trend is still underestimated in CMIP6 models after FD reddening, which suggests that a higher level of reddening may be needed, or that the CMIP6 models are underestimating or missing processes that are needed to simulate such a negative trend.

## 7.4 Distribution of NAO extreme trends in future climate projections

The distribution of maximum NAO trends is sensitive to the total block length, i.e. the period over which the trend series is computed, as well as the window length for which the individual trend values are estimated (Chapter 3). The future period is constrained to be 75-years long (2024-2098), so to make direct comparisons to the distribution of maximum trends in the historical period the

most recent 75-year block available from the observed HadSLP2r dataset is chosen: 1947-2021 (as in Section 7.3.1). Figure 7.7 shows the raw CMIP6 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). All four SSP experiments fail to simulate a maximum 31-year NAO trend as large in magnitude as that observed for the window 1963-1993 when using the raw model output. The distributions of maximum NAO trends for the four future SSP experiments do not show a noticeable shift in location of the centre of the distribution compared to the historical distribution, but the SSP585 distribution generally has the fattest upper tail while the SSP126 distribution has the least fat upper tail (Figure 7.7). The biggest differences in the distribution of future time mean trends ($\bar{Z}$) were also for experiments SSP585 (positive median) and SSP126 (negative median) (Figure 7.5), so it is these contrasting scenarios which will be the focus of the rest of this chapter.

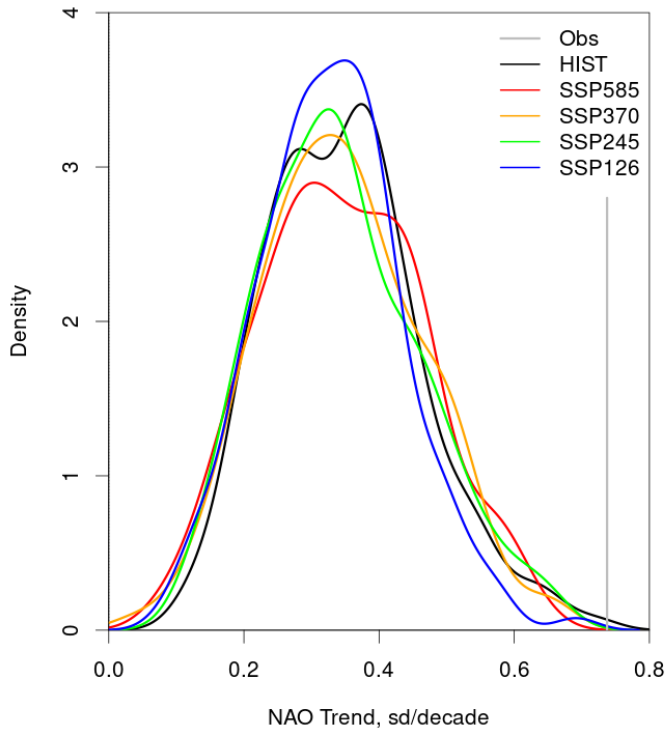


**Figure 7.7 Historical and future distributions of maximum 31-year NAO trends.** The probability density function (kernel density, see Appendix B.6) of maximum 31-year NAO trends in the 75-year period 2024-2098 for the four different CMIP6 SSP future experiments using raw model output (coloured curves). The distribution of historical CMIP6 maximum trends is shown for the 75-year period 1947-2021 (black curve), using SSP585 beyond 2014 and the maximum observed (HadSLP2r) trend for this period is shown in grey (vertical line).

Figure 7.8 uses box and whisker plots to display the raw and FD reddened ($d_R$ = 0.24) CMIP6 distribution of maximum 31-year NAO trends with the multi-model ensemble median value shown as a central line within a box representing the empirical 25th to 75th percentiles and whiskers representing the total range. After FD reddening with $d_R$ = 0.24, the historical and future CMIP6 distributions of maximum NAO trends are shifted in location towards more positive values than for the raw CMIP6 simulations and the variance is increased leading to a widening of the distribution (Figure 7.8).

The CMIP6 exceedance probabilities for 31-year block maximum NAO trends relative to the observed threshold (0.737 sd/decade) are shown in Table 7.2 for

the historical experiments (block 1947-2021). Note that the exceedance probabilities for the historical experiments are smaller here (Table 7.2) than in Chapter 6 (Table 6.1) as the distribution of maximum NAO trends is sensitive to the total block length within which the trend series is computed (75 years in Table 7.2, 144 years in Table 6.1). For the raw CMIP6 experiments in the recent historical period, none of the ensemble members have a maximum trend greater than the maximum observed (0.737 sd/decade). The application of the FD reddening method increases this CMIP6 probability estimate to around 10% (Table 7.2, Figure 7.8), similar to the results from Chapter 6 (which used a longer historical period and larger multi-model ensemble). Even after FD reddening of the CMIP6 models, the observed maximum trend is still estimated to be a *very unlikely* occurrence (IPCC likelihood scale, Appendix B.2) in the historical period.

| | P(Max Trend > 0.737) | | P(Min Trend < -0.737) | |
|---|---|---|---|---|
| Experiment | FD (0.15) | FD (0.24) | FD (0.15) | FD (0.24) |
| Historical | 3.03% | 10.3% | 2.42% | 6.67% |
| SSP126 | 0.606% | 2.42% | 4.85% | 10.3% |
| SSP585 | 3.03% | 6.67% | 1.21% | 4.24% |

**Table 7.2 CMIP6 Extreme NAO trend exceedance probabilities.** Exceedance probabilities (as percentages) for CMIP6 maximum and minimum 31-year NAO trends in the 75-year periods 1947-2021 (historical) and 2024-2098 (future SSP), relative to +/-0.737 sd/decade respectively (the magnitude of the maximum observed 31-year NAO trend). Probabilities are shown using the FD reddened model output with difference parameter $d_R$ = 0.15 or 0.24 as specified in brackets.

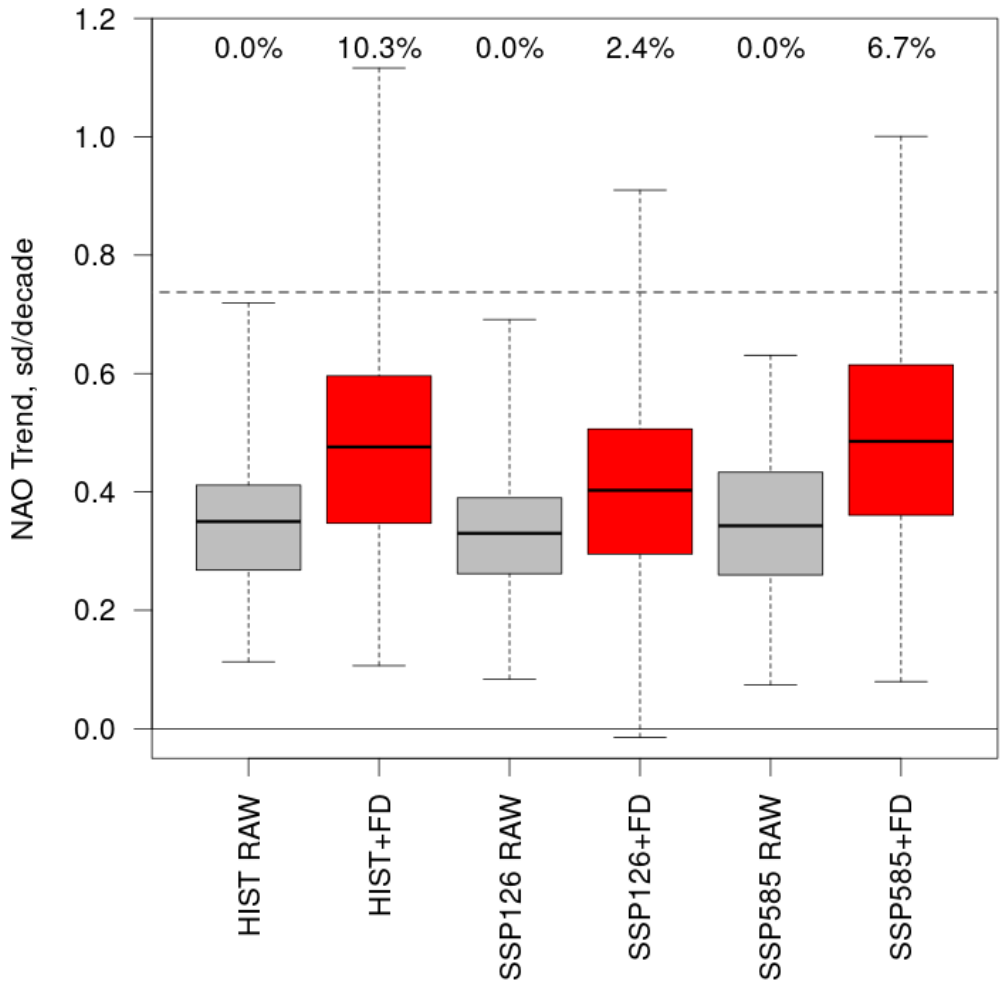


**Figure 7.8 Effect of reddening on distribution of future maximum NAO trends.** Box and whisker plots of the distribution of maximum 31-year NAO trends in the 75-year period 2024-2098 for the SSP126 and SSP585 future experiments using raw model output (grey box plots) and after applying the fractional difference ("+FD") reddening method with $d_R$ = 0.24 (red box plots), displaying the median line in the centre of the 25 to 75 percentile box and total range whiskers. Box and whisker plots for the historical (HIST) CMIP6 maximum trends are shown for the 75-year period 1947-2021, using SSP585 beyond 2014. The observed maximum 31-year trend is shown as the dashed horizontal line, and the percentage of model members exceeding this threshold is displayed above each box plot.

Under future scenario SSP585, the FD reddened distribution of maximum NAO trends is similar to that of the historical period for the same set of models (Figure 7.8). Using the uncertainty range for $d_R$ of (0.15, 0.24) from Chapter 6, the likelihood of seeing a trend of similar magnitude to the 1963-1993 observed NAO trend in the future is 3 to 7%, which is similar to the estimate of 3 to 10% for the recent 75-year historical period (Figure 7.8; Table 7.2). The occurrence of such large magnitude NAO trends in the future could potentially lead to decades where the climate warming signal is considerably enhanced in the northern extra-tropics (Iles and Hegerl, 2017), thus it is concerning that the exceedance probability of such occurrences is greatly underestimated when using raw CGCM output (Figure 7.8).

For scenario SSP126, where the time mean NAO trend is slightly negative, the FD reddened distribution of maximum NAO trends has a total range that is similar to that of scenario SSP585, but the whole distribution is shifted towards slightly smaller maximum trend values (Figure 7.8). This shift in the SSP126 maximum trend distribution leads to exceedance probabilities (relative to threshold 0.737 sd/decade) that are smaller than for SSP585: 0.6% to 2% (SSP126) compared to 3% to 7% (SSP585) for $d_R$ in the range 0.15 to 0.24 (Figure 7.8; Table 7.2). This shows that the future radiative forcing scenario can cause a small shift in the distribution of maximum NAO trends, related to the shift in time mean trend ($\bar{Z}$), while keeping the ensemble spread fairly unchanged.

To consider the reversed case where an extreme negative NAO trend might partially offset the climate warming signal, the future CMIP6 distribution of *minimum* trends can be assessed. In this case the probability of seeing a negative 31-year trend of absolute magnitude comparable to the maximum observed (-0.737 sd/decade), using FD reddened CMIP6 experiments, is higher for the future SSP126 experiments (exceedance probability 5% to 10% using $d_R$ in the range 0.15 to 0.24) than for the future SSP585 experiments (probability 1% to 4%). This relates to the distribution of $\bar{Z}$ having a slightly negative average value for SSP126 and a slightly positive average value for SSP585 experiments (Table 7.1). The partial enhancement of northern extra-tropical regional climate warming discussed in Iles and Hegerl (2017) is more likely

under SSP585 than SSP126 whereas the partial offsetting of regional climate warming is more likely under SSP126 than SSP585. At the time of writing, the SSP585 scenario seems more realistic than the SSP126 scenario. For the up-coming 75 years under this “business as usual” scenario, the probability of extreme positive 31-year NAO trends large enough to enhance regional climate warming is similar to the probability estimated for the previous 75 years. Without the FD reddening method these probabilities would be greatly underestimated.

## 7.5 Near-term NAO trend projections

It is of interest to make near-term forecasts of NAO trends specifically for the next 31-years, as well as predictions of extreme trends over the coming decades. For forecasts of regional climate change, a key timescale is out to the middle of the 21$^{st}$ century, for example the UK Climate Projections (UKCP, Lowe et al., 2018) issue forecasts 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 7.9 uses box and whisker plots to show the CMIP6 distribution of 31-year NAO trends for the specific future period 2024-2054 (the nearest 31-years in the future relative to the completion date of this thesis) for future scenarios SSP126 and SSP585 using raw model output (grey boxes) and FD reddened output with difference parameter $d_R$ = 0.24 (red boxes). 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.

The distribution for the maximum observed NAO trend window 1963-1993 is also shown, using the CMIP6 historical experiments. The raw CMIP6 historical experiments for the 1963-1993 NAO trend have an ensemble median close to zero and the ensemble range does not encompass the observed trend value (Figure 7.9). FD reddening with the upper estimate of $d_R$ = 0.24 is needed to widen the distribution sufficiently to encompass the observed trend, though still only one member out of 165 has a trend exceeding this observed value (similar

to the Chapter 6 results of 2 members out of 538 for the full CMIP5+6 ensemble). It is assumed that this level of FD reddening will also be needed for realistic future projections of NAO trends.

For future scenario SSP585, the raw ensemble mean trend (2024-2054) is very close to zero and the FD reddening method (with $d_R$ = 0.24) reduces the magnitude of the mean trend by 91% so it is even closer to zero (Table 7.3). For scenario SSP126 the raw ensemble mean trend for 2024-2054 is slightly negative and this is accentuated after applying the FD reddening method, inflating the magnitude of the ensemble mean trend by 33% (Table 7.3). The SSP585 and SSP126 distributions have a similar ensemble range for the raw data. Combining the SSP126 with SSP585 experiments, the multi-model ensemble total range of possible NAO trend values for 2024-2054 is (-0.901, 0.847) sd/decade after FD reddening and this range is 52.4% greater than for the raw CMIP6 model output. These near-term projections of NAO trends for 2024-2054 are not overly sensitive to the choice of radiative forcing scenario, with ensemble mean trends close to zero and large uncertainty across the ensembles, though the SSP126 experiments predict slightly more negative trends compared to the SSP585 experiments.

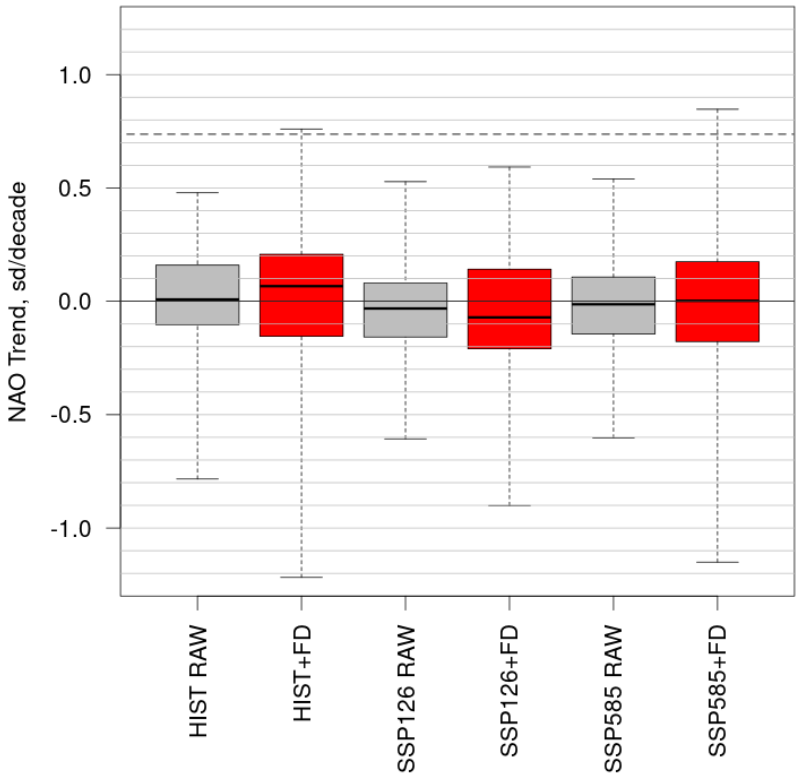


**Figure 7.9 Near-term NAO trends.** Box and whisker plots for CMIP6 historical ("HIST") 1963-1993 NAO trends and for CMIP6 future 2024-2054 NAO trends using scenarios SSP126 and SSP585. Grey box plots use the raw CMIP6 model output (median line in centre of 25 to 75 percentile box; total range whiskers). Red box plots use CMIP6 model output after applying the fractional difference reddening method ("+FD") with $d_R$ = 0.24. The observed 1963-1993 NAO trend is shown as the dashed horizontal line.

| Model | Ensemble Mean | 95% P.I. | Ensemble Range |
|---|---|---|---|
| SSP585 Raw | -0.00977 | (-0.374, 0.379) | (-0.603, 0.540) |
| SSP585 FD | -0.000853 | (-0.518, 0.547) | (-1.15, 0.847) |
| SSP585 % change | -91.3% | 41.4% | 74.7% |
| SSP126 Raw | -0.0402 | (-0.481, 0.362) | (-0.607, 0.528) |
| SSP126 FD | -0.0534 | (-0.632, 0.458) | (-0.901, 0.592) |
| SSP126 % change | 32.7% | 29.3% | 31.6% |

**Table 7.3. Near-term NAO trend projections.** The ensemble mean, 95% prediction interval (P.I.) and range of near-term projections for the 2024-2054 NAO trend in CMIP6 SSP585 future experiments for raw model output and for FD reddened output with $d_R$ = 0.24 (in units of sd/decade). The percentage increase (or decrease) in the mean and the length of the P.I. and range are shown for the reddened CGCMs relative to the raw CGCM values.

In the interest of equating the magnitude of projected NAO trends to actual changes in MSLP, the near-term NAO trend projection is shown in Figure 7.10 for the alternative NAO *anomaly* index definition based on the simple difference between pressure at the Azores and Iceland, leading to NAO trend units of hPa/decade (Equation 5.2). In this case the observed 1963-1993 NAO trend is displayed for C20C (5.59 hPa/decade) rather than HadSLP2r as the latter is known to underestimate MSLP variability (e.g. Semenov et al., 2008; Section 5.4.1). Figure 7.10 shows similar results to that of the original standardised NAO index so the FD reddened SSP126 and SSP585 experiments are again combined. These CMIP6 projections for the coming 31-years show a large range of possible NAO trends due to natural internal variability (-7.41 to 7.68 hPa/decade) that could have considerable impacts on regional climate trends, despite showing no discernible externally forced NAO trend (ensemble mean -0.131 hPa/decade, i.e. -0.393 hPa over the 31-years). As for the *standardised* NAO index, the ensemble mean trend for the *anomaly* index is relatively unchanged compared to the raw estimate (-0.154 hPa/decade) while the range has been inflated by 69% (raw range -4.20 to 4.73 hPa/decade). The large underestimation of the plausible range of near-term future NAO trends when using raw CMIP6 model output is concerning as this could cause the range of

plausible near-term northern hemisphere regional winter climate trends to also be greatly underestimated.

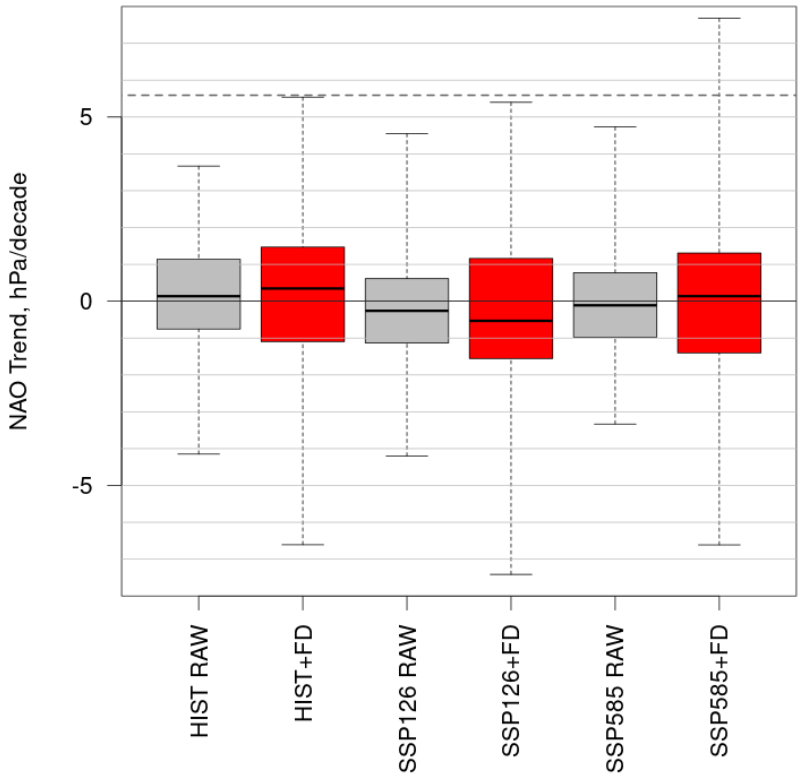


**Figure 7.10 Anomaly index near-term future NAO trends.** As Figure 7.9 but using the *anomaly* based NAO index (without standardisation) so that units are in hPa/decade, and the observed maximum trend 1963-1993 is based on the C20C MSLP data (5.59 hPa/decade). Box and whisker plots for CMIP6 historical ("HIST") 1963-1993 NAO trends and for CMIP6 future 2024-2054 NAO trends using scenarios SSP126 and SSP585. Grey box plots use the raw CMIP6 model output (median line in centre of 25 to 75 percentile box; total range whiskers). Red box plots use CMIP6 model output after applying the fractional difference reddening method ("+FD") with $d_R$ = 0.24. The observed 1963-1993 NAO trend is shown as the dashed horizontal line.

## 7.6 Summary

The FD reddening post-processing method from Chapter 6 has been applied, with $d_R$ in the range 0.15 to 0.24, to quantify the likelihood of extreme multi-decadal NAO trends in the future. The CMIP6 projections show that:

- The CMIP6 models of the NAO index show a small non-stationary response to external forcings in the future, with increased radiative forcing scenarios leading to a systematic increase in the long-term NAO trends, though the magnitude is underestimated by the raw CMIP6 NAO index series consistent with the lack of autocorrelation. This leads to an associated increase in the likelihood of extreme positive multi-decadal NAO trends under the higher radiative forcing scenarios.
- The long-term trend for 1950-2020 is unusually positive relative to the raw CMIP6 models (shown for window 1951-2020 in Blackport and Fyfe, 2022). However, this window has been pre-selected as an extreme window - it contains the maximum 71-year NAO trend in the historical record. Treating this as a block maximum trend (in the 1862-2021 block), the likelihood of the observed magnitude of trend is 27% to 49% when using the FD reddened model output ($d_R$ = 0.15 to 0.24), i.e. not especially unusual. In this case the underestimation of maximum 71-year NAO trends in the raw CMIP6 models is consistent with an underestimation of the autocorrelation.
- The minimum observed 71-year NAO trend (1903-1973) is more unusual than the recent maximum trend, with likelihood 4% to 12% using the FD reddened models ($d_R$ = 0.15 to 0.24). This could mean that a higher level of reddening is required, or that the CMIP6 models are underestimating or missing processes needed to simulate such a negative trend.
- In the future (2024-2098) under the “business as usual scenario” (SSP585), the CMIP6 distribution of extreme NAO trends is similar to that of the historical experiments in the recent past (1947-2021). The probability of a future maximum 31-year NAO trend larger than the maximum observed trend (0.737 sd/decade) in the coming 75-years is 3% to 7% when using the FD reddened model output ($d_R$ = 0.15 to 0.24), while the probability of a minimum 31-year trend less than -0.737 sd/decade is 1% to 4%. In the next 75 years under SSP585, there

is more chance of a period where the NAO positive trend partially enhances regional climate warming than of a period where the NAO negative trend partially offsets the regional warming signal.

- Near term projections of the next 31 years (2024-2054) show a large range of possible NAO trends due to natural internal variability (-7.41 to 7.68 hPa/decade), noting that for specific single window trends (such as the 1963-1993 window), application of the FD reddening method with $d_R$ = 0.24 is needed for the CMIP6 models to fully represent the observed multi-decadal NAO variability. On this timescale, results are not particularly sensitive to the choice of radiative forcing scenario, and there is no discernible externally forced NAO trend in the models.

Use of *raw* CMIP6 model output greatly underestimates the plausible range of near-term future NAO trends and extreme multi-decadal NAO trends. The relationship of multi-decadal NAO trends to northern hemisphere regional winter temperature and European precipitation means that the range of plausible multi-decadal regional winter temperature and precipitation trends are also likely to be greatly underestimated which will have a concerning impact on the fidelity of near-term climate projections in these regions. The FD reddening method is a simple tool to improve the multi-decadal variability of the NAO in future projections.

# 8. Conclusions

This chapter is a summary of the main findings in this thesis in relation to the key questions from Chapter 1, with some suggestions for future research.

## 8.1. Summary

### 1. How can we better diagnose and quantify extreme trends in climate indices?

Stochastic processes are a useful tool for quantifying extreme trends in climate indices as the observational record is too short to estimate the distribution of extreme trends directly. Chapter 3 defined a covariance approach to mathematically determine the distribution of moving window trends for an index series that can be represented as a first order stationary process such as an Auto-Regressive Moving Average (ARMA) process. This approach is also applicable to time series that consist of a stationary part and a deterministic signal such as a long-term linear trend as the trend distribution is simply shifted to be centred around the long-term trend instead of zero. For the distribution of block maxima trends, it was found that a simulation method was necessary as the moving window trend series has very high autocorrelation (due to the overlapping windows) so it is not possible to use mathematical derivations of the appropriate generalised extreme value distribution. Results show that for moderate levels of autocorrelation in the original index time series ($\rho < 0.4$ is typical for atmospheric seasonal mean climate indices) the variance of moving window trends increases with this level of autocorrelation, leading to an associated increase in exceedance probabilities of extreme trends.

Chapter 4 applied trend analysis methods from Chapter 3 to the NAO index case study using gridded observation and reanalysis datasets of mean sea level pressure (MSLP) to calculate the NAO index. Modelling methods, such as stochastic models or GCMs, are required to quantify the distribution of extreme NAO trends as the observed record is too short to estimate the distribution of trends from the observed series directly. In this chapter, fitting first order auto-regressive (AR(1)) and fractional difference (FD) stochastic processes to the

observed NAO index series shows some weak year-to-year memory which is shown to have a considerable effect on the distribution of trends.

The maximum 31-year NAO trend in the observed record occurs in the window 1963-1993 (0.737 sd/decade). Exceedance probability estimates for 31-year trends relative to this threshold are:

- A 1 in 1000 chance for the single window 1963-1993 NAO trend, using the derived moving window trend distribution from Chapter 3 for the fitted stochastic processes.
- A 1 in 20 chance for the block maxima NAO trend in the observed record length of 144 years (1862-2005), estimated using stochastic simulations from the fitted stochastic processes with weak year-to-year memory.
- A 1 in 200 chance for the block maxima NAO trend in the observed record length of 144 years (1862-2005), estimated using stochastic simulations from the fitted white noise processes (with no year-to-year memory).

When considering a trend such as the 1963-1993 NAO trend, it is important to admit that it has been specially identified due to its unusual nature. If it is treated as a randomly chosen single window, then the exceedance probability is greatly reduced compared to when it is treated as a block maxima trend. Ignoring the autocorrelation in the NAO index also leads to a considerable underestimation of exceedance probabilities. If it is assumed that the NAO behaviour will remain stationary into the future, then the block maxima probability will remain the same for the next 144-year block. Even in the context of block maxima trends, the observed 1963-1993 NAO trend is termed a *very unlikely* occurrence (IPCC likelihood scale, Appendix B.2) with respect to the fitted stochastic models which suggests there may be a role for drivers or mechanisms that are not included in these simple stochastic models of internal variability.

## 2. How realistically do General Circulation Models simulate extreme trends in the NAO?

Chapter 5 uses the Coupled Model Intercomparison Project 5 and 6 (CMIP5+6) large ensemble of Coupled General Circulation Model (CGCM) historical simulations to quantify the distribution of maximum 31-year NAO trends. These CGCMs considerably underestimate the magnitude of multi-decadal NAO trends, as found in other recent studies (Bracegirdle et al., 2018; Bracegirdle, 2022; Blackport and Fyfe, 2022; Schurer, 2023). The CGCM exceedance probability estimate for 31-year trends relative to the 1963-1993 observed trend threshold is:

- A 1 in 200 chance for the block maxima NAO trend in the observed record length of 144 years (1862-2005).

This CGCM exceedance probability estimate is only about 10% of the magnitude estimated using the fitted stochastic processes with year-to-year memory. The CGCM exceedance probabilities are much closer to the white noise process estimates, and the CGCM underestimation of multi-decadal NAO trends is consistent with a lack of autocorrelation in the modelled NAO index.

## 3. Can post-processing of GCM simulations improve the representation of extreme trends in the NAO?

Chapter 6 proposes useful new post-processing reddening methods to improve the representation of extreme trends in the NAO by increasing the autocorrelation in the CGCMs, using the known behaviour of AR(1) and FD processes. Increasing the autocorrelation of the CGCM NAO index series to match that of the observed NAO index series shifts the distribution of extreme trends to match those of the fitted stochastic processes from Chapter 4. Reddening methods widen the CGCM distribution of moving window trends and increase the variability of the ensemble mean. For the specific window 1963-1993, a relatively large increase in difference parameter is required ($d_R$ = 0.24) for the FD reddened CGCMs to encompass the observed trend. This level of reddening leads to:

- A 42% increase in the ensemble range for the 31-year NAO trend in the specific 1963-1993 window.
- A 45% increase in the total ensemble standard deviation of historical 31-year moving trends over the period 1862-2005, becoming comparable with observations.
- A 98% increase in the standard deviation of the ensemble mean moving window trend series over the period 1862-2005.

Compared to stochastic processes, reddened CGCMs have the advantage of including dynamical properties related to the NAO response to the external forcing.

## 4. How might extreme trends in the NAO change in the future?

Chapter 7 finds that CMIP6 models show a systematic increase in long-term NAO trends for future scenarios with higher radiative forcing and this has a small effect on the distribution of 31-year trends. It is assumed that the FD reddening method ($d_R$ in range 0.15 to 0.24) is necessary for realistic future NAO projections as well as for historical simulations. The reddened CMIP6 block maxima exceedance probability estimates for 31-year trends greater than the 1963-1993 observed trend thresholds are then:

- 3-7% in the next 75-years with a high radiative forcing (2024-2098, SSP585)
- 1-2% in the next 75-years with a low radiative forcing (2024-2098, SSP126)
- 3-10% for the previous 75-years (1947-2021)

Probability estimates for the next 75-years under a *high* radiative forcing scenario are similar to those for the previous 75-years whereas probability estimates for the next 75-years under a *low* radiative forcing scenario are considerably lower.

In contrast, near-term projections of the NAO trend for the next 31-years (2024-2054) are relatively insensitive to the forcing scenario, showing no externally forced trend in the models but with a large level of uncertainty related to natural internal variability:

- -7.41 to 7.68 hPa/decade is the ensemble range for the 31-year trend in the NAO anomaly point index using FD reddened future projections.
- -4.20 to 4.73 hPa/decade is the ensemble range for the raw future projections

This range of plausible NAO trends using FD reddened CGCMs encompasses the observed maximum 31-year NAO trend (5.59hPa/decade) so could greatly increase (or decrease) regional climate change signals in the northern hemisphere, as the 1960s to 1990s NAO trend has been shown to account for at least half of the winter warming in the northern hemisphere extra-tropics in that period (Scaife et al., 2005). This NAO influence is underestimated when using the raw CGCM output.

## 8.2. Future research

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., 2022a). Having developed post-processing reddening methods to improve the multi-decadal variability of the NAO in CGCMs, an important next step is to assess how this might influence future projections of atmospheric multi-decadal variability in the northern hemisphere, e.g. regional temperature and precipitation trends. Comparisons could be made with other methods to improve models of NAO multi-decadal variability using particle filter methods (Schurer et al., 2023) or synthetic observationally-constrained ensembles (O'Reilly et al., 2021). It would also be interesting to see if the autocorrelation and multi-decadal trends for other atmospheric variables are similarly underestimated in CGCMs and if so, to test the application of the reddening method directly on other useful climate indices. If multiple climate indices are being considered, it will be important to make sure that these indices are still dynamically consistent after post-processing, or perhaps they could be jointly post-processed.

Future work should also investigate the cause of the lack of autocorrelation in CGCMs, which 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). A topic of great interest in the last decade is the issue of a signal-to-noise paradox in CGCMs. This was 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), leading 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 lack of autocorrelation or persistence for the NAO in CGCMs has been shown to be closely linked to the signal-to-noise paradox (Zhang et al., 2019; Strommen and Palmer, 2018; Strommen, 2020). The FD reddening method to increase year-to-year persistence was shown in Chapter 6 to almost double the standard deviation of the ensemble mean trend series for historical CGCM experiments while only inflating the total ensemble standard deviation by 45%, i.e. increasing the signal-to-noise ratio for the historical CGCM trend series. It would thus be interesting to see how reddening methods might influence the ratio of predictable components for the NAO index series and whether similar adjustments give rise to realistic results on both interannual and multidecadal timescales. Finally, there is still a lack of understanding as to the cause of the signal-to-noise paradox. Seeking to explain why the autocorrelation of the NAO index is underestimated in CGCMs may help with this quest.

# Appendices

# A. Tables of CMIP5 and CMIP6 Models

| Institution | Model name | Realisations |
|---|---|---|
| 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 ripf code f1; f2 | 2; 1 |
| CAS | CAS-ESM2-0; FGOALS-f3-L; FGOALS-g3 | 4; 3; 6 |
| CCCma | CanESM5 ripf code p1; p2; CanESM5-CanOE | 25; 40; 3 |
| CCCR-IITM | IITM-ESM | 1 |
| CMCC | CMCC-CM2-HR4; CMCC-CM2-SR5; CMCC-ESM2-1 | 1; 1; 1 |
| CNRM-CERFACS | CNRM-CM6-1; CNRM-CM6-1-HR; CNRM-ESM2-1 | 30; 1; 10 |
| CSIRO-ARCCSS | ACCESS-CM2 | 3 |
| E3SM-Project | E3SM-1-0; E3SM-1-1 | 5; 1 |
| EC-Earth-Consortium | EC-Earth3; EC-Earth3-CC; EC-Earth3-Veg; EC-Earth3-Veg-LR | 22; 1; 8; 3 |
| FIO-QLNM | FIO-ESM-2-0 | 3 |
| HAMMOZ-Consortium | MPI-ESM-1-2-HAM | 2 |
| INM | INM-CM4-8; INM-CM5-0 | 1; 10 |
| IPSL | IPSL-CM6A-LR | 32 |
| KIOST | KIOST-ESM | 1 |
| MIROC | MIROC6; MIROC-ES2L | 50; 10 |
| MOHC | HadGEM3-GC31-LL; HadGEM3-GC31-MM; UKESM1-0-LL ripf code f2; f3 | 4; 3; 13; 3 |
| MPI-M | MPI-ESM1-2-HR; MPI-ESM1-2-LR | 10; 10 |
| MRI | MRI-ESM2-0 | 5 |
| NASA-GISS | GISS-E2-1-G-CC; GISS-E2-1-G ripf code p1f1; p1f2; p1f3; p3f1; p5f1; GISS-E2-1-H ripf code p1f1; p1f2; p3f1; p5f1 | 1; 12; 8; 5; 7; 6; 10; 5; 5; 5 |
| NCAR | CESM2; CESM2-FV2; CESM2-WACCM; | 3; 1; 2; 1 |

| | | |
|---|---|---|
| | CESM2-WACCM-FV2 | |
| NCC | NorCPM1; NorESM2-LM; NorESM2-MM | 2; 2; 1 |
| NIMS-KMA | KACE-1-0-G | 2 |
| NOAA-GFDL | GFDL-CM4; GFDL-ESM4 | 1; 3 |
| NUIST | NESM3 | 5 |
| SNU | SAM0-UNICON | 1 |
| THU | CIESM | 3 |
| UA | MCM-UA-1-0 ripf code f1; f2 | 1; 1 |

**Table A.1**: CMIP6 CGCM historical experiments used to model the NAO with columns showing the institution, model names and the number of realisations (ensemble members). 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).

| Institution | Model name | Realisations |
|---|---|---|
| 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 |

**Table A.2**: CMIP5 CGCM historical experiments used to model the NAO with columns showing the institution, model name and the number of realisations (ensemble members) as in Table A.1. All experiments are from the CMIP5 database https://esgf-node.llnl.gov/search/cmip5 (Taylor et al., 2012).

| 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; FGOALS-f3-L; FGOALS-g3 | 1; 1; 2 |
| CCCma | CanESM5 ripf code p1; p2; CanESM5-CanOE | 25; 25; 3 |
| CMCC | CMCC-CM2-SR5; CMCC-ESM2-1 | 1; 1 |
| CNRM-CERFACS | CNRM-CM6-1-HR | 1 |
| CSIRO | ACCESS-ESM1-5 | 40 |
| CSIRO-ARCCSS | ACCESS-CM2 | 3 |
| EC-Earth-Consortium | EC-Earth3-Veg-LR | 3 |
| INM | INM-CM4-8; INM-CM5-0 | 1; 1 |
| IPSL | IPSL-CM6A-LR | 6 |
| MIROC | MIROC6; MIROC-ES2L | 3; 10 |
| MOHC | UKESM1-0-LL ripf code f2 | 5 |
| MPI-M | MPI-ESM1-2-LR | 10 |
| MRI | MRI-ESM2-0 | 1 |
| NASA-GISS | GISS-E2-1-G ripf code p1f2;p3f1; p5f1 | 3; 4; 4 |
| NCC | NorESM2-LM; NorESM2-MM | 1; 1 |
| NIMS-KMA | KACE-1-0-G | 2 |
| NOAA-GFDL | GFDL-ESM4 | 1 |
| UA | MCM-UA-1-0 ripf code f2 | 1 |

**Table A.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) as in Table A.1. This is a subset of the models from Table A.1 that are available for the historical experiments and all four SSP scenario experiments assessed in Chapter 7 (SSP126, SSP245, SSP370, SSP585). All experiments are from the CMIP6 database https://esgf-index1.ceda.ac.uk/search/cmip6-ceda (Eyring et al., 2016).

# B. Statistical methods and code references

## B.1 Stochastic simulations

Stochastic simulations are generated using the R function **fracdiff.sim** from the R package **fracdiff** (Maechler, 2020).

## B.2 IPCC Likelihood scale

The Intergovernmental Panel on Climate Change (IPCC) has a likelihood scale (Mastrandrea et al., 2010, Table 1) that is used as a convenient interpretation of exceedance probabilities:

| | |
|---|---|
| 0 - 1% | Exceptionally unlikely |
| 0 – 10% | Very unlikely |
| 0 – 33% | Unlikely |
| 33 – 66% | About as likely as not |

## B.3 Significance test for data statistics

A standard hypothesis test for the mean ($\bar{x}$) of an ensemble of predictions is to test whether it is likely to come from a distribution with population mean μ = 0, i.e. H0: μ = 0 vs. H1: μ ≠ 0. To test for this, a 95% confidence interval can be defined as $\bar{x} \pm 1.96s/\sqrt{n}$ where *s* is the sample standard deviation and *n* is the ensemble size (e.g. Wilks, 2006). This interval assumes that $\bar{x}$ has a Gaussian distribution, which is generally a satisfactory assumption due to the central limit theorem. If the confidence interval does not contain zero, i.e. $|\bar{x}| > 1.96s/\sqrt{n}$ then the null hypothesis H0 is rejected (at the 5% level in this case) and the ensemble mean is often described as being “significantly different to zero”. An extension of this method is to test whether a data statistic (such as the mean or autocorrelation parameter) is likely to have come from a known distribution such as that derived from a stochastic model. To test for this, a large number of sample statistics can be generated from the stochastic model to create an empirical distribution which can then be used to compare with the data statistic in the same way as the Gaussian distribution method above.

## B.4 Generalised extreme value distribution

To estimate the generalised extreme value distribution parameters for a sample, the maximum likelihood method is applied using the R function **fgev** from the R package **evd** (Stephenson, 2002).

## B.5 Empirical Orthogonal Function

Empirical Orthogonal Function (EOF) patterns are calculated using the Python library **eofs** (Dawson, 2016).

## B.6 Kernel density

The kernel density is chosen as a simple method to display the empirical distributions of trends in this thesis as it enables multiple distributions to be visualised more clearly on one figure than when histograms are used, but it is not used to mathematically compare the different distributions. The R function **density** is used from the **stats** package (R Core Team, 2020) with default parameters, namely the choice of smoothing kernel method is set as “gaussian” and the method to compute the bandwidth is set as “nrd0” which refers to the “rule of thumb” method from Silverman (1986). Kernel density estimation is a non-parametric method to estimate the probability density function from a sample of data by fitting a kernel, in this case a gaussian kernel, to each data point and combining these to form a smooth probability density function with standard characteristics (non-negative, integrates to 1).

## B.7 Prediction interval ellipse

A 95% prediction interval ellipse is used in Figure 5.6 to display the 2-dimensional characteristics of the distribution of points in the scatter plot. The R function **ellipse** is used from the **mixtools** package (Benaglia et al., 2009) with parameter alpha set to be 0.05 such that 95% of the scatter plot points are included within the ellipse. The ellipse represents a bivariate normal density contour based on the covariance matrix of the original data.

# Bibliography


Allan, R. and Ansell, T. (2006). A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004. *Journal of Climate*, 19(22), pp.5816–5842. doi:https://doi.org/10.1175/jcli3937.1.

Ambaum, M.H.P., Hoskins, B.J. and Stephenson, D.B. (2001). Arctic Oscillation or North Atlantic Oscillation? *Journal of Climate*, 14(16), pp.3495–3507. doi:https://doi.org/10.1175/1520-0442(2001)014%3C3495:aoonao%3E2.0.co;2.

Arias, P.A., and Coauthors (2021) Technical summary. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Eds., Cambridge University Press, pp. 33-144, doi:10.1017/9781009157896.002.

Baldwin, M.P. (2001). Stratospheric Harbingers of Anomalous Weather Regimes. *Science*, 294(5542), pp.581–584. doi:https://doi.org/10.1126/science.1063315.

Bartlett, M.S. (1946). On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Supplement to the Journal of the Royal Statistical Society*, 8(1), p.27. doi:https://doi.org/10.2307/2983611.

Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D. (2009). mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software*, 32(6), pp.1-29. URL http://www.jstatsoft.org/v32/i06/

Blackport, R. and Fyfe, J.C. (2022). Climate models fail to capture strengthening wintertime North Atlantic jet and impacts on Europe. *Science Advances*, 8(45). doi:https://doi.org/10.1126/sciadv.abn3112.

Bracegirdle, T.J. (2022). Early-to-Late Winter 20th Century North Atlantic Multidecadal Atmospheric Variability in Observations, CMIP5 and CMIP6. *Geophysical Research Letters*, 49(11). doi:https://doi.org/10.1029/2022gl098212.

Bracegirdle, T.J., Lu, H., Eade, R. and Woollings, T. (2018). Do CMIP5 Models Reproduce Observed Low-Frequency North Atlantic Jet Variability? *Geophysical Research Letters*, 45(14), pp.7204–7212. doi:https://doi.org/10.1029/2018gl078965.

Chandler, R.B. and Scott, E. (2011). *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. *Wiley eBooks*. Wiley. doi:https://doi.org/10.1002/9781119991571.

Chiodo, G., García-Herrera, R., Calvo, N., Vaquero, J.M., Añel, J.A., Barriopedro, D. and Matthes, K. (2016). The impact of a future solar minimum on climate change projections in the Northern Hemisphere. Environmental Research Letters, 11(3), p.034015. doi:https://doi.org/10.1088/1748-9326/11/3/034015.

Cochrane, D. and Orcutt, G. (1949). Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. *Journal of the American Statistical Association*, 44, pp.32–61.

Cohen, J. and Entekhabi, D. (1999). Eurasian snow cover variability and northern hemisphere climate predictability. *Geophysical Research Letters*, 26(3), pp.345–348. doi:https://doi.org/10.1029/1998gl900321.

Cohen, J., Screen, J.A., Furtado, J.C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J. and Jones, J. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, [online] 7(9), pp.627–637. doi:https://doi.org/10.1038/ngeo2234.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*. London: Springer London. doi:https://doi.org/10.1007/978-1-4471-3675-0.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M.C., Kruger, A.C., Marshall, G.J. and Maugeri, M. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654), pp.1–28. doi:https://doi.org/10.1002/qj.776.

Cubasch, U. and Coauthors (2001) Projections of future climate change. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Xiaosu D

(eds) Climate change 2001, the scientific basis. Cambridge University Press, Cambridge UK pp 525–582

Davini, P. and Cagnazzo, C. (2013). On the misinterpretation of the North Atlantic Oscillation in CMIP5 models. Climate Dynamics, 43(5-6), pp.1497–1511. doi:https://doi.org/10.1007/s00382-013-1970-y.

Davis, R.A., Mikosch, T. and Zhao, Y. (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications*, 123(7), pp.2575–2602. doi:https://doi.org/10.1016/j.spa.2013.03.014.

Dawson, A. (2016). eofs: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data. Journal of Open Research Software, 4(1). doi:https://doi.org/10.5334/jors.122.

Delworth, T.L. and Zeng, F. (2016). The Impact of the North Atlantic Oscillation on Climate through Its Influence on the Atlantic Meridional Overturning Circulation. *Journal of Climate*, 29(3), pp.941–962. doi:https://doi.org/10.1175/jcli-d-15-0396.1.

Deser, C., Hurrell, J.W. and Phillips, A.S. (2017). The role of the North Atlantic Oscillation in European climate projections. *Climate Dynamics*, 49(9-10), pp.3141–3157. doi:https://doi.org/10.1007/s00382-016-3502-z.

Deser, C., Tomas, R., Alexander, M. and Lawrence, D. (2010). The Seasonal Atmospheric Response to Projected Arctic Sea Ice Loss in the Late Twenty-First Century. *Journal of Climate*, 23(2), pp.333–351. doi:https://doi.org/10.1175/2009jcli3053.1.

Deser, C., Walsh, J.E. and Timlin, M.S. (2000). Arctic Sea Ice Variability in the Context of Recent Atmospheric Circulation Trends. *Journal of Climate*, 13(3), pp.617–633. doi:https://doi.org/10.1175/1520-0442(2000)013%3C0617:asivit%3E2.0.co;2.

Driscoll, S., Bozzo, A., Gray, L.J., Robock, A. and Stenchikov, G. (2012). Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. *Journal of Geophysical Research: Atmospheres*, 117(D17), p.n/a-n/a. doi:https://doi.org/10.1029/2012jd017607.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L. and Robinson, N. (2014). Do seasonal-to-decadal climate predictions

underestimate the predictability of the real world? *Geophysical Research Letters*, 41(15), pp.5620–5628. doi:https://doi.org/10.1002/2014gl061146.

Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R.J. and Taylor, K.E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), pp.1937–1958. doi:https://doi.org/10.5194/gmd-9-1937-2016.

Feldstein, S.B. (2000). The Timescale, Power Spectra, and Climate Noise Properties of Teleconnection Patterns. *Journal of Climate*, 13(24), pp.4430–4440. doi:https://doi.org/10.1175/1520-0442(2000)013%3C4430:ttpsac%3E2.0.co;2.

Feldstein, S.B. (2002). The Recent Trend and Variance Increase of the Annular Mode. *Journal of Climate*, 15(1), pp.88–94. doi:https://doi.org/10.1175/1520-0442(2002)015%3C0088:trtavi%3E2.0.co;2.

Fereday, D., Chadwick, R., Knight, J. and Scaife, A.A. (2018). Atmospheric Dynamics is the Largest Source of Uncertainty in Future Winter European Rainfall. *Journal of Climate*, [online] 31(3), pp.963–977. doi:https://doi.org/10.1175/JCLI-D-17-0048.1.

Francis, J.A. and Vavrus, S.J. (2012). Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, 39(6), p.n/a-n/a. doi:https://doi.org/10.1029/2012gl051000.

Franzke, C. (2009). Multi-scale analysis of teleconnection indices: climate noise and nonlinear trend analysis. *Nonlinear Processes in Geophysics*, 16(1), pp.65–76. doi:https://doi.org/10.5194/npg-16-65-2009.

Gastineau, G. and Frankignoul, C. (2015). Influence of the North Atlantic SST Variability on the Atmospheric Circulation during the Twentieth Century. *Journal of Climate*, 28(4), pp.1396–1416. doi:https://doi.org/10.1175/jcli-d-14-00424.1.

Ge, Q., Wang, H., Zheng, J., Rutishauser, T. and Dai, J. (2014). A 170 year spring phenology index of plants in eastern China. *Journal Of Geophysical Research: Biogeosciences*, 119(3), pp.301–311. doi:https://doi.org/10.1002/2013jg002565.

Ghatak, D., Frei, A., Gong, G., Stroeve, J. and Robinson, D. (2010). On the emergence of an Arctic amplification signal in terrestrial Arctic snow extent.

*Journal of Geophysical Research: Atmospheres*, 115(D24). doi:https://doi.org/10.1029/2010jd014007.

Gillett, N.P., Allen, M. and Williams, K.D. (2002). The role of stratospheric resolution in simulating the Arctic Oscillation response to greenhouse gases. *Geophysical Research Letters*, 29(10), pp.138–4. doi:https://doi.org/10.1029/2001gl014444.

Gillett, NP., Graf, HF., & Osborn, TJ. (2003). Climate change and the North Atlantic oscillation. In JW. Hurrell, Y. Kushnir, G. Ottersen, & M. Visbeck (Eds.), *North Atlantic Oscillation: Climate Significance and Environmental Impact* (pp. 193-209). AGU.

Gray, L.J., Scaife, A.A., Mitchell, D.M., Osprey, S., Ineson, S., Hardiman, S., Butchart, N., Knight, J., Sutton, R. and Kodera, K. (2013). A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *Journal of Geophysical Research: Atmospheres*, 118(24), pp.13, 405–13, 420. doi:https://doi.org/10.1002/2013jd020062.

Greatbatch, R.J. (2000). The North Atlantic Oscillation. *Stochastic Environmental Research and Risk Assessment*, 14(4), pp.0213–0242. doi:https://doi.org/10.1007/s004770000047.

Hamilton, J.D. (1994). *Time series analysis*. Princeton, N.J.: Princeton University Press.

Hanna, E., Cropper, T.E., Jones, P.D., Scaife, A.A. and 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. *International Journal of Climatology*, 35(9), pp.2540–2554. doi:https://doi.org/10.1002/joc.4157.

Hermanson, L., Bilbao, R., Dunstone, N., Ménégoz, M., Ortega, P., Pohlmann, H., Robson, J.I., Smith, D.M., Strand, G., Timmreck, C., Yeager, S. and Danabasoglu, G. (2020). Robust Multiyear Climate Impacts of Volcanic Eruptions in Decadal Prediction Systems. *Journal of Geophysical Research: Atmospheres*, 125(9). doi:https://doi.org/10.1029/2019jd031739.

Hilmer, M. and Jung, T. (2000). Evidence for a recent change in the link between the North Atlantic Oscillation and Arctic Sea ice export. *Geophysical Research Letters*, 27(7), pp.989–992. doi:https://doi.org/10.1029/1999gl010944.

Hoerling, M.P. (2001). Tropical Origins for Recent North Atlantic Climate Change. *Science*, 292(5514), pp.90–92. doi:https://doi.org/10.1126/science.1058582.

Hoerling, M.P., Hurrell, J.W., Xu, T., Bates, G.T. and Phillips, A.S. (2004). Twentieth century North Atlantic climate change. Part II: Understanding the effect of Indian Ocean warming. *Climate Dynamics*, 23(3-4), pp.391–405. doi:https://doi.org/10.1007/s00382-004-0433-x.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68(1), pp.165–176. doi:https://doi.org/10.1093/biomet/68.1.165.

Hurrell, J.W. (1995). Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, 269(5224), pp.676–679. doi:https://doi.org/10.1126/science.269.5224.676.

Hurrell, J.W. (2001). The North Atlantic Oscillation. *Science*, 291(5504), pp.603–605. doi:https://doi.org/10.1126/science.1058761.

Hurrell, J. W., Kushnir, Y. , Ottersen, G. and Visbeck, M. (2003). An Overview of the North Atlantic Oscillation. In JW. Hurrell, Y. Kushnir, G. Ottersen, & M. Visbeck (Eds.), *North Atlantic Oscillation: Climate Significance and Environmental Impact* (pp. 1-36). AGU.

Hurrell, J.W., Hoerling, M.P., Phillips, A.S. and Xu, T. (2004). Twentieth century north atlantic climate change. Part I: assessing determinism. *Climate Dynamics*, 23(3-4), pp.371–389. doi:https://doi.org/10.1007/s00382-004-0432-y.

Hurrell, J.W. and Van Loon, H. (1997). Decadal variations in climate associated with the north atlantic oscillation. *Climatic Change*, [online] 36(3/4), pp.301–326. doi:https://doi.org/10.1023/a:1005314315270.

Iles, C. and Hegerl, G. (2017). Role of the North Atlantic Oscillation in decadal temperature trends. *Environmental Research Letters*, 12(11), p.114010. doi:https://doi.org/10.1088/1748-9326/aa9152.

Ineson, S., Scaife, A.A., Knight, J.R., Manners, J.C., Dunstone, N.J., Gray, L.J. and Haigh, J.D. (2011). Solar forcing of winter climate variability in the Northern Hemisphere. *Nature Geoscience*, 4(11), pp.753–757. doi:https://doi.org/10.1038/ngeo1282.

Jerez, S., Trigo, R.M., Vicente-Serrano, S.M., Pozo-Vázquez, D., Lorente-Plazas, R., Lorenzo-Lacruz, J., Santos-Alamillos, F. and Montávez, J.P. (2013). The Impact of the North Atlantic Oscillation on Renewable Energy

Resources in Southwestern Europe. *Journal of Applied Meteorology and Climatology*, [online] 52(10), pp.2204–2225. doi:https://doi.org/10.1175/JAMC-D-12-0257.1.

Jones, P.D., Jonsson, T. and Wheeler, D. (1997). Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *International Journal of Climatology*, 17(13), pp.1433–1450. doi:https://doi.org/10.1002/(sici)1097-0088(19971115)17:13%3C1433::aid-joc203%3E3.0.co;2-p.

Keeley, S.P.E., Sutton, R.T. and Shaffrey, L.C. (2009). Does the North Atlantic Oscillation show unusual persistence on intraseasonal timescales? *Geophysical Research Letters*, 36(22). doi:https://doi.org/10.1029/2009gl040367.

Kidston, J., Scaife, A.A., Hardiman, S.C., Mitchell, D.M., Butchart, N., Baldwin, M.P. and Gray, L.J. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, 8(6), pp.433–440. doi:https://doi.org/10.1038/ngeo2424.

Kim, M.-K. and McCarl, B.A. (2005). The Agricultural Value Of Information On The North Atlantic Oscillation: Yield And Economic Effects. *Climatic Change*, 71(1-2), pp.117–139. doi:https://doi.org/10.1007/s10584-005-5928-x.

Kinnison, R.R. (1985). *Applied Extreme Value Statistics*. MacMillan, New York.

Klavans, J.M., Cane, M.A., Clement, A.C. and Murphy, L.N. (2021). NAO predictability from external forcing in the late 20th century. npj Climate and Atmospheric Science, 4(1). doi: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. *Geophysical Research Letters*, 44(11), pp.5749–5757. doi:https://doi.org/10.1002/2017gl074016.

Kuzmina, S.I., Bengtsson, L., Johannessen, O.M., Drange, H., Bobylev, L.P. and Miles, M.W. (2005). The North Atlantic Oscillation and greenhouse-gas forcing. *Geophysical Research Letters*, 32(4), p.n/a-n/a. doi:https://doi.org/10.1029/2004gl021064.

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, S.L., Pean, 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: Cambridge University Press, pp. 553–672. doi: 10.1017/9781009157896.006.

Lowe J.A. 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. and Wanner, H. (1999). Reconstruction of monthly NAO and EU indices back to AD 1675. *Geophysical Research Letters*, 26(17), pp.2745–2748. doi:https://doi.org/10.1029/1999gl900576.

Luterbacher, J., Xoplaki, E., Dietrich, D., Jones, P.D., Davies, T.D., Portis, D., Gonzalez-Rouco, J.F., von Storch, H., Gyalistras, D., Casty, C. and Wanner, H. (2001). Extending North Atlantic Oscillation reconstructions back to 1500. *Atmospheric Science Letters*, 2(1-4), pp.114–124. doi:https://doi.org/10.1006/asle.2002.0047.

Maechler, M. (2020). fracdiff: Fractionally Differences ARIMA aka ARFIMA(P,d,q) Models. R package version 1.5-1. URL https://CRAN.R-project.org/package=fracdiff

Marshall, A.G., Scaife, A.A. and Ineson, S. (2009). Enhanced Seasonal Prediction of European Winter Warming following Volcanic Eruptions. *Journal of Climate*, 22(23), pp.6168–6180. doi:https://doi.org/10.1175/2009jcli3145.1.

Mastrandrea, M.D., Field, C.B., Stocker, T.F., Ottmar Edenhofer, Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P., Plattner, G.-K., Yohe, G.W. and Zwiers, F.W. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC).

Matti, C., Pauling, A., Küttel, M. and Wanner, H. (2008). Winter precipitation trends for two selected European regions over the last 500 years and their possible dynamical background. *Theoretical and Applied Climatology*, 95(1-2), pp.9–26. doi:https://doi.org/10.1007/s00704-007-0361-x.

McKenna, C.M., Bracegirdle, T.J., Shuckburgh, E.F., Haynes, P.H. and Joshi, M.M. (2018). Arctic Sea Ice Loss in Different Regions Leads to Contrasting Northern Hemisphere Impacts. *Geophysical Research Letters*, 45(2), pp.945–954. doi:https://doi.org/10.1002/2017gl076433.

McKenna, C.M. and Maycock, A.C. (2022). The Role of the North Atlantic Oscillation for Projections of Winter Mean Precipitation in Europe. *Geophysical Research Letters*, 49(19). doi:https://doi.org/10.1029/2022gl099083.

Meehl, G.A., Goddard, L., Murphy, J., Stouffer, R.J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M.A., Greene, A.M., Hawkins, E., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D. and Stockdale, T. (2009). Decadal Prediction. *Bulletin of the American Meteorological Society*, 90(10), pp.1467–1486. doi:https://doi.org/10.1175/2009bams2778.1.

Menary, M.B., Kuhlbrodt, T., Ridley, J., Andrews, M.B., Dimdore-Miles, O.B., Deshayes, J., Eade, R., Gray, L., Ineson, S., Mignot, J., Roberts, C.D., Robson, J., Wood, R.A. and Xavier, P. (2018). Preindustrial Control Simulations With HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, 10(12). doi:https://doi.org/10.1029/2018ms001495.

Molteni, F., Stockdale, T.N. and Vitart, F. (2015). Understanding and modelling extra-tropical teleconnections with the Indo-Pacific region during the northern winter. *Climate Dynamics*, 45(11-12), pp.3119–3140. doi:https://doi.org/10.1007/s00382-015-2528-y.

Mori, M., Watanabe, M., Shiogama, H., Inoue, J. and Kimoto, M. (2014). Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. *Nature Geoscience*, 7(12), pp.869–873. doi:https://doi.org/10.1038/ngeo2277.

Mosedale, T.J., Stephenson, D.B., Collins, M. and Mills, T.C. (2006). Granger Causality of Coupled Climate Processes: Ocean Feedback on the North Atlantic Oscillation. *Journal of Climate*, 19(7), pp.1182–1194. doi:https://doi.org/10.1175/jcli3653.1.

Nie, Y., Scaife, A.A., Ren, H.-L., Comer, R.E., Andrews, M.B., Davis, P.J. and Martin, N. (2019). Stratospheric initial conditions provide seasonal predictability of the North Atlantic and Arctic Oscillations. *Environmental*

*Research Letters*, 14(3), pp.034006–034006. doi:https://doi.org/10.1088/1748-9326/ab0385.

O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K. and Sanderson, B.M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, [online] 9(9), pp.3461–3482. doi:https://doi.org/10.5194/gmd-9-3461-2016.

O'Reilly, C.H., Befort, D.J., Weisheimer, A., Woollings, T., Ballinger, A. and Hegerl, G. (2021). Projections of northern hemisphere extratropical climate underestimate internal variability and associated uncertainty. *Communications Earth & Environment*, 2(1). doi:https://doi.org/10.1038/s43247-021-00268-7.

Omrani, N.-E., Keenlyside, N., Bader, J. and Manzini, E. (2014). Stratosphere key for wintertime atmospheric response to warm Atlantic decadal conditions. *Climate Dynamics*, 42(3-4), pp.649–663. doi:https://doi.org/10.1007/s00382-013-1860-3.

Omrani, N.-E., Keenlyside, N., Matthes, K., Boljka, L., Davide Zanchettin, Jungclaus, J. and Lubis, S.W. (2022). Coupled stratosphere-troposphere-Atlantic multidecadal oscillation and its importance for near-future climate projection. *npj climate and atmospheric science*, 5(1). doi:https://doi.org/10.1038/s41612-022-00275-1.

Osborn, T.J., Briffa, K.R., Tett, S.F.B., Jones, P.D. and Trigo, R.M. (1999). Evaluation of the North Atlantic Oscillation as simulated by a coupled climate model. *Climate Dynamics*, 15(9), pp.685–702. doi:https://doi.org/10.1007/s003820050310.

Otterå, O.H. (2008). Simulating the effects of the 1991 Mount Pinatubo volcanic eruption using the ARPEGE atmosphere general circulation model. *Advances in Atmospheric Sciences*, 25(2), pp.213–226. doi:https://doi.org/10.1007/s00376-008-0213-3.

Percival, D.B. and Rothrock, D.A. (2005). 'Eyeballing' Trends in Climate Time Series: A Cautionary Note. *Journal of Climate*, 18(6), pp.886–891. doi:https://doi.org/10.1175/jcli-3300.1.

Peterson, K.A., Greatbatch, R.J., Lu, J., Lin, H. and Derome, J. (2002). Hindcasting the NAO using diabatic forcing of a simple AGCM. *Geophysical*

*Research Letters*, 29(9), pp.50–150–4. doi:https://doi.org/10.1029/2001gl014502.

Pinto, J.G. and Raible, C.C. (2012). Past and recent changes in the North Atlantic oscillation. *Wiley Interdisciplinary Reviews: Climate Change*, 3(1), pp.79–90. doi:https://doi.org/10.1002/wcc.150.

Pozo-Vázquez, D., Esteban-Parra, M.J., Rodrigo, F.S. and Castro-Díez, Y. (2001). A study of NAO variability and its possible non-linear influences on European surface temperature. *Climate Dynamics*, 17(9), pp.701–715. doi:https://doi.org/10.1007/s003820000137.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Raible, C.C., Stocker, T.F., Masakazu Yoshimori, Renold, M., Beyerle, U., Casty, C. and Jürg Luterbacher (2005). Northern Hemispheric Trends of Pressure Indices and Atmospheric Circulation Patterns in Observations, Reconstructions, and Coupled GCM Simulations. *Journal of Climate*, 18(19), pp.3968–3982. doi:https://doi.org/10.1175/jcli3511.1.

Reichler, T., Kim, J., Manzini, E. and Kröger, J. (2012). A stratospheric connection to Atlantic climate variability. *Nature Geoscience*, 5(11), pp.783–787. doi:https://doi.org/10.1038/ngeo1586.

Robock, A. and Mao, J. (1992). Winter warming from large volcanic eruptions. *Geophysical Research Letters*, 19(24), pp.2405–2408. doi:https://doi.org/10.1029/92gl02627.

Robson, J., Sutton, R., Lohmann, K., Smith, D. and Palmer, M.D. (2012). Causes of the Rapid Warming of the North Atlantic Ocean in the Mid-1990s. *Journal of Climate*, 25(12), pp.4116–4134. doi:https://doi.org/10.1175/jcli-d-11-00443.1.

Roxy, M.K., Ritika, K., Terray, P. and Masson, S. (2014). The Curious Case of Indian Ocean Warming*,+. *Journal of Climate*, 27(22), pp.8501–8509. doi:https://doi.org/10.1175/jcli-d-14-00471.1.

Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M. and Wallace, E. (2014). Skillful long-range prediction

of European and North American winters. *Geophysical Research Letters*, 41(7), pp.2514–2519. doi:https://doi.org/10.1002/2014gl059637.

Scaife, A.A., Comer, R.E., Dunstone, N., Knight, J., Smith, D.J., MacLachlan, C., Martin, N., Peterson, K., Rowlands, D., Carroll, E.J., Belcher, S.E. and Slingo, J. (2017). Tropical rainfall, Rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, 143(702), pp.1–11. doi:https://doi.org/10.1002/qj.2910.

Scaife, A.A., Folland, C.K., Alexander, L.V., Moberg, A. and Knight, J.R. (2008). European Climate Extremes and the North Atlantic Oscillation. *Journal of Climate*, 21(1), pp.72–83. doi:https://doi.org/10.1175/2007jcli1631.1.

Scaife, A.A., Ineson, S., Knight, J.R., Gray, L., Kodera, K. and Smith, D.M. (2013). A mechanism for lagged North Atlantic climate response to solar variability. *Geophysical Research Letters*, 40(2), pp.434–439. doi:https://doi.org/10.1002/grl.50099.

Scaife, A.A., Karpechko, A.Yu., Baldwin, M.P., Brookshaw, A., Butler, A.H., Eade, R., Gordon, M., MacLachlan, C., Martin, N., Dunstone, N. and Smith, D. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, 17(1), pp.51–56. doi:https://doi.org/10.1002/asl.598.

Scaife, A.A., Knight, J.R., Vallis, G.K. and Folland, C.K. (2005). A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophysical Research Letters*, 32(18), p.n/a-n/a. doi:https://doi.org/10.1029/2005gl023226.

Scaife, A.A., Kucharski, F., Folland, C.K., Kinter, J., Brönnimann, S., Fereday, D., Fischer, A.M., Grainger, S., Jin, E.K., Kang, I.S., Knight, J.R., Kusunoki, S., Lau, N.C., Nath, M.J., Nakaegawa, T., Pegion, P., Schubert, S., Sporyshev, P., Syktus, J. and Yoon, J.H. (2009). The CLIVAR C20C project: selected twentieth century climate events. *Climate Dynamics*, 33(5), pp.603–614. doi:https://doi.org/10.1007/s00382-008-0451-1.

Scaife, A.A. and Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, 1(1). doi:https://doi.org/10.1038/s41612-018-0038-4.

Schurer, A., Hegerl, G.C., Goosse, H., Bollasina, M., England, M.H., Smith, D. and Tett, S. (2023). Role of multi-decadal variability of the winter North Atlantic oscillation on northern hemisphere climate. *Environmental*

*Research Letters*, 18(4), pp.044046–044046. doi:https://doi.org/10.1088/1748-9326/acc477.

Screen, J.A. (2017). Simulated Atmospheric Response to Regional and Pan-Arctic Sea Ice Loss. *Journal of Climate*, 30(11), pp.3945–3962. doi:https://doi.org/10.1175/jcli-d-16-0197.1.

Screen, J.A., Deser, C., Smith, D.M., Zhang, X., Blackport, R., Kushner, P.J., Oudar, T., McCusker, K.E. and Sun, L. (2018). Consistency and discrepancy in the atmospheric response to Arctic sea-ice loss across climate models. *Nature Geoscience*, 11(3), pp.155–163. doi:https://doi.org/10.1038/s41561-018-0059-y.

Screen, J.A., Simmonds, I., Deser, C. and Tomas, R. (2013). The Atmospheric Response to Three Decades of Observed Arctic Sea Ice Loss. *Journal of Climate*, 26(4), pp.1230–1248. doi:https://doi.org/10.1175/jcli-d-12-00063.1.

Seabrook, M., Smith, D.M., Dunstone, N.J., Eade, R., Hermanson, L., Scaife, A.A. and Hardiman, S.C. (2023). Opposite Impacts of Interannual and Decadal Pacific Variability in the Extratropics. Geophysical Research Letters, 50(2). doi:https://doi.org/10.1029/2022gl101226.

Semenov, V.A., Latif, M., Jungclaus, J.H. and Park, W. (2008). Is the observed NAO variability during the instrumental record unusual? *Geophysical Research Letters*, 35(11). doi:https://doi.org/10.1029/2008gl033273.

Sévellec, F., Fedorov, A.V. and Liu, W. (2017). Arctic sea-ice decline weakens the Atlantic Meridional Overturning Circulation. *Nature Climate Change*, 7(8), pp.604–610. doi:https://doi.org/10.1038/nclimate3353.

Shepherd, T.G. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7(10), pp.703–708. doi:https://doi.org/10.1038/ngeo2253.

Shi, Y., Zhai, P. and Jiang, Z. (2016). Multi-sliding time windows based changing trend of mean temperature and its association with the global-warming hiatus. *Journal of Meteorological Research*, 30(2), pp.232–241. doi:https://doi.org/10.1007/s13351-016-5093-3.

Shindell, D.T., Miller, R.L., Schmidt, G.A. and Pandolfo, L. (1999). Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature*, 399(6735), pp.452–455. doi:https://doi.org/10.1038/20905.

Shindell, D.T., Schmidt, G.A., Mann, M.E., Rind, D. and Waple, A. (2001). Solar Forcing of Regional Climate Change During the Maunder Minimum.

Science, 294(5549), pp.2149–2152. doi:https://doi.org/10.1126/science.1064363.

Shpakova, R.N., Kusatov, K.I. and S.K. Mustafin (2020). Spatiotemporal Trends in Changes in the River Water Contents in the Sakha Republic (Yakutia). *IOP conference series*, 459(5), pp.052062–052062. doi:https://doi.org/10.1088/1755-1315/459/5/052062.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simpson, I.R., Deser, C., McKinnon, K.A. and Barnes, E.A. (2018). Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures. *Journal of Climate*, 31(20), pp.8313–8338. doi:https://doi.org/10.1175/jcli-d-18-0168.1.

Smith, D., Gillett, N.P., Simpson, I.R., Athanasiadis, P., Baehr, J., Bethke, I., Bilge, T.A., Bonnet, R., Boucher, O., Findell, K.L., Gastineau, G., Silvio Gualdi, Hermanson, L., L. Ruby Leung, Mignot, J., Müller, W.A., Osprey, S., Odd Helge Otterå, Persad, G. and Scaife, A.A. (2022a). Attribution of multi-annual to decadal changes in the climate system: The Large Ensemble Single Forcing Model Intercomparison Project (LESFMIP). *Frontiers in climate*, 4. doi:https://doi.org/10.3389/fclim.2022.955414.

Smith, D.M., Eade, R., Andrews, M.B., Ayres, H., Clark, A., Chripko, S., Deser, C., Dunstone, N.J., García-Serrano, J., Gastineau, G., Graff, L.S., Hardiman, S.C., He, B., Hermanson, L., Jung, T., Knight, J., Levine, X., Magnusdottir, G., Manzini, E. and Matei, D. (2022b). Robust but weak winter atmospheric circulation response to future Arctic sea ice loss. *Nature Communications*, [online] 13(1). doi:https://doi.org/10.1038/s41467-022-28283-y.

Smith, D.M., Scaife, A.A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., Bilbao, R., Borchert, L.F., Caron, L.-P. ., Counillon, F., Danabasoglu, G., Delworth, T., Doblas-Reyes, F.J., Dunstone, N.J., Estella-Perez, V., Flavoni, S., Hermanson, L., Keenlyside, N., Kharin, V. and Kimoto, M. (2020). North Atlantic climate far more predictable than models imply. *Nature*, 583(7818), pp.796–800. doi:https://doi.org/10.1038/s41586-020-2525-0.

Smith, D.M., Scaife, A.A., Eade, R. and Knight, J.R. (2014). Seasonal to decadal prediction of the winter North Atlantic Oscillation: emerging

capability and future prospects. *Quarterly Journal of the Royal Meteorological Society*, 142(695), pp.611–617. doi:https://doi.org/10.1002/qj.2479.

Stenchikov, G. (2002). Arctic Oscillation response to the 1991 Mount Pinatubo eruption: Effects of volcanic aerosols and ozone depletion. *Journal of Geophysical Research*, 107(D24). doi:https://doi.org/10.1029/2002jd002090.

Stenchikov, G., Hamilton, K., Robock, A., Ramaswamy, V. and Schwarzkopf, M.D. (2004). Arctic oscillation response to the 1991 Pinatubo eruption in the SKYHI general circulation model with a realistic quasi-biennial oscillation. *Journal of Geophysical Research: Atmospheres*, 109(D3), p.n/a-n/a. doi:https://doi.org/10.1029/2003jd003699.

Stenchikov, G., Hamilton, K., Stouffer, R.J., Robock, A., Ramaswamy, V., Santer, B. and Graf, Hans-F. (2006). Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. *Journal of Geophysical Research*, 111(D7). doi:https://doi.org/10.1029/2005jd006286.

Stephenson, A.G. (2002). evd: Extreme Value Distributions. *R News*, 2(2), pp.31-32. URL https://CRAN.R-project.org/doc/Rnews/

Stephenson, D. and Pavan, V. (2003). The North Atlantic Oscillation in coupled climate models: a CMIP1 evaluation. *Climate Dynamics*, 20(4), pp.381–399. doi:https://doi.org/10.1007/s00382-002-0281-5.

Stephenson, D.B., Pavan, V. and Bojariu, R. (2000). Is the North Atlantic Oscillation a random walk? *International Journal of Climatology*, 20(1), pp.1–18. doi:https://doi.org/10.1002/(sici)1097-0088(200001)20:1%3C1::aid-joc456%3E3.0.co;2-p.

Stephenson, D.B., Pavan, V., Collins, M., Junge, M.M. and Quadrelli, R. (2006). North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment. *Climate Dynamics*, 27(4), pp.401–420. doi:https://doi.org/10.1007/s00382-006-0140-x.

Strommen, K. (2020). Jet latitude regimes and the predictability of the North Atlantic Oscillation. Quarterly Journal of the Royal Meteorological Society, 146(730), pp.2368–2391. doi:https://doi.org/10.1002/qj.3796.

Strommen, K. and Palmer, T. (2018). Signal and noise in regime systems: A hypothesis on the predictability of the North Atlantic Oscillation. Quarterly

Journal of the Royal Meteorological Society, 145(718), pp.147–163. doi:https://doi.org/10.1002/qj.3414.

Strong, C., Magnusdottir, G. and Stern, H. (2009). Observed Feedback between Winter Sea Ice and the North Atlantic Oscillation. *Journal of Climate*, 22(22), pp.6021–6032. doi:https://doi.org/10.1175/2009jcli3100.1.

Sun, L., Deser, C. and Tomas, R.A. (2015). Mechanisms of Stratospheric and Tropospheric Circulation Response to Projected Arctic Sea Ice Loss. *Journal of Climate*, 28(19), pp.7824–7845. doi:https://doi.org/10.1175/jcli-d-15-0169.1.

Sun, C., Li, J. and Jin, F.-F. (2015). A delayed oscillator model for the quasi-periodic multidecadal variability of the NAO. *Climate Dynamics*, 45(7-8), pp.2083–2099. doi:https://doi.org/10.1007/s00382-014-2459-z.

Suo, L., Gao, Y., Guo, D. and Bethke, I. (2017). Sea-ice free Arctic contributes to the projected warming minimum in the North Atlantic. *Environmental Research Letters*, 12(7), pp.074004–074004. doi:https://doi.org/10.1088/1748-9326/aa6a5e.

Swingedouw, D., Terray, L., Cassou, C., Voldoire, A., Salas-Mélia, D. and Servonnat, J. (2011). Natural forcing of climate during the last millennium: fingerprint of solar variability. Climate Dynamics, 36(7-8), pp.1349–1364. doi:https://doi.org/10.1007/s00382-010-0803-5.

Taylor, K., Stouffer, R. and Meehl, G. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, [online] 93, pp.485–498. doi:https://doi.org/10.1175/BAMS-D-11-00094.1.

Thompson, D.W.J., Barnes, E.A., Deser, C., Foust, W.E. and Phillips, A.S. (2015). Quantifying the Role of Internal Climate Variability in Future Climate Trends. *Journal of Climate*, 28(16), pp.6443–6456. doi:https://doi.org/10.1175/jcli-d-14-00830.1.

Thornton, H.E., Scaife, A.A., Hoskins, B.J. and Brayshaw, D.J. (2017). The relationship between wind power, electricity demand and winter weather patterns in Great Britain. *Environmental Research Letters*, 12(6), p.064017. doi:https://doi.org/10.1088/1748-9326/aa69c6.

Uvo, C.B. and Berndtsson, R. (2002). North Atlantic Oscillation; a Climatic Indicator to Predict Hydropower Availability in Scandinavia. *Hydrology Research*, 33(5), pp.415–424. doi:https://doi.org/10.2166/nh.2002.0016.

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.J. and Rose, S.K. (2011). The representative concentration pathways: an overview. *Climatic Change*, [online] 109(1-2), pp.5–31. doi:https://doi.org/10.1007/s10584-011-0148-z.

Walker, G. and Bliss, E. (1932). World weather V. *Memoirs of the Royal Meteorological Society*, [online] 4(36), pp.53–84. Available at: https://www.rmets.org/sites/default/files/papers/ww5.pdf.

Wang, X., Li, J., Sun, C. and Liu, T. (2017). NAO and its relationship with the Northern Hemisphere mean surface temperature in CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 122(8), pp.4202–4227. doi:https://doi.org/10.1002/2016jd025979.

Wang, Y.-H. ., Magnusdottir, G., Stern, H., Tian, X. and Yu, Y. (2012). Decadal variability of the NAO: Introducing an augmented NAO index. *Geophysical Research Letters*, 39(21), p.n/a-n/a. doi:https://doi.org/10.1029/2012gl053413.

Wanner, H., Brönnimann, S., Casty, C., Gyalistras, D., Luterbacher, J., Schmutz, C., Stephenson, D.B. and Xoplaki, E. (2001). North Atlantic Oscillation – Concepts And Studies. *Surveys in Geophysics*, 22(4), pp.321–381. doi:https://doi.org/10.1023/a:1014217317898.

Wilks, D.S. (2006). *Statistical methods in the atmospheric sciences*. Second ed. Amsterdam: Elsevier.

Woodward, W.A. and Gray, H.L. (1993). Global Warming and the Problem of Testing for Trend in Time Series Data. *Journal of Climate*, [online] 6(5), pp.953–962. doi:https://doi.org/10.1175/1520-0442(1993)006%3C0953:GWATPO%3E2.0.CO;2.

Woollings, T., Franzke, C., Hodson, D.L.R., Dong, B., Barnes, E.A., Raible, C.C. and Pinto, J.G. (2015). Contrasting interannual and multidecadal NAO variability. *Climate Dynamics*, 45(1-2), pp.539–556. doi:https://doi.org/10.1007/s00382-014-2237-y.

Woollings, T., Hannachi, A., Hoskins, B. and Turner, A. (2010). A Regime View of the North Atlantic Oscillation and Its Response to Anthropogenic Forcing. *Journal of Climate*, 23(6), pp.1291–1307. doi:https://doi.org/10.1175/2009jcli3087.1.

Wunsch, C. (1999). The Interpretation of Short Climate Records, with Comments on the North Atlantic and Southern Oscillations. *Bulletin of the American Meteorological Society*, 80(2), pp.245–255. doi:https://doi.org/10.1175/1520-0477(1999)080%3C0245:tioscr%3E2.0.co;2.

Ye, K., Messori, G., Chen, D. and Woollings, T. (2022). An NAO-dominated mode of atmospheric circulation drives large decadal changes in wintertime surface climate and snow mass over Eurasia. *Environmental Research Letters*, 17(4), p.044025. doi:https://doi.org/10.1088/1748-9326/ac592f.

Yu, B. and Lin, H. (2016). Tropical Atmospheric Forcing of the Wintertime North Atlantic Oscillation. *Journal of Climate*, 29(5), pp.1755–1772. doi:https://doi.org/10.1175/jcli-d-15-0583.1.

Zanardo, S., Nicotina, L., Hilberts, A.G.J. and Jewson, S.P. (2019). Modulation of Economic Losses From European Floods by the North Atlantic Oscillation. Geophysical Research Letters, 46(5), pp.2563–2572. doi:https://doi.org/10.1029/2019gl081956.

Zhang, W., Kirtman, B.P., Siqueira, L., Clement, A.C. and Xia, J. (2021). Understanding the signal-to-noise paradox in decadal climate predictability from CMIP5 and an eddying global coupled model. *Climate Dynamics*, 56(9-10), pp.2895–2913. doi:https://doi.org/10.1007/s00382-020-05621-8.

Zuo, J., Ren, H.-L., Li, W. and Wang, L. (2016). Interdecadal Variations in the Relationship between the Winter North Atlantic Oscillation and Temperature in South-Central China. *Journal of Climate*, 29(20), pp.7477–7493. doi:https://doi.org/10.1175/jcli-d-15-0873.1.

Zwiers, F.W. (1987). A Potential Predictability Study Conducted with an Atmospheric General Circulation Model. *Monthly Weather Review*, 115(12), pp.2957–2974. doi:https://doi.org/10.1175/1520-0493(1987)115%3C2957:appscw%3E2.0.co;2