Received: 17 January 2024 | Revised: 8 July 2024 | Accepted: 13 July 2024

DOI: 10.1002/qj.4825

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Large model biases in the Pacific centre of the Northern Annular Mode due to exaggerated variability of the Aleutian Low

Simon H. Lee[1,2] | **Lorenzo M. Polvani**[1,3,4]

[1]Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, USA

[2]School of Earth and Environmental Sciences, University of St Andrews, St Andrews, UK

[3]Department of Earth and Environmental Sciences, Columbia University, New York, New York, USA

[4]Lamont–Doherty Earth Observatory, Columbia University, Palisades, New York, USA

Correspondence
Simon H. Lee, School of Earth and Environmental Sciences, University of St Andrews, St Andrews, UK.
Email: shl21@st-andrews.ac.uk

Funding information
National Science Foundation, Division of Atmospheric and Geospace Sciences, Grant/Award Number: 1914569

Abstract

The Northern Annular Mode (NAM) is traditionally defined as the leading empirical orthogonal function (EOF) of mean sea-level pressure (MSLP) anomalies during winter. Previous studies have shown that the Pacific centre-of-action of the NAM is typically more amplified in models than in reanalysis. Here, we analyse the NAM in hindcasts from nine seasonal prediction models over 1993/1994–2016/2017. In all the models, the Pacific centre-of-action is much larger than in reanalysis over that period, during which the NAM and the North Atlantic Oscillation (NAO) are almost indistinguishable. As a result, the NAM in the models is correlated with Aleutian Low variability around four times more strongly than in reanalysis. We show that this discrepancy can be explained primarily by the amplitude of Aleutian Low variability, which is on average 17% higher in models than in reanalysis, with a secondary effect from a stronger correlation between the Aleutian Low and NAO. When the NAM is computed using zonally averaged MSLP, the Aleutian Low amplitude does not influence the pattern directly. Instead, the amplitude of the Pacific centre-of-action is governed primarily by the correlation between the Aleutian Low and NAO, reducing the apparent Pacific biases in models. While the two methods yield almost identical results in reanalysis, the large Aleutian Low biases result in differences when applied to model data. Modifying the MSLP statistically to alter the Aleutian Low amplitude reveals that the spatial pattern of the traditionally defined NAM is highly sensitive to Aleutian Low variability, even without modifying the correlation between the Aleutian Low and NAO. Hence, the NAM in models may not be as biased as the traditional method would suggest. We therefore conclude that the traditional EOF method is unsuitable for defining the NAM in the presence of highly amplified Aleutian Low variability, and encourage the use of the zonal-mean method.

KEYWORDS

Aleutian Low, annular mode, Arctic Oscillation, biases, climate variability, North Atlantic Oscillation

1 | INTRODUCTION

The Northern Annular Mode (NAM) is the principal mode of atmospheric variability in the extratropical Northern Hemisphere, representing fluctuations in mass between the Arctic and the midlatitudes (Thompson & Wallace, 2000). It is also known as the Arctic Oscillation (AO) after the terminology used upon its introduction by Thompson and Wallace (1998), and both terms remain in use.[1] In the troposphere, the NAM is associated with variability in the latitude of the polar/eddy-driven jets, with same-signed centres-of-action over the midlatitude Atlantic and Pacific (respectively co-located with the climatological Azores High and Aleutian Low) and an opposite-signed centre-of-action extending over the Arctic from the high-latitude Atlantic (co-located with the climatological Iceland Low). In the North Atlantic, the NAM closely resembles the North Atlantic Oscillation (NAO) (e.g., Ambaum *et al.*, 2001). In the stratosphere, the NAM describes variability in the strength of the polar vortex, and is closely related to the strength of the zonal-mean zonal winds at 60°N (Baldwin & Thompson, 2009). Thus, particular attention has been devoted to the NAM as the primary diagnostic of boreal winter stratosphere–troposphere coupling (Baldwin *et al.*, 2003; Baldwin & Dunkerton, 1999, 2001). The NAM exerts a strong influence on the monthly-to-seasonal climate of the Northern Hemisphere (Thompson & Wallace, 2001), while decadal fluctuations in the NAM can locally mask or amplify regional climate change (Butler *et al.*, 2023).

Despite its prominence, the extent to which the NAM represents a meaningful mode of "annular" tropospheric variability has been the focus of much discussion in the literature. This is partly driven by the NAM emerging through empirical orthogonal function (EOF) analysis, which does not a priori capture physical modes (e.g., Dommenget & Latif, 2002). Typically, the NAM is defined as the leading EOF of mean sea-level pressure (MSLP) or geopotential height anomalies poleward of 20°N (e.g., Limpasuvan & Hartmann, 1999; Thompson & Wallace, 1998). Baldwin and Thompson (2009) argued that the NAM in the free troposphere is represented better by the leading EOF of *zonal-mean* geopotential height anomalies; EOFs in the free troposphere in latitude–longitude space instead capture a mix of variability associated with the NAM and Pacific–North American (PNA) patterns (Quadrelli & Wallace, 2004). Nevertheless, the leading EOF of free-tropospheric geopotential height anomalies is often still considered to be the NAM (e.g., L'Heureux *et al.*, 2017). Non-EOF methods have also been proposed, based primarily on standardised anomalies of polar-cap average geopotential heights (Gerber & Martineau, 2018) or simple two-point differences in zonal-mean MSLP (Li & Wang, 2003).

The Pacific centre of the NAM has received notable attention, primarily because the extent to which the NAM and NAO are distinct hinges on its existence. Shortly after the introduction of the NAM, Deser (2000) and Ambaum *et al.* (2001) pointed to the poor correlation between MSLP fluctuations in the Pacific and Atlantic centres as an argument against the NAM paradigm. However, Wallace and Thompson (2002) demonstrated that the correlations of the full-field MSLP would be larger were it not for the existence of the PNA-like second EOF of Northern Hemisphere MSLP, which features *opposite-signed* MSLP fluctuations between the Pacific and Atlantic. Later, Feldstein and Franzke (2006) concluded that the NAO and NAM were indistinguishable; they asserted that the associated variability was not entirely confined regionally, nor truly annular. Debate on the relevance of the NAM versus the NAO persists (Huth & Beranová, 2021).

Moreover, the spatial pattern of the NAM in the Pacific appears to be non-stationary. Zhao and Moore (2009) suggested that the Pacific Decadal Oscillation (PDO) influences the Pacific centre of the NAM. They found that the traditional NAM structure only appears when the PDO is positive, consistent with amplified Aleutian Low variability (Minobe & Mantua, 1999); otherwise, the NAM was indistinguishable from the NAO. Zhao and Moore (2009) remarked that such non-stationarity may have contributed to uncertainty in the difference between the NAM and NAO, because the difference is sensitive to the period analysed. Shi and Nakamura (2014) examined 100 years of reanalysis data, finding that the correlation between Pacific and Atlantic variability—and hence the "axial symmetry" of the NAM—was influenced by Aleutian Low variability generating downstream wave propagation (Honda *et al.*, 2001, 2005). Similar multi-decadal variability in the Pacific centre was also reported by Gong *et al.* (2018), with comparatively little fluctuation in the Atlantic centre.

Notwithstanding debate regarding the interpretation of the NAM, it remains a dominant pattern of variability and is thus a key component of the climate system, which models should reproduce faithfully. Across multiple generations of climate models contributing to the Coupled Model Intercomparison Project (CMIP), a notable and persistent exaggeration of the Pacific centre-of-action of the NAM has been widely documented (see, e.g., Miller *et al.* (2006) for CMIP3, Gong *et al.* (2017) for CMIP5, Coburn and Pryor (2021) for CMIP6, and Fasullo *et al.* (2020) for a multi-generation review). In an analysis of CMIP5 historical simulations, Gong *et al.* (2017) proposed that the exaggerated Pacific centre was due to biased coupling between North Pacific variability (which they termed the "North

Pacific Mode") and the NAM. They proposed a statistical correction technique to extract a representative NAM pattern by linearly removing the North Pacific variability correlated with the NAM. Subsequently, Gong *et al.* (2019) found that the amplitude of the Pacific centre-of-action of the NAM in CMIP5 models was modulated by the stratospheric polar vortex mean state: models with a stronger vortex exhibited greater coupling between Pacific and Atlantic variability, exaggerating the Pacific centre of the NAM. Mechanistically, Gong *et al.* (2019) attributed this process to wave reflection off the strong vortex, following Sun and Tan (2013), which communicates variability between the two basins and is enhanced in models with a stronger vortex. Similarly, Cai *et al.* (2022) found that the Pacific centre of the NAM is stronger in both reanalysis and model simulations when the Quasi-Biennial Oscillation (QBO) is westerly. They attributed this variability to the stronger stratospheric polar vortex during westerly QBO years (i.e., the polar route of the QBO teleconnections: Holton and Tan (1980)), which induces enhanced wave reflection and Pacific–Atlantic coupling similar to the mechanism of Sun and Tan (2013).

While the representation of the NAM has been assessed across several generations of uninitialised CMIP-class models, relatively fewer studies have performed an equivalent intercomparison in *initialised* climate models used for seasonal prediction. This may be partly because verification studies usually use the projection onto the EOF obtained from reanalysis (e.g., Lee *et al.*, 2020; Stockdale *et al.*, 2015). However, several studies across generations of seasonal prediction models have documented large biases in the amplitude of the Pacific centre-of-action of the NAM, similar to those found in CMIP models (Furtado *et al.*, 2021; Kang *et al.*, 2014; Riddle *et al.*, 2013). These studies indicate that even initialised models exhibit Pacific NAM biases, meaning that such biases must develop fairly quickly (i.e., on subseasonal time-scales).

Hence, assessing the representation of the NAM in a suite of state-of-the-art seasonal climate models is a key goal of the present study. In particular, we investigate the sensitivity of the NAM pattern to the method used to compute it. Our results show a dominant role for large model biases in the amplitude of Aleutian Low variability, which confound the NAM pattern when it is defined as the leading EOF of hemisphere-wide MSLP anomalies. This has implications for the computation and interpretation of the NAM in models where Pacific variability biases are widespread. Indeed, we question whether or not the NAM is truly as biased as the leading EOF would suggest.

The remainder of the article is laid out as follows. In Section 2, we provide details on the model hindcast and reanalysis data used in this study. Then, in Section 3, we assess the representation of the NAM, NAO, and Aleutian Low (AL) in the models. In Section 4, we build and test a hypothesis of what governs the involvement of Pacific variability in the NAM, and its method dependence. A discussion of the implications of our findings and our conclusions follow in Section 5.

2 | DATA AND METHODS

2.1 | Model and reanalysis data

Our analysis is based on monthly-averaged hindcasts (reforecasts for dates in the past) from nine different models that contribute to the Copernicus Climate Change Service (C3S) suite of seasonal prediction systems. Key details are provided in Table 1. Hereafter, we refer to each model by the name of the modelling centre (except in the case of Environment and Climate Change Canada (ECCC), where we refer to the low-resolution CanCM4i as "ECCC-lo" and the higher-resolution GEM5-NEMO as "ECCC-hi" to distinguish between the two models that contribute to the ECCC CanSIPSv2.1 prediction system).

Only data from the common hindcast period is used, which covers the 24 winters from 1993/1994 to 2016/2017. We analyse an extended winter season from November–March (NDJFM), and hence we use only hindcasts nominally initialised on October 1. This allows a period at the start for individual ensemble members to diverge from the initial state (consistent with the use of data from ≥1 month ahead in seasonal forecasts), while it is constrained by the maximum available six-month lead time. Note that the initialisation and ensemble generation procedures vary among the models, but this is unlikely to be relevant for the purposes of this study. For verification, we use the ERA5 reanalysis (Hersbach *et al.*, 2020) from the European Centre for Medium-Range Weather Forecasts (ECMWF). The large-scale fields analysed here are unlikely to differ noticeably between modern reanalyses in this time period. All data are obtained on a common 1° latitude–longitude grid, and all anomalies are defined as departures from the average for each month across 1993/1994–2016/2017 (i.e., with the lead-time-dependent climatology subtracted).

2.2 | Statistical analysis

The leading patterns of MSLP variability are extracted using EOF analysis with the Python package *eofs* (Dawson, 2016). EOF analysis is performed across all ensemble members over NDJFM, with input data weighted by the square-root of cosine latitude. The resultant principal

TABLE 1 Details of the nine seasonal prediction models used in this study.

Centre	Model	Horizontal resolution	Vertical levels	Model top	Ensemble size
CMCC	SPS3.5	0.5°	L46	0.2 hPa	40
DWD	GCFS 2.1	T127	L95	0.01 hPa	30
ECCC-lo	CanCM4i	T63	L35	1 hPa	10
ECCC-hi	GEM5-NEMO	1.1°	L85	0.1 hPa	10
ECMWF	SEAS5	Tco319	L91	0.01 hPa	25
JMA	CPS3	TL319	L100	0.01 hPa	10
Météo-France	System 8	TL359	L137	0.01 hPa	25
NCEP	CFSv2	T128	L64	0.02 hPa	24
UKMO	GloSea6-GC3.2	N216	L85	85 km	28

Abbreviations: CMCC, Centro Euro-Mediterraneo sui Cambiamenti Climatici; DWD, Deutscher Wetterdienst; ECCC, Environment and Climate Change Canada; ECMWF, European Centre for Medium-Range Weather Forecasts; JMA, Japan Meteorological Agency; NCEP, National Centers for Environmental Prediction; UKMO, United Kingdom Met Office.

component (PC) time series are then standardised, and the EOF patterns are expressed as the linear regression of the standardised PC time series with the input variable. We also define the "amplitude" A of an EOF as the square-root of its eigenvalue λ (i.e., the standard deviation σ of the PC time series); this is equivalent to the square root of the (square-root cosine weighted) sum-of-squares of the regression pattern. For ease of interpreting any model biases, we normalise the amplitudes by the equivalent ERA5 statistic (i.e., $A = \sqrt{\lambda_{\text{model}}/\lambda_{\text{ERA5}}}$), such that $A = 1$ indicates the modelled amplitude matches that in ERA5.

We assess three modes of variability: the NAM, the NAO, and the AL. We use two different methods to compute the NAM, which we refer to as the "traditional NAM" and "zonal-mean NAM". The traditional NAM is defined as the leading EOF of monthly NDJFM MSLP anomalies across the Northern Hemisphere poleward of 20°N, following, for example, Thompson and Wallace (2000). The zonal-mean NAM is defined similarly, except the MSLP anomalies are first zonally averaged prior to computing the EOF, following Baldwin and Thompson (2009). Similar results can also be obtained when using a shorter winter season (e.g., DJF or DJFM); we opt for NDJFM for a larger sample size. The NAO is defined similarly but for the North Atlantic (20–80°N, 90°W–40°E), following, for example, Hurrell (1995), while the AL is defined over the North Pacific (20–70°N, 120°E–120°W) following Gong *et al.* (2017, 2018). The sign of all EOFs is defined such that a positive loading is associated with a positive MSLP anomaly in the North Pacific and a negative MSLP anomaly over the Arctic, following convention for the NAM (note that this results in a positive AL index representing a *weaker* Aleutian low-pressure system). Structural biases aside, the NAM, NAO, and AL patterns all emerge as the leading EOF in all the models (as defined by their pattern correlation with the equivalent ERA5 EOFs).

Statistical significance is assessed throughout the article at the two-sided 95% confidence level. This is computed using the 2.5th and 97.5th percentiles of a distribution obtained by bootstrapping the sample statistic with replacement 10,000 times. In addition, p-values are computed from the bootstrapped distribution as twice the probability of the sample statistic having the opposite sign (i.e., equivalent to the inverse of the confidence interval within which zero would lie).

3 | REPRESENTATION OF THE NAM, NAO, AND AL

3.1 | The NAM

We begin by showing the regression pattern of the traditional NAM in Figure 1. In ERA5 (Figure 1j), the NAM explains 21.7% of the monthly MSLP variance in the domain, similar to previous studies (e.g., Thompson & Wallace, 2000); note that some studies compute the EOF on the seasonal-mean field (e.g., Gong *et al.*, 2017) and thus report a higher variance fraction. Atlantic variability resembling the NAO is dominant; only minimal Pacific variability is seen, with a small Pacific centre-of-action peaking at 2 hPa· σ^{-1} west of British Columbia. The NAM over this period is therefore a less annular, more NAO-like pattern than that first introduced by Thompson and Wallace (1998), but resembles NAM patterns obtained during periods of low Pacific–Atlantic coupling (Gong *et al.*, 2018; Shi & Nakamura, 2014).

In contrast, in *all* the models analysed here (Figure 1a–i), the Pacific centre-of-action is much larger

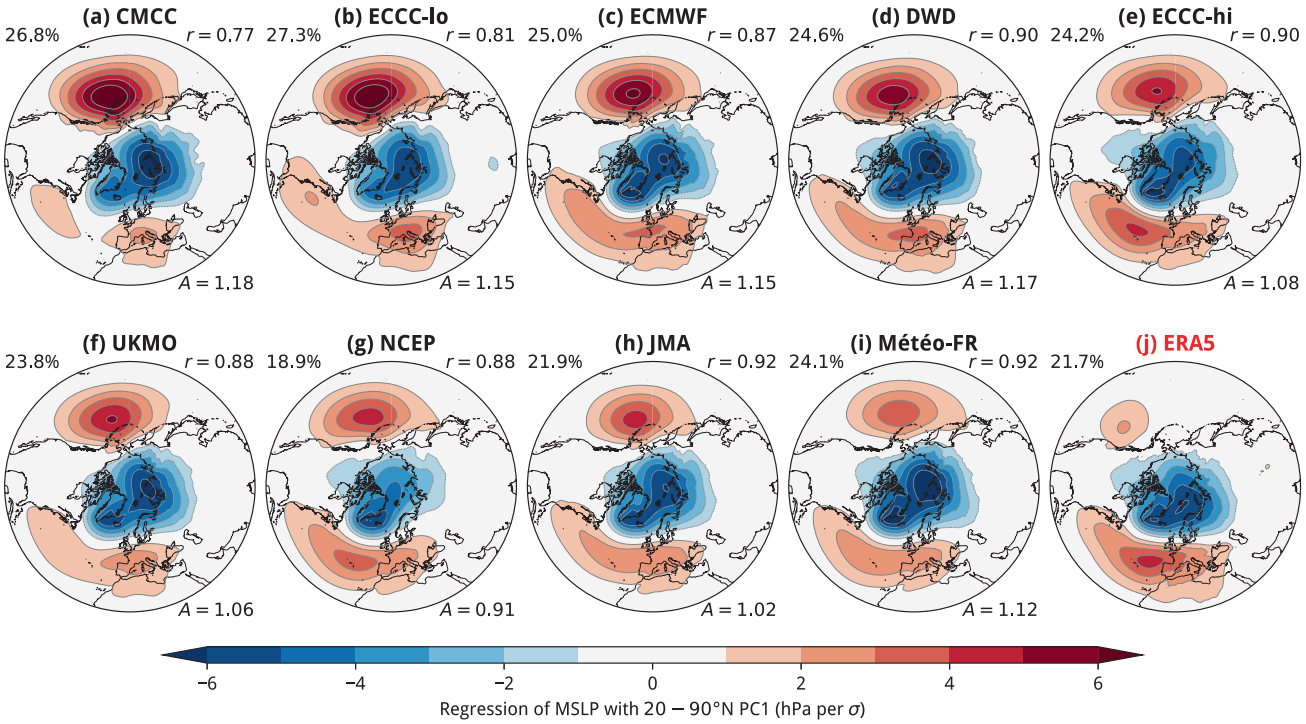

FIGURE 1 Regression pattern of the standardised PC1 of 20–90°N monthly NDJFM MSLP anomalies in (a–i) nine C3S models and (j) ERA5 over 1993/1994–2016/2017. In the top left of all panels, the percentage of total variability explained by the EOF is shown. In panels (a–i), the cosine-latitude weighted pattern correlation (r) of the map with the ERA5 map in (j) is shown on the top right, and the ratio of the standard deviation of the PC time series in each model to that in ERA5 is shown on the bottom right (A). [Colour figure can be viewed at wileyonlinelibrary.com]

than in ERA5 over the same period, consistent with previous studies of climate models. The fraction of explained variance is also greater than in ERA5 in seven of the models, with the largest fraction (27.3%) in the ECCC-lo model (Figure 1b); in the National Centers for Environmental Prediction (NCEP) model, the leading EOF explains only 18.9% of the variance, while the explained variance fraction in the Japan Meteorological Agency (JMA) model matches that in ERA5. Similarly, the amplitude of the pattern is high-biased versus ERA5 in all but NCEP ($A = 0.91$), with biases ≥15% in the ECCC-lo, Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), ECMWF, and Deutscher Wetterdienst (DWD) models. Notably, in all the models, the amplitude of the Pacific variability exceeds the equivalent in the Atlantic, as already documented for climate models (e.g. Gong *et al.*, 2017); in ECCC-lo and CMCC, it is the largest of the three centres-of-action. These two models also exhibit the lowest pattern correlation with ERA5 ($r = 0.81$ and 0.77, respectively). In addition to the exaggerated Pacific centre-of-action, the low pattern correlation in these two models is consistent with their lack of a traditional NAO-like structure in the Atlantic, with no clear Iceland Low centre (instead, the high-latitude centre is displaced toward the Laptev Sea). In these two models, there is a second Atlantic centre located near Bermuda, and the structure as a whole shows some similarity with the mixed PNA–NAM pattern of the free-tropospheric EOF1 in Baldwin and Thompson (2009).

It would be simple to assume that such large discrepancies between the structure of the leading EOF in the models and in reanalysis indicate a large, fundamental bias in the representation of "annular variability" in the models. However, this assumes that the leading EOF in latitude–longitude space always captures the NAM cleanly. As we shall see, such an assumption is not necessarily valid.

An alternative method for calculating the NAM is first to zonally average the MSLP field, prior to taking an EOF. The resultant PC time series—the zonal-mean NAM—can then be regressed against MSLP anomalies in latitude–longitude space, as shown in Figure 2. This zonal-mean method is usually applied when computing the NAM on multiple pressure levels. However, Baldwin and Thompson (2009) showed that, at least in reanalysis, the zonal-mean method yields the same pattern at the surface as the traditional latitude–longitude definition, and is thus interchangeable. Our results agree for ERA5:

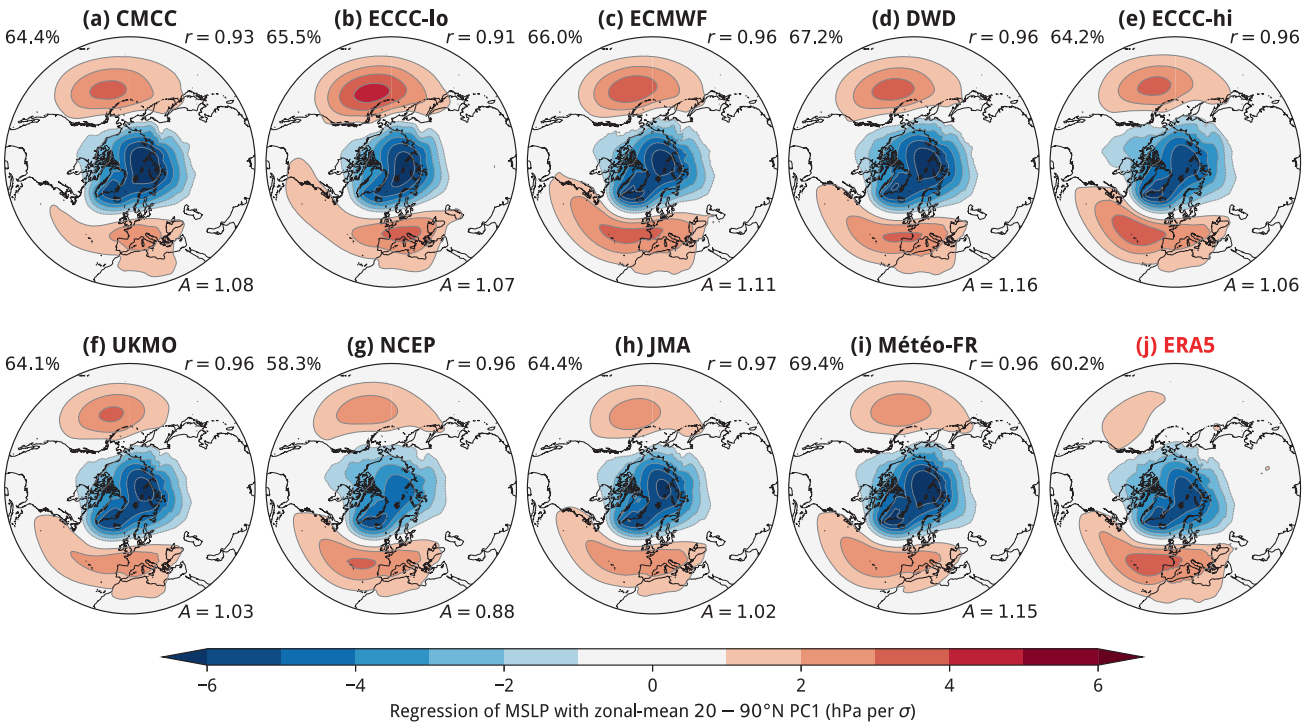

FIGURE 2 As in Figure 1, but for the standardised PC1 of zonal-mean MSLP anomalies poleward of 20°N. The percentage of total variance refers only to the variance of the zonal-mean anomalies. [Colour figure can be viewed at wileyonlinelibrary.com]

the regression pattern of the zonal-mean NAM (Figure 2j) closely resembles the traditional NAM (cf. Figure 1j), with only a weak Pacific centre.

In contrast, while the Pacific centre remains systematically larger than in ERA5 in all the models, it is *not as large* as in the equivalent traditional NAM (cf. Figure 1a–i). For example, in ECCC-lo (Figure 2b), the Pacific centre of the zonal-mean NAM peaks at 4 hPa, versus the 7-hPa peak in the traditionally defined NAM (cf. Figure 1b). This is not because the spatial pattern of the zonal-mean NAM is necessarily zonally symmetric, as evidenced by differences in the magnitudes of the Azores and Aleutian centres across the models and ERA5. The pattern correlations with ERA5 are improved in all models versus the traditional NAM definition, and all are above 0.90; notably, CMCC increases from $r = 0.77$ to $r = 0.93$. In three models (NCEP, JMA, and Météo-France), the Pacific centre is only minimally high-biased (unlike in the traditional NAM), although in NCEP the amplitude of the pattern as a whole is low-biased ($A = 0.88$). While the zonal-mean NAM in these three models features a larger Pacific centre than in the equivalent reanalysis, it does not differ as markedly from the (traditionally defined) NAM defined in reanalysis over longer periods (e.g., Thompson & Wallace, 1998).

Hence, the exaggeration of the Pacific centre of the NAM is sensitive to the method: while the spatial pattern obtained by either method is almost identical when applied to reanalysis, the same cannot be said for the models. The Pacific centre in the models is larger than reanalysis in either method, but *much* larger in the traditional method. We now seek to understand (1) why the traditional NAM in the models differs from reanalysis and (2) why the zonal-mean NAM differs markedly from the traditional NAM in *only* the models. To do so, we first assess the variability in the Atlantic (the NAO) and Pacific (the AL) separately.

3.2 | The NAO

In Figure 3, we show the spatial pattern of the NAO as the regression of the PC time series with MSLP across the same domain as the NAM. The pattern in ERA5 across the whole hemisphere (Figure 3j) closely resembles both definitions of the NAM, but is associated with slightly less Pacific variability. In the models, the structure of the NAO within the North Atlantic domain is similar to that in ERA5. However, in CMCC (Figure 3a) and ECCC-lo (Figure 3b), the southern centre-of-action is displaced eastwards and centred over France, rather than the Azores (similar to the structure seen in either NAM method; cf. Figures 1 and 2). This displaced centre means that non-EOF methods for defining the NAO that are based on the variability in reanalysis (e.g., Hurrell & Deser, 2010) do not necessarily

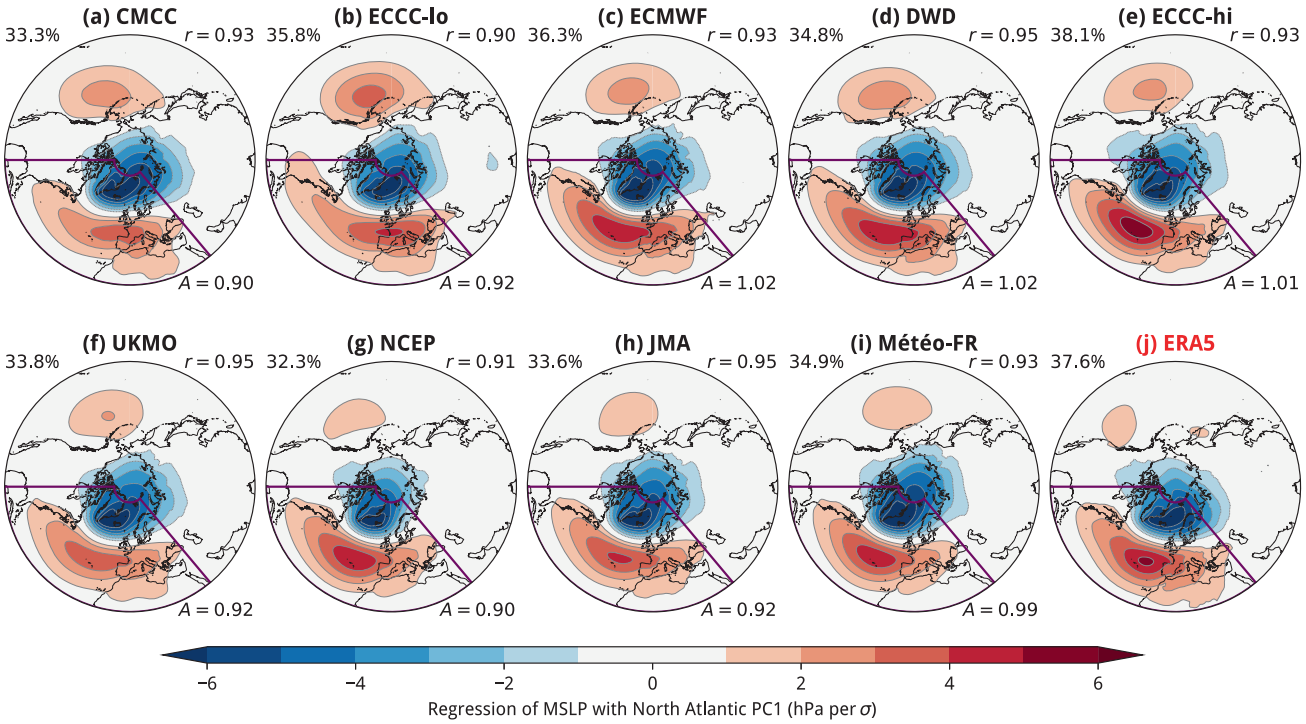

FIGURE 3 As in Figures 1 and 2, but for the standardised PC1 of MSLP anomalies in the North Atlantic (delineated by the polygon). The percentage of total variance refers only to the variance within the EOF domain. [Colour figure can be viewed at wileyonlinelibrary.com]

capture the same dominant pattern of North Atlantic variability in these models. Overall, the modelled amplitude of the NAO is generally close to or lower than that in ERA5, with the largest bias in the CMCC and NCEP models (10% smaller). Similarly, the fraction of explained variance is close to or smaller than that in ERA5 (38%), ranging from 32% in NCEP to 38% in ECCC-hi.

Outside the Atlantic sector, the NAO in the models is associated with slightly more Pacific variability than in ERA5. This difference is largest in ECCC-lo (with a peak of $\sim 3\,\text{hPa}\cdot\sigma^{-1}$ in the Pacific; Figure 3b), while it is small in the NCEP (Figure 3g), JMA (Figure 3h), and Météo-France (Figure 3i) models. However, the hemisphere-wide regression pattern is not identical to either the traditional NAM (Figure 1) or the zonal-mean NAM (Figure 2)—in other words, both definitions of the NAM capture more Pacific variability than that which is concurrently associated with the NAO.

3.3 | The AL

We next examine the regression pattern and amplitude of the AL (Figure 4), the leading EOF of MSLP in the North Pacific. The AL constitutes one node of the PNA pattern; when its index is regressed on mid-tropospheric geopotential height, the canonical PNA wave train (e.g., Wallace & Gutzler, 1981) emerges (not shown). The structure of the AL within the North Pacific domain in all models closely matches that in ERA5, consisting of a single centre-of-action centred near the Aleutian Islands. However, two key differences are evident: (1) in all models, the amplitude of the AL exceeds that in ERA5, with a maximum bias of +36% in the CMCC model (Figure 4a) and a minimum of +4% in the Météo-France model (Figure 4i); and (2) in all models, the AL is more strongly associated with variability over the Arctic and North Atlantic than in ERA5. This follows from the stronger link to Pacific variability seen in the model NAO regression patterns (cf. Figure 3), and indicates a stronger correlation between the dominant patterns of variability in both basins.

Consistent with the high-biased amplitudes, the fraction of explained variance is also higher than in ERA5 (36.5%) in all but the Météo-France model (despite the amplitude being slightly larger in that model, indicating greater variance in higher-order EOFs). In the CMCC model, the AL is not only the leading mode, but explains the majority (51%) of the variance. We also verified that the high-biased AL amplitude is not an artefact of EOF analysis by computing the standard deviation of the MSLP anomalies; these are similarly high-biased in the AL region in all the models (not shown).

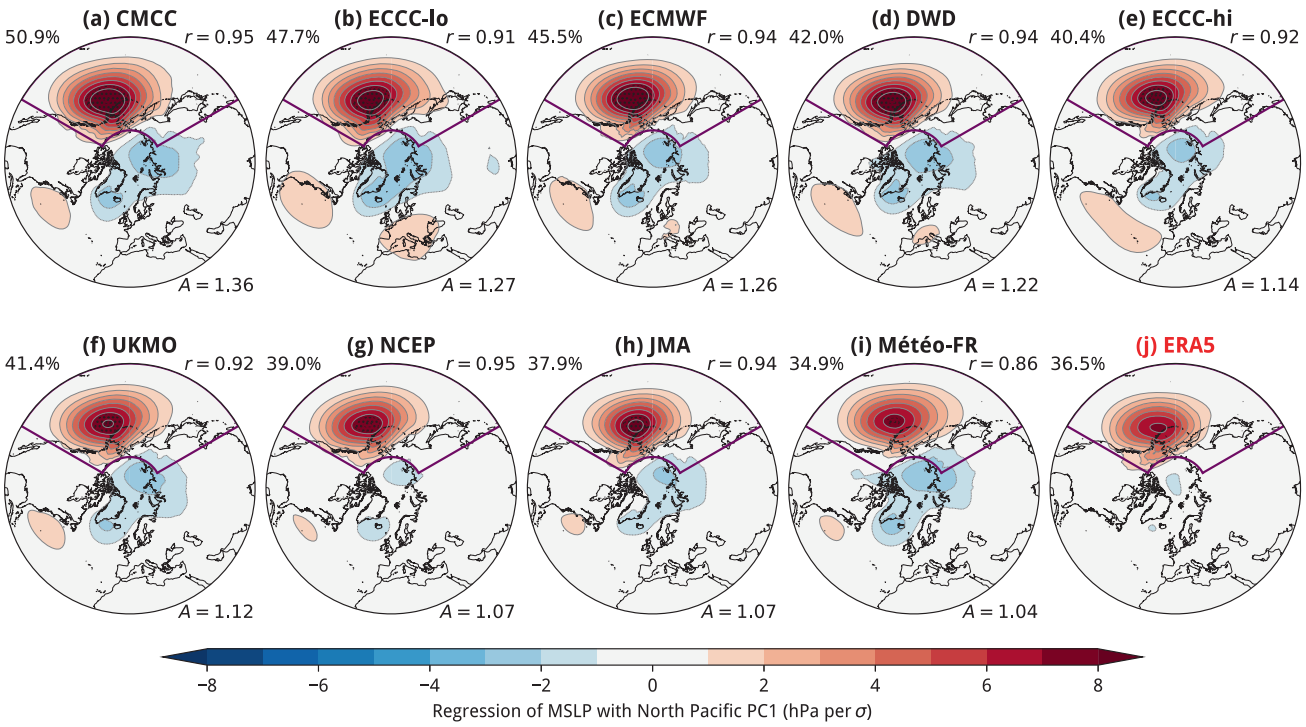

FIGURE 4 As in Figures 1–3, but for the standardised PC1 of MSLP anomalies in the North Pacific (delineated by the polygon). The percentage of total variance refers only to the variance within the EOF domain. Regions where the regression coefficient exceeds the maximum in ERA5 are stippled. [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | Time series correlations

Having examined the spatial patterns of the EOFs, we now quantify the correlation between their associated PC time series (Figure 5; note that, because all indices have unit standard deviation, the Pearson correlation is equivalent to the slope of the linear regression) to determine the extent to which the different EOFs capture distinct patterns of variability. In ERA5, the NAO and the traditional NAM (Figure 5a) are almost indistinguishable ($r = 0.92, p < 0.01$), similar to the correlation reported by Deser (2000) over 1947–1997 (0.95). All nine models show a lower temporal correlation between the two indices; in CMCC, the correlation is only 0.70. This difference is consistent with the different spatial patterns of the NAM and NAO shown in Figures 1 and 3. In contrast, the correlation between the NAO and zonal-mean NAM (Figure 5b) is effectively identical to that in ERA5 in all models ($r \approx 0.83$). This is ∼0.10 less than the equivalent for the traditional NAM in ERA5, indicating a somewhat greater distinction between the NAO and NAM than is obtained without zonal averaging.

As suggested by the spatial patterns in Figure 1, the AL is 3–5 times more strongly correlated with the NAM (Figure 5c) in the models than in ERA5 (in which $r = 0.15, p = 0.10$). The correlation exceeds 0.50 in all models; in ECCC-lo, the AL–NAM correlation ($r = 0.79$) matches the NAO–NAM correlation, while in CMCC, the AL and NAM are more strongly correlated ($r = 0.76$) than the NAM and NAO ($r = 0.70$). Such large biases are not found when analysing the zonal-mean NAM (Figure 5d); while the correlations in all models exceed that in ERA5, they are all *less* than 0.50, with all but CMCC less than 0.40. Thus, while the traditional NAM definition shows a weaker NAO–NAM link and substantially stronger AL–NAM link in models than in reanalysis, the zonal-mean definition shows no biases in the NAO–NAM link and a moderately stronger AL–NAM link. Importantly, even if it is high-biased versus ERA5, the correlation between the AL and zonal-mean NAM is systematically much lower (∼50% smaller) than the correlation between the NAO and zonal-mean NAM (cf. Figure 5d with 5b). Consequently, the zonal-mean NAM in all the models remains rooted in NAO variability, which is consistent with observations over the 20th century (Gong *et al.*, 2018).

The correlation between the AL and NAO (Figure 5e) represents the physical link between variability in the Atlantic and Pacific, as the two EOFs are computed over independent regions. It is minimal and non-significant in ERA5 ($r = 0.08, p = 0.39$) but 2–5 times larger in the models, albeit lying within the ERA5 confidence interval in NCEP, JMA, and Météo-France, and it does not exceed

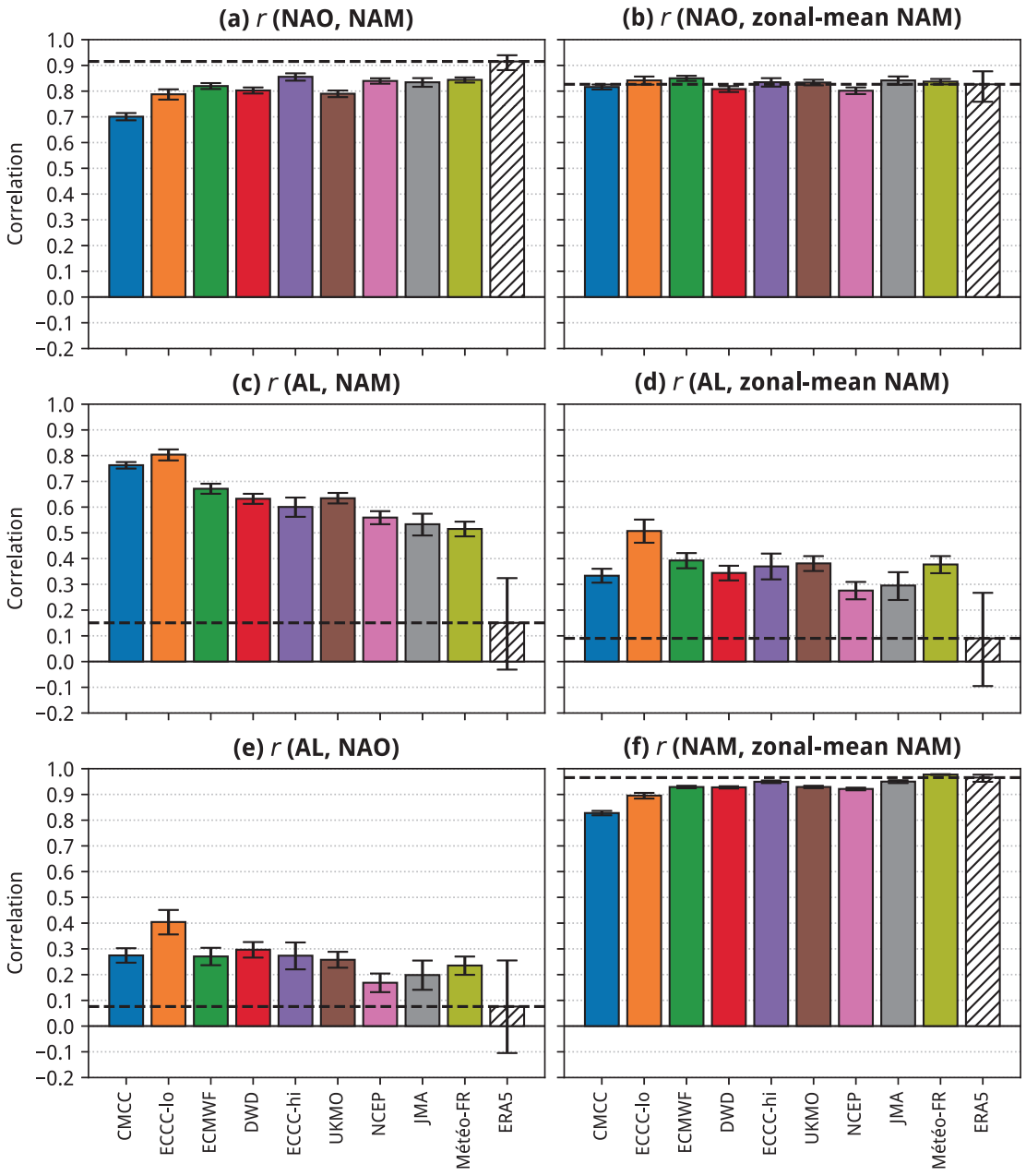

FIGURE 5 Pearson correlations between the monthly time series of (a) the NAO and the NAM, (b) the NAO and the zonal-mean NAM, (c) the Aleutian Low and the NAM, (d) the Aleutian Low and the zonal-mean NAM, (e) the Aleutian Low and the NAO, and (f) the NAM and zonal-mean NAM in the nine C3S models (solid bars) and ERA5 (hatched bar) over NDJFM 1993/1994–2016/2017. Error bars show 95% confidence intervals obtained by 10,000 bootstrap resamples with replacement. [Colour figure can be viewed at wileyonlinelibrary.com]

0.30 except in ECCC-lo ($r = 0.40, p < 0.01$). In all models, the correlation between the AL and NAO is similar to, but slightly less (~30% smaller) than, the correlation between the AL and zonal-mean NAM, but substantially less (~60% smaller) than the correlation with the traditional NAM. Therefore, the correlation between the AL and the traditional NAM cannot simply be equated to the correlation between the AL and NAO. This does not appear to be the case for the correlation between the AL and zonal-mean NAM, which more closely matches the correlation between the leading modes in the two basins. Finally, Figure 5f shows that the two definitions of the NAM are temporally indistinguishable in ERA5 ($r = 0.97$); in the models, the correlation is either close to ERA5 or somewhat lower (particularly for CMCC, in which $r = 0.83$).

To summarise our analysis so far, we have identified three systematic biases in the nine seasonal models.

1. The amplitude of AL variability is on average 17% greater than in reanalysis.
2. The correlation between the AL and the traditional NAM is around four times stronger than in reanalysis, and around twice as strong as the correlation with the zonal-mean NAM.
3. The correlation between the AL and NAO is around three times stronger than in reanalysis.

In the next section, we demonstrate that the biased amplitude of the AL variability plays a dominant role in the pattern of the traditionally defined NAM, and in distinguishing it from the zonal-mean NAM.

4 | THE ROLE OF AL VARIANCE

Given the systematic exaggeration of AL variability in seasonal prediction models of the present generation (Figure 4), we hypothesise that AL amplitude biases play a key role in NAM biases. To further this hypothesis, Figure 6a shows the average Pacific sector loading (computed over 20–70°N, 120°E–120°W; i.e., the same domain over which the AL is defined) in the traditional NAM regression pattern as a function of AL amplitude. The large spread across all nine models and ERA5 is explained extremely well by the AL amplitude ($r = 0.96$, $p < 0.01$). Next, we relate the temporal correlations presented in Figure 5 to the AL amplitude. The inter-model spread in the AL–NAM correlation is also explained extremely well by the AL amplitude (Figure 6b; $r = 0.90$, $p < 0.01$). When including ERA5, the correlation remains high, but is slightly reduced ($r = 0.81, p < 0.01$); in other words, the NAM in ERA5 is less well-correlated with the AL than would be expected from a simple linear extrapolation from the models based solely on AL variance. By considering Figure 5a and 5b together, we also note that the AL–NAM correlation and the intensity of the Pacific centre of the NAM are themselves well correlated ($r = 0.91$, $p < 0.01$ across the models and ERA5).

In a similar but opposite sense, the extent to which the NAO and NAM are distinct (Figure 6c) increases as a function of AL amplitude across the models and ERA5 ($r = -0.83$, $p < 0.01$). We also find a good correlation between the AL–NAO correlation and AL amplitude in the models and ERA5 (Figure 6d; $r = 0.74$, $p < 0.01$): as AL amplitude increases, so does the correlation between variability in the two basins. This has previously been shown in observations and climate models, arising from enhanced downstream wave propagation (Gong *et al.*, 2019; Pinto *et al.*, 2011; Shi & Nakamura, 2014; Sun & Tan, 2013). Thus, in models with a higher AL amplitude, the traditional NAM is more strongly correlated with the AL and more distinct from the NAO, *and* the AL and NAO are more strongly correlated.

Figure 6e,f shows equivalent analyses to Figure 6b,c, but using the zonal-mean NAM definition. We find no comparably strong dependence on AL amplitude; the inter-model spread in the correlation between the AL and zonal-mean NAM is only weakly explained by the AL amplitude ($r = 0.39$, $p = 0.06$), while the correlation between the NAO and zonal-mean NAM does not depend on the AL amplitude. Hence, unlike the traditional definition, the zonal-mean NAM does not appear to be directly sensitive to the amplitude of AL variability. The modest increase in correlation between the AL and zonal-mean NAM in models with a higher AL amplitude would be expected as an indirect effect of the higher AL–NAO correlation at high AL amplitude (Figure 5c). Finally, we note that the difference in correlation between the NAO and the two NAM definitions (i.e., Figure 5c minus 5f) also depends on AL amplitude ($r = -0.83$, $p < 0.01$), such that the traditional NAM is more NAO-like than the zonal-mean NAM at low AL amplitude (as in ERA5) and less NAO-like at high AL amplitude.

4.1 | A statistical model

The correlation analyses in Figure 6 suggest that the extent to which the traditional NAM is correlated with the AL depends strongly on the AL amplitude, while the same is not true for the zonal-mean NAM. We now test this with a statistical model based on perturbing the MSLP field in ERA5 linearly over the 24-year hindcast period.

First, we construct an artificial quantity, $MSLP'$, by linearly adjusting the component of Northern Hemisphere MSLP variability associated with the AL by a factor C:

$$MSLP'(t,x,y) = MSLP(t,x,y) + C \times AL(t) \times \beta\{AL(t), MSLP(t,x,y)\}, \quad (1)$$

where $AL(t)$ is the standardised AL PC time series and β is a linear regression function (i.e., $\beta\{AL(t), MSLP(t,x,y)\}$ is the map shown in Figure 4j). If C is positive, the AL amplitude is increased, while negative values of C reduce the amplitude; the AL is entirely regressed out with $C = -1$ and doubled in amplitude for $C = 1$. For each value of C, we recompute the AL, NAM, and NAO indices using the $MSLP'$ field, and then compute the correlation between the time series. The results are shown in Figure 7 (thick line). For convenience, in this analysis we use the term "NAM" to refer to the leading EOF of $MSLP'$, regardless of whether it resembles an annular mode.

For decreases or small increases ($\sim +0.1$) in amplitude of the AL, we find only a modest change in the Pacific centre of the traditional NAM (Figure 7a) and the correlation

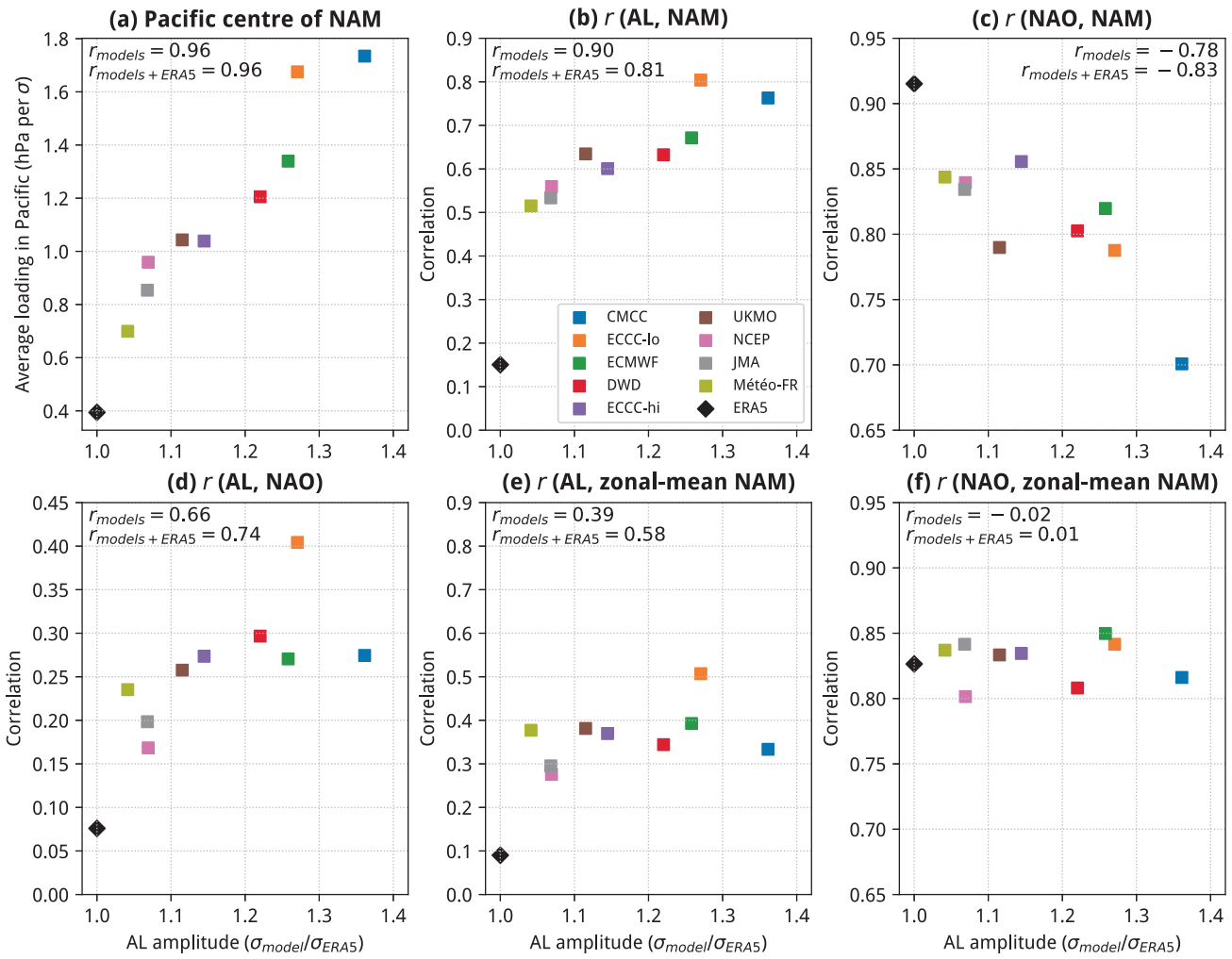

FIGURE 6 Scatter plots showing, as a function of the Aleutian Low amplitude, (a) the average Pacific sector (20–70°N, 120°E–120°W) loading in the NAM regression pattern, and (b–f) the time series correlation between (b) the Aleutian Low and the NAM, (c) the NAO and the NAM, (d) the Aleutian Low and the NAO, (e) the Aleutian Low and the zonal-mean NAM, and (f) the NAO and the zonal-mean NAM. In each panel, the correlation across the nine models (r_{models}) and the correlation across models and ERA5 ($r_{\text{models + ERA5}}$) are given. The AL amplitudes are normalised by the amplitude in ERA5. Note that the *y*-axes differ between subplots and are shared in only the second and third columns. [Colour figure can be viewed at wileyonlinelibrary.com]

between the AL and the traditional NAM (Figure 7b). However, further increases in AL amplitude lead to a steep and nonlinear increase in intensity of the Pacific centre of the NAM and the correlation between the NAM and the AL. We note that, as AL amplitude increases, the intensity of the Pacific centre of the NAM evolves very similarly to the AL–NAM correlation, indicating that the two are closely related. The correlation with the NAO (Figure 7c) is slightly less sensitive to AL amplitude, remaining largely unchanged for values of C below ~1.25, but then undergoes a rapid decline. Thus, for AL amplitudes between ~1.1 and 1.3 times the 1994–2017 value, the traditionally defined NAM incorporates primarily NAO variability alongside some AL variability, while at amplitudes >1.35 the AL–NAM correlation exceeds the AL–NAO correlation (similar to what is seen in the CMCC model). In contrast, we find no such extreme sensitivity of the zonal-mean NAM to the AL amplitude: there is only a minimal increase (~0.10) in the intensity of the Pacific centre (Figure 7d) and the correlation with the AL (Figure 7e), while the correlation between the zonal-mean NAM and NAO is insensitive to the AL amplitude (Figure 7f). Importantly, these changes occur without modifying the AL–NAO correlation directly (only a minimal increase of 0.06 occurs across the full range of AL amplitudes shown here). Hence, the extent to which AL variability is captured by the traditionally defined NAM is *not* solely representative of the physical linkage between the basins (i.e., the AL–NAO correlation), but can simply be a statistical effect that is dependent on the method used.

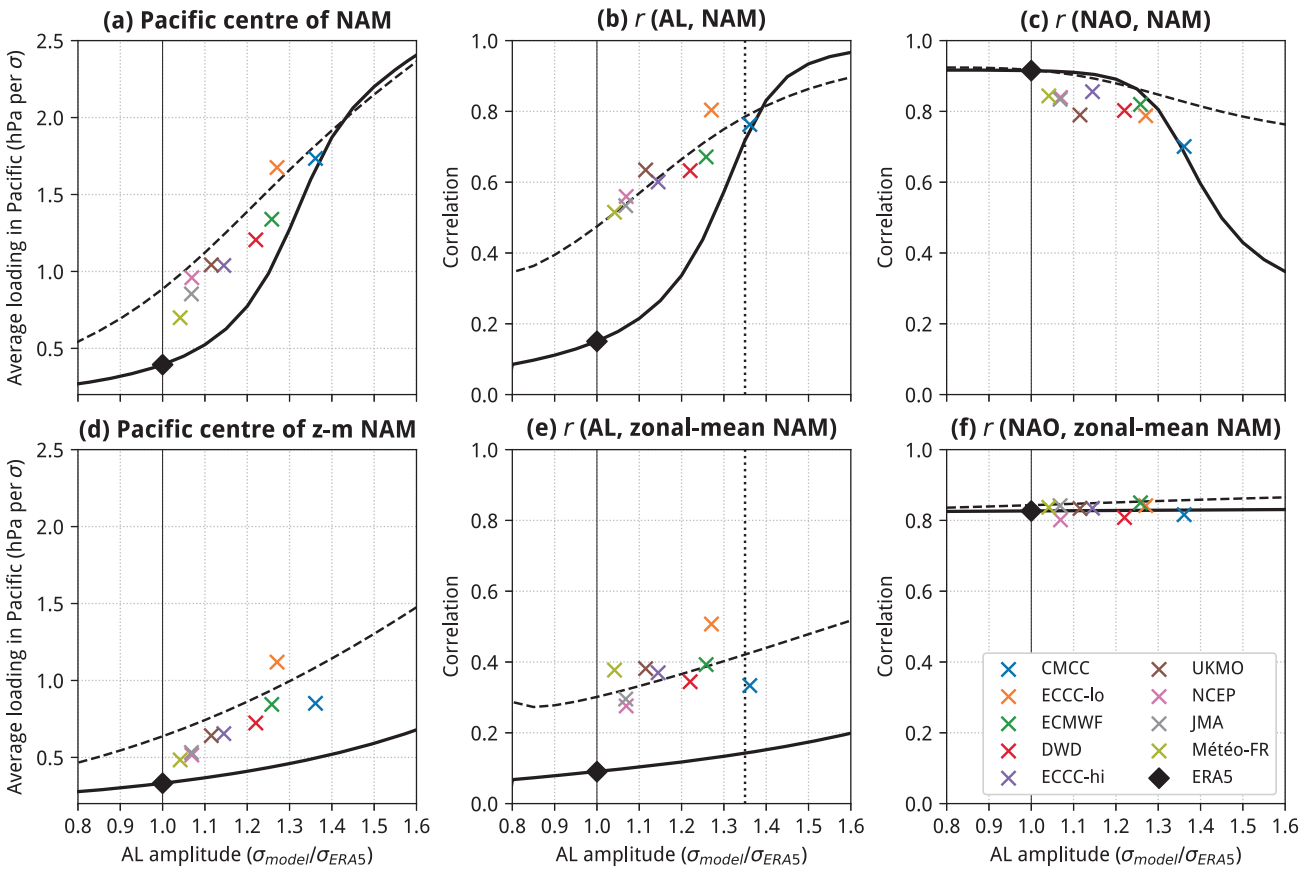

FIGURE 7 Results from the statistical model. As a function of AL amplitude, (a) shows the average Pacific sector loading in the NAM regression pattern, (b) shows the correlation between the NAM and AL, (c) shows the correlation between the NAM and NAO, (d) shows the average Pacific sector loading in the zonal-mean NAM regression pattern, (e) shows the correlation between the zonal-mean NAM and AL, and (f) shows the correlation between the zonal-mean NAM and NAO. Crosses show the values for the nine C3S models. The thick line shows the results for the raw ERA5 field, while the dashed line shows the results for the ERA5 field modified to the multi-model median AL–NAO correlation ($r = 0.27$). The vertical dotted line in (b) and (e) indicates the amplitude at which the traditional NAM becomes more correlated with the AL than the NAO. [Colour figure can be viewed at wileyonlinelibrary.com]

It is also apparent from Figure 7 that statistically modifying the AL amplitude alone does not fully explain the model correlations: the AL in the models is more strongly correlated with both the traditional NAM and zonal-mean NAM than can be explained by only changing the amplitude. This is expected, given that the statistical model does not incorporate the physical effects of enhanced AL amplitude; namely, the increased downstream wave propagation and enhanced AL–NAO correlation (as suggested by Figure 6c). To approximate this effect, we first adjust the MSLP field to the multi-model median AL–NAO correlation ($r = 0.27$). We do so by adding a component of NAO variability that is correlated with the AL:

$$MSLP''(t,x,y) = MSLP(t,x,y) + \Delta r \times AL(t) \times \beta\{NAO(t), MSLP(t,x,y)\}, \quad (2)$$

where $\Delta r = 0.27 - 0.08 = 0.18$ is the difference between the specified and raw AL–NAO correlation (for small values of Δr).[2] We then perform the same AL amplitude adjustment as in Equation (1). It should be noted that this test serves only as an approximation of the expected physical effects, in part because it does not incorporate a dependence of the AL–NAO correlation on the AL amplitude. Nevertheless, the results (shown in the dashed black lines in Figure 7) fit the models more closely than those obtained by solely perturbing the AL amplitude. The overall story does not change, with a large sensitivity to AL amplitude remaining for the traditional NAM versus the zonal-mean method. The imposed AL–NAO correlation reduces the extent to which the correlation between the NAO and traditional NAM decreases (Figure 7c), but the AL–NAM correlation still exceeds the NAO–NAM correlation at large AL amplitudes unless the zonal-mean method is used. We note that

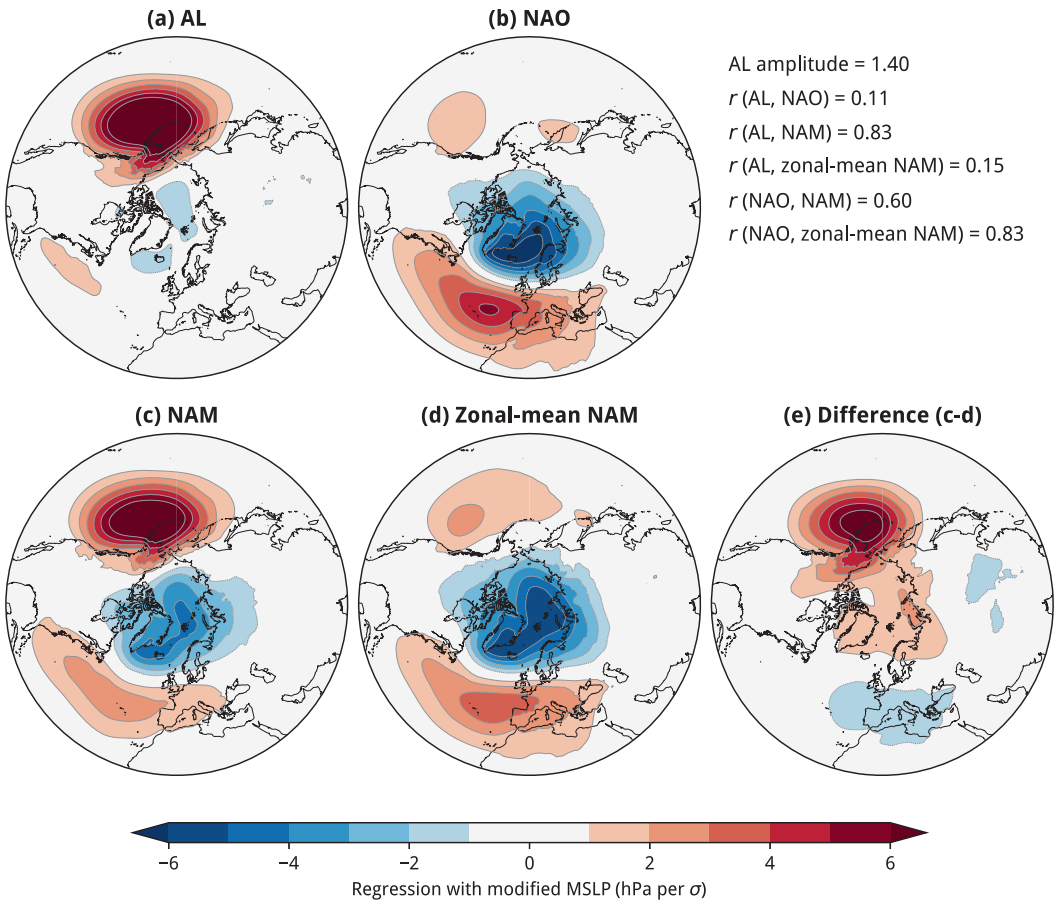

FIGURE 8 Regression of (a) AL, (b) NAO, (c) traditionally defined NAM, (d) zonal-mean NAM, and (e) the difference between the two definitions of the NAM (c minus d), computed using the ERA5 MSLP with the AL amplitude increased to 1.4 times the 1994–2017 value. Correlations between the indices are also shown at the top right. [Colour figure can be viewed at wileyonlinelibrary.com]

the models show a slightly weaker NAO–NAM correlation than expected from our statistical model, which might be due to a slight low bias in NAO amplitude (cf. Figure 3), which is not included here.

To visualise the results of our statistical model better, Figure 8 shows regression maps of the EOFs obtained from the perturbed MSLP anomalies with the AL amplitude set to 1.40 (at which point the AL–NAM correlation exceeds the NAO–NAM correlation). The correlation between the AL and NAO is not modified directly, and so neither the AL (Figure 8a) nor NAO (Figure 8b) show large signatures in the opposing basins. As expected from the correlations in Figure 7, the traditionally defined NAM (Figure 8c) shows a highly amplified Pacific centre, reminiscent of that seen in the model NAM patterns and very different from the ERA5 NAM (cf. Figure 1). The Pacific centre is also stronger than the Atlantic centre, consistent with the model biases. Furthermore, no such highly exaggerated Pacific centre exists in the the zonal-mean NAM (Figure 8d). The difference between the two (Figure 8e) shows clearly that, when AL amplitude is increased, the traditional NAM captures more AL variability at the expense of the NAO.

Finally, we use a similar approach to correct the MSLP field statistically in each model. As an approximation of the dynamical effect of the biased AL amplitude, we first subtract the additional component of correlated AL–NAO variability above that in ERA5 (following Equation 2). Similar results are obtained if we subtract only the component that is linearly explained by the AL amplitude (per the correlation in Figure 6c). We then adjust the AL amplitude itself to match that in ERA5 (following Equation 1) and recompute the traditionally defined NAM. The results are shown in Figure 9. By comparison with Figure 1, the amplitude of the Pacific centre is markedly reduced and matches much more closely both the zonal-mean NAM for each model (Figure 2) and the ERA5 NAM (all models show an increase in the pattern correlation). The largest difference is seen in the CMCC model: the pattern correlation with ERA5 increases from $r = 0.77$ to 0.92 and the twin Atlantic centres-of-action vanish. That such a large change occurs in CMCC is consistent with

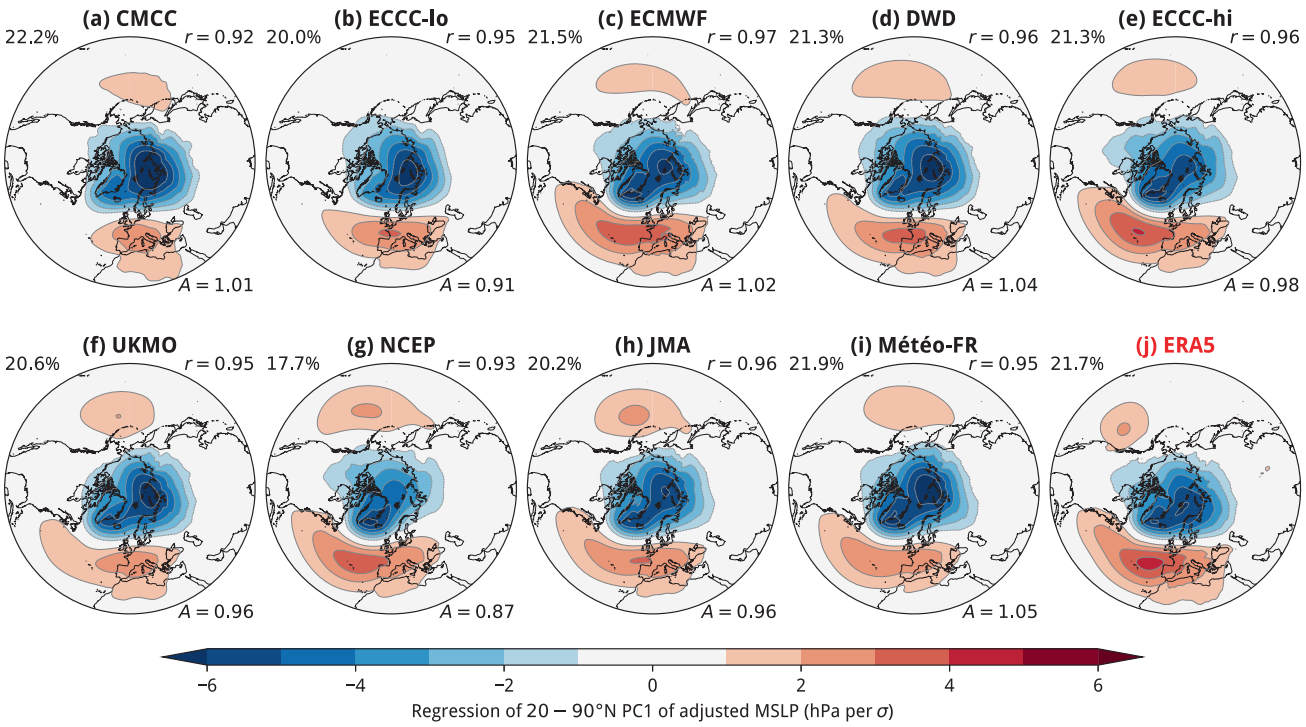

FIGURE 9 As in Figure 1, except computed using the MSLP field with the AL–NAO correlation and AL amplitude adjusted to match ERA5. [Colour figure can be viewed at wileyonlinelibrary.com]

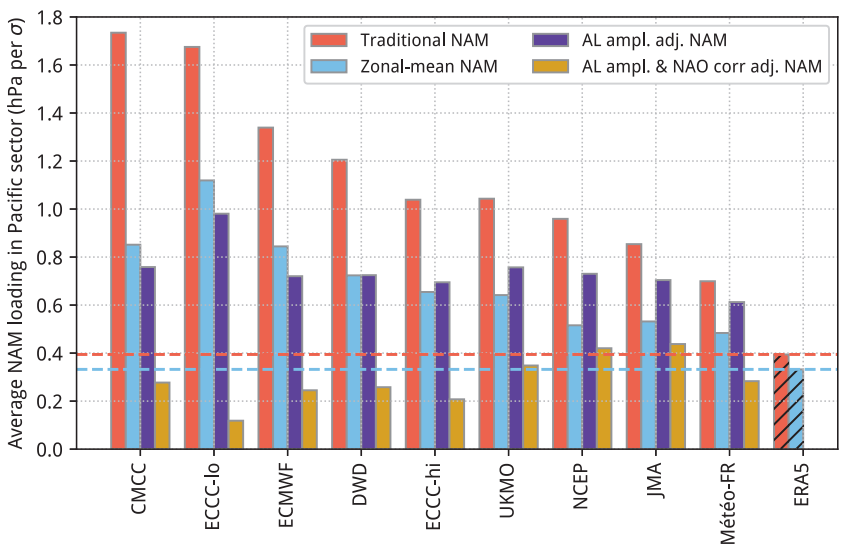

FIGURE 10 Average Pacific sector loading in the NAM regression pattern for the traditional NAM (leftmost bars; cf. Figure 1), the zonal-mean NAM (second from left; cf. Figure 2), the traditional NAM with the AL amplitude adjusted to match ERA5 (second from right), and the traditional NAM with the AL amplitude and the AL–NAO correlation adjusted to match ERA5 (rightmost bars; cf. Figure 9). [Colour figure can be viewed at wileyonlinelibrary.com]

it possessing the largest AL amplitude bias of the nine models.

We summarise our findings in Figure 10, where we show the intensity of the Pacific centre of the NAM for four variants: (1) the traditional NAM computed on the raw field (i.e., Figure 1), (2) the zonal-mean NAM (i.e., Figure 2), (3) the traditional NAM computed on the field with only the AL amplitude adjusted, and (3) the traditional NAM computed on the field with both the AL amplitude and AL–NAO correlation adjusted (i.e., Figure 9). Adjusting solely the AL amplitude yields a Pacific centre of similar magnitude to the one for the zonal-mean NAM, decreasing the magnitude by up to 56% in the case of CMCC. This confirms our hypothesis that the AL amplitude plays a direct role in the difference between the two definitions of the NAM. The adjustment of the AL–NAO correlation then brings the Pacific centre even closer to observations; some differences still remain, likely

because this only approximates the true dynamical influence. Overall, this confirms that the excess AL variance plays a key role in generating the extremely large Pacific centre of the NAM that is only seen when computing the EOF in latitude–longitude space.

The high sensitivity of the traditional NAM to AL amplitude demonstrated herein can be explained by first noting that, in ERA5, the AL is almost indistinguishable from the second EOF of Northern Hemisphere MSLP poleward of 20°N (see, e.g., Wallace & Thompson, 2002): the correlation between their respective PC time series is 0.93 ($p < 0.01$). As a result (and recalling that the AL and NAO are not significantly correlated in ERA5 over the hindcast period), the leading two hemisphere-wide EOFs (uncorrelated by construction) represent effectively the same variability as the NAO and AL. It then follows that the increasingly large component of AL variability captured by the NAM (EOF1) at higher AL amplitudes is equivalent to incorporating an increasingly large component of the original EOF2, thus indicating a rotation of EOF1 in the direction of the original EOF2. The leading zonal-mean EOF does not exhibit a similar rotation, because EOF2 of the zonal mean is not as strongly correlated with the AL ($r = 0.49$, $p < 0.01$) and the separation between the leading two zonal-mean EOFs is much greater (37 percentage points versus 10 percentage points).

4.2 | Comparison with historical variability

Finally, we compare the results obtained for the models and ERA5 over the 1993/1994–2016/2017 period with 60 equivalent 24-year rolling periods over the full ERA5 dataset (i.e., 1940/1941–1963/1964 through to 1999/2000–2022/2023). For each 24-year period, we subtract the corresponding 24-year climatology, then compute the NAM, zonal-mean NAM, NAO, and AL, and the correlations between their time series (as in Figure 5). We perform this analysis in order to assess whether the models are biased relative to only this particular 24-year period (and whether it was a particularly unusual such period) or lie outside historical variability altogether. The results are shown in Figure 11. Although the choice of 24 years is determined purely by the length of the common hindcast period in our model database, previous analyses of decadal variability in the NAM and AL have used similar-length windows (e.g., Gong *et al.*, 2019; Shi & Nakamura, 2014). Further, while we recognise that the representation of MSLP variability in the earlier part of the ERA5 record is more uncertain due to sparse observations, a thorough quantification of such uncertainty is beyond the scope of this article.

In all but three of the models (Météo-France, JMA, and NCEP), the amplitude of the AL lies above the maximum

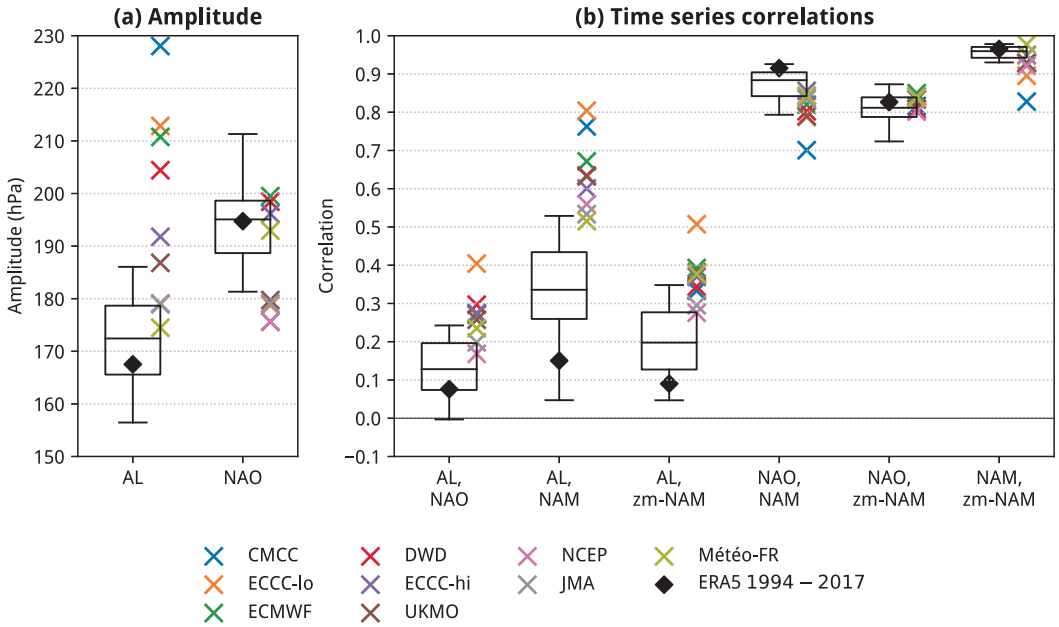

FIGURE 11 For (a) AL and NAO amplitude and (b) correlations between the time series of different indices, box plots show the distribution of statistics computed over all 60 24-year periods in ERA5 1940/1941–2022/2023, with diamonds corresponding to 1993/1994–2016/2017. Crosses show the statistics for each of the C3S models during 1993/1994–2016/2017. [Colour figure can be viewed at wileyonlinelibrary.com]

in all 24-year periods in ERA5 (Figure 11a), which is never more than 10% above the 1994–2017 value. Recall that our statistical model (Figure 7) indicated that AL amplitude perturbations within 10% of the 1994–2017 value do not yield large changes in the AL–NAM correlation, and so historical variability appears not to have spanned a sufficiently large range (as we have found in the models) to demonstrate the sensitivity of the traditional NAM definition. The AL amplitude during 1994–2017 was below the median, but not exceptionally low (31st percentile); in contrast, the 1994–2017 NAO amplitude was close to the median, while the amplitude in the models lies well within, or in the case of the ECCC-lo, CMCC, NCEP, United Kingdom Met Office (UKMO), and JMA models, slightly below the distribution. Notably, in all but three of the models (ECCC-hi, JMA, and Météo-France), the amplitude of the AL exceeds the amplitude of the NAO; in contrast, the amplitude of the NAO exceeds the amplitude of the AL in *all* 60 24-year periods in ERA5. The inter-model spread in AL amplitude also far exceeds the inter-model spread in the NAO, whereas the two have a very similar spread across the ERA5 time periods.

Consistent with the amplitude biases, the correlation between the AL and NAO (Figure 11b) exceeds the maximum of the ERA5 distribution in all but the Météo-France, JMA, and NCEP models, while the 1994–2017 correlation was relatively low (close to the 25th percentile). However, apart from ECCC-lo, the correlations lie within ~ 0.05 of the maximum ERA5 value. Most strikingly, the correlation between the AL and the traditional NAM exceeds the historical 24-year maximum in all but the Météo-France model, albeit only minimally in JMA and NCEP (while the ERA5 1994–2017 correlation was unusually low, at the 10th percentile). This explains why the exaggerated Pacific centre of the traditional NAM in the models is not reminiscent of any observed NAM pattern, because such a high correlation between the AL and NAM has no obvious historical counterpart. Furthermore, as suggested by our analyses thus far, the discrepancies in the zonal-mean NAM are not quite so extreme: although the model correlations with the AL are biased high relative to the ERA5 value, they lie either within the observed range (for CMCC, NCEP, JMA, and DWD) or marginally outside (within 0.05), and only the ECCC-lo model lies well outside the range of observed values (as for the AL–NAO correlation). A similar picture emerges for the correlations with the NAO and the correlation between the NAM and zonal-mean NAM. In the case of the latter, while it remains highly correlated in all models ($r > 0.80$), there is a weaker agreement between the two than the minimum in ERA5 (in which $r > 0.92$).

It is also clear that the NAM—regardless of the method used—has historically remained very strongly correlated with the NAO ($r > 0.70$), with less variability compared with the correlation with the AL. This is consistent with the NAO-dominant conclusions of Gong *et al.* (2018), and highlights the sensitivity of the Pacific centre of the NAM to the period under consideration—particularly when the traditional definition is used. Figure 11 also shows that, while larger than for the NAO, the historical spread in the correlation of the AL with the zonal-mean NAM is less than that with the traditional NAM ($\Delta r = 0.30$ versus 0.48). Hence, on 20–30 year time-scales, we assert that the zonal-mean NAM is a more robust method in both models *and* observations.

As for the correlation across the models shown in Figure 6, we find a positive correlation between the amplitude of the AL and the AL–NAO correlation ($r = 0.44, p = 0.13$) and the AL–NAM correlation ($r = 0.55, p = 0.06$). Similarly, the correlation between the amplitude of the AL and the intensity of the Pacific centre of the NAM is 0.60 ($p = 0.03$).[3] Although subject to other sources of variability and not covering a similarly large range to the models, this result further supports our statistical model (Figure 7) and the role of AL variance in modulating the Pacific centre of the NAM in reanalysis.

Figure 11 also suggests that the model biases cannot simply be attributed to unusually low AL variability and AL–NAO/AL–NAM correlations during the 1994–2017 hindcast period, since most of the models lie outside the historical range (i.e., such values would have been incorrect for *any* 24-year period). However, the AL amplitude and the correlations between the AL and both definitions of the NAM were all low or unusually low during 1994–2017 (although none were exceptional). Longer hindcast periods, or hindcasts initialised in different phases of multi-decadal variability, would help in determining whether the models are able to capture the observed variability (as proposed by Weisheimer *et al.* (2020) in terms of predictive skill). For example, it is unclear whether these models exhibit an offset from variability that they can capture (such that their AL amplitude is equal to the true amplitude plus a constant), or whether they cannot capture the multi-decadal variability in AL amplitude correctly at all.

Nevertheless, given that the hindcasts consist of multiple realisations of the same 24-year period, we also tested whether the ERA5 AL amplitude could have occurred as a single realisation within each ensemble. To do so, we randomly picked a single ensemble member from each year to create a synthetic 24-year time series and computed its standard deviation. We repeated this process 50,000 times for each model. In four models, we find a modest to low probability of obtaining an AL time series with equal to or less than the ERA5 variability (29% for Météo-France, 14% for JMA and NCEP, 7% for UKMO, <2% for the rest). These

are also the models for which the overall statistics lie closest to the 1994–2017 ERA5 value and within the range of historical variability. Noting that the probability obtained for Météo-France is similar to the rank of the 1994–2017 amplitude within the 60 24-year samples from ERA5, it is possible that the observed amplitude was simply unlikely to occur. Although beyond the scope of this study, similar large-ensemble subsampling techniques could shed light on factors that modulate the amplitude of AL variability.

5 | SUMMARY AND CONCLUSIONS

In this study, we have assessed the representation of the surface NAM in hindcasts for 24 extended winters from nine operational seasonal prediction models. Confirming previous studies with uninitialised and initialised climate models, we found the Pacific centre of the NAM to be systematically larger than that in ERA5 over the same time period (Figure 1). We also found that this bias is larger when the NAM is computed following the traditional definition of EOF1 of Northern Hemisphere MSLP anomalies, and that it is smaller when the NAM is computed on the zonally averaged field (Figure 2), despite nearly-identical results when both methods are applied to ERA5.

By considering the leading EOFs in the Atlantic and Pacific separately, we identified a systematic exaggeration of the amplitude of EOF1 in the Pacific by 4%–36% (Figure 4), representing excessive AL/PNA-like variability. We further found that the inter-model spread in the amplitude of the AL explained the majority of the inter-model spread in the correlation of the traditionally defined NAM with the AL (81%) and NAO (61%), while no strong relationship was found for the zonal-mean NAM (Figure 6). This suggested a key role of AL amplitude in the structure of the traditionally defined NAM, which we tested with a simple statistical model in which we specified the amplitude of the AL artificially. Our statistical model showed a large direct dependence of the traditionally defined NAM on the AL amplitude, but no large dependence of the zonal-mean NAM (Figure 7). Importantly, we found that this effect can occur without altering the AL–NAO correlation.

However, we also identified a modest relationship between the AL–NAO correlation and AL amplitude (explaining 44% of the inter-model spread). Unlike the purely statistical effect of increased AL variance, this points to a dynamical pathway which physically links some variability between the two basins, consistent with previous studies (e.g., Shi & Nakamura, 2014; Sun & Tan, 2013). Correcting for the biased AL amplitude alone results in the Pacific centre of the NAM in the models matching the (unadjusted) equivalent for the zonal-mean NAM (Figure 10). Hence, the large differences between the traditional and zonal-mean NAM seen only in the models can be attributed to the statistical effect of the exaggerated AL amplitude on the EOF structure. Additionally correcting for the higher correlation between the AL and NAO—which we attribute, at least in part, to the dynamical effect of enhanced AL variability—results in the Pacific centre of the NAM becoming close to, or smaller than, that in ERA5, and a marked improvement in pattern correlations (Figure 9). Thus, we conclude that the traditional NAM definition effectively penalises the AL amplitude bias twice.

Such a large sensitivity of the traditional NAM definition to AL amplitude has not been previously reported, likely because the AL amplitude in most of the models lies well outside the range of observed variability. As a result, the model biases actually serve to further our understanding of the statistical techniques used to identify annular variability. Previous reanalysis-based studies (e.g., Baldwin & Thompson, 2009) have suggested that the two methods are interchangeable for the surface NAM, but we have shown that this conclusion cannot be extended to models in which AL variability is highly amplified. In such circumstances, the traditionally defined NAM cannot be interpreted as necessarily representing the same mode of climate variability as in reanalysis to date, and we therefore strongly encourage the use of the more robust zonal-mean method. The zonal-mean method is also more well suited to identifying the free-tropospheric NAM and diagnosing stratosphere–troposphere coupling (Baldwin & Thompson, 2009), and thus has additional potential benefits beyond those described here. One could even avoid EOF analysis altogether, by defining the NAM as the standardised anomalies of the 40–60°N minus 60–90°N average MSLP (e.g. Bett *et al.*, 2023). Because this non-EOF method is based around the two nodes of the zonal-mean NAM, it behaves almost identically.

AL variability of the magnitude generated by some of the models has not been seen thus far in reanalysis. However, some future climate projections show an increase in the variability and/or intensity of the AL (Gan *et al.*, 2017; Giamalaki *et al.*, 2021; Hamouda *et al.*, 2021), with important consequences for how the NAM is computed and interpreted in projections of future climate. For example, in a study of CMIP5 models under the Representative Concentration Pathway 8.5 scenario in the 23rd century, Hamouda *et al.* (2021) concluded that the NAM and NAO "decouple" (i.e., become less strongly correlated) due to faster warming in the North Pacific versus the North Atlantic. Our results, however, suggest that such decoupling is likely to result from the spatial EOF method chosen to compute the NAM under the influence of increased

AL variability, as AL variability clearly increases in their model analyses (their fig. 1c). The "decoupled" NAM identified by Hamouda *et al.* (2021), and its reduced correlation with the NAO, is in fact somewhat similar to the one found herein for some of the seasonal models (particularly ECCC-lo and CMCC). Thus, to some extent a "decoupled" NAM can be seen in model simulations of *historical* variability, but only when a specific method is chosen (i.e., spatial EOF analysis). We also note that CMIP5 models exaggerate the Pacific centre of the NAM in historical simulations, presenting a significant challenge in drawing conclusions about future changes. Together, this underscores the challenge of interpreting EOFs as representing modes of climate variability directly, and of diagnosing the physical variability of the climate system.

While our results indicate a dominant role for AL variability biases, this may not be the only factor governing biases in the correlation between the basins. Gong *et al.* (2019) concluded that the strength of the stratospheric polar vortex plays a key role in modulating Pacific–Atlantic coupling, and thus the Pacific centre of the NAM, in their analysis of 27 CMIP5 models. However, we find no such large dependence in these seasonal models (although we are limited to eight models by the absence of stratospheric-level data from NCEP in the C3S database). Considering the diagnostic of 50-hPa 65°N zonal-mean zonal winds used in Gong *et al.* (2019), we find only a minimal and non-significant correlation between the vortex strength and the AL–NAO correlation ($r = 0.28$, $p = 0.55$). Nevertheless, we do not rule out the physical mechanism proposed by Gong *et al.* (2019)—noting also the much smaller number of models considered in our study—but instead suggest that stratospheric vortex biases do not play a dominant role in AL–NAO coupling in these seasonal models during the 24-year hindcast period.

In conclusion, while large model biases in AL variability exist, we assert that the NAM in current-generation seasonal prediction models is in fact not as biased as may first appear. Under the presence of large AL variability, the zonal-mean method is more well suited for defining the NAM. Addressing the tendency for models to exaggerate AL variability will likely solve, or at least markedly reduce, the persistent NAM biases seen across models. However, the ubiquity of this bias suggests that solving it may be challenging. Model resolution (either vertical or horizontal) may play a role (as suggested by Gong *et al.* (2019)); we find the largest biases in the models with the lowest vertical resolution (ECCC-lo and CMCC) and the smallest in those with the highest vertical resolution (JMA and Météo-France), but it is unclear whether this difference occurs directly due to the resolution or other model improvements. In this article, we have focused on assessing bulk statistics across NDJFM, but analysis of the intraseasonal evolution of the AL–NAO and AL–NAM coupling (following, e.g., Honda *et al.*, 2001; Ayarzagüena *et al.*, 2018) may also yield further insight into the source of the model biases. Further, the North Pacific atmospheric response to the El Niño–Southern Oscillation projects strongly onto the AL/PNA patterns, and has recently been shown to be too weak in both subseasonal (Garfinkel *et al.*, 2022) and seasonal (Williams *et al.*, 2023) models. Assessing whether this is related to the biased AL variability is a potential avenue for future work.

ACKNOWLEDGEMENTS

The authors acknowledge funding from National Science Foundation grant AGS-1914569 to Columbia University. All computations were performed on the Ginsburg HPC research cluster at Columbia University. We acknowledge the Copernicus Climate Change Service for the provision of the ERA5 reanalysis (https://doi.org/10.24381/cds.f17050d7) and seasonal model data (https://doi.org/10.24381/cds.68dd14c3), which are freely available from the Copernicus Climate Data Store. The model data are available with the following system identifiers: ECCC-lo, 2; CMCC, 35; NCEP, 2; ECCC-hi, 3; UKMO, 601; ECMWF, 5; DWD, 21; Météo-France, 8. Further specific information is available online at https://confluence.ecmwf.int/display/CKB/C3S+Seasonal+Forecasts. We thank three reviewers and the Associate Editor for their constructive comments on our article.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on the Copernicus Climate Data Store at https://doi.org/10.24381/cds.f17050d7 and https://doi.org/10.24381/cds.68dd14c3.

ENDNOTES

[1]Here, we use the term "NAM" exclusively, including in discussions of previous studies that used the AO terminology.

[2]Equation (2) assumes the correlation and regression coefficients are approximately equivalent (as is the case for time series with close to unit variance). Large values of Δr increase the variance substantially, at which point this assumption is no longer valid.

[3]To account for the substantial autocorrelation in these time series, the three p-values reported in this paragraph are obtained using an adjusted bootstrap test in which input data points are separated by 5 years (at which point the autocorrelation is not significantly different from zero).

ORCID

Simon H. Lee https://orcid.org/0000-0003-0986-0093

Lorenzo M. Polvani https://orcid.org/0000-0003-4775-8110

REFERENCES

Ambaum, M.H.P., Hoskins, B.J. & Stephenson, D.B. (2001) Arctic oscillation or North Atlantic oscillation? *Journal of Climate*, 14(16), 3495–3507. Available from: https://doi.org/10.1175/1520-0442(2001)014<3495:AOONAO>2.0.CO;2

Ayarzagüena, B., Ineson, S., Dunstone, N.J., Baldwin, M.P. & Scaife, A.A. (2018) Intraseasonal effects of El niño–Southern Oscillation on North Atlantic climate. *Journal of Climate*, 31(21), 8861–8873. Available from: https://doi.org/10.1175/JCLI-D-18-0097.1

Baldwin, M.P. & Dunkerton, T.J. (1999) Propagation of the arctic oscillation from the stratosphere to the troposphere. *Journal of Geophysical Research: Atmospheres*, 104(D24), 30937–30946. Available from: https://doi.org/10.1029/1999JD900445

Baldwin, M.P. & Dunkerton, T.J. (2001) Stratospheric harbingers of anomalous weather regimes. *Science*, 294(5542), 581–584. Available from: https://doi.org/10.1126/science.1063315

Baldwin, M.P., Stephenson, D.B., Thompson, D.W.J., Dunkerton, T.J., Charlton, A.J. & O'Neill, A. (2003) Stratospheric memory and skill of extended-range weather forecasts. *Science*, 301(5633), 636–640. Available from: https://doi.org/10.1126/science.1087143

Baldwin, M.P. & Thompson, D.W.J. (2009) A critical comparison of stratosphere–troposphere coupling indices. *Quarterly Journal of the Royal Meteorological Society*, 135(644), 1661–1672. Available from: https://doi.org/10.1002/qj.479

Bett, P.E., Scaife, A.A., Hardiman, S.C., Thornton, H.E., Shen, X., Wang, L. et al. (2023) Using large ensembles to quantify the impact of sudden stratospheric warmings and their precursors on the North Atlantic Oscillation. *Weather and Climate Dynamics (Pembroke, Ont.)*, 4(1), 213–228. Available from: https://doi.org/10.5194/wcd-4-213-2023

Butler, A.H., Karpechko, A.Y. & Garfinkel, C.I. (2023) Amplified decadal variability of extratropical surface temperatures by stratosphere-troposphere coupling. *Geophysical Research Letters*, 50(16), e2023GL104607. Available from: https://doi.org/10.1029/2023GL104607

Cai, Q., Chen, W., Chen, S., Ma, T. & Garfinkel, C.I. (2022) Influence of the quasi-biennial oscillation on the spatial structure of the wintertime arctic oscillation. *Journal of Geophysical Research: Atmospheres*, 127(8), e2021JD035564. Available from: https://doi.org/10.1029/2021JD035564

Coburn, J. & Pryor, S.C. (2021) Differential credibility of climate modes in CMIP6. *Journal of Climate*, 34(20), 8145–8164. Available from: https://doi.org/10.1175/JCLI-D-21-0359.1

Dawson, A. (2016) eofs: A library for EOF analysis of meteorological, oceanographic, and climate data. *Journal of Open Research Software*, 4(1), e14. Available from: https://doi.org/10.5334/jors.122

Deser, C. (2000) On the teleconnectivity of the "Arctic Oscillation". *Geophysical Research Letters*, 27(6), 779–782. Available from: https://doi.org/10.1029/1999GL010945

Dommenget, D. & Latif, M. (2002) A cautionary note on the interpretation of EOFs. *Journal of Climate*, 15(2), 216–225. Available from: https://doi.org/10.1175/1520-0442(2002)015<0216:ACNOTI>2.0.CO;2

Fasullo, J.T., Phillips, A.S. & Deser, C. (2020) Evaluation of leading modes of climate variability in the CMIP archives. *Journal of Climate*, 33(13), 5527–5545. Available from: https://doi.org/10.1175/JCLI-D-19-1024.1

Feldstein, S.B. & Franzke, C. (2006) Are the North Atlantic oscillation and the Northern Annular Mode distinguishable? *Journal of the Atmospheric Sciences*, 63(11), 2915–2930. Available from: https://doi.org/10.1175/JAS3798.1

Furtado, J.C., Cohen, J., Becker, E.J. & Collins, D.C. (2021) Evaluating the relationship between sudden stratospheric warmings and tropospheric weather regimes in the NMME phase-2 models. *Climate Dynamics*, 56, 2321–2338. Available from: https://doi.org/10.1007/s00382-020-05591-x

Gan, B., Lixin, W., Jia, F., Li, S., Cai, W., Nakamura, H. et al. (2017) On the response of the Aleutian low to greenhouse warming. *Journal of Climate*, 30(10), 3907–3925. Available from: https://doi.org/10.1175/JCLI-D-15-0789.1

Garfinkel, C.I., Chen, W., Li, Y., Schwartz, C., Yadav, P. & Domeisen, D. (2022) The winter North Pacific teleconnection in response to ENSO and the MJO in operational subseasonal forecasting models is too weak. *Journal of Climate*, 35(24), 8013–8030. Available from: https://doi.org/10.1175/JCLI-D-22-0179.1

Gerber, E.P. & Martineau, P. (2018) Quantifying the variability of the annular modes: reanalysis uncertainty vs. sampling uncertainty. *Atmospheric Chemistry and Physics*, 18(23), 17099–17117. Available from: https://doi.org/10.5194/acp-18-17099-2018

Giamalaki, K., Claudie Beaulieu, S.A., Henson, A.P.M., Kassem, H. & Faranda, D. (2021) Future intensification of extreme Aleutian low events and their climate impacts. *Scientific Reports*, 11(1), 18395. Available from: https://doi.org/10.1038/s41598-021-97615-7

Gong, H., Wang, L., Chen, W., Chen, X. & Nath, D. (2017) Biases of the wintertime Arctic Oscillation in CMIP5 models. *Environmental Research Letters*, 12(1), 014001. Available from: https://doi.org/10.1088/1748-9326/12/1/014001

Gong, H., Wang, L., Chen, W. & Nath, D. (2018) Multidecadal fluctuation of the wintertime Arctic Oscillation pattern and its implication. *Journal of Climate*, 31(14), 5595–5608. Available from: https://doi.org/10.1175/JCLI-D-17-0530.1

Gong, H., Wang, L., Chen, W., Renguang, W., Zhou, W., Liu, L. et al. (2019) Diversity of the wintertime Arctic Oscillation pattern among CMIP5 models: Role of the stratospheric polar vortex. *Journal of Climate*, 32(16), 5235–5250. Available from: https://doi.org/10.1175/JCLI-D-18-0603.1

Hamouda, M.E., Pasquero, C. & Tziperman, E. (2021) Decoupling of the Arctic Oscillation and North Atlantic Oscillation in a warmer climate. *Nature Climate Change*, 11(2), 137–142. Available from: https://doi.org/10.1038/s41558-020-00966-8

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Holton, J.R. & Tan, H.-C. (1980) The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *Journal of the Atmospheric Sciences*, 37(10), 2200–2208. Available from: https://doi.org/10.1175/1520-0469(1980)037<2200:TIOTEQ>2.0.CO;2

Honda, M., Kushnir, Y., Nakamura, H., Yamane, S. & Zebiak, S.E. (2005) Formation, mechanisms, and predictability of the Aleutian–Icelandic low seesaw in ensemble AGCM simulations.

Journal of Climate, 18(9), 1423–1434. Available from: https://doi.org/10.1175/JCLI3353.1

Honda, M., Nakamura, H., Ukita, J., Kousaka, I. & Takeuchi, K. (2001) Interannual seesaw between the Aleutian and Icelandic lows. Part I: Seasonal dependence and life cycle. *Journal of Climate*, 14(6), 1029–1042. Available from: https://doi.org/10.1175/1520-0442(2001)014<1029:ISBTAA>2.0.CO;2

Hurrell, J.W. (1995) Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 269(5224), 676–679. Available from: https://doi.org/10.1126/science.269.5224.676

Hurrell, J.W. & Deser, C. (2010) North Atlantic climate variability: the role of the North Atlantic Oscillation. *Journal of Marine Systems*, 79(3-4), 231–244. Available from: https://doi.org/10.1016/j.jmarsys.2008.11.026

Huth, R. & Beranová, R. (2021) How to recognize a true mode of atmospheric circulation variability. *Earth and Space Science*, 8(3), e2020EA001275. Available from: https://doi.org/10.1029/2020EA001275

Kang, D., Lee, M.-I., Im, J., Kim, D., Kim, H.-M., Kang, H.-S. et al. (2014) Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting systems. *Geophysical Research Letters*, 41(10), 3577–3585. Available from: https://doi.org/10.1002/2014GL060011

L'Heureux, M.L., Tippett, M.K., Kumar, A., Butler, A.H., Ciasto, L.M., Ding, Q. et al. (2017) Strong relations between ENSO and the Arctic Oscillation in the North American multimodel ensemble. *Geophysical Research Letters*, 44(22), 11–654. Available from: https://doi.org/10.1002/2017GL074854

Lee, S.H., Lawrence, Z.D., Butler, A.H. & Karpechko, A.Y. (2020) Seasonal forecasts of the exceptional Northern Hemisphere winter of 2020. *Geophysical Research Letters*, 47(21), e2020GL090328. Available from: https://doi.org/10.1029/2020GL090328

Li, J. & Wang, J.X.L. (2003) A modified zonal index and its physical sense. *Geophysical Research Letters*, 30(12), 1632. Available from: https://doi.org/10.1029/2003GL017441

Limpasuvan, V. & Hartmann, D.L. (1999) Eddies and the annular modes of climate variability. *Geophysical Research Letters*, 26(20), 3133–3136. Available from: https://doi.org/10.1029/1999GL010478

Miller, R.L., Schmidt, G.A. & Shindell, D.T. (2006) Forced annular variations in the 20th century intergovernmental panel on climate change fourth assessment report models. *Journal of Geophysical Research: Atmospheres*, 111(D18), D18101. Available from: https://doi.org/10.1029/2005JD006323

Minobe, S. & Mantua, N. (1999) Interdecadal modulation of interannual atmospheric and oceanic variability over the North Pacific. *Progress in Oceanography*, 43(2-4), 163–192. Available from: https://doi.org/10.1016/S0079-6611(99)00008-7

Pinto, J.G., Reyers, M. & Ulbrich, U. (2011) The variable link between PNA and NAO in observations and in multi-century CGCM simulations. *Climate Dynamics*, 36, 337–354. Available from: https://doi.org/10.1007/s00382-010-0770-x

Quadrelli, R. & Wallace, J.M. (2004) A simplified linear framework for interpreting patterns of Northern Hemisphere wintertime climate variability. *Journal of Climate*, 17(19), 3728–3744.

Riddle, E.E., Butler, A.H., Furtado, J.C., Cohen, J.L. & Kumar, A. (2013) CFSv2 ensemble prediction of the wintertime Arctic Oscillation. *Climate Dynamics*, 41, 1099–1116. Available from: https://doi.org/10.1007/s00382-013-1850-5

Shi, N. & Nakamura, H. (2014) Multi-decadal modulations in the Aleutian-Icelandic Low seesaw and the axial symmetry of the Arctic Oscillation signature, as revealed in the 20th century reanalysis. *Tellus A: Dynamic Meteorology and Oceanography*, 66(1), 22660. Available from: https://doi.org/10.3402/tellusa.v66.22660

Stockdale, T.N., Molteni, F. & Ferranti, L. (2015) Atmospheric initial conditions and the predictability of the Arctic Oscillation. *Geophysical Research Letters*, 42(4), 1173–1179. Available from: https://doi.org/10.1002/2014GL062681

Sun, J. & Tan, B. (2013) Mechanism of the wintertime Aleutian low–Icelandic low seesaw. *Geophysical Research Letters*, 40(15), 4103–4108. Available from: https://doi.org/10.1002/grl.50770

Thompson, D.W.J. & Wallace, J.M. (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, 25(9), 1297–1300. Available from: https://doi.org/10.1029/98GL00950

Thompson, D.W.J. & Wallace, J.M. (2000) Annular modes in the extratropical circulation. Part I: Month-to-month variability. *Journal of Climate*, 13(5), 1000–1016. Available from: https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2

Thompson, D.W.J. & Wallace, J.M. (2001) Regional climate impacts of the Northern Hemisphere annular mode. *Science*, 293(5527), 85–89. Available from: https://doi.org/10.1126/science.1058958

Wallace, J.M. & Gutzler, D.S. (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review*, 109(4), 784–812. Available from: https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

Wallace, J.M. & Thompson, D.W.J. (2002) The Pacific center of action of the Northern Hemisphere annular mode: Real or artifact? *Journal of Climate*, 15(14), 1987–1991. Available from: https://doi.org/10.1175/1520-0442(2002)015<1987:TPCOAO>2.0.CO;2

Weisheimer, A., Befort, D.J., MacLeod, D., Palmer, T., O'Reilly, C. & Strømmen, K. (2020) Seasonal forecasts of the twentieth century. *Bulletin of the American Meteorological Society*, 101(8), E1413–E1426. Available from: https://doi.org/10.1175/BAMS-D-19-0019.1

Williams, N.C., Scaife, A.A. & Screen, J.A. (2023) Underpredicted ENSO teleconnections in seasonal forecasts. *Geophysical Research Letters*, 50(5), e2022GL101689. Available from: https://doi.org/10.1029/2022GL101689

Zhao, H. & Moore, G.W.K. (2009) Temporal variability in the expression of the Arctic Oscillation in the North Pacific. *Journal of Climate*, 22(11), 3110–3126. Available from: https://doi.org/10.1175/2008JCLI2611.1

How to cite this article: Lee, S.H. & Polvani, L.M. (2024) Large model biases in the Pacific centre of the Northern Annular Mode due to exaggerated variability of the Aleutian Low. *Quarterly Journal of the Royal Meteorological Society*, 1–20. Available from: https://doi.org/10.1002/qj.4825