Large sampling uncertainty when diagnosing the ‘eddy feedback parameter’ and its role in the signal-to-noise paradox

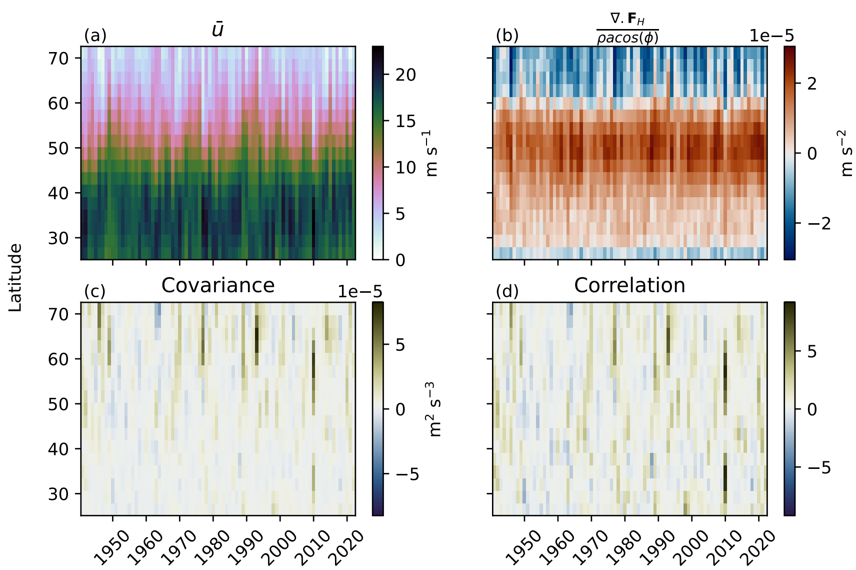
Leo Saffin[1], Amanda Maycock[2], Christine M. McKenna[2], and Rémy Bonnet[3]

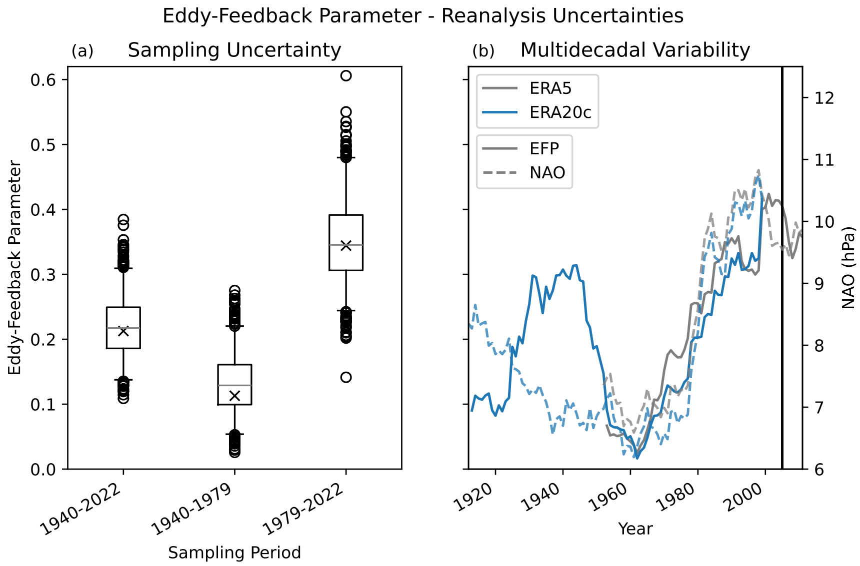
[1]University of Reading
[2]University of Leeds
[3]IPSL/CNRS

February 28, 2024

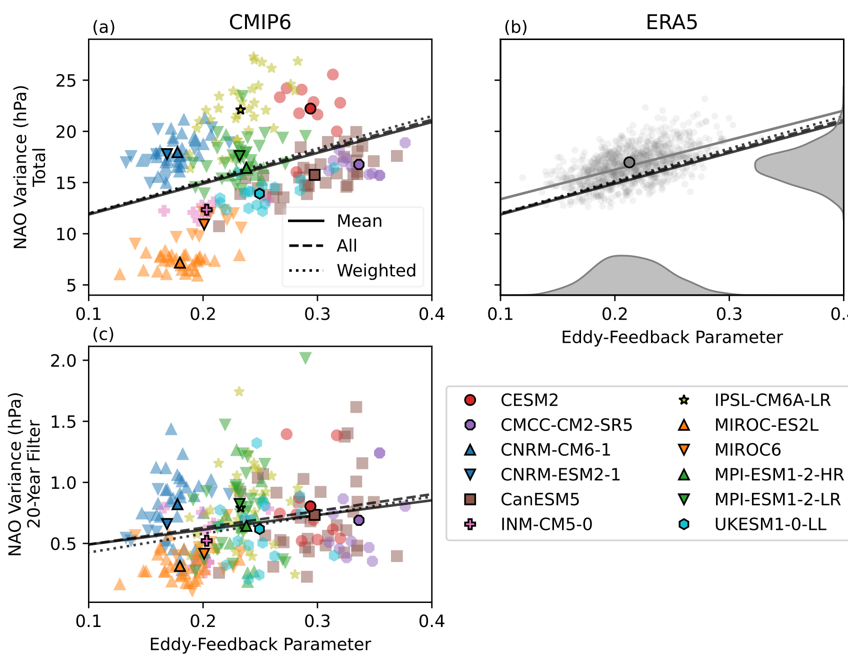
Abstract

A too-weak eddy feedback in models has been proposed to explain the signal-to-noise paradox in seasonal-to-decadal forecasts of the winter Northern Hemisphere. We show that the “eddy feedback parameter’ (EFP) used in previous studies is sensitive to sampling and multidecadal variability. When these uncertainties are accounted for, the EFP diagnosed from CMIP6 historical simulations generally falls within the reanalysis uncertainty. We find the EFP is not independent of the sampled North Atlantic Oscillation (NAO). Within the same dataset, a sample containing larger NAO variability will show a larger EFP, suggesting that the link between eddy feedbacks and the signal-to-noise paradox could be due to sampling effects with the EFP. An alternative measure of eddy feedback, the barotropic energy generation rate, is less sensitive to sampling errors and delineates CMIP6 models that have weak, strong, or unbiased eddy feedbacks, but shows little relation to NAO variability.





Eddy-Feedback Parameter - Reanalysis Uncertainties
(a) Sampling Uncertainty
(b) Multidecadal Variability
Eddy-Feedback Parameter
Sampling Period
1940-2022
1940-1979
1979-2022
ERA5
ERA20c
EFP
NAO
Year
NAO (hPa)



(a) CMIP6
(b) ERA5
(c)
NAO Variance (hPa)
Total
NAO Variance (hPa)
20-Year Filter
Eddy-Feedback Parameter
Mean
All
Weighted
CESM2
CMCC-CM2-SR5
CNRM-CM6-1
CNRM-ESM2-1
CanESM5
INM-CM5-0
IPSL-CM6A-LR
MIROC-ES2L
MIROC6
MPI-ESM1-2-HR
MPI-ESM1-2-LR
UKESM1-0-LL

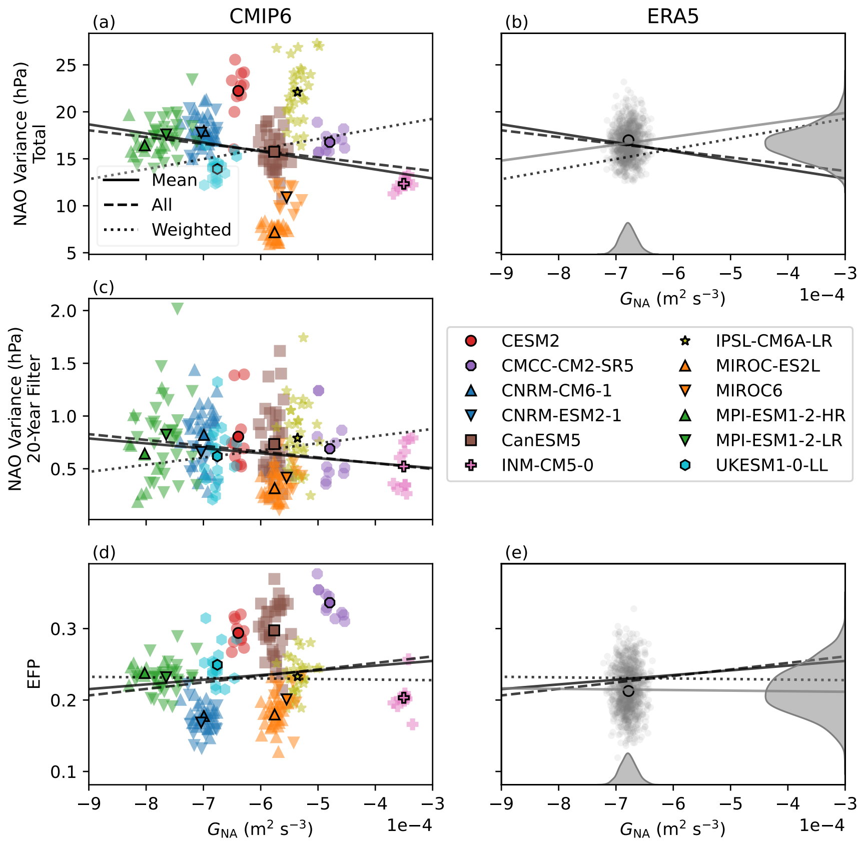

(a)
CMIP6
(b)
ERA5
NAO Variance (hPa)
Total
Mean
All
Weighted
(c)
NAO Variance (hPa)
20-Year Filter
(d)
EFP
(e)
G_{NA} (m^2 s^{-3})
1e−4
CESM2
CMCC-CM2-SR5
CNRM-CM6-1
CNRM-ESM2-1
CanESM5
INM-CM5-0
IPSL-CM6A-LR
MIROC-ES2L
MIROC6
MPI-ESM1-2-HR
MPI-ESM1-2-LR
UKESM1-0-LL

fig1_eddy-feedback-parameter_calculation_steps.png.

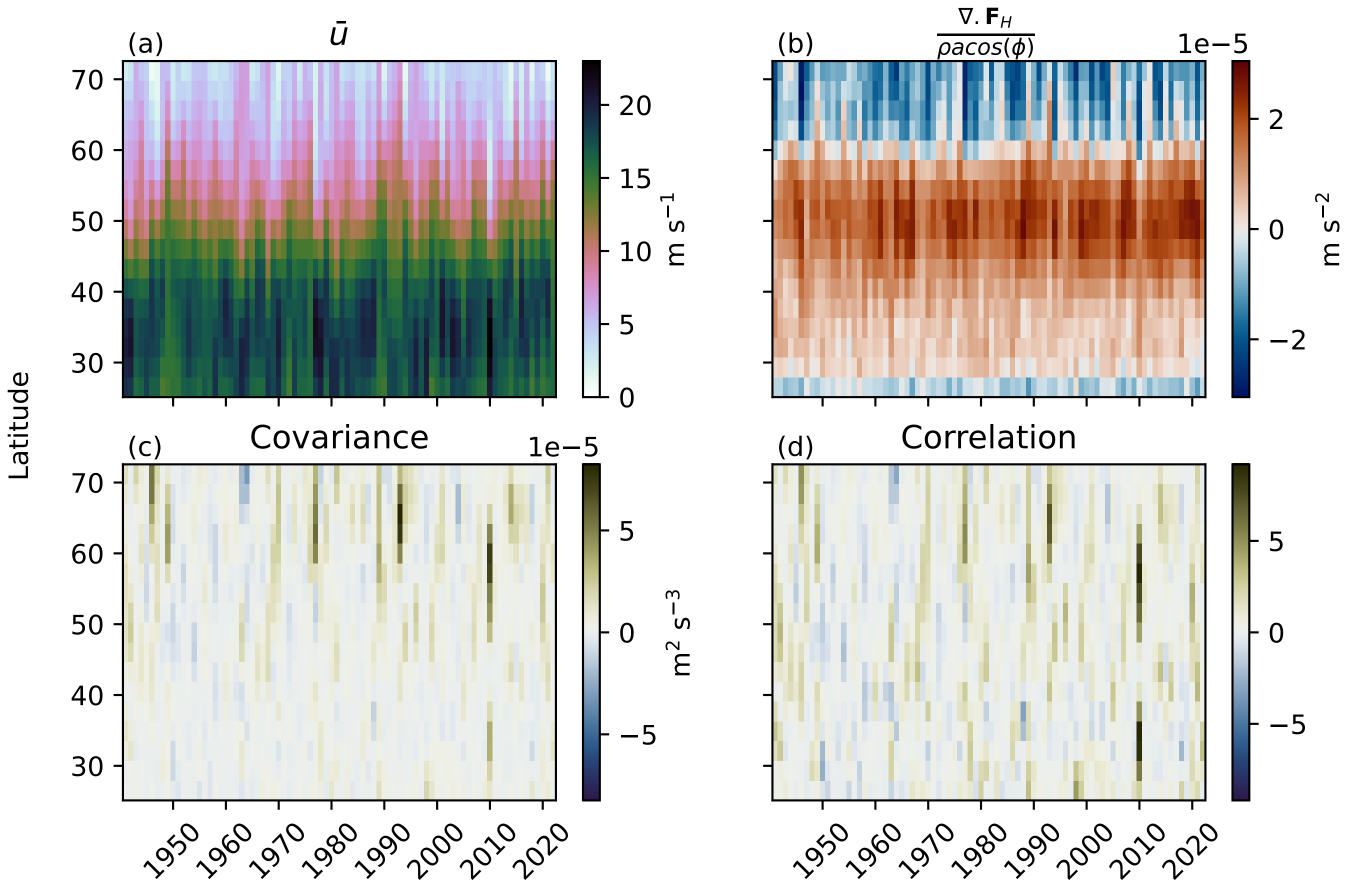
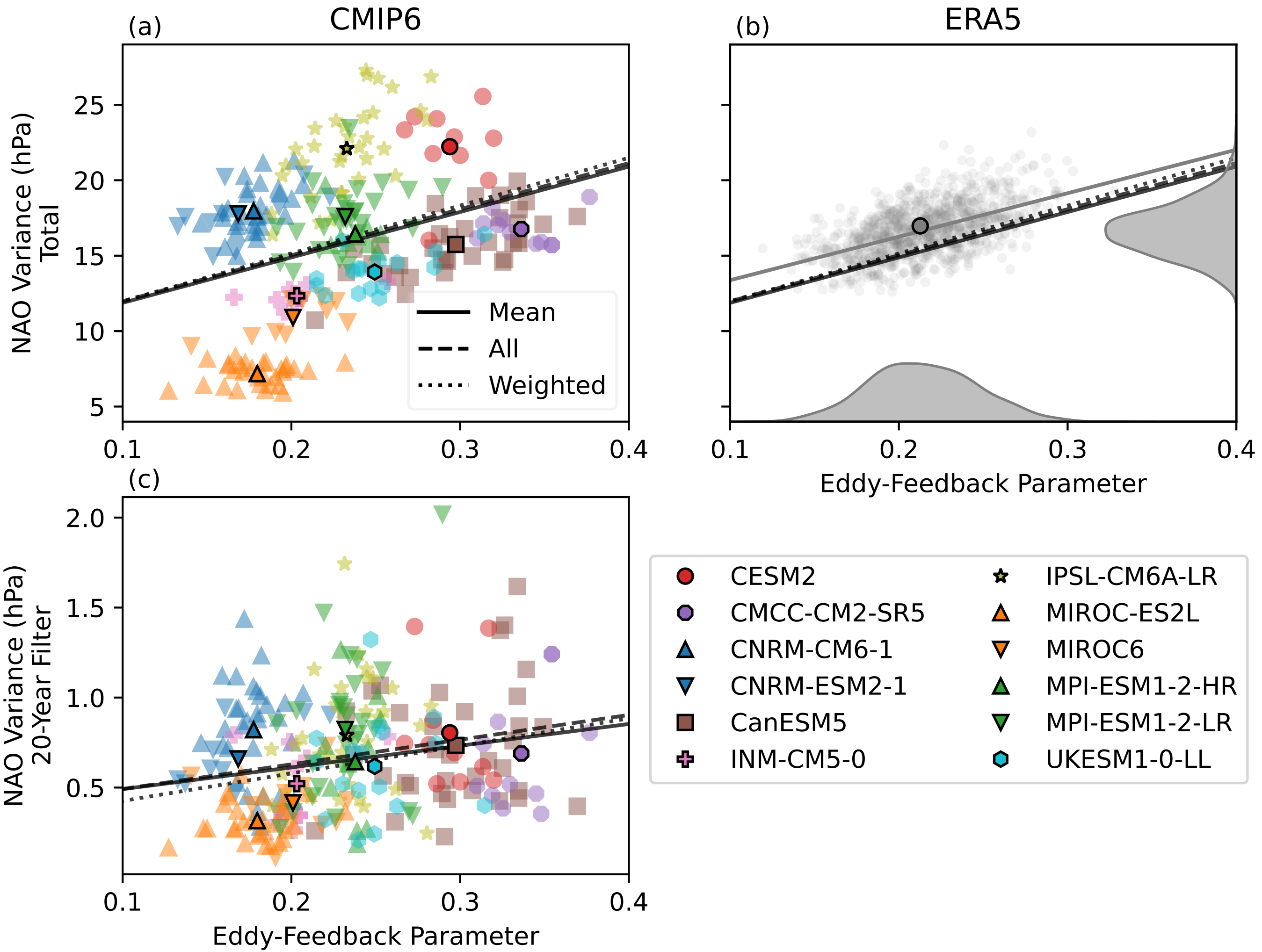
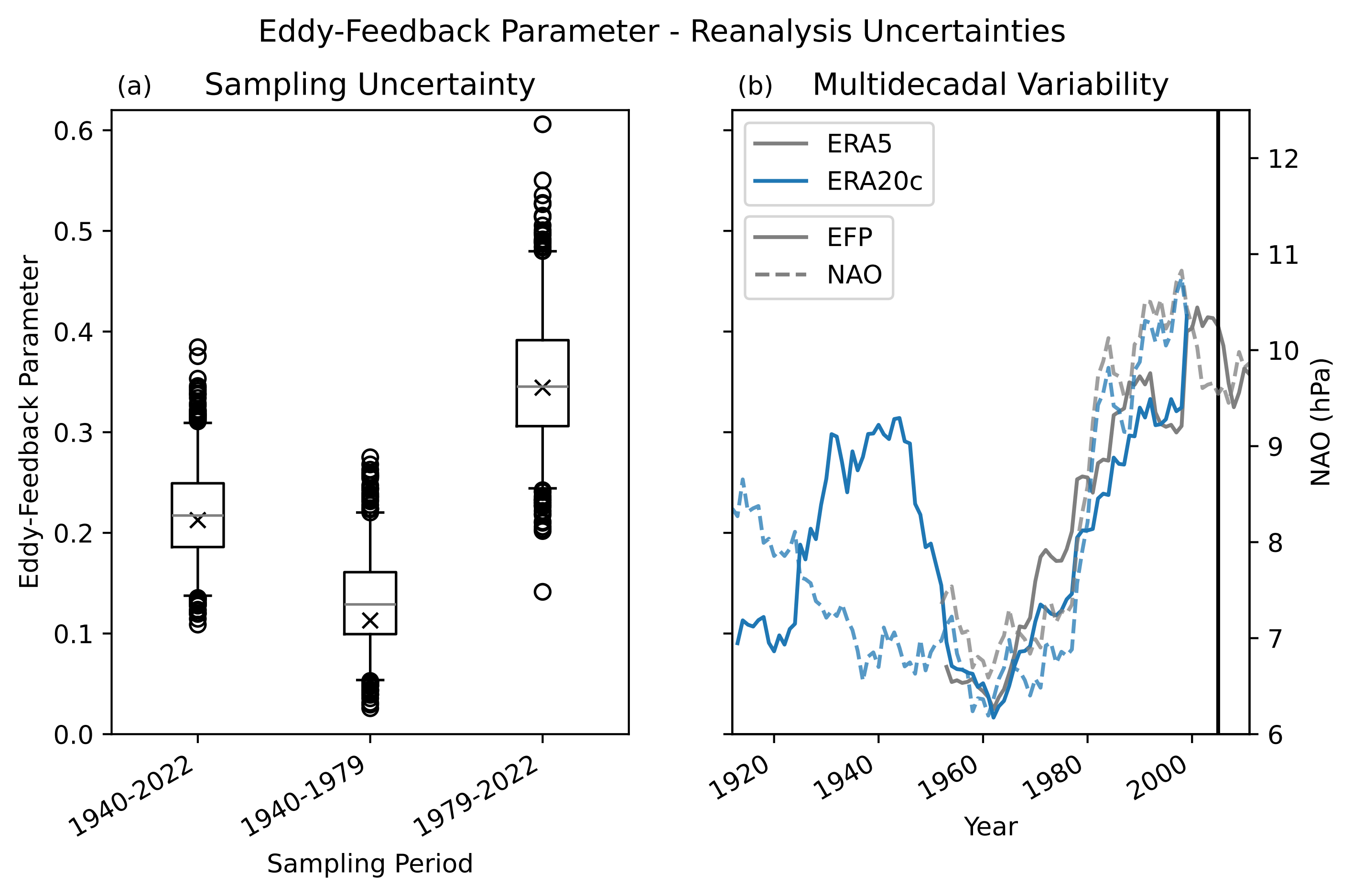

(a)
$\bar{u}$
m s^{-1}
(b)
$\frac{\nabla.\mathbf{F}_H}{\rho a cos(\phi)}$
1e−5
m s^{-2}
(c)
Covariance
1e−5
m^2 s^{-3}
(d)
Correlation
Latitude
1950
1960
1970
1980
1990
2000
2010
2020

fig3_efp_nao_correlation_cmip6_DJF_1850-2014.png.



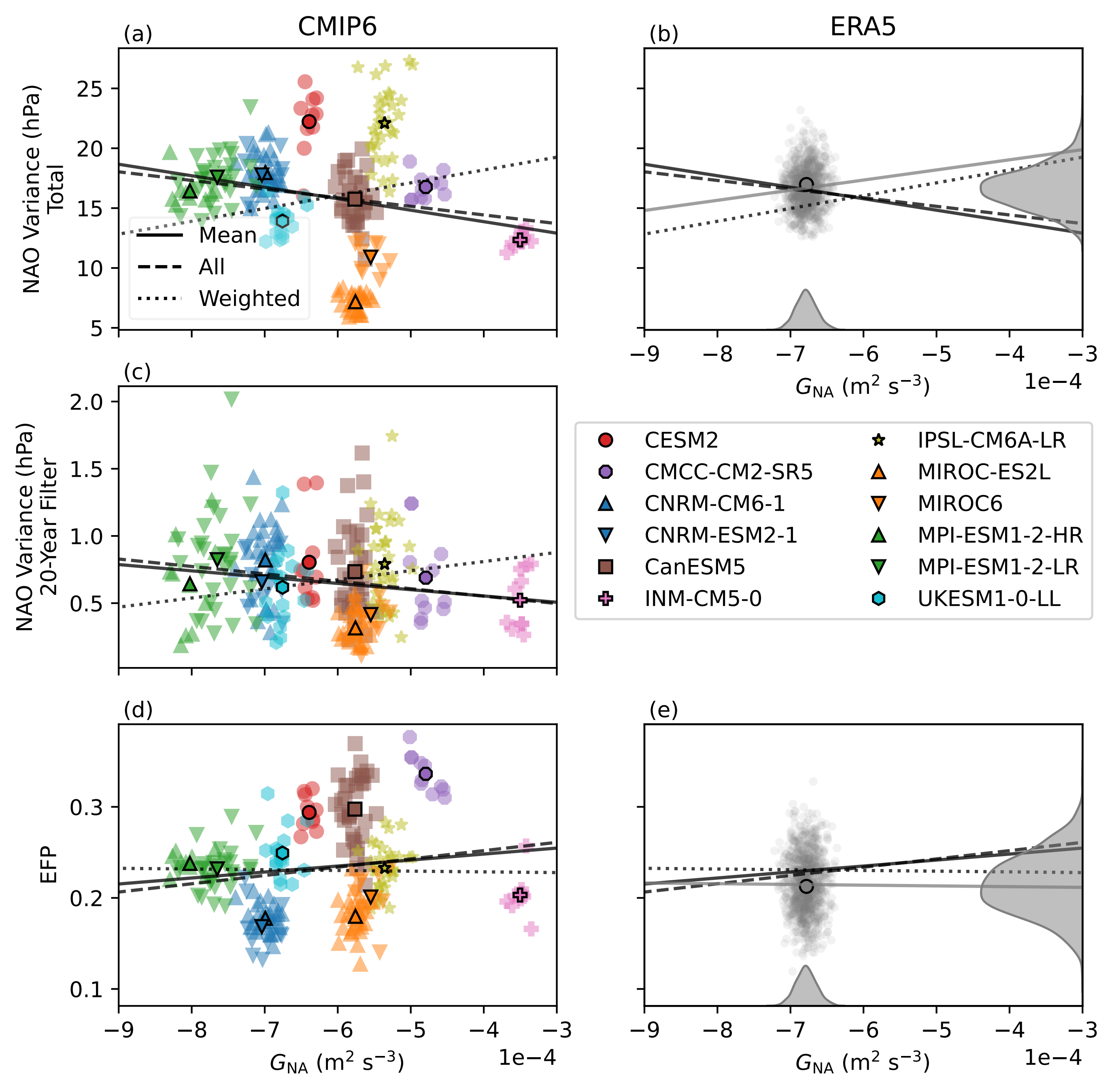
(a)
CMIP6
(b)
ERA5
(c)
NAO Variance (hPa)
Total
NAO Variance (hPa)
20-Year Filter
Eddy-Feedback Parameter
Mean
All
Weighted
CESM2
CMCC-CM2-SR5
CNRM-CM6-1
CNRM-ESM2-1
CanESM5
INM-CM5-0
IPSL-CM6A-LR
MIROC-ES2L
MIROC6
MPI-ESM1-2-HR
MPI-ESM1-2-LR
UKESM1-0-LL

fig2_efp_reanalysis_uncertainties.png.



Eddy-Feedback Parameter - Reanalysis Uncertainties
(a) Sampling Uncertainty
Eddy-Feedback Parameter
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1940-2022
1940-1979
1979-2022
Sampling Period
(b) Multidecadal Variability
ERA5
ERA20c
EFP
NAO
1920
1940
1960
1980
2000
Year
NAO (hPa)
6
7
8
9
10
11
12

fig4_lefp-in-box-cmip6_vs_efp_and_nao.png.



(a)
CMIP6
(b)
ERA5
NAO Variance (hPa)
Total
Mean
All
Weighted
(c)
NAO Variance (hPa)
20-Year Filter
(d)
(e)
EFP
G_{NA} (m^2 s^{-3})
1e−4
CESM2
CMCC-CM2-SR5
CNRM-CM6-1
CNRM-ESM2-1
CanESM5
INM-CM5-0
IPSL-CM6A-LR
MIROC-ES2L
MIROC6
MPI-ESM1-2-HR
MPI-ESM1-2-LR
UKESM1-0-LL

Large sampling uncertainty when diagnosing the 'eddy feedback parameter' and its role in the signal-to-noise paradox

Leo Saffin[1,2], Christine M. McKenna[1,3], Rémy Bonnet[4], and Amanda C. Maycock[1]

[1]School of Earth and Environment, University of Leeds, Leeds, UK
[2]Department of Meteorology, University of Reading, Reading, UK
[3]JBA Consulting, Warrington, UK
[4]CECI, Université de Toulouse, CERFACS/CNRS, Toulouse, France

Key Points:

- The 'eddy feedback parameter' is a highly non-stationary quantity, making reanalysis and model comparisons problematic on short time periods
- Sampling uncertainty in the eddy feedback parameter from reanalysis data is comparable to the intermodel spread across models
- Barotropic energy generation rate is a more stable quantity, but does not explain model spread in North Atlantic climate variability

Corresponding author: Leo Saffin, `l.saffin@reading.ac.uk`

Abstract

Model forecasts on seasonal-to-decadal timescales have recently been shown to have significant skill in predicting the North Atlantic Oscillation (NAO, a large-scale pattern of variability). However, these forecasts are undermined by signal-to-noise ratios that are lower than expected given the skill, meaning the models are underconfident. This problem is known as the "signal-to-noise paradox'. Previous work has shown that models underestimate the strength of feedback from atmospheric eddies onto the midlatitude circulation, but models with a stronger eddy feedback suffer less from the signal-to-noise paradox. However, we find that the "eddy feedback parameter' (EFP) used in these studies exhibits large sampling uncertainty that has not previously been taken into account. When accounting for this sampling uncertainty, the EFP in models is generally consistent with reanalysis data. Furthermore, across samples, the EFP correlates with the variability of the NAO, meaning they are not independent, which makes the EFP problematic for understanding the causes of the signal-to-noise paradox. Samples with larger NAO variability are diagnosed with a larger EFP, even within the same dataset. An alternative measure of eddy feedback is less sensitive to sampling and better identifies models which have weak, strong, or unbiased eddy feedbacks.

Plain Language Summary

Model forecasts on seasonal-to-decadal timescales have recently been shown to have significant skill in predicting the North Atlantic Oscillation (NAO, a large-scale pattern of variability). However, these forecasts are undermined by signal-to-noise ratios that are lower than expected given the skill, meaning the models are underconfident and larger ensembles of simulations are needed to be able to extract the predictable signal. This problem is known as the "signal-to-noise paradox". Previous work has shown that models tend to underestimate the strength of feedback from atmospheric eddies onto the midlatitude circulation, but models with a stronger eddy feedback suffer less from the signal-to-noise paradox, suggesting that more confident predictions would be possible if eddy feedbacks in models were improved. However, we find that the "eddy feedback parameter" (EFP) used in these studies exhibits large sampling uncertainty that has not previously been taken into account. When accounting for this sampling uncertainty, the EFP in models is generally consistent with reanalysis data, rather than being too weak. Furthermore, across samples, the EFP correlates with the variability of the NAO, meaning they are not independent. The lack of independence between the EFP and the NAO makes the EFP problematic for understanding the causes of the signal-to-noise paradox. What could have been interpreted as models with a stronger eddy feedback giving stronger NAO variability, is actually a result of samples with larger NAO variability being diagnosed with a larger EFP, even within the same dataset. We test an alternative measure of eddy feedback and find it is much less sensitive to sampling issues than the EFP, finding no systematic model bias but better distinguishing which models have weak, strong, or unbiased eddy feedbacks.

1 Introduction

The winter North Atlantic Oscillation (NAO) has been shown to be predictable on seasonal (Scaife et al., 2014) and decadal (Smith et al., 2019) timescales. However, the predictable NAO signal in models (variability of the ensemble mean) is weaker than expected given the skill, meaning forecasts are underconfident (Scaife & Smith, 2018). This underconfidence occurs despite models having a relatively good representation of total NAO variability and has been coined the signal-to-noise paradox (Scaife et al., 2014; Scaife & Smith, 2018). This underconfidence could be a manifestation of a too-large component of forecast noise or a too-weak predictable signal (Eade et al., 2014; Scaife & Smith, 2018).

Several studies have investigated whether predictable NAO signals are poorly captured in models, including the representation of teleconnections from the tropics to the North Atlantic (O'Reilly et al., 2019; Williams et al., 2023), the response to Arctic sea ice anomalies (Smith et al., 2022), the response to North Atlantic sea surface temperature (SST) anomalies (Simpson et al., 2018), the response to solar cycle variability (Gray et al., 2013; Scaife et al., 2014) and the response to predictable tropical stratospheric variability (Andrews et al., 2019).

There are currently two main hypotheses to explain the NAO signal-to-noise problem.

1. Weak air-sea coupling in the North Atlantic. This has been shown to contribute to an underestimation of winter North Atlantic eddy-driven jet variability on multidecadal timescales (Simpson et al., 2018; Bracegirdle et al., 2018) and summer NAO variability on decadal timescales (Ossó et al., 2020).
2. Weak eddy feedbacks in midlatitudes. Eddy momentum fluxes can act to reinforce the zonal-mean flow and increase the persistence of jets (Lorenz & Hartmann, 2001, 2003) and the NAO is known to be driven by momentum forcing from synoptic and stationary eddies (Luo et al., 2007). Smith et al. (2022) introduced the "eddy feedback parameter" (EFP) to quantify the relationship between eddy forcing and the midlatitude jet (see Section 2.2.1). Smith et al. (2022) showed the EFP in present day climate correlated with the amplitude of the midlatitude zonal wind response to projected Arctic sea ice loss across a set of climate models. They showed that models underestimated the EFP compared to reanalyses and used an emergent constraint approach to derive a constrained spread of the modeled jet shift. Hardiman et al. (2022) found that models with a weaker EFP (further from reanalysis) generally have less skill and worse signal-to-noise errors for predicting the Northern hemisphere winter circulation.

Most of the work on the NAO signal-to-noise problem has focused on seasonal-to-decadal timescales; it remains an open question as to whether similar issues manifest in multidecadal projections of the NAO including externally forced trends (McKenna & Maycock, 2021). The initial motivation of this work was to test the eddy feedback hypothesis in climate simulations by examining whether the EFP is related to multidecadal NAO variability. However, we found that our results were strongly affected by sampling issues with the EFP not accounted for in past studies. In this study, we address the sampling uncertainty in the EFP within reanalysis and climate model datasets, as well as the inherent relationship between the EFP and NAO characteristics within a sample. The EFP is based on zonal-mean data and does not separate the timescales of eddies and the mean flow. Therefore, we also analyze a spatially-resolved diagnostic of eddy feedback, the barotropic energy generation rate (Mak & Cai, 1989), which allows us to investigate the relationship between North-Atlantic eddy feedback and NAO variability using time-filtered data.

This study is laid out as follows: Section 2 describes the datasets used in the study and methods for quantifying eddy-mean flow feedback, Section 3 presents the results and Section 4 presents a summary of the key findings.

2 Methods

2.1 Datasets

Climate model data is taken from phase 6 of the Coupled Model Intercomparison Project (CMIP6) (Eyring et al., 2016). We use the historical experiment (1850-2014) from 12 CMIP6 models that provide the required variables (monthly-mean mean-sea-level pressure, and daily-mean zonal (u) and meridional (v) wind on pressure levels) for at least 10 ensemble members (see supplement Table S6). We select models that provide large ensembles in order to quantify sampling effects and the role of internal variability in calculating the EFP and its relationship with the NAO. Diagnostics are calculated from data regridded to the coarsest resolution climate model (CanESM5, roughly 2.8°). All diagnostics are for Northern hemisphere winter (DJF) with the year labeled by the JF (e.g. 2009/10 is labeled 2010).

We use the ERA5 (Hersbach et al., 2020) and ERA20C (Poli et al., 2016) reanalysis datasets. The back extension of ERA5 covers the period 1940 to 1978 and the standard ERA5 covers 1979 to present. ERA20C covers 1900-2010 and only assimilates surface pressure and surface marine wind observations. For ERA5 and ERA20C winds, we aggregate six hourly data (00, 06, 12, 18) to daily means to provide an equivalent comparison to the CMIP6 data. We also use monthly-mean mean-sea-level pressure data from 20CRv3 (Slivinski et al., 2019), a longer timescale reanalysis that only assimilates surface pressure, and HadSLP (Allan & Ansell, 2006), a gridded dataset produced from surface pressure observations, to calculate NAO timeseries in the supplement.

2.2 Diagnostics

2.2.1 Eddy feedback parameter

Smith et al. (2022) defined the eddy feedback parameter (EFP) as the squared correlation coefficient (r^2) between the DJF-mean zonal-mean zonal wind ($\bar{u}$) and the DJF-mean of the horizontal component of the Eliassen-Palm flux (EP-flux) divergence, calculated as a function of latitude and pressure, and then averaged over 25-72°N, and 200-600 hPa. Hardiman et al. (2022) used a similar formulation, but calculated the EFP at a single level (500 hPa) and only included the quasi-geostropic component of EP-flux divergence, expressed as a zonal acceleration: eq. 1 from Hardiman et al. (2022),

$$\frac{\nabla.\mathbf{F}_H}{\rho a cos(\phi)} = -\frac{1}{a cos^2\phi}\frac{d(\overline{u'v'}cos^2\phi)}{d\phi}, \tag{1}$$

where ρ is density, ϕ is latitude, a is Earth's radius. Overbars represent a zonal mean, and primes represent local deviations from the zonal mean. Here, we calculate the EFP following Hardiman et al. (2022). The differences in methodology for calculating the EFP can give a different absolute value, but give similar results for the uncertainty (see supplement Fig. S1).

2.2.2 Barotropic energy generation rate

The barotropic energy generation rate (G) diagnoses the exchange of energy between eddies and the large-scale flow based on an energy equation for the ageostrophic perturbation flow in quasi-geostrophic dynamics (Mak & Cai, 1989). If (U, V) describes the large-scale geostrophic wind and (u', v') the eddies, then the barotropic energy generation rate is given by

$$G = \mathbf{E} \cdot \mathbf{D}, \tag{2}$$

where

$$\mathbf{E} = cos(\phi)\left(\frac{1}{2}(v'^2 - u'^2), -u'v'\right), \tag{3}$$

is the E-vector, which describes the elongation of the eddy, and

$$\mathbf{D} = \frac{1}{a cos(\phi)}\left(\frac{\partial U}{\partial \lambda} - \frac{\partial V cos(\phi)}{\partial \phi}, \frac{\partial V}{\partial \lambda} + \frac{\partial U cos(\phi)}{\partial \phi}\right), \tag{4}$$

is the deformation of the large-scale flow (Mak & Cai, 1989), where λ is latitude. Note that we use the spherical coordinate version of these equations from Fukotomi and Yasunari (2002). We diagnose G using daily-mean winds at 250 hPa that are separated into a high frequency (2-6 day) eddy component and a slowly varying (> 10 day) large-scale component using Lanczos filters with a window of 61 days.

In comparison to the EFP, G is spatially-resolved, giving a measure of the local energy exchange. To provide a comparison with the EFP and relate G to NAO variability, we average G over a box in the North Atlantic (60°–25°W, 30°–45°N) giving G_{NA}. This region is where the models and reanalysis show climatological negative values (see supplement Fig. S2), indicating exchange of energy from the eddies to the large scale flow.

2.2.3 North Atlantic Oscillation index

The NAO index is calculated as the difference in DJF area-averaged mean-sea-level pressure between a southern box (90°W–60°E, 20°N–55°N) and a northern box (90°W–60°E, 55°N–90°N) following Stephenson et al. (2006). From the NAO timeseries we calculate variance. Multidecadal NAO variance is also calculated by first applying a 20-year running mean.

The NAO has not been detrended, which could lead to a overestimation of NAO variance in the CMIP6 models compared to ERA5 because we are retaining longer-timescale variability. However, multidecadal variability is only a small part of the total NAO variance (see section 3.3), so the difference in NAO variance due to including these longer timescales is small.

2.3 Statistics

To estimate sampling uncertainty, we recalculate the EFP in ERA5 by resampling winters with replacement (bootstrapping) using the same sample size as the input dataset (e.g. for 1940-2022, each sample is 82 years), repeating 1000 times. We also recalculate the EFP, NAO variance, and G_{NA} in ERA5 in the same way, but with a sample size matching the historical simulation length (164 years) to compare with the CMIP6 simulations. Each diagnostic is calculated using the same sample years, allowing us to assess relationships between these diagnostics due to sampling. Relationships between variables are estimated using linear least squares regression.

3 Results

3.1 Uncertainties in reanalysis derived eddy feedback parameter

In this section, we show how the EFP is affected by sampling uncertainty and multidecadal variability. Figure 1 shows the calculation of the EFP in ERA5 broken into constituent steps. Figure 1a and 1b show the DJF-mean input variables as a function of latitude and year: $\bar{u}$ and the acceleration of $\bar{u}$ diagnosed from the quasigeostropic component of the horizontal EP-flux divergence. The EFP is calculated by calculating the correlation coefficient (r) between these two variables at each latitude and then averaging r^2 across latitudes. r is defined as the covariance of two variables normalized by their standard deviations. To understand how different years and latitudes contribute to the EFP, Fig. 1c shows the anomalies of the input variables, relative to the time mean at each latitude, multiplied together, so the time mean is the covariance as a function of latitude. Figure 1d shows the same, but normalized by the standard deviations of the input variables at each latitude, so the time mean is r as a function of latitude.

Figure 1 reveals two potential issues with the EFP:

1. Calculating r at each latitude and then taking a spatial average overemphasizes latitudes with weaker variability. This can be seen by comparing Fig. 1c and 1d: anomalies are weaker closer to the equator for the covariance but have a larger contribution to r because the standard deviation at those latitudes is smaller.
2. A single outlier season can make a large contribution to the EFP (e.g. 2009/2010 in Fig. 1d). This undermines comparisons of the EFP in reanalysis data and cli-

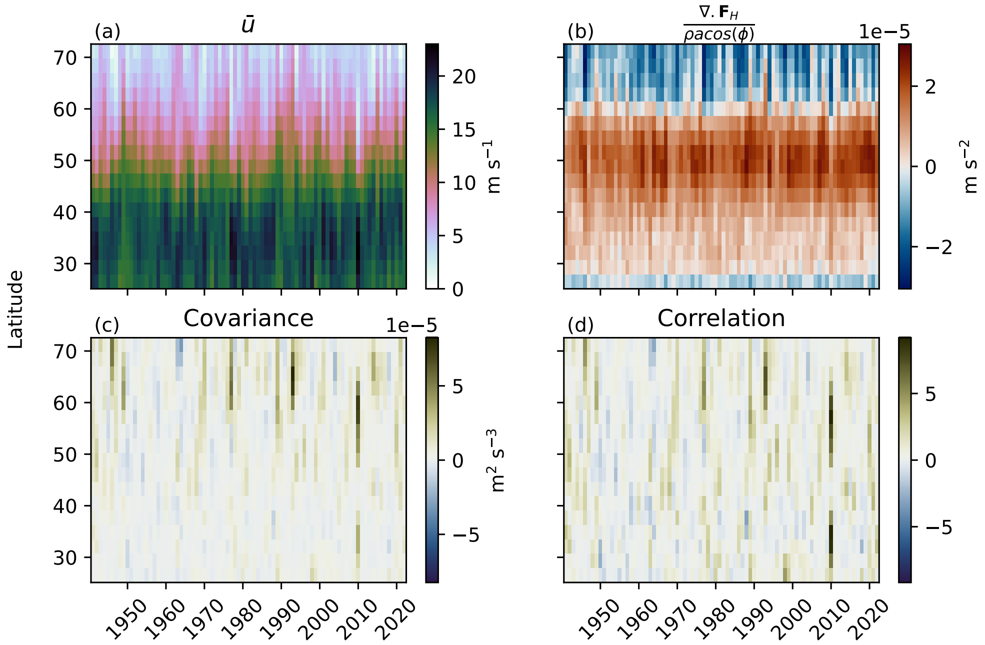


Figure 1. Calculation of the EFP using ERA5. Variables used to calculate the EFP as a DJF mean, (a) zonal-mean zonal wind and (b) the acceleration of the zonal-mean zonal wind diagnosed from the quasi-geostrophic component of the horizontal EP-flux divergence. (c) The product of anomalies of (a) and (b), where the anomalies are calculated against the time mean (mean across rows) by latitude. (d) shows the same as (c), but normalized by the standard deviations, at each latitude, of (a) and (b). The time mean of (c) and (d) give the covariance and correlation as a function of latitude, respectively.

mate models when they do not span a common period and do not sample the same internal variability. For example, if a model with inherently weak eddy feedback happens to simulate a season like 2009/2010, it may appear to have a larger EFP than a model with a strong eddy feedback that by chance does not simulate a season like 2009/2010.

Building on point 2, to quantify the sampling uncertainty we recalculate the EFP by sampling years from ERA5 with replacement (see Section 2.3). Figure 2a shows results with the resampling period varied to show the dependence of "observed" EFP on time period: the full ERA5 period (1940-2022); the pre-satellite backward extension period only (1940-1979); and the satellite period only (1979-2022). In all cases, the sampling uncertainty in the EFP ($\approx$ 0.2-0.3) is comparable to the median value based on the 95% confidence interval. This sampling effect represents a substantial uncertainty that has not been acknowledged in previous studies (e.g., Smith et al. (2022); Hardiman et al. (2022); Screen et al. (2022)).

Figure 2a also shows the EFP is dependent on time period: the satellite period has a larger EFP than the pre-satellite back extension period, with no overlap of the 95% intervals. To better understand the dependence of EFP on time period, we calculate the EFP using a rolling 23-year window (consistent with the 1993-2016 period used in Hardiman et al. (2022)). ERA5 shows a systematic increasing trend in the 23-year EFP (Fig. 2b)). A long-term trend in the EFP could be spurious if the reanalysis is poorly constrained by observations and behaves more like the underlying atmospheric model further back

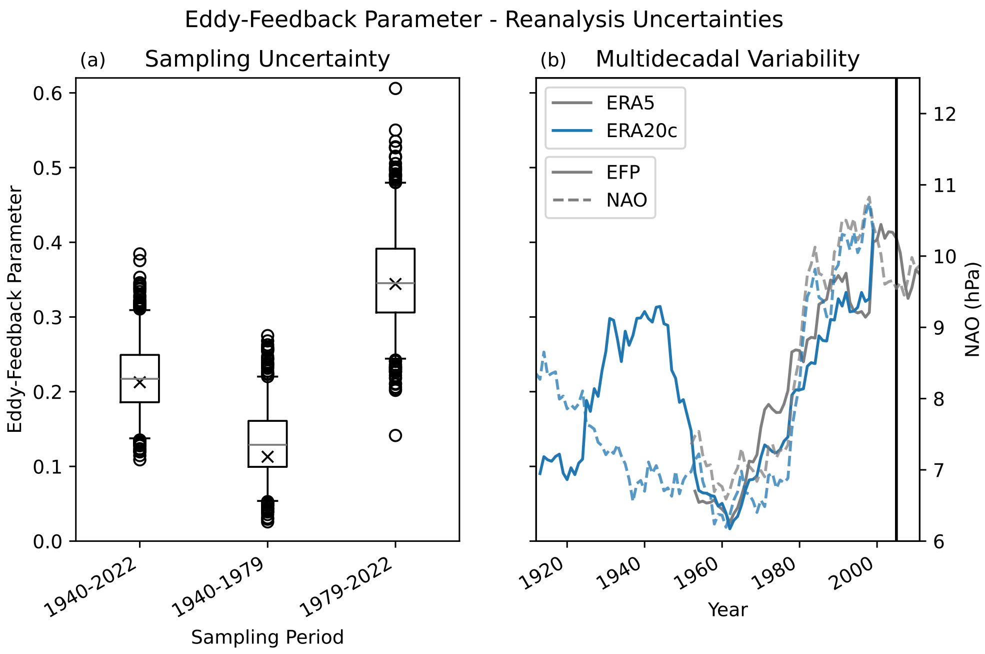


Figure 2. Uncertainties in the EFP identified from reanalysis data. (a) The EFP calculated by resampling ERA5 over different periods. Orange lines show the median, crosses show the EFP from the original set of years, boxes show the 25-75% range, whiskers show the 2.5-97.5% range, and circles show points outside this range. (b) The EFP calculated over 23-year rolling windows and 23-year running mean NAO for ERA5 and ERA20C data. The x-axis shows the middle year in each sample. The vertical line is for 1993-2016, the years used in Hardiman et al. (2022).

in time. Figure. 2b also shows the EFP from ERA20C, which extends back to 1900. Longer-term reanalyses that only assimilate a limited set of surface observations, such as ERA20C, have been shown to produce unrealistic trends as the density of the observation network evolves with time (Krueger et al., 2013; Oliver, 2016; Befort et al., 2016; Bloomfield et al., 2018). However, ERA20C actually shows a larger EFP in the 1930s/1940s when there is less observation data and reproduces the increase in EFP over the late 20th century. This shows the apparent EFP trend is unlikely to be due to an intrinsic bias of weak eddy feedback in the model that produces ERA5 and instead is related to multidecadal variability in the input parameters.

Interestingly, the increase in EFP over the late 20th century closely mirrors the positive trend in the NAO over this period, though this common temporal behavior does not appear in the earlier period covered only by ERA20C. It makes sense that the NAO and EFP are related. Eddy-driven jet latitude is related to the NAO (Woollings et al., 2010) and NAO predictability (Parker et al., 2019; Strommen, 2020), and zonal-mean zonal wind is one of the inputs to the EFP calculation. The EFP calculation also emphasizes large seasonal deviations in jet latitude. For example, winter 2009/10 had a strongly southward shifted jet and negative NAO (Santos et al., 2013). Figure 1d showed how the shift in jet in 2009/10 is emphasized in the correlation calculation and Fig. 2b shows a step increase of almost 0.1 when 2009/10 is included in the rolling window.

The time period used by Hardiman et al. (2022) (1993-2016) is very close to the maximum EFP over the entire 20th century due to the inclusion of 2009/10 and the coincidence with a "high phase" of multidecadal variability. The results in this section show that previous studies have likely overestimated the long-term mean EFP in reanalysis data.

3.2 Comparison of climate models and reanalysis eddy feedback parameter

We next address the comparison of EFP in climate models with reanalysis data in the context of the sampling uncertainties described in the previous section. Figure 3 shows the range of EFP calculated from the CMIP6 ensembles (a, c) and from repeatedly sampling 164 years from ERA5 with replacement (b), as well as the relationship with NAO variance (discussed in the following section). In contrast to previous results, we do not find that the EFP is weaker in models than in reanalysis. The EFP diagnosed from CMIP6 models is generally within the uncertainty from ERA5, with some models potentially having too large EFP (CanESM5, CESM2, CMCC-CM2-SR5). If we only considered the EFP and its associated uncertainty from the satellite period of ERA5 (Fig. 2a), then we would conclude that some CMIP6 models underestimate the EFP. This highlights the importance of considering longer-timescale variability, as well as sampling uncertainty, when quantifying the EFP and the limitation of using the EFP as a diagnostic for model performance.

3.3 Relationship between the eddy feedback parameter and the North Atlantic oscillation

Section 3.1 highlighted a relationship between long-term variations in the EFP and the NAO. We next show how this relationship can lead to correlations that should be interpreted as a sample with larger NAO variability giving a larger EFP, rather than stronger eddy feedbacks leading to stronger NAO variability. Most CMIP6 models capture NAO variability well (Fig. 3a) compared to ERA5 (Fig. 3b). Only MIROC-ES2L is systematically too weak. Some models are potentially too weak (MIROC6, INM-CM5-0) or too strong (IPSL-CM6A-LR, CESM2), but produce ensemble members within the range of ERA5 uncertainty.

The lines in Fig. 3 show linear regressions calculated from the data in each panel in different ways:

1. For “ERA5” (gray line in Fig. 3b) the regression is across the bootstrap samples. Because EFP and NAO variance are calculated using the same sets of sample years, this tells us how the EFP relates to NAO variability purely due to sampling.
2. “Mean” is the regression across the ensemble mean points of all models. This relates to model biases and is what would typically be used for emergent constraints (e.g. Smith et al. (2022))
3. “Weighted” is a weighted average regression across all models. For each model, a regression is calculated across ensemble members. The average slope and intercept are then calculated from these individual model regressions, weighted by the number of ensemble members for each model. This indicates whether a sampling relationship between EFP and NAO variability is present, on average, in individual models.
4. “All” is the regression across all ensemble members of all models with each sample treated independently. This gives a mix between “Mean” and “Weighted”.

The full set of results from the linear regressions are given in the supplement (Tables S1-S5). Note that many of the individual model regressions in 3) are not significant due to low sample sizes and the p-value test is less meaningful for the “ERA5” and “All” regressions because the points are not independent. However, the analysis is intended to

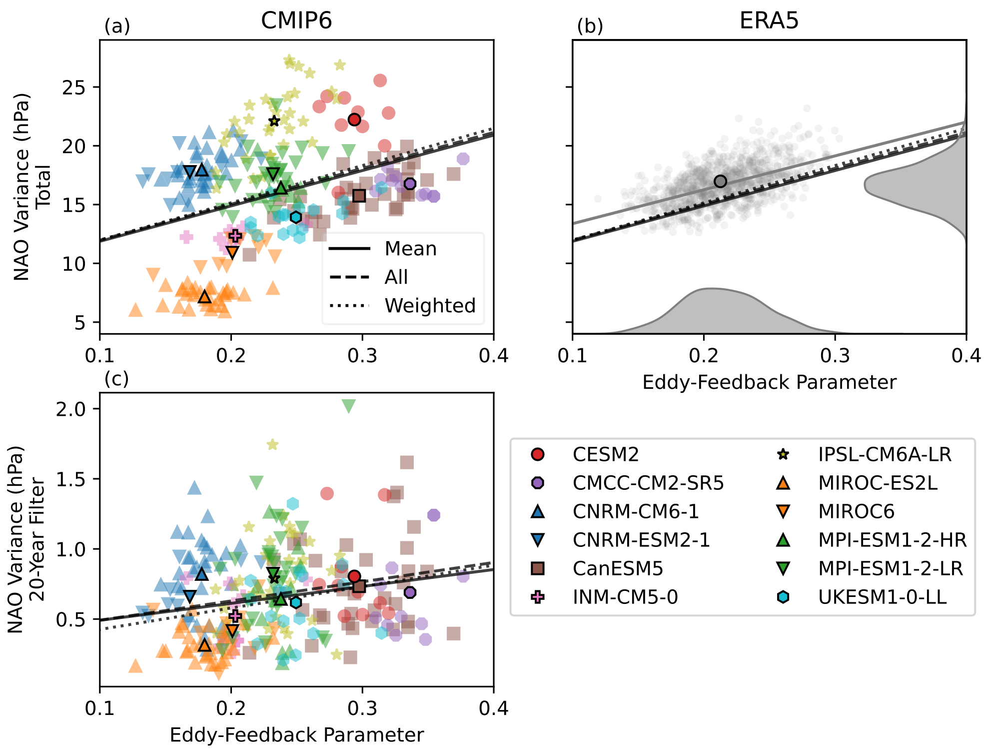


Figure 3. The relationship between the EFP and NAO variance for (a, c) CMIP6 historical simulations (1850-2014) and (b) ERA5 (full period, 1940-2022). (a) The EFP and NAO variance for CMIP6 ensemble members and mean for each model ensemble (outlined symbols). (b) EFP and NAO variance calculated using 164 years sampled from ERA5 with replacement (repeated 1000 times). The outlined dot shows the EFP and NAO variance for the full ERA5 data. (c) The same as (a), but for NAO variance calculated after applying a 20-year running-mean filter. The lines on each subfigure show linear regressions calculated from each set of data in the subfigures (see text for details). The lines from (a) are duplicated in (b) for comparison.

show how sampling issues with the EFP can produce spurious relationships with the NAO rather than identifying significant relationships.

All three regressions in Figure 3a show a similar relationship between EFP and NAO variance and are well reproduced by sampling ERA5 (r=0.34-0.55). This means that the across model relationship between the EFP and NAO variance ("Mean"), which could have been interpreted as physically related model biases, is most likely an extension of the sampling relationship found in ERA5: a model with stronger NAO variability is diagnosed with a larger EFP.

Although total NAO variability is relatively well represented for models exhibiting the signal-to-noise paradox (Scaife & Smith, 2018), weak multidecadal NAO variability Bracegirdle (2022); Bonnet et al. (2024) could be evidence of signal-to-noise issues in climate models. However, similar relationships are found when multidecadal variability is isolated (Fig. 3c), suggesting this is still only identifying sampling relationships. We haven't estimated the reanalysis relationship between the EFP and multidecadal NAO variance because ERA5 is too short for sampling and longer-timescale reanalyses give less consistent values of NAO further back in time (see supplement Figs. S3 and S4).

3.4 Alternative measure of eddy feedback

We next show that an alternative measure of eddy feedback targeted at the North Atlantic (G_{NA}, see section 2.2.2) suffers much less from the sampling issues identified for the EFP. Figure 4 shows G_{NA} for ERA5 and the CMIP6 ensembles and its relationship to NAO variance and the EFP. G_{NA} is better able to identify models that are weak (CanESM5, CESM2, IPSL-CM6A-LR, CMCC-CM2-SR5, INM-CM5-0, MIROC6, MIROC-ES2L), strong (MPI-ESM1-2-LR/HR), or unbiased (UKESM1-0-LL, CNRM-CM6-1, CNRM-ESM2-1) compared to ERA5 due to having much smaller sampling uncertainty.

The sampling relationship between G_{NA} and NAO variability in ERA5 is much weaker (r=0.07) in contrast to that of the EFP and NAO variability (r=0.55). Furthermore, the relationship differs from the (nonsignificant) across model relationship. Similar results are found for multidecadal NAO variability. Interestingly G_{NA} shows no sampling relationship to the EFP and very little relationship across the models used here (Fig 4d,e). This suggests that either the EFP is capturing different aspects of eddy feedback, due to G_{NA} being more localized, or that the EFP is a poor measure of eddy feedback due to the sampling issues shown in earlier.

4 Conclusions

Previous studies have suggested that seasonal prediction systems and free running climate models systematically underestimate Northern hemisphere midlatitude eddy feedbacks (Smith et al., 2022; Screen et al., 2022), and that this bias may explain the signal-to-noise paradox (Scaife et al., 2019; Hardiman et al., 2022). However, we find that the eddy feedback parameter (EFP) used by Smith et al. (2022), Screen et al. (2022), and Hardiman et al. (2022) exhibits large sampling uncertainty which can impede model-reanalysis comparisons and makes determining physical mechanisms difficult.

We have shown that the EFP is sensitive to individual outlier seasons and also exhibits strong multidecadal variability. This makes the EFP problematic to interpret as an intrinsic property of a model or the real world because very large sample sizes are needed to produce an estimate with sufficiently small uncertainties. Previous published estimates of the EFP in modern reanalysis data are close to the maximum value derived within the 1940-2022 period because of the pronounced effect of an outlier season (2009/10) and the phasing of multidecadal variability in the EFP. When sampling uncertainty is taken into account, the EFP in CMIP6 historical simulations is largely consistent with ERA5.

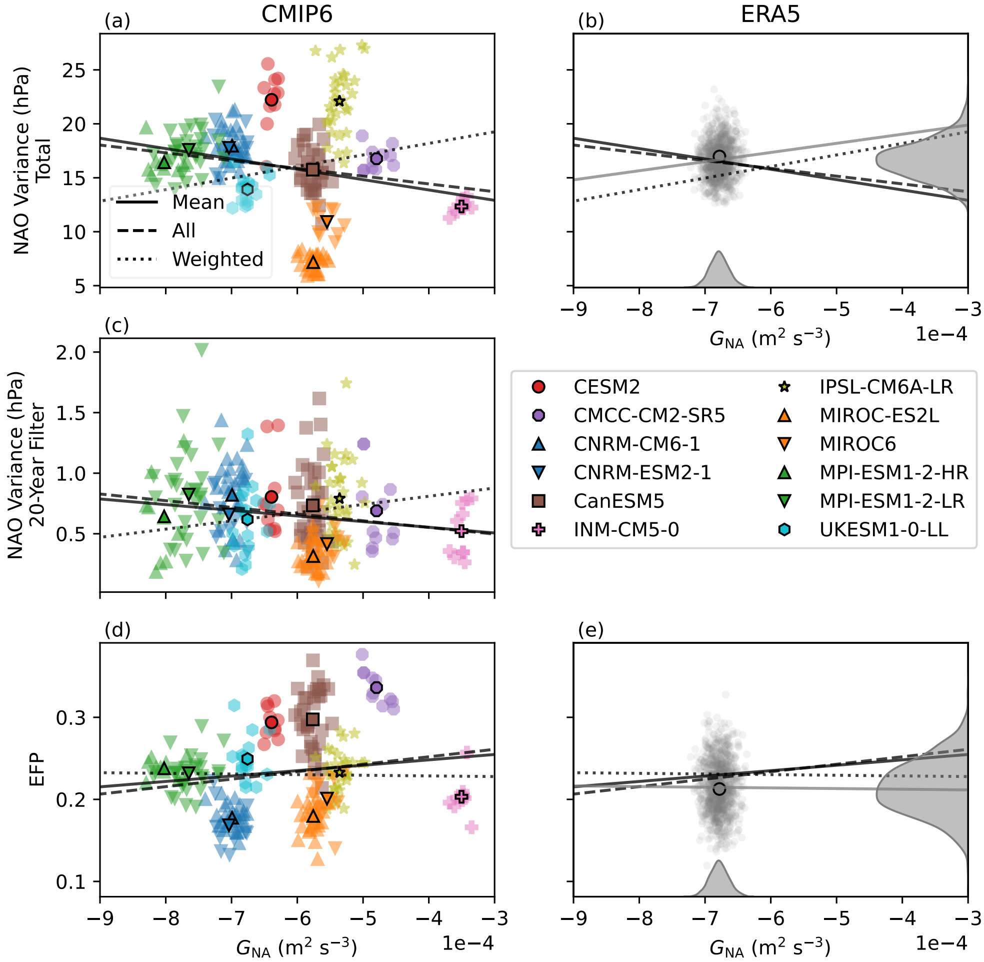


Figure 4. The same as Fig. 3, but with North-Atlantic DJF-mean barotropic energy generation rate (G_{NA}) on the x-axis instead of EFP and extra panels (d) and (e) with EFP on the y-axis for CMIP6 models and ERA5, respectively.

Previous results using the EFP as an emergent constraint (Smith et al., 2022; Screen et al., 2022) should have much larger error bars to account for these sampling uncertainties. These uncertainties may be the reason that Screen et al. (2022) found that the reanalysis EFP in the Southern Hemisphere is roughly in the middle of the model values, while the Northern Hemisphere EFP appeared too weak in models.

We have also shown that the sample EFP correlates with sample NAO variability and this can lead to spurious across-model correlations between the EFP and NAO variability. The across model correlation could have been interpreted as a stronger model eddy feedback causing stronger NAO variability, but is actually due to a sample with stronger NAO variability being diagnosed with a stronger EFP because the EFP and NAO are not independent. The relation between the EFP and NAO makes sense because both variables have an underlying relationship with jet latitude. For example, winter 2009/10 had an anomalously southward shifted jet and negative NAO (Santos et al., 2013) and makes the largest single contribution to the EFP in ERA5. It could be argued that models with stronger eddy feedbacks would produce more years like 2009/10; however, it is clear that we need a much larger sample of data than is available for reanalyses to determine if this is the case.

We have investigated another measure of eddy feedback, the barotropic energy generation rate G, which more cleanly separates eddy forcing and mean flow terms and can be calculated locally for the North Atlantic region (G_{NA}). G_{NA} shows much smaller sampling uncertainty than the EFP and a much weaker sampling relationship with NAO variability, suggesting that it is better at describing intrinsic properties of the models and reanalysis. We find no systematic bias in G_{NA}, but G_{NA} does better distinguish which models are too weak, too strong or unbiased.

In summary, our results raise questions about previous interpretations that weak eddy feedbacks can explain the signal-to-noise paradox. Firstly, we find that models do not systematically underestimate eddy feedbacks when accounting for sampling uncertainty in the EFP or using an alternative, better constrained, diagnostic (G_{NA}). Secondly, the diagnosed EFP from a sample is dependent on the sample NAO variability, which makes it difficult to interpret differences associated with the EFP as being caused by eddy feedbacks rather than some confounding variable. Therefore previous results should be re-examined with a diagnostic of eddy feedback that is more robust to climate variability and where clearer causality can be determined, such as the barotropic energy generation rate or a more in depth lead-lag approach that formally isolates the feedback of eddies on the mean flow (e.g., Lorenz and Hartmann (2001)).

Data availability

The ERA5 reanalysis data is available from the Copernicus Climate Data Store. ERA20C was accessed from the NCAR research data archive. The CMIP6 data used in the study was accessed from the Earth System Grid Federation. The diagnostics calculated for each CMIP6 simulation are given in the supplement (Table S6). The code used for calculating the diagnostics is available at `github.com/leosaffin/constrain` and the code for further processing and making the figures is available at `github.com/leosaffin/eddy_feedback`.

Acknowledgments

We gratefully acknowledge funding for this work by the EU Horizon 2020 project CONSTRAIN (GA no. 820829). We are grateful to Doug Smith and Steven Hardiman for constructive discussion about the work.

References

Allan, R., & Ansell, T. (2006). A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004. *Journal of Climate*, *19*(22), 5816–5842. Retrieved from `https://journals.ametsoc.org/view/journals/clim/19/22/jcli3937.1.xml` doi: https://doi.org/10.1175/JCLI3937.1

Andrews, M. B., Knight, J. R., Scaife, A. A., Lu, Y., Wu, T., Gray, L. J., & Schenzinger, V. (2019). Observed and simulated teleconnections between the stratospheric quasi-biennial oscillation and northern hemisphere winter atmospheric circulation. *Journal of Geophysical Research: Atmospheres*, *124*(3), 1219-1232. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029368` doi: https://doi.org/10.1029/2018JD029368

Befort, D. J., Wild, S., Kruschke, T., Ulbrich, U., & Leckebusch, G. C. (2016, nov). Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. *Atmospheric Science Letters*, *17*(11), 586–595. Retrieved from `https://doi.org/10.1002/asl.694` doi: https://doi.org/10.1002/asl.694

Bloomfield, H. C., Shaffrey, L. C., Hodges, K. I., & Vidale, P. L. (2018). A critical assessment of the long-term changes in the wintertime surface Arctic Oscillation and Northern Hemisphere storminess in the ERA20C reanalysis. *Environmental Research Letters*, *13*(9), 94004. Retrieved from `https://dx.doi.org/10.1088/1748-9326/aad5c5` doi: 10.1088/1748-9326/aad5c5

Bonnet, R., McKenna, C., & Maycock, A. (2024, jan). Model spread in multidecadal NAO variability connected to stratosphere-troposphere coupling. *EGUsphere*, *2024*, 1–24. Retrieved from `https://egusphere.copernicus.org/preprints/2024/egusphere-2023-3103/https://egusphere.copernicus.org/preprints/2024/egusphere-2023-3103/egusphere-2023-3103.pdf` doi: 10.5194/egusphere-2023-3103

Bracegirdle, T. J. (2022, jun). Early-to-Late Winter 20th Century North Atlantic Multidecadal Atmospheric Variability in Observations, CMIP5 and CMIP6. *Geophysical Research Letters*, *49*(11), e2022GL098212. Retrieved from `https://doi.org/10.1029/2022GL098212` doi: https://doi.org/10.1029/2022GL098212

Bracegirdle, T. J., Lu, H., Eade, R., & Woollings, T. (2018, jul). Do CMIP5 Models Reproduce Observed Low-Frequency North Atlantic Jet Variability? *Geophysical Research Letters*, *45*(14), 7204–7212. Retrieved from `https://doi.org/10.1029/2018GL078965https://onlinelibrary.wiley.com/doi/10.1029/2018GL078965` doi: 10.1029/2018GL078965

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*(15), 5620–5628. Retrieved from `https://doi.org/10.1002/2014GL061146` doi: https://doi.org/10.1002/2014GL061146

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016, may). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, *9*(5), 1937–1958. Retrieved from `https://gmd.copernicus.org/articles/9/1937/2016/https://gmd.copernicus.org/articles/9/1937/2016/gmd-9-1937-2016.pdf` doi: 10.5194/gmd-9-1937-2016

Fukotomi, Y., & Yasunari, T. (2002). Tropical-Extratropical Interaction Associated with the 10-25-day Oscillation over the Western Pacific during the Northern Summer. *Journal of the Meteorological Society of Japan. Ser. II*, *80*(2), 311–331. Retrieved from `http://www.jstage.jst.go.jp/article/jmsj/80/2/80_2_311/_article` doi: 10.2151/jmsj.80.311

Gray, L. J., Scaife, A. A., Mitchell, D. M., Osprey, S., Ineson, S., Hardiman, S.,

… Kodera, K. (2013, dec). A lagged response to the 11 year solar cycle in observed winter Atlantic/European weather patterns. *Journal of Geophysical Research: Atmospheres*, *118*(24), 13,405–413,420. Retrieved from https://doi.org/10.1002/2013JD020062 doi: https://doi.org/10.1002/2013JD020062

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Comer, R., Nie, Y., & Ren, H.-L. (2022). Missing eddy feedback may explain weak signal-to-noise ratios in climate predictions. *npj Clim. Atmos. Sci.*, *5*(1), 57. Retrieved from https://doi.org/10.1038/s41612-022-00280-4 doi: 10.1038/s41612-022-00280-4

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, *146*(730), 1999–2049. Retrieved from https://doi.org/10.1002/qj.3803 doi: https://doi.org/10.1002/qj.3803

Krueger, O., Schenk, F., Feser, F., & Weisse, R. (2013). Inconsistencies between Long-Term Trends in Storminess Derived from the 20CR Reanalysis and Observations. *Journal of Climate*, *26*(3), 868–874. Retrieved from https://journals.ametsoc.org/view/journals/clim/26/3/jcli-d-12-00309.1.xml doi: https://doi.org/10.1175/JCLI-D-12-00309.1

Lorenz, D. J., & Hartmann, D. L. (2001). Eddy–Zonal Flow Feedback in the Southern Hemisphere. *Journal of the Atmospheric Sciences*, *58*(21), 3312–3327. Retrieved from https://journals.ametsoc.org/view/journals/atsc/58/21/1520-0469{_}2001{_}058{_}3312{_}ezffit{_}2.0.co{_}2.xml doi: https://doi.org/10.1175/1520-0469(2001)058⟨3312:EZFFIT⟩2.0.CO;2

Lorenz, D. J., & Hartmann, D. L. (2003). Eddy–Zonal Flow Feedback in the Northern Hemisphere Winter. *Journal of Climate*, *16*(8), 1212–1227. Retrieved from https://journals.ametsoc.org/view/journals/clim/16/8/1520-0442_2003_16_1212_effitn_2.0.co_2.xml doi: https://doi.org/10.1175/1520-0442(2003)16⟨1212:EFFITN⟩2.0.CO;2

Luo, D., Gong, T., & Diao, Y. (2007). Dynamics of eddy-driven low-frequency dipole modes. part iii: Meridional displacement of westerly jet anomalies during two phases of nao. *Journal of the Atmospheric Sciences*, *64*(9), 3232 - 3248. Retrieved from https://journals.ametsoc.org/view/journals/atsc/64/9/jas3998.1.xml doi: https://doi.org/10.1175/JAS3998.1

Mak, M., & Cai, M. (1989). Local Barotropic Instability. *J. Atmos. Sci.*, *46*(21), 3289–3311. Retrieved from https://journals.ametsoc.org/view/journals/atsc/46/21/1520-0469{_}1989{_}046{_}3289{_}lbi{_}2{_}0{_}co{_}2.xml doi: https://doi.org/10.1175/1520-0469(1989)046⟨3289:LBI⟩2.0.CO;2

McKenna, C. M., & Maycock, A. C. (2021). Sources of uncertainty in multimodel large ensemble projections of the winter north atlantic oscillation. *Geophysical Research Letters*, *48*(14), e2021GL093258. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL093258 (e2021GL093258 2021GL093258) doi: https://doi.org/10.1029/2021GL093258

Oliver, E. C. J. (2016, aug). Blind use of reanalysis data: apparent trends in Madden–Julian Oscillation activity driven by observational changes. *International Journal of Climatology*, *36*(10), 3458–3468. Retrieved from https://doi.org/10.1002/joc.4568 doi: https://doi.org/10.1002/joc.4568

O'Reilly, C. H., Weisheimer, A., Woollings, T., Gray, L. J., & MacLeod, D. (2019, jan). The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Q. J. R. Meteorol. Soc.*, *145*(718), 131–146. Retrieved from https://doi.org/10.1002/qj.3413 doi: https://doi.org/10.1002/qj.3413

Ossó, A., Sutton, R., Shaffrey, L., & Dong, B. (2020). Development, Amplification, and Decay of Atlantic/European Summer Weather Patterns Linked to Spring

North Atlantic Sea Surface Temperatures. *J. Clim.*, *33*(14), 5939–5951. Retrieved from https://journals.ametsoc.org/view/journals/clim/33/14/JCLI-D-19-0613.1.xml doi: https://doi.org/10.1175/JCLI-D-19-0613.1

Parker, T., Woollings, T., Weisheimer, A., O'Reilly, C., Baker, L., & Shaffrey, L. (2019, aug). Seasonal Predictability of the Winter North Atlantic Oscillation From a Jet Stream Perspective. *Geophys. Res. Lett.*, *46*(16), 10159–10167. Retrieved from https://doi.org/10.1029/2019GL084402 doi: https://doi.org/10.1029/2019GL084402

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., ... Fisher, M. (2016). ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, *29*(11), 4083–4097. Retrieved from https://journals.ametsoc.org/view/journals/clim/29/11/jcli-d-15-0556.1.xml doi: https://doi.org/10.1175/JCLI-D-15-0556.1

Santos, J. A., Woollings, T., & Pinto, J. G. (2013). Are the Winters 2010 and 2012 Archetypes Exhibiting Extreme Opposite Behavior of the North Atlantic Jet Stream? *Mon. Weather Rev.*, *141*(10), 3626–3640. Retrieved from https://journals.ametsoc.org/view/journals/mwre/141/10/mwr-d-13-00024.1.xml doi: https://doi.org/10.1175/MWR-D-13-00024.1

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., ... Williams, A. (2014). Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, *41*(7), 2514–2519. Retrieved from https://doi.org/10.1002/2014GL059637 doi: https://doi.org/10.1002/2014GL059637

Scaife, A. A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., ... Vidale, P. L. (2019). Does increased atmospheric resolution improve seasonal climate predictions? *Atmospheric Science Letters*, *20*(8), e922. Retrieved from https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.922 doi: https://doi.org/10.1002/asl.922

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Clim. Atmos. Sci.*, *1*(1), 28. Retrieved from https://doi.org/10.1038/s41612-018-0038-4 doi: 10.1038/s41612-018-0038-4

Screen, J. A., Eade, R., Smith, D. M., Thomson, S., & Yu, H. (2022, dec). Net Equatorward Shift of the Jet Streams When the Contribution From Sea-Ice Loss Is Constrained by Observed Eddy Feedback. *Geophys. Res. Lett.*, *49*(23), e2022GL100523. Retrieved from https://doi.org/10.1029/2022GL100523 doi: https://doi.org/10.1029/2022GL100523

Simpson, I. R., Deser, C., McKinnon, K. A., & 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), 8313–8338. Retrieved from https://journals.ametsoc.org/view/journals/clim/31/20/jcli-d-18-0168.1.xml doi: 10.1175/JCLI-D-18-0168.1

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., ... Wyszyński, P. (2019, oct). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908. Retrieved from https://doi.org/10.1002/qj.3598 doi: https://doi.org/10.1002/qj.3598

Smith, D. M., Eade, R., Andrews, M. B., Ayres, H., Clark, A., Chripko, S., ... Walsh, A. (2022). Robust but weak winter atmospheric circulation response to future Arctic sea ice loss. *Nat. Commun.*, *13*(1), 727. Retrieved from https://doi.org/10.1038/s41467-022-28283-y doi: 10.1038/s41467-022-28283-y

Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., ... Yang, X. (2019). Robust skill of decadal climate predictions. *npj Clim. Atmos. Sci.*, *2*(1), 13. Retrieved from https://doi.org/10.1038/

s41612-019-0071-yhttps://www.nature.com/articles/s41612-019-0071-y doi: 10.1038/s41612-019-0071-y

Stephenson, D. B., Pavan, V., Collins, M., Junge, M. M., Quadrelli, R., & Groups, P. C. M. (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), 401–420. Retrieved from https://doi.org/10.1007/s00382-006-0140-x doi: 10.1007/s00382-006-0140-x

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

Williams, N. C., Scaife, A. A., & Screen, J. A. (2023, mar). Underpredicted ENSO Teleconnections in Seasonal Forecasts. *Geophysical Research Letters*, *50*(5), e2022GL101689. Retrieved from https://doi.org/10.1029/2022GL101689 doi: https://doi.org/10.1029/2022GL101689

Woollings, T., Hannachi, A., & Hoskins, B. (2010, apr). Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 856–868. doi: 10.1002/qj.625

Supporting Information for 'Large sampling uncertainty when diagnosing the 'eddy feedback parameter' and its role in the signal-to-noise paradox"

Leo Saffin[1,2], Christine M. McKenna[1,3], Rémy Bonnet[4], and Amanda C. Maycock[1]

[1]School of Earth and Environment, University of Leeds, Leeds, UK

[2]Department of Meteorology, University of Reading, Reading, UK

[3]JBA Consulting, Warrington, UK

[4]CECI, Université de Toulouse, CERFACS/CNRS, Toulouse, France

Contents of this file

1. Text S1
2. Figures S1 to S4
3. Tables S1 to S5

Additional Supporting Information (Files uploaded separately)

1. Caption for large table S6 (uploaded as separate excel file)

Introduction The supporting information in this document includes a comparison of using different pressure levels to calculate the EFP (Fig. S1), maps of barotropic energy generation rate for each model and ERA5 (Fig. S2), various estimates of NAO variance from different reanalysis datasets (Figs. S3 and S4), the caption for Table S1 uploaded separately, and the full results of the linear regressions from Figs. 3 and 4 (Tables S2-S6).

Table S1. Table S6 shows the CMIP6 historical simulations used in this study listed by model and variant, as well as the calculated EFP, NAO variance, multidecadal NAO variance, and G_{NA} for each simulation.

References

Hardiman, S. C., Dunstone, N. J., Scaife, A. A., Smith, D. M., Comer, R., Nie, Y., & Ren, H.-L. (2022). Missing eddy feedback may explain weak signal-to-noise ratios in climate predictions. *npj Clim. Atmos. Sci.*, *5*(1), 57. Retrieved from `https://doi.org/10.1038/s41612-022-00280-4` doi: 10.1038/s41612-022-00280-4

Smith, D. M., Eade, R., Andrews, M. B., Ayres, H., Clark, A., Chripko, S., … Walsh, A. (2022). Robust but weak winter atmospheric circulation response to future Arctic sea ice loss. *Nat. Commun.*, *13*(1), 727. Retrieved from `https://doi.org/10.1038/s41467-022-28283-y` doi: 10.1038/s41467-022-28283-y

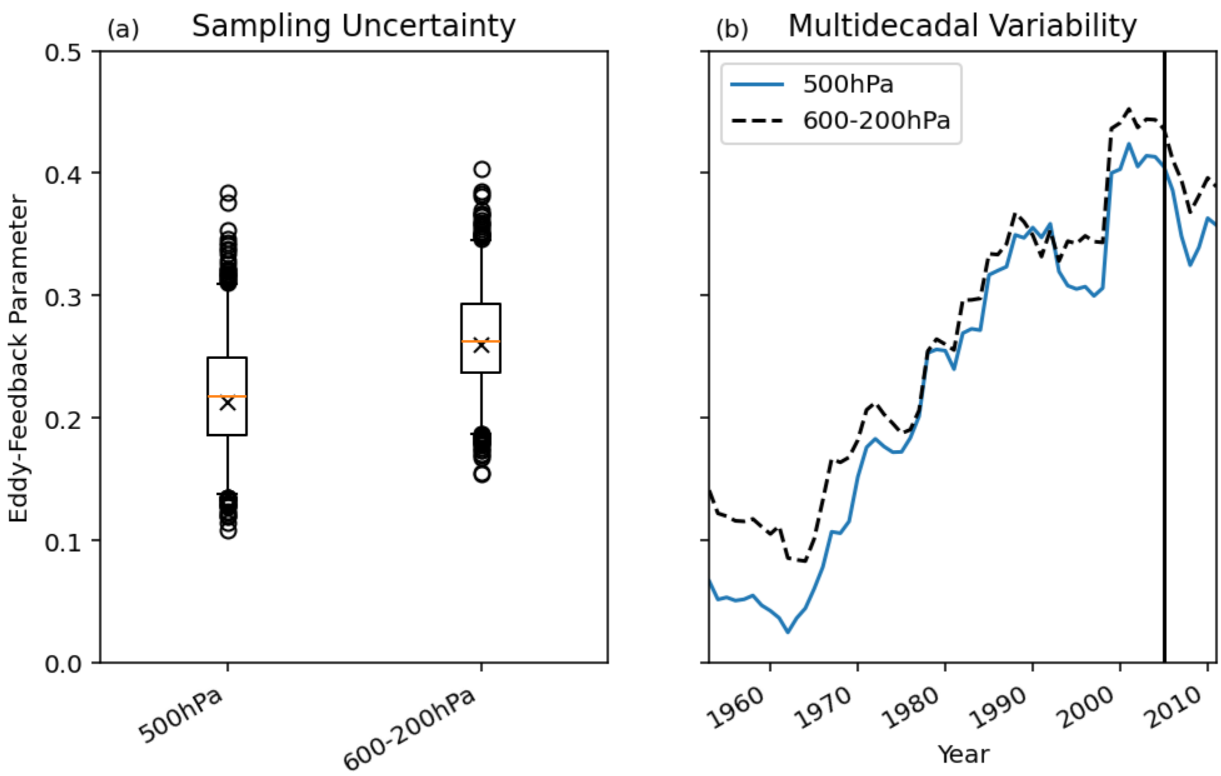


Figure S1. Uncertainties in the EFP calculation in ERA5 for (a) sampling and (b) multidecadal variability, calculated in the same way as Fig. 2 but comparing averaging r^2 at 500 hPa (following Hardiman et al. (2022) and averaging r^2 from 600-200 hPa (following Smith et al. (2022)), here using data at every 50 hPa. The 600-200 hPa average generally shows larger EFP but similar uncertainty.

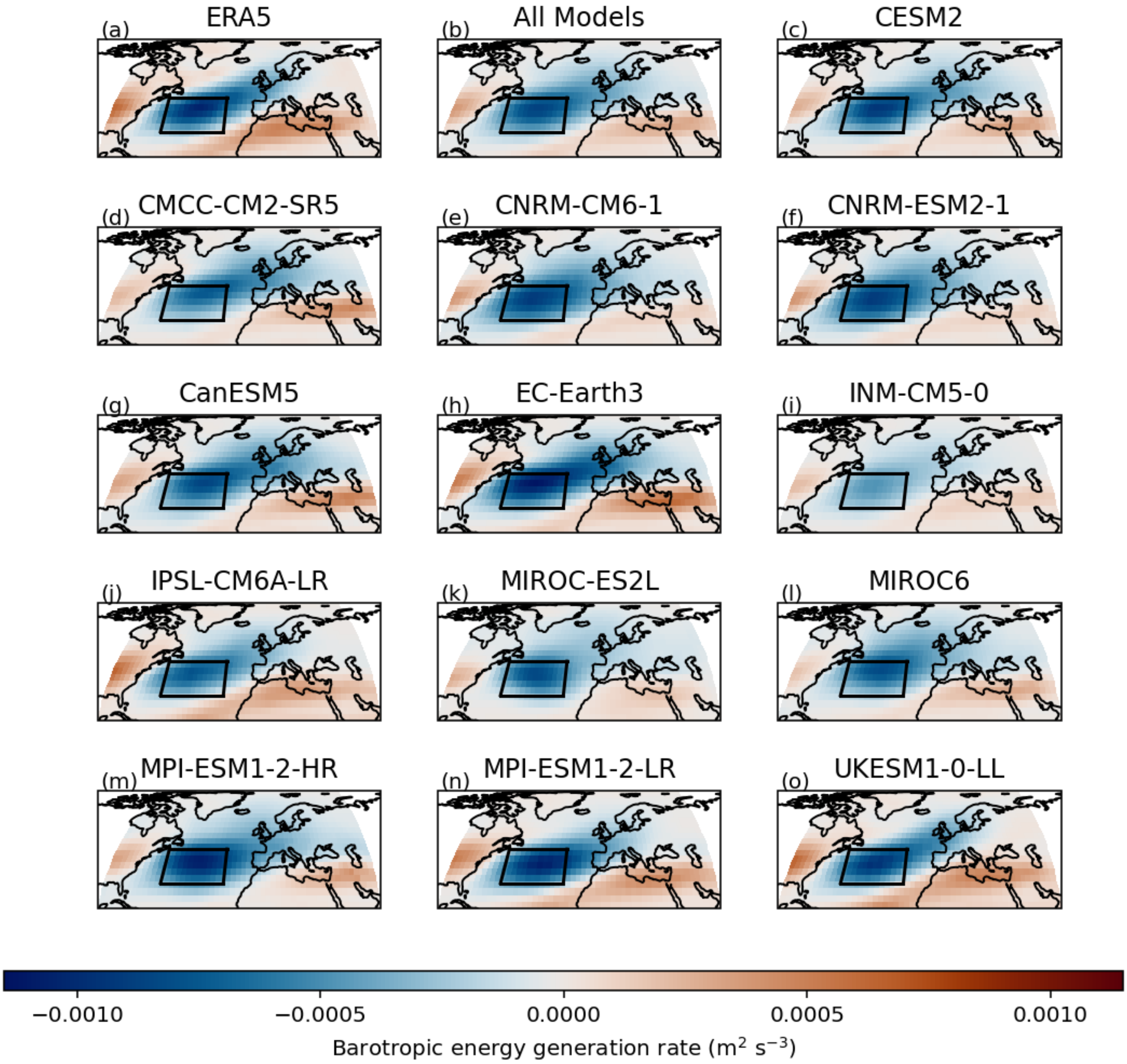


Figure S2. The DJF mean barotropic energy generation rate (G) for (a) ERA5 (1940-2022), (b) CMIP6 multi-model mean for historical simulations (1850-2014) and (c)-(o) CMIP6 ensemble mean for each model for historical simulations. The black box in each figure covers 60°–25°W, 30°–45°N and is the averaging region used in the main paper to reduce G to a single number, G_{NA}. The box is designed to capture the regional minimum while excluding positive regions to avoid cancellation.

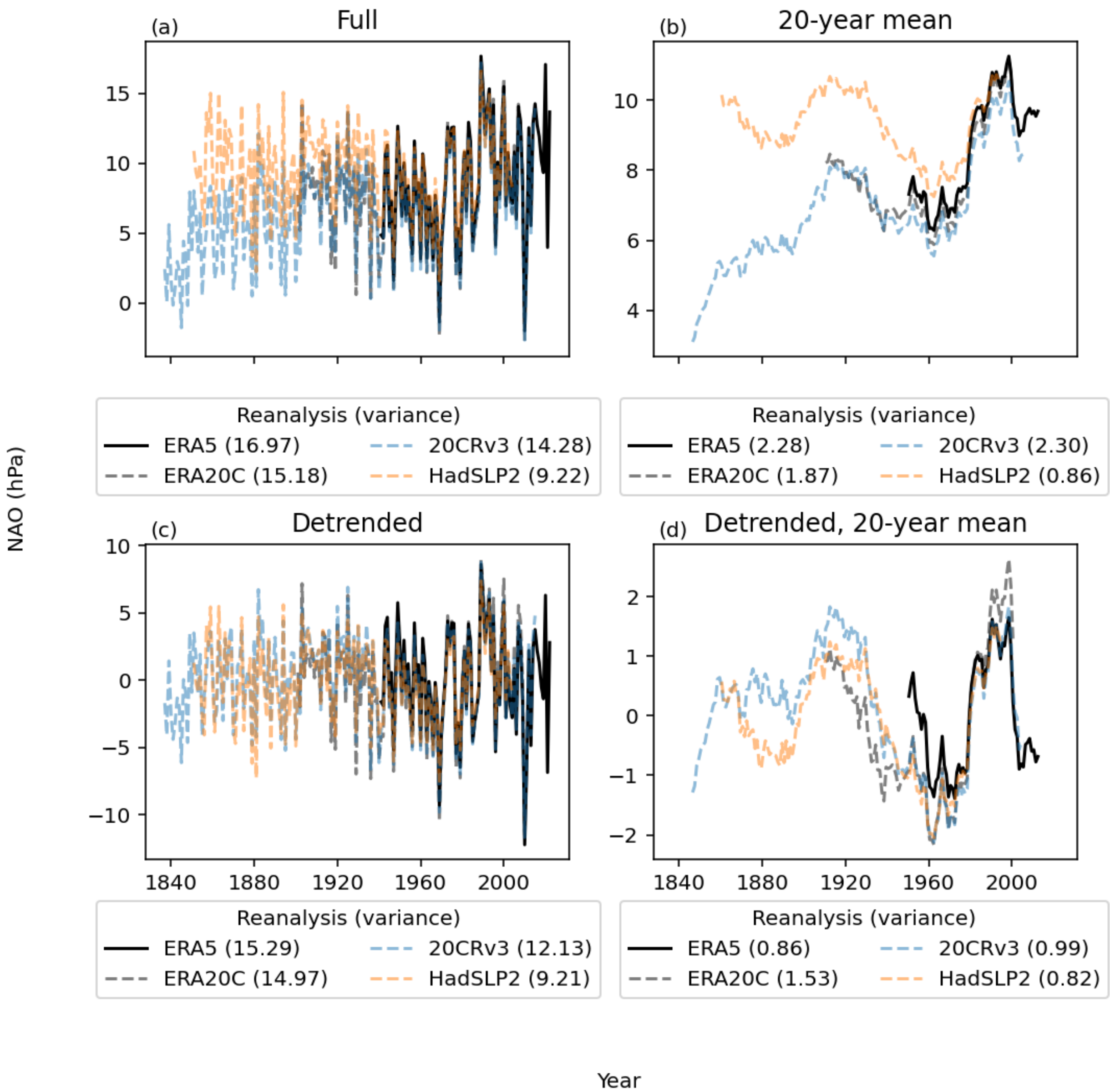


Figure S3. Timeseries of the NAO (a, b) and 20-year running mean NAO (c, d) calculated using different reanalysis datasets, and the influence of detrending these timeseries (c, d). The numbers in each panel show the NAO variance for the timeseries in that panel for each dataset.

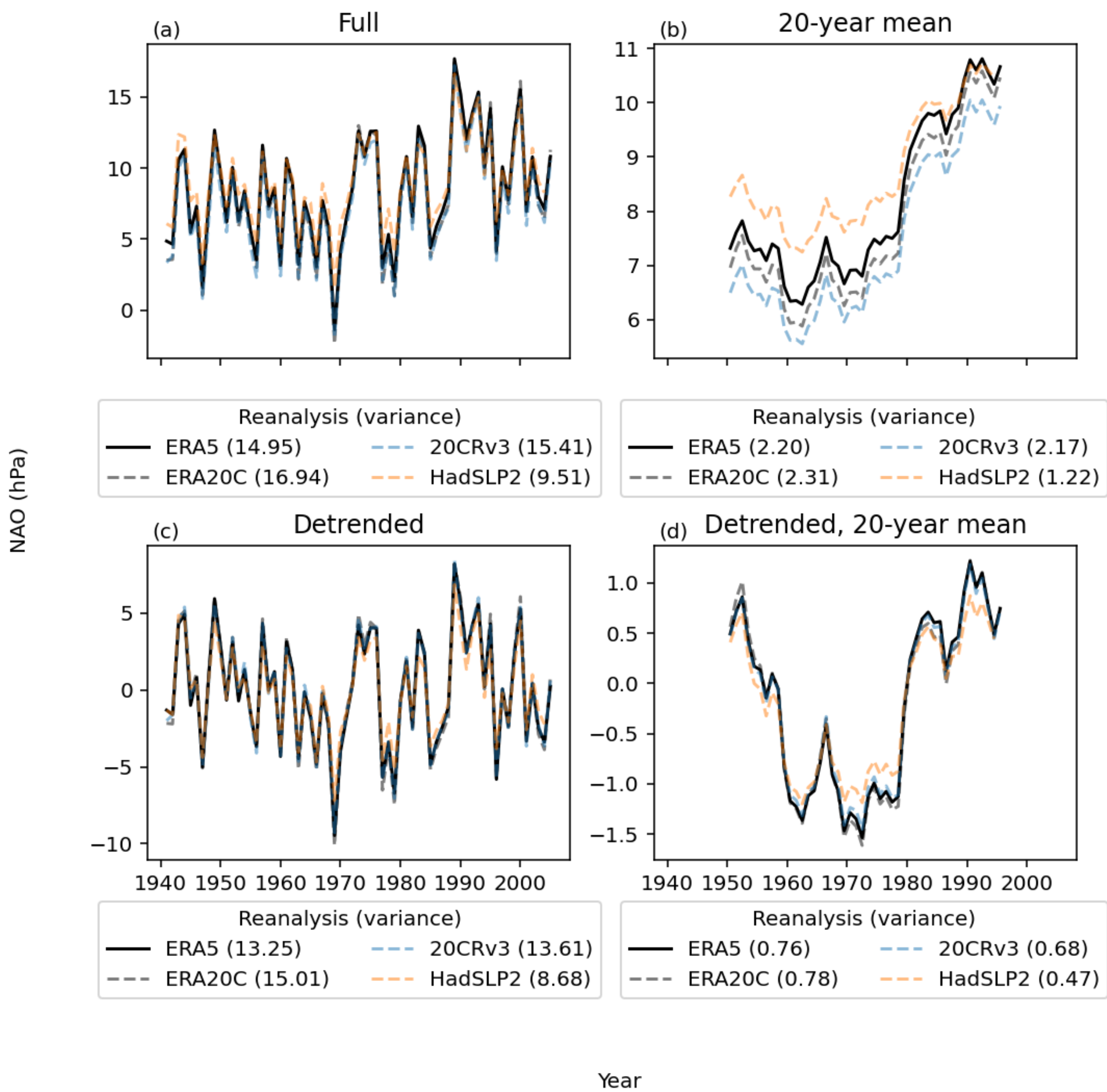


Figure S4. The same as Fig. S3, but only using data in the period that is common to all reanalyses (1900-2010).

Table S2. Linear regression results for EFP vs. total NAO variance.

	Slope	Intercept	Correlation	p-value
ERA5	28.83	10.49	0.55	6.7e-79
CESM2	7.22	20.10	0.05	0.89
CMCC-CM2-SR5	0.75	16.52	0.015	0.97
CNRM-CM6-1	55.62	8.09	0.5	0.0052
CNRM-ESM2-1	30.97	12.56	0.51	0.13
CanESM5	36.69	4.83	0.68	7.8e-06
INM-CM5-0	15.27	9.24	0.57	0.082
IPSL-CM6A-LR	74.84	4.67	0.63	7.7e-05
MIROC-ES2L	5.57	6.18	0.16	0.4
MIROC6	31.62	4.55	0.71	0.022
MPI-ESM1-2-HR	-53.21	29.09	-0.49	0.15
MPI-ESM1-2-LR	37.47	8.90	0.42	0.026
UKESM1-0-LL	26.34	7.35	0.56	0.019
Mean	30.10	8.86	0.37	0.24
All	30.53	8.92	0.34	8.1e-08
Weighted	31.68	8.81	0.42	-

Table S3. Linear regression results for EFP vs. mulitdecadal NAO variance.

	Slope	Intercept	Correlation	p-value
CESM2	-2.44	1.52	-0.14	0.7
CMCC-CM2-SR5	6.50	-1.50	0.42	0.19
CNRM-CM6-1	-2.33	1.23	-0.12	0.54
CNRM-ESM2-1	2.91	0.17	0.42	0.23
CanESM5	1.68	0.24	0.19	0.28
INM-CM5-0	1.62	0.19	0.17	0.64
IPSL-CM6A-LR	2.24	0.27	0.18	0.33
MIROC-ES2L	0.63	0.20	0.12	0.52
MIROC6	-0.46	0.51	-0.074	0.84
MPI-ESM1-2-HR	0.63	0.49	0.02	0.96
MPI-ESM1-2-LR	6.24	-0.62	0.36	0.057
UKESM1-0-LL	0.41	0.52	0.039	0.88
Mean	1.21	0.37	0.39	0.21
All	1.37	0.35	0.22	0.0006
Weighted	1.53	0.27	0.14	-

Table S4. Linear regression results for G_{NA} vs. total NAO variance.

	Slope	Intercept	Correlation	p-value
ERA5	8461.79	22.42	0.073	0.022
CESM2	131289.14	106.14	0.38	0.28
CMCC-CM2-SR5	8617.46	20.90	0.14	0.68
CNRM-CM6-1	-14521.78	7.82	-0.14	0.47
CNRM-ESM2-1	20153.73	31.97	0.23	0.52
CanESM5	-36974.19	-5.57	-0.24	0.16
INM-CM5-0	40523.43	26.54	0.65	0.04
IPSL-CM6A-LR	43468.05	45.39	0.23	0.2
MIROC-ES2L	-4334.54	4.68	-0.07	0.71
MIROC6	-23431.50	-2.09	-0.31	0.38
MPI-ESM1-2-HR	-16241.35	3.40	-0.3	0.41
MPI-ESM1-2-LR	29384.11	40.07	0.39	0.042
UKESM1-0-LL	28850.67	33.41	0.35	0.17
Mean	-9590.76	10.03	-0.28	0.38
All	-7204.36	11.55	-0.16	0.015
Weighted	10707.70	22.47	0.075	-

Table S5. Linear regression results for G_{NA} vs. multidecadal NAO variance.

	Slope	Intercept	Correlation	p-value
CESM2	2963.38	2.70	0.069	0.85
CMCC-CM2-SR5	-7903.51	-3.10	-0.44	0.18
CNRM-CM6-1	1225.89	1.68	0.065	0.73
CNRM-ESM2-1	-2613.77	-1.18	-0.26	0.46
CanESM5	271.12	0.89	0.011	0.95
INM-CM5-0	11251.15	4.46	0.5	0.14
IPSL-CM6A-LR	-613.47	0.46	-0.03	0.87
MIROC-ES2L	-547.56	0.00	-0.061	0.75
MIROC6	557.26	0.72	0.053	0.88
MPI-ESM1-2-HR	4283.84	4.08	0.27	0.45
MPI-ESM1-2-LR	167.79	0.95	0.011	0.95
UKESM1-0-LL	3939.05	3.28	0.21	0.42
Mean	-467.55	0.37	-0.36	0.25
All	-551.83	0.33	-0.18	0.0067
Weighted	678.36	1.08	0.021	-

Table S6. Linear regression results for G_{NA} vs. EFP.

	Slope	Intercept	Correlation	p-value
ERA5	-8.08	0.21	-0.0037	0.91
CESM2	-79.61	0.24	-0.033	0.93
CMCC-CM2-SR5	-1009.61	-0.15	-0.86	0.00073
CNRM-CM6-1	-262.64	-0.01	-0.27	0.14
CNRM-ESM2-1	-357.42	-0.08	-0.25	0.49
CanESM5	-4.72	0.29	-0.0017	0.99
INM-CM5-0	174.96	0.26	0.075	0.84
IPSL-CM6A-LR	357.42	0.42	0.22	0.22
MIROC-ES2L	157.37	0.27	0.09	0.64
MIROC6	44.54	0.23	0.026	0.94
MPI-ESM1-2-HR	-0.31	0.24	-0.00062	1
MPI-ESM1-2-LR	204.11	0.39	0.24	0.22
UKESM1-0-LL	-160.93	0.14	-0.09	0.73
Mean	65.74	0.27	0.16	0.62
All	90.96	0.29	0.18	0.0059
Weighted	-7.83	0.23	-0.019	-